BMC Genomics

BioMed Central

Research article

Open Access

Comparison of mixed-model approaches for association mapping in rapeseed, potato, sugar beet, maize, and Arabidopsis

Benjamin Stich and Albrecht E Melchinger*

Address: Department of Applied Genetics and Plant Breeding, Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, 70593, Stuttgart, Germany

Email: Benjamin Stich - stich@mpiz-koeln.mpg.de; Albrecht E Melchinger* - melchinger@uni-hohenheim.de

* Corresponding author

Published: 27 February 2009

BMC Genomics 2009, **10**:94 doi:10.1186/1471-2164-10-94

Received: 3 June 2008
Accepted: 27 February 2009

This article is available from: http://www.biomedcentral.com/1471-2164/10/94

Abstract

Background: In recent years, several attempts have been made in plant genetics to detect QTL by using association mapping methods. The objectives of this study were to (i) evaluate various methods for association mapping in five plant species and (ii) for three traits in each of the plant species compare the T_{opt}, the restricted maximum likelihood (REML) estimate of the conditional probability that two genotypes carry at the same locus alleles that are identical in state but not identical by descent. In order to compare the association mapping methods based on scenarios with realistic estimates of population structure and familial relatedness, we analyzed phenotypic and genotypic data of rapeseed, potato, sugar beet, maize, and Arabidopsis. For the same reason, QTL effects were simulated on top of the observed phenotypic values when examining the adjusted power for QTL detection.

Results: The correlation between the T_{opt} values identified using REML deviance profiles and profiles of the mean of squared difference between observed and expected *P* values was 0.83.

Conclusion: The mixed-model association mapping approaches using a kinship matrix, which was based on T_{opt}, were more appropriate for association mapping than the recently proposed QK method with respect to the adherence to the nominal α level and the adjusted power for QTL detection. Furthermore, we showed that T_{opt} differs considerably among the five plant species but only marginally among different traits.

Background

Artificially induced variations, such as mutations, have been successfully used for gene identification in genetic and physiological studies [1]. Development of DNA markers, however, has made it possible to study the naturally occuring allelic variation underlying complex traits [2,3]. In many plant species, the approaches for detecting quantitative trait loci (QTL) relied so far on segregating populations derived from crosses between inbred lines. These QTL detection procedures, commonly referred to as linkage mapping, have major limitations, that include high costs [4] and a poor resolution in detecting QTL. Moreover, with biparental crosses of inbred lines only two alleles at any given locus can be studied simultaneously [5]. Association mapping methods, which have been successfully applied in human genetics to detect genes coding for human diseases [6], promise to overcome these limitations [7]. Therefore, in recent years several attempts have been made in a plant genetics context to detect QTL by using such methods [7-10].

Linkage disequilibrium (LD) in linkage mapping populations is caused by genetic linkage [9]. In contrast, LD in association mapping populations can also be the consequence of population structure, relatedness, genetic drift, and selection [5,11]. Therefore, the success of association mapping efforts depends on the ability to separate LD due to linkage from LD due to other causes. To correct for LD caused by population structure, linear models accounting for sub-population effects [8] or a logistic regression ratio test [12,13] were proposed. Owing to the large germplasm sets required for dissecting complex traits [14], the probability of including partially related individuals increases. This applies in particular when genotypes selected from plant breeding populations are used for association mapping [7,9,13]. However, the above-mentioned approaches fail to adhere to the nominal α level if the germplasm set analyzed comprises related individuals [13].

The recently proposed QK mixed-model for association mapping promises to correct for LD caused by population structure and familial relatedness [15]. The authors demonstrated the suitability of their new method for association mapping in two allogamous species, humans and maize. The suitability of the QK method, however, has to be evaluated in plant species with different reproduction systems covering a wide range of population structure and familial relatedness.

In contrast to coancestry coefficients calculated from pedigree records, marker-based kinship estimates may account for the effects of deviations from expected parental contributions to progeny due to selection or genetic drift [16]. Therefore, marker-based kinship estimates might be more appropriate for association mapping approaches than coancestry coefficients calculated from pedigree records [15,17]. A difficulty with calculation of marker-based kinship estimates is the definition of unrelated individuals [18]. The marker-based kinship matrix might be determined based on the definition that random pairs of genotypes are unrelated [15] or that pairs of genotypes are unrelated if they have no allele in common [17]. However, both definitions seem to be arbitrary. Recently, it was proposed to estimate by restricted maximum likelihood (REML) T_{opt}, the conditional probability that marker alleles are alike in state, given that they are not identical by descent [19], using genotypic and phenotypic data [20]. However, no study compared this estimation of unrelated individuals among plant species with different reproduction systems as well as for various phenotypic traits.

The objectives of our study were to evaluate various methods for association mapping with respect to their adherence to the nominal α level and the adjusted power for QTL detection based on (i) empirical data sets and (ii) computer simulations in five plant species with different reproduction systems. Second, we compared T_{opt} for three traits in each of the plant species.

Methods

With computer simulations it is hardly possible to simulate data sets showing a population structure and familial relatedness comparable to that of empirical data sets. Nevertheless, to compare association mapping methods with respect to their adherence to the nominal a level based on scenarios with realistic estimates of population structure and familial relatedness, we analyzed phenotypic and genotypic data of rapeseed, potato, sugar beet, maize, and Arabidopsis. For the same reason, QTL effects were simulated on top of the observed phenotypic values when examining the adjusted power for QTL detection.

Plant materials, phenotypic data, and molecular markers

In each of the five plant species, with the exception of *Arabidopsis thaliana*, we selected three traits with different genetic complexity (presumably low, medium, and high). Detailed descriptions of the examined data sets are available as Additional file 1.

*Rapeseed (*Brassica napus *L.)*

We studied a total of n = 136 rapeseed inbreds, proprietary to Norddeutsche Pflanzenzucht Hans-Georg Lembke KG (Holt-see, Germany). The entries were evaluated for thousand kernel weight (TKW; g), oil content (OC; %), and oil yield (OY; t/ha). All entries were fingerprinted with m = 59 genome-wide distributed simple sequence repeat markers by Saaten-Union Resistenzlabor GmbH (Hovedissen, Germany) following standard protocols.

*Potato (*Solanum tuberosum *L.)*

Our study was based on the phenotypic and genotypic data evaluated earlier [21]. Briefly, the n = 184 tetraploid potato clones from the breeding programs of Böhm-Nordkartoffel Agrarproduktion OHG (Lüneburg, Germany) and Saka-Ragis Pflanzenzucht GbR (Windeby, Germany) were evaluated for *Globodera pallida* St. resistance (GPR) [22]. Our statistical analyses were based on the square root of the number of visible nematode cysts. Furthermore, the area under the disease progress curve [23] was used as measure for *P. infestans* resistance (PIR). In addition, plant maturity (PM) was evaluated in uninfected plants, using a 1 to 9 scale (1 = very early, 9 = very late). All entries were fingerprinted with m = 31 genome-wide distributed simple sequence repeat markers [21] by the potato genome analysis group of the Max Planck Institute for Plant Breeding Research (Cologne, Germany). For 21 markers the allele dosage was scored based on relative band intensities.

*Sugar beet (*Beta vulgaris *L.)*

We analyzed a total of n = 178 sugar beet inbreds of the pollen parent heterotic pool, proprietary to KWS SAAT AG (Einbeck, Germany). The test-cross progenies of these entries with an inbred of the seed parent heterotic pool were evaluated in a series of plant breeding trials. Data were recorded for amino nitrogen (AN) [24], beet yield (BY), and corrected sugar yield (CSY) [25] in % of the mean performance of checks. All entries were fingerprinted with 59 simple sequence repeat markers and 41 single nucleotide polymorphism markers (m = 100), both randomly distributed across the sugar beet genome. The fingerprinting was done by KWS SAAT AG following standard protocols.

*Maize (*Zea mays *L.)*

Our study was based on the phenotypic and genotypic data analyzed earlier [15]. In short, the n = 277 maize inbreds representing worldwide genetic diversity were evaluated for ear height (EH; cm), ear diameter (ED; cm), and days to pollen shed (DPS). For all inbreds, genotypic data of m = 553 genome-wide distributed single nucleotide polymorphism markers was available.

Arabidopsis thaliana *L.*

Our study was based on the n = 95 *Arabidopsis thaliana* L. inbreds for which phenotypic information was available [17]. These inbreds represent world-wide genetic diversity of Arabidopsis. We examined the normalized gene expression of *FLOWERING LOCUS C* (FLC) and *FRIGIDA* (FRI) as well as the number of days from germination to first opening of flowers under long day conditions with vernalisation treatment (LDV). For these inbreds, resequencing data of m = 876 genome-wide distributed short fragments was available [26]. To reduce the computational load, we used only the central single nucleotide polymorphism marker of each fragment.

The anonymised data sets of rapeseed, potato, and sugar beet are available upon request from the authors.

Statistical analyses

The empirical type I error rate of association-mapping approaches based on adjusted entry means (two-step approaches) is only slightly higher than that of approaches in which the phenotypic data analysis and the association analysis were performed in one step (one-step approaches) [20]. Therefore, in a first step we analyzed the phenotypic data and calculated adjusted entry means (rapeseed, potato) or entry means (sugar beet, maize, and Arabidopsis) M_i for each individual under consideration (Additional file 2). These estimates were then used in a second step for the association analyses.

Association analyses

For each of the five plant species, nine different statistical models (Table 1), which were described in detail previously [20], were used to calculate the P value for the association of each of the m marker loci with each of the three phenotypic traits. The entries of four of the five plant species in our study were homozygous inbred lines (Table 2) and, thus, no inferences can be made about dominance effects. Furthermore, for potato, di-, tri, and tetragenic effects [27] were neglected in our study.

The first model was an ANOVA model of the form:

$$M_i = \mu + \boldsymbol{\alpha}'\mathbf{X}_i + e_i$$

where α were the effects of allele substitution of the marker under study, $\mathbf{x}_i$ a column vector with the number of copies of the corresponding alleles, and e_i the residual.

Table 1: Methods used for association mapping and the corresponding statistical models.

Method	Statistical model	Population structure matrix **D**	Kinship matrix **K**
ANOVA	$M_i = \mu + \boldsymbol{\alpha}'\mathbf{X}_i + e_i$	-	-
K	$M_i = \mu + \boldsymbol{\alpha}'\mathbf{X}_i + g_i^* + e_i$	-	SPAGeDi
Q_1K	$M_i = \mu + \boldsymbol{\alpha}'\mathbf{X}_i + \sum_{u=1}^{z} D_{iu}v_u + g_i^* + e$	STRUCTURE; ΔK criterion	SPAGeDi
Q_2K	$M_i = \mu + \boldsymbol{\alpha}'\mathbf{X}_i + \sum_{u=1}^{z} D_{iu}v_u + g_i^* + e$	STRUCTURE; Log likelihood	SPAGeDi
PK	$M_i = \mu + \boldsymbol{\alpha}'\mathbf{X}_i + \sum_{u=1}^{z} D_{iu}v_u + g_i^* + e$	Principal components; explaining simultaneously 25% of the variance	SPAGeDi
K_T	$M_i = \mu + \boldsymbol{\alpha}'\mathbf{X}_i + g_i^* + e_i$	-	$K_{Tij} = \frac{S_{ij}-1}{1-T} + 1$; T = 0,0.025, ..., 0.975
Q_1K_T	$M_i = \mu + \boldsymbol{\alpha}'\mathbf{X}_i + \sum_{u=1}^{z} D_{iu}v_u + g_i^* + e$	STRUCTURE; ΔK criterion	T = 0,0.025, ..., 0.975
Q_2K_T	$M_i = \mu + \boldsymbol{\alpha}'\mathbf{X}_i + \sum_{u=1}^{z} D_{iu}v_u + g_i^* + e$	STRUCTURE; Log likelihood	T = 0,0.025, ..., 0.975
PK_T	$M_i = \mu + \boldsymbol{\alpha}'\mathbf{X}_i + \sum_{u=1}^{z} D_{iu}v_u + g_i^* + e$	Principal components; explaining simultaneously 25% of the variance	T = 0,0.025, ..., 0.975

For a detailed definition of the statistical models and description of the different methods see Materials and Methods.

Table 2: Description of the examined data sets.

Parameter	Rapeseed	Potato	Sugar beet	Maize	Arabidopsis
n	136	184	178	277	95
Entry type	Inbred line	Non-inbred clone	Inbred line	Inbred line	Inbred line
Phenotypic data					
Trait 1	Thousand kernel weight	Resistance to *G. pallida*	Amino nitrogen	Ear height	Norm. gene expression of *FLC*
Abbrev.	TKW	GPR	AN	EH	FLC
Unit	g	$\sqrt{\text{No. of nematode cysts}}$	%	cm	%
h^2	0.78	0.98	0.89	-	-
Range M_i	3.0–4.6	0.4–19.5	71.1–226.2	8–136	0.021–6.270
Trait 2	Oil content	Resistance to *P. infestans*	Beet yield	Ear diamter	Norm. gene expression of *FRI*
Abbrev.	OC	PIR	BY	ED	FRI
Unit	%	Area under disease progress curve	%	mm	%
h^2	0.81	0.77	0.90	-	-
Range M_i	46.1–51.7	-6.4–165.1	84.8–113.6	23.7–46.4	0.211–4.386
Trait 3	Oil yield	Plant maturity	Corrected sugar yield	Days to pollen shed	Flowering time
Abbrev.	OY	PM	CSY	DPS	LDV
Unit	t/ha	Rating scale 1 to 9	%	No. of days	No. of days
h^2	0.50	0.94	0.81	-	-
Range M_i	2.2–3.0	3.4–9.5	87.8–108.7	54.5–82.5	18.7–55.7
Genotypic data					
Type of markers	SSRs	SSRs	SSRs & SNPs	SNPs	SNPs
m	59	31	100	553	876
Avg. allele freq.	0.37	0.18	0.30	0.50	0.50

M_i is the adjusted entry mean (rapeseed and potato) or entry mean (sugar beet, maize, and Arabidopsis) of the *i*th genotype calculated over all environments.

The statistical model underlying our mixed-model association mapping approaches was:

$$M_i = \mu + \boldsymbol{\alpha}'\mathbf{x}_i + \sum_{u=1}^{z} D_{iu}v_u + g_i^* + e_i,$$

where v_u was the effect of the *u*th column of the population structure matrix **D** and g_i^* was the residual genetic effect of the *i*th entry. The matrix **D**, which comprised *z* linear independent columns, differed among the examined mixed-model association mapping methods (Table 1), which is why it is described in the paragraphs on the individual methods. The variances of the random effects $\mathbf{g}^* = \{g_1^*, \ldots, g_n^*\}$ and $\mathbf{e} = \{e_1, \ldots, e_n\}$ were assumed to be $\mathrm{Var}(\mathbf{g}^*) = 2\mathbf{K}\sigma^2_{g^*}$ and $\mathrm{Var}(\mathbf{e}) = \mathbf{R}\sigma^2_r$, where **K** was a $n \times n$ matrix of kinship coefficients that define the degree of genetic covariance between all pairs of entries. $\sigma^2_{g^*}$ was the residual genetic variance and σ^2_r the residual variance, both estimated by REML. **R** was an $n \times n$ matrix in which the off-diagonal elements were 0 and the diagonal elements were reciprocals of the number of phenotypic observations underlying each entry mean or adjusted entry mean [15].

The K method was based on the above described mixed-model with the difference that it did not include any v_u effects (Table 1). The kinship matrix **K** was calculated based on all marker data using the software package SPAGeDi [28], where negative kinship values between entries were set to 0.

The Q_1K and Q_2K methods were based on the above described mixed-model. For these two methods, the population structure matrices $\mathbf{Q}_1$ and $\mathbf{Q}_2$, which were calculated using software STRUCTURE [29] and described in the following paragraphs, were used as **D** matrix. In our investigations, the set of *n* entries was analyzed by setting *z* from 0 to 14 in each of five repetitions. For each run of STRUCTURE, the burn-in time as well as the iteration number for the Markov Chain Monte Carlo algorithm were set to 100 000 [30].

For the $\mathbf{Q}_1$ matrix, the number of sub-populations was estimated based on the ad-hoc criterion ΔK [31]. In contrast, for the $\mathbf{Q}_2$ matrix, we used the run with the highest log likelihood to and the lowest number of sub-populations [32]. The $z + 1$ columns of both, the $\mathbf{Q}_1$ and $\mathbf{Q}_2$ matrix, add up to one and, thus, only the first *z* columns were used as **D** matrix of the Q_1K and Q_2K method, respectively, to achieve linear independence. The Q_1K and Q_2K methods were based on the same kinship matrix **K** as used for the K method.

We used the first p principal components of an allele frequency matrix as **D** matrix of the PK method (Table 1) [17]. p was chosen in such a way that the explained variance of the first p principal components was about 25%. The PK method was based on the same kinship matrix **K** as used for the K method.

The Q_1K_T, Q_2K_T, PK_T, and K_T methods were based on a matrix $\mathbf{K_T}$ which was calculated according to:

$$K_{Tij} = max\left(1-\left(\frac{1-S_{ij}}{1-T}\right), 0\right),$$

where S_{ij} is the proportion of marker loci with shared variants between inbreds i and j [20]. We examined T = 0, 0.025, ..., 0.975 to obtain a REML estimate of T, which is the conditional probability that marker alleles are alike in state, given that they are not identical by descent.

Measures for comparison of association mapping methods

The mean squared difference (MSD) between observed and expected P values of all marker loci was calculated as measure for the adherence to the nominal α level [20]. High MSD values indicate that the empirical type I error rate of these approaches is considerably higher than the nominal α level. Computer simulations were performed based on a bivariate beta-distribution [33] to examine which difference in MSD values between two association mapping methods could be expected purely by chance [20]. For each trait of each plant species, we investigated five pairs of association mapping approaches (i) Q_1K/ANOVA, (ii) Q_1K/K, (iii) Q_1K/Q_2K, (iv) Q_1K/PK, and (v) $Q_1K/Q_1K_{T_{opt}}$.

For each of the five plant species, the Pearson correlation coefficient between the observed P values of all association mapping methods was calculated for the trait with medium genetic complexity.

Power simulations

The power to detect a biallelic QTL of interest, which explained a fraction of the phenotypic variance and was in complete LD with one marker locus, was examined as described in detail previously [20]. Briefly, the QTL effect G_r, calculated as r = 0.1 multiplied by the standard deviation of the vector of adjusted entry means $\mathbf{m} = (M_1, M_i, ..., M_n)$ of the n entries, was assigned in consecutive simulation runs to each of the detected marker alleles whereas all other alleles were assigned the genotypic effect 0. In each simulation run, the phenotypic value of each entry i was calculated by summing up the QTL effect of the alleles and the adjusted entry mean M_i. All association mapping methods were run on the phenotypic values of the entries to determine whether the QTL can be detected. To adjust the association mapping methods for their different empirical type I error rates, we calculated the adjusted power as the proportion of QTL detected, based on the nominal α for which the empirical type I error rate α^* was 0.05. In addition to r = 0.1, we examined r = 0.4, 0.7, ..., 1.9. The percentage (π) of the total phenotypic variation explained by a QTL effect G_r was calculated [15].

All mixed-model calculations were performed with ASReml release 2.0 [34].

Results

For each trait examined in the current study, considerable variation was observed for the entry means or adjusted entry means M_i (Table 2). The total number of marker alleles detected for rapeseed, potato, sugar beet, maize, and Arabidopsis was 331, 158, 176, 1106, and 1752, respectively. The average allele frequency ranged from 0.18 for potato to 0.50 for maize and Arabidopsis.

The model-based approach of STRUCTURE revealed $z + 1$ = two, two, two, five, and six sub-populations for rapeseed, potato, sugar beet, maize, and Arabidopsis, respectively, when using the ad-hoc criterion ΔK. In contrast, based on SBC, the number of sub-populations revealed by STRUCTURE was 11, 15, 10, 15, and 5. For rapeseed, potato, sugar beet, maize, and Arabidopsis, the minimum number of principal components p explaining simultaneously 25% of the variance was 4, 5, 4, 13, and 8, respectively.

The MSD between observed and expected P values of the K approach ranged from 0.0002 (maize, ED) to 0.0604 (potato, PM) and was considerably lower than that of the ANOVA approach ranging from 0.0004 (Arabidopsis, FRI) to 0.1928 (potato, GPR) (Table 3). For the Q_1K and Q_2K methods, the MSD values were of similar size and varied between 0.0002 (maize, DPS) and 0.0389 (potato, PM). The MSD value of the PK method ranged from 0.0002 (maize, DPS) to 0.0422 (potato, PM).

For all plant species, traits, and mixed-model approaches examined, considerably different values of REML-based deviance as well as MSD were observed for the examined levels of T (Additional file 3). The optimum threshold T_{opt}, identified based on deviance profiles, ranged from 0.450 to 0.925 (Table 4). By comparison, the threshold T_{opt},

Table 3: Mean of squared differences (MSD) between observed and expected *P* values for various association mapping methods in five plant species.

Method		Rapeseed	
	TKW	OC	OY
ANOVA	0.0624	0.0326	0.0523
K	0.0098	0.0053	0.0016
Q_1K	0.0021	0.0047	0.0061
Q_2K	0.0013	0.0010	0.0192
PK	0.0008	0.0007	0.0026
		Potato	
	GPR	PIR	PM
ANOVA	0.1928	0.0947	0.1534
K	0.0499	0.0162	0.0604
Q_1K	0.0122	0.0179	0.0389
Q_2K	0.0181	0.0017	0.0063
PK	0.0189	0.0183	0.0422
		Sugar beet	
	AN	BY	CSY
ANOVA	0.1526	0.1625	0.1533
K	0.0136	0.0191	0.0173
Q_1K	0.0060	0.0239	0.0051
Q_2K	0.0118	0.0167	0.0081
PK	0.0090	0.0137	0.0065
		Maize	
	EH	ED	DPS
ANOVA	0.0333	0.0147	0.0909
K	0.0003	0.0002	0.0006
Q_1K	0.0002	0.0003	0.0002
Q_2K	0.0002	0.0002	0.0003
PK	0.0003	0.0005	0.0002
		Arabidopsis	
	FLC	FRI	LDV
ANOVA	0.0040	0.0004	0.0070
K	0.0006	0.0022	0.0013
Q_1K	0.0026	0.0033	0.0017
Q_2K	0.0019	0.0022	0.0013
PK	0.0021	0.0034	0.0018

For abbreviations of the analyzed traits see Table 2. For a detailed definition of the statistical models and description of the different methods see Materials and Methods.

identified based on MSD profiles, ranged from 0.275 to 0.975. The correlation between the T_{opt} values identified using these two criteria was 0.83 (Additional file 4). The MSD values observed for the mixed- model approaches, which were based on the $\mathbf{K}_{T_{opt}}$ matrix, were lower than that observed for the approaches which were based on the **K** matrix (Table 3; Table 4; Fig. 1).

The 95% quantile of differences in MSD calculated for the five pairs of association methods Q1K/ANOVA, Q1K/K, Q1K/Q2K, Q1K/PK, and Q1K/ $Q_1K_{T_{opt}}$ was highest for potato and ranged from 0.0041 to 0.0114 (Additional file 5). For Arabidopsis, the 95% quantile of differences in MSD was lowest and varied from 0.0001 to 0.0004.

The slopes of the power curve were flat for small as well as large genetic effects, whereas for genetic effects of medium size the slope was steep (Fig. 2). For most traits under consideration, the adjusted power of the $Q_1K_{T_{opt}}$, $Q_2K_{T_{opt}}$, and $PK_{T_{opt}}$ methods was slightly higher across all examined sizes of genetic effects than those of the Q_1K, Q_2K, and PK methods. In comparison with the other association mapping methods, the ANOVA method showed the lowest adjusted power to detect QTL across all examined sizes of genetic effects for all traits and plant species except potato (PIR).

The Pearson correlation coefficient between the observed *P* values of all examined association mapping methods ranged from -0.05 to 0.99 (Additional file 6)

Discussion

Assumptions underlying the comparison of association mapping approaches using empirical data sets

Simulation of data sets mimicing the population structure and familial relatedness of empirical data sets is hardly possible. However, only with such data sets a reliable assessment of the performance of different association mapping approaches is possible. Therefore, our study was based on empirical data sets.

Investigations on the type I error rate and on the adjusted power to detect QTL of association mapping approaches using empirical data sets require that the examined marker loci are unlinked to polymorphisms controlling the expression of the trait under consideration. In the present study, this assumption seems to be reasonable as for the five plant species examined the available marker density was considerably lower than that required for genome-wide association mapping. Similarly to other studies comparing association mapping approaches based on empirical data [15,17], however, we cannot rule out the possibility that some markers might be linked to functional polymorphisms of the traits under consideration.

In accordance with previous studies [15,17], we used the same markers for estimation of population structure as well as familial relatedness as were used for calculating the MSD between observed and expected *P* values. Theoretical considerations suggest that MSD values calculated in this way might underestimate the MSD values for markers which are not included in the estimation of population structure and familial relatedness such as markers in candidate genes. However, our computer simulations on the Arabidopsis

Table 4: *T* values for which the lowest deviance or the lowest mean of squared differences between observed and expected *P* values were found for various association mapping methods in five plant species.

Mixed-model method	deviance	MSD	deviance	MSD	deviance	MSD
			Rapeseed			
	TKW		OC		OY	
K_T	0.725	0.350 (0.0068)	0.800	0.475 (0.0006)	0.750	0.400 (0.0011)
Q_1K_T	0.775	0.700 (0.0039)	0.800	0.425 (0.0007)	0.775	0.700 (0.0011)
Q_2K_T	0.700	0.825 (0.0009)	0.800	0.850 (0.0007)	0.900	0.900 (0.0128)
PK_T	0.725	0.700 (0.0006)	0.800	0.725 (0.0004)	0.750	0.750 (0.0019)
			Potato			
	GPR		PIR		PM	
K_T	0.525	0.500 (0.0065)	0.625	0.550 (0.0034)	0.475	0.475 (0.0082)
Q_1K_T	0.600	0.600 (0.0054)	0.625	0.550 (0.0033)	0.475	0.525 (0.0086)
Q_2K_T	0.600	0.600 (0.0091)	0.625	0.500 (0.0010)	0.625	0.525 (0.0031)
PK_T	0.575	0.550 (0.0121)	0.625	0.550 (0.0048)	0.475	0.475 (0.0153)
			Sugar beet			
	AN		BY		CSY	
K_T	0.575	0.325 (0.0022)	0.475	0.300 (0.0022)	0.475	0.350 (0.0012)
Q_1K_T	0.575	0.325 (0.0023)	0.450	0.300 (0.0059)	0.475	0.375 (0.0006)
Q_2K_T	0.475	0.325 (0.0029)	0.475	0.275 (0.0021)	0.575	0.350 (0.0009)
PK_T	0.475	0.300 (0.0019)	0.350	0.300 (0.0034)	0.475	0.350 (0.0009)
			Maize			
	EH		ED		DPS	
K_T	0.575	0.450 (0.0002)	0.575	0.575 (0.0001)	0.600	0.575 (0.0002)
Q_1K_T	0.575	0.500 (0.0001)	0.725	0.575 (0.0003)	0.600	0.575 (0.0001)
Q_2K_T	0.875	0.475 (0.0003)	0.725	0.525 (0.0001)	0.600	0.525 (0.0001)
PK_T	0.875	0.475 (0.0002)	0.725	0.525 (0.0001)	0.600	0.600 (0.0001)
			Arabidopsis			
	FLC		FRI		LDV	
K_T	0.875	0.875 (0.0004)	0.825	0.975 (0.0007)	0.875	0.900 (0.0034)
Q_1K_T	0.875	0.975 (0.0023)	0.875	0.800 (0.0028)	0.875	0.900 (0.0020)
Q_2K_T	0.875	0.950 (0.0006)	0.875	0.800 (0.0018)	0.875	0.900 (0.0017)
PK_T	0.875	0.775 (0.0015)	0.925	0.925 (0.0025)	0.900	0.900 (0.0011)

For the latter measure, the observed mean of squared differences are given in parentheses. For abbreviations of the analyzed traits see Table 2. For a detailed definition of the statistical models and description of the different methods see Materials and Methods.

dataset, in which the half of the available markers were used for estimation of population structure and familial relatedness and the remaining markers for calculation of the MSD values, suggested that this underestimation is negligible (data not shown). This result indicates that association mapping methods, for which we observed MSD values close to zero, will also adhere to the nominal α level in empirical association mapping experiments.

Our power simulations assumed a QTL allele which is in complete LD with one marker allele. This assumption allows the comparison of results from various plant species irrespective of the available number of markers. However, it maximizes the power for QTL detection. In most empirical studies no markers are available which are in complete LD with the QTL. Therefore, for such studies, a lower power for QTL detection is expected depending on

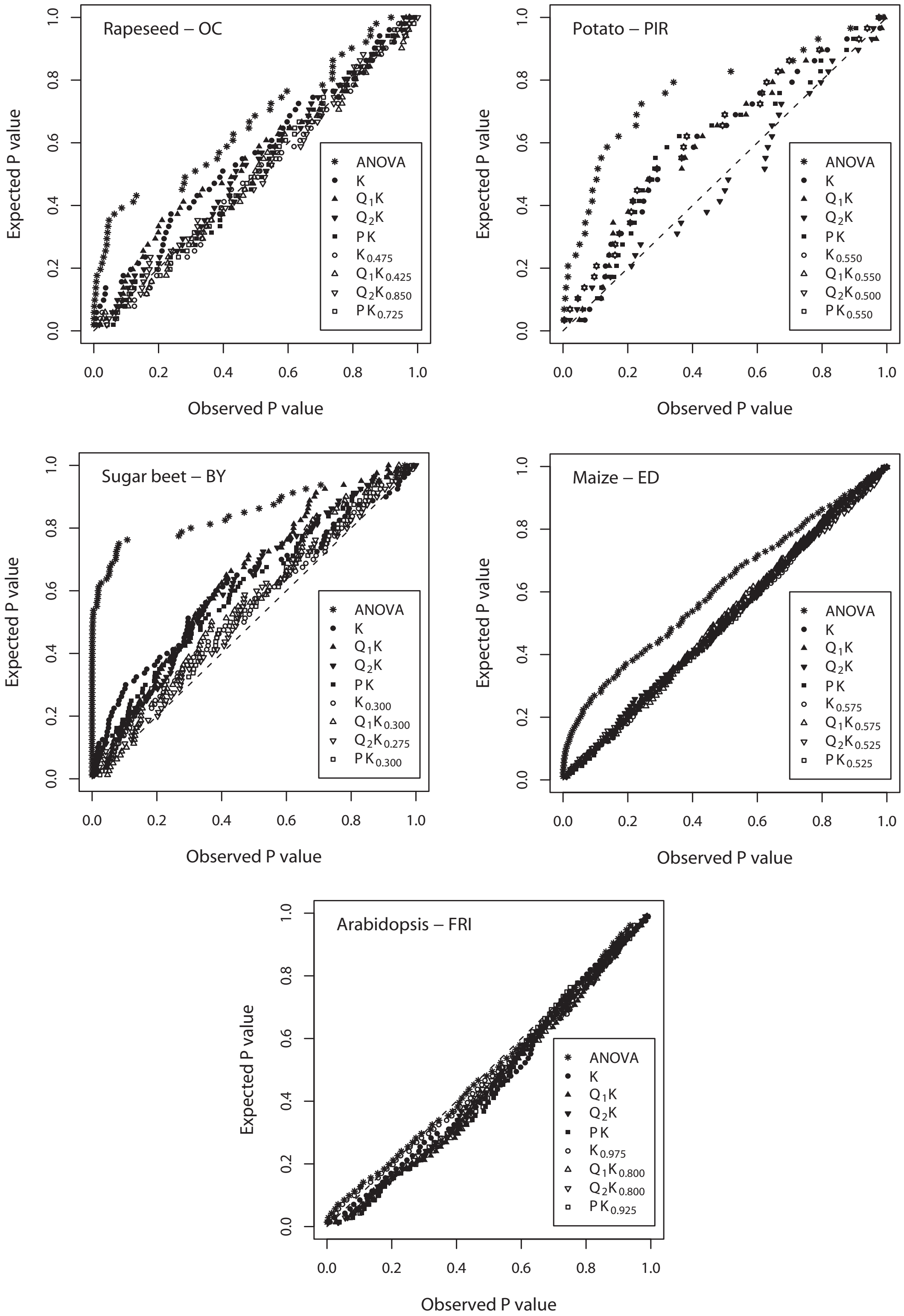

Figure 1

Plot of observed *vs.* expected *P* values for the nine different association mapping methods. For maize, every fifth, and for Arabidopsis, every eigth *P* value was plotted to increase the clarity of the plot. For each of the five plant species, the result of the trait with medium genetic complexity is presented.

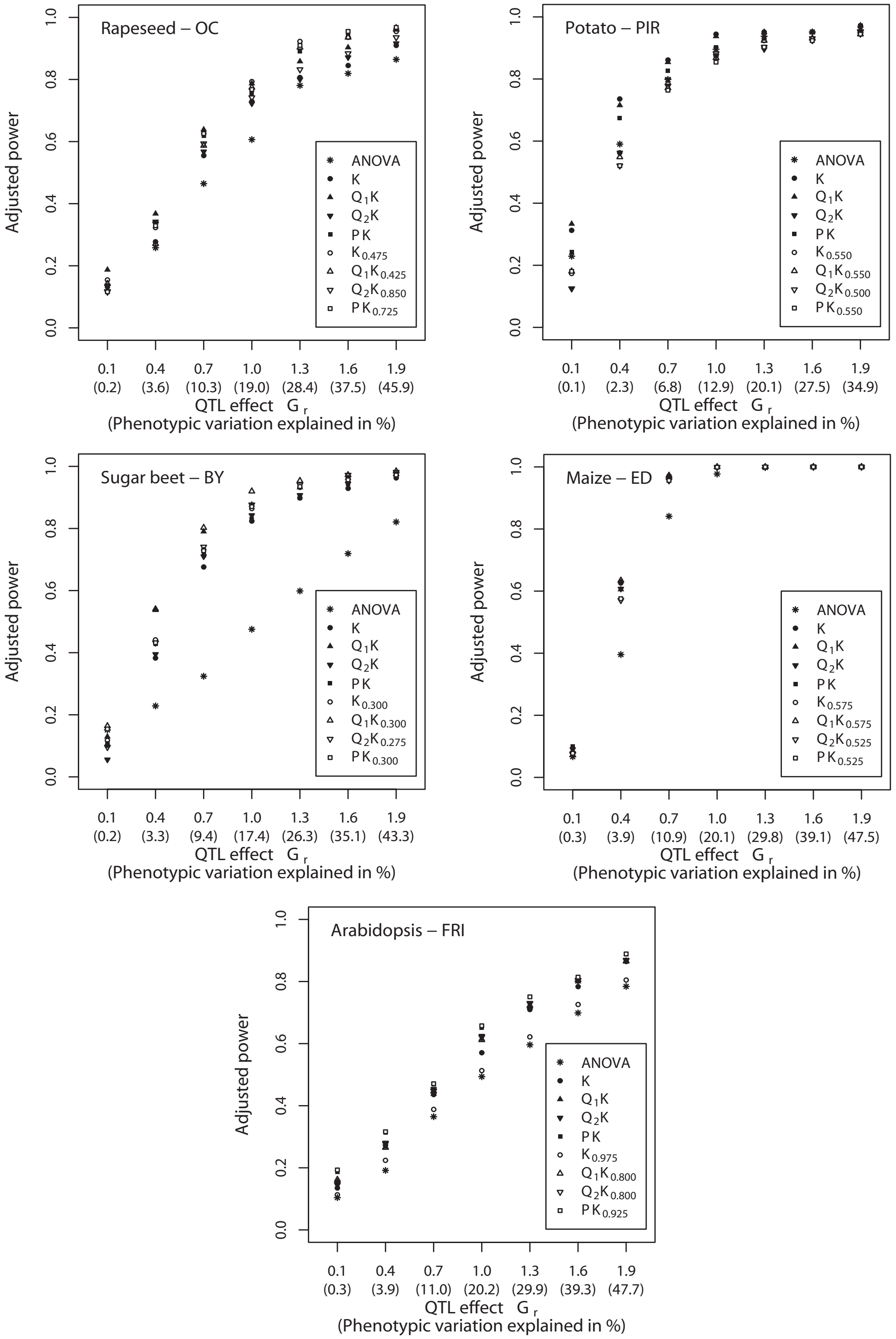

Figure 2

Adjusted power to detect quantitative trait loci (QTL) for the nine different association mapping methods depending on the size of the QTL effect G_r. The percentage of phenotypic variation explained by a QTL was calculated for the average allele frequency (see Table 2). For each of the five plant species, the result of the trait with medium genetic complexity is presented.

the extent of LD between marker and QTL. A further factor influencing the detection of the QTL of interest, which was neglected in our power simulations, are additional QTL that are linked to the QTL of interest. Incomplete LD between marker and QTL as well as additional linked QTL are expected to alter the power of QTL detection in all association mapping methods to a similar extent. Therefore, no influence on our conclusions regarding the ranking of various methods for association mapping is expected with respect to the assumptions made in our power simulations.

Comparison of various association mapping approaches

ANOVA approach

A frequently used method for association mapping in a plant genetics context is the ANOVA approach [10]. This approach was therefore used in our study as reference method. Under the assumption that the random marker loci in our study are unlinked to the polymorphisms controlling the expression of the traits under consideration, association mapping methods that adhere to the nominal α level show a uniform distribution of P values, *i.e.*, a MSD value close to zero. With the exception of the normalized gene expression data of the FRI gene in Arabidopsis, we observed a non-uniform distribution of P values in the ANOVA approach of all traits (Table 3). This finding is in accordance with the results of previous studies [15,17,20] and indicates that the ANOVA approach is inappropriate for association mapping in the examined plant species, because the resulting proportion of spurious marker-phenotype associations is considerably higher than the nominal type I error rate.

QK approach

The recently proposed QK mixed-model association mapping method promises to correct for multiple levels of relatedness [15]. The MSD between observed and expected P values found for the Q_1K and Q_2K methods of all examined traits was considerably lower than that observed for the ANOVA approach (Table 3). Furthermore, this difference in MSD values was considerably larger than the 95% quantile observed based on the computer simulations (Additional file 5). These findings suggest the advantage of the Q_1K and Q_2K methods over the ANOVA method for association mapping not only in maize and Arabidopsis for which similar results were previously reported [15,17] but also in rapeseed, potato, and sugar beet.

For estimation of the number of sub-populations using STRUCTURE [29], ΔK, an ad hoc criterion related to the second order rate of change in the log likelihood of data, was proposed [31]. In other studies, the number of sub-populations z+1 was chosen in such a way that a further increase in z did not considerably improve the log likelihood of data [35]. We used these two criteria to estimate the number of sub-populations for the $\mathbf{Q}_1$ and $\mathbf{Q}_2$ matrices.

For some traits, we observed a smaller MSD value for the Q_1K than for the Q_2K method, whereas the opposite was true for the other traits (Table 3). Furthermore, with few exceptions, these differences were smaller than the corresponding 95% quantiles observed in our computer simulations on the correlated beta-distribution (Additional file 5). These findings demonstrate that the association mapping models based on the two population structure matrices, $\mathbf{Q}_1$ and $\mathbf{Q}_2$, are equally appropriate for association mapping with respect to (i) adherence to the nominal α level as well as (ii) the adjusted power for QTL detection.

Despite promising results for the Q_1K and Q_2K association mapping approaches, these methods have several drawbacks, as previously discussed [20]. Therefore, we examined two association mapping methods which were not based on the population structure matrix from STRUCTURE. For the PK mixed-model association mapping approach, the $\mathbf{Q}_1$ or $\mathbf{Q}_2$ matrix from STRUCTURE was replaced by a matrix comprising p principal components (Table 1). In contrast, the K method was based on a mixed-model which does not include any v_u effects.

PK approach

The MSD between observed and expected P values, which was found for this method, was similar to those observed for the Q_1K and Q_2K methods (Table 3). Furthermore, all three methods yielded a similar adjusted power of QTL detection across the examined plant species (Fig. 2). These findings were in accordance with those of previous studies [17,20], suggesting that the PK approach is a promising alternative to the Q_1K and Q_2K methods.

K approach

For the K approach, we observed for most examined traits a higher MSD value than for the mixed-model methods Q_1K, Q_2K, and PK. The opposite result was observed with respect to the adjusted power of QTL detection (Fig. 2).

These results indicated that the K approach was less appropriate for association mapping than the approaches based on the integration of fixed effects in the statistical model. This conclusion may be explained by the fact that the software package SPAGeDi [28] used for calculation of the kinship coefficients assumes that random pairs of individuals of the germplasm set under consideration are unrelated and assigns them a kinship coefficient of 0. This definition of unrelated individuals results in a kinship matrix for which a large number of pair-wise kinship estimates are negative. It was proposed to replace these negative values by 0, because such pairs of individuals are less related than random pairs of individuals [15]. This approach, however, ignores information on the structure of unrelated individuals, which was captured in the kinship matrix, and consequently necessitates the inclusion of fixed effects in the mixed-model. Therefore, we examined mixed-model association mapping approaches which are based on **K** matrices calculated for different thresholds *T* [20].

Approaches based on K matrices calculated for different values of T

The values of T_{opt} calculated for the current data sets using the REML approach, which might also be used to infer the probability of identity by descent for genotypes with no pedigree information available, were not always identical with those identified based on the MSD profiles (Table 4). Across all plant species, traits, and association mapping methods, however, the correlation between the T_{opt} value identified based on both approaches was 0.83 (Additional file 4). This result suggested that for association mapping approaches the T_{opt} value might be identified using the REML approach because it is associated with a lower computational load. The REML-based deviance, used to estimate T_{opt}, however, can only be compared among models which are based on the same set of fixed effects. Therefore, we used the MSD between observed and expected *P* values for comparison of the Q_1K_T, Q_2K_T, PK_T, and K_T method and furthermore used the T_{opt} values identified based on this criterion.

The MSD values observed for the association mapping approaches based on the T_{opt} value, were considerably lower than that of the corresponding association mapping approaches based on the **K** matrix from SPAGeDi, for all examined plant species and traits (Table 4). Furthermore, the adjusted power observed for the former approaches was for most examined traits higher than that observed for the latter approaches. These findings suggest that methods based on a kinship matrix calculated for the T_{opt} value are more appropriate for association mapping than the corresponding association mapping approaches which are based on the **K** matrix from SPAGeDi. Nevertheless, the MSD values observed for the association mapping methods, which include fixed effects such as the $Q_1K_{T_{opt}}$, $Q_2K_{T_{opt}}$, or $PK_{T_{opt}}$, were lower than that of the $K_{T_{opt}}$. Therefore, in our study the $Q_1K_{T_{opt}}$, $Q_2K_{T_{opt}}$ or $PK_{T_{opt}}$ are the most appropriate methods for association mapping.

Comparison of the properties of association mapping approaches among plant species and traits

MSD values

The MSD values observed for potato and sugar beet across all association mapping methods were considerably higher than those for maize and Arabidopsis, whereas those for rapeseed were of medium size (Table 3). This may be due to the low number of random molecular markers available in our study for potato, sugar beet, and rapeseed. Thereby, not very precise estimation of population structure is possible which in turn increases the MSD values.

To examine this issue in more detail, random markers were selected in replicated simulation runs from maize and Arabidopsis linkage maps in such a way that the total number of alleles of the selected markers corresponds to those observed for the other three species. All association mapping methods were then run with these markers. Our results (data not shown) suggested that the low number of random molecular markers for potato, sugar beet, and rape seed only partially explains the observed differences in MSD values.

Another factor that explains the observed difference in MSD values among the plant species is the difference in the extent of population structure and relatedness present in the examined genetic materials. This difference in population structure and relatedness may partly be due to the fact that the entries of the examined plant species differ in their origin. While the Arabidopsis entries were selected from natural populations, the entries of the other four plant species were chosen from plant breeding programs. Because entries selected from plant breeding programs have a complex ancestry, the extent of population structure and relatedness in such germplasm sets is expected to be higher than in germplasm sets consisting of entries selected from natural populations.

In addition, the difference in the extent of population structure and relatedness between rapeseed, potato, sugar beet, and maize can be explained by the different sampling strategies underlying the examined genetic materials. The entries of the maize data set represent world-wide genetic diversity, whereas the genetic materials of rapeseed, potato, and sugar beet were sampled from commercial plant breeding programs. Theoretical considerations suggest that this increases the probability of including partially related entries.

Furthermore, the difference in the extent of population structure and relatedness between rapeseed, potato, sugar beet, and maize may partly be due to the different reproduction systems and types of varieties bred in a particular crop. For entries from hybrid breeding programs [11] such as sugarbeet and maize, distinct sub-poulations are expected. In contrast, when line or clonal varieties are bred, as in the case of rapeseed and potato, no distinct sub-populations are expected to develop as population structure is disregarded when choosing the parents of a cross. Nevertheless, this procedure is expected to generate diverse levels of familial relatedness [36].

Adjusted power for QTL detection

Across all examined statistical methods for association mapping, considerable differences in the adjusted power for QTL detection were observed for the five examined plant species (Fig. 2). The adjusted power is influenced by (i) the size of the QTL effect G_r, (ii) the extent of LD between marker allele and QTL allele, (iii) the number of entries *n*, (iv) the QTL allele frequency, and (v) the heritability of the trait under consideration. Our power simulations assumed the same QTL effects for all plant species and a QTL allele which is in complete LD with one marker allele. These two factors cannot contribute to the observed difference in adjusted power for QTL detection among the examined plant species.

High adjusted power for the maize data set with its high number of entries and a low adjusted power for the Arabidopsis data set with a low number of entries indicated that differences in the number of entries *n* have a large influence on the observed differences in adjusted power among the examined plant species. This explanation is supported by results of previous studies [37]. In contrast, the small difference in adjusted power for QTL detection between sugar beet and potato data sets, which comprised a similar number of entries but differed in their average allele frequency, suggested that variation in this factor caused only small differences in the adjusted power.

In our study, heritability estimates were only available for two plant species and, thus, no inferences can be made about the contribution of this factor to differences in the adjusted power for QTL detection. However, results from previous studies suggested that increasing heritability has the potential to considerably increase the power for QTL detection [14].

T_{opt}

The optimum T values identified in our study differed considerably among the various plant species (Table 4). This finding may be due to the difference in the extent of population structure and familial relatedness among the examined plant species as described above. The influence of population structure and familial relatedness on the optimum *T* value can be explained by the fact that lower values for *T* reduce the number of negative pair-wise kinship estimates in the kinship matrix $\mathbf{K}_T$. Thereby the use of information concerning the structure of unrelated individuals, which was comprised in the kinship matrix $\mathbf{K}_T$, is improved and decreases the MSD values.

In comparison with the large differences among the optimum *T* values identified for different plant species, differences in the optimum *T* values for different traits of the same species were only small (Table 4). This finding might be explained by the fact that differences in the optimum *T* values identified for different traits of the same plant species can only be due to differences in the extent of population structure and relatedness for the traits under consideration generated by natural or artificial selection. Therefore, one optimum *T* value might be calculated across all traits of one species to improve the precision of this value. However, this requires further research on the standard error of the optimum *T* values.

Conclusion

Our study suggests that the QK method [15] is not only appropriate for association mapping in humans, maize, and Arabidopsis but also in rapeseed, potato, and sugar beet. Furthermore, our results indicate that the estimation of the number of sub-populations based on the two criteria, ΔK and SBC, results in different numbers of sub-populations. Nevertheless, the association mapping models which are based on these two population structure matrices are equally appropriate with respect to adherence to the nominal α level as well as the adjusted power for QTL detection. Furthermore, we recommend replacing the **K** matrix of the Q_1K, Q_2K, and PK approach by a $\mathbf{K}_T$ matrix, which is based on a REML estimate of the conditional probability that two inbreds carry alleles at the same locus which are identical in state but not identical by descent and, thus, increase the adherence to the nominal α level. Finally, we showed that the T_{opt} value estimated in this way differs considerably among the five plant species but only a little for the different traits within species.

Abbreviations

AN: amino nitrogen; BY: beet yield; CSY: corrected sugar yield; ED: ear diameter; EH: ear height; FLC: *FLOWERING LOCUS C*; FRI: *FRIGIDA*; GPR: *Globodera pallida* resistance; LD: linkage disequilibrium; LDV: long day conditions with vernalisation treatment; MSD: mean of squared difference; OC: oil content; OY: oil yield; PIR: *Phytophthora infestans* resistance; PM: plant maturity; QTL: quantitative trait locus; REML: restricted maximum likelihood; SBC: Schwarz Bayesian criterion; TKW: thousand kernel weight.

Authors' contributions

BS designed the project and analyzed the data. BS and AEM wrote the manuscript.

Additional material

Additional file 1

Plant materials, phenotypic data, and molecular markers description: *Description of the plant materials, phenotypic data, and molecular markers used for the study.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-10-94-S1.pdf]

Additional file 2

Phenotypic data analyses. *Description of the statistical analyses of the phenotypic data.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-10-94-S2.pdf]

Additional file 3

Comparison of four different mixed-model association mapping methods. *Mean of squared differences between observed and expected* P *values for four different mixed-model association mapping methods depending on the threshold* T.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-10-94-S3.pdf]

Additional file 4

***Comparison of two methods for estimation of the threshold* T.** *Optimum values for threshold* T *identified based on mean of squared differences between observed and expected* P *values plotted versus optimum* T *values identified based on deviance for the four mixed-model association mapping methods of the five plant species and three traits.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-10-94-S4.pdf]

Additional file 5

Difference in mean square differences between pairs of association mapping methods expected purely by chance. *Ninety-five % quantile of the difference of the mean square differences between observed and expected* P *values for five pairs of association mapping approaches determined based on a bivariate beta-distribution.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-10-94-S5.pdf]

Additional file 6

Comparison of the results of various association mapping methods. *Pearson correlation coefficient between the observed* P *values for various association mapping methods.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-10-94-S6.pdf]

Acknowledgements

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. We thank the breeding companies Norddeutsche Pflanzenzucht Hans-Georg Lembke KG (Holtsee, Germany), Böhm-Nordkartoffel Agrarproduktion OHG (Lüneburg, Germany), Saka-Ragis Pflanzenzucht GbR (Windeby, Germany), and KWS SAAT AG (Einbeck, Germany) as well as the potato genome analysis group of the Max Planck Institute for Plant Breeding Research (Cologne, Germany) for providing phenotypic and genotypic data. The authors appreciate the editorial work of Prof. Dr. B.S. Dhillon and Dr. J. Muminovi&#x0107;, whose suggestions considerably improved the style of the manuscript. The authors thank three anonymous reviewers for their valuable suggestions.

References

1. Yano M: **Genetic and molecular dissection of naturally occuring variation.** *Curr Opin Plant Biol* 2001, **4:**130-135.
2. Tanksley SD: **Mapping polygenes.** *Annu Rev Genet* 1993, **27:**205-233.
3. Paterson AH: **Molecular dissection of quantitative traits: progress and prospects.** *Genome Res* 1995, **5:**321-333.
4. Parisseaux B, Bernardo R: **In silico mapping of quantitative trait loci in maize.** *Theor Appl Genet* 2004, **109:**508-514.
5. Flint-Garcia SA, Thornsberry JM, Buckler ES: **Structure of linkage disequilibrium in plants.** *Annu Rev Plant Biol* 2003, **54:**357-374.
6. Willer CJ, Sanna S, Jackson AU, Scuteri A, Bonnycastle LL, Clarke R, Heath SC, Timpson NJ, Najjar SS, Stringham HM, Strait J, Duren WL, Maschio A, Busonero F, Mulas A, Albai G, Swift AJ, Morken MA, Narisu N, Bennett D, Parish S, Shen H, Galan P, Meneton P, Hercberg S, Zelenika D, Chen WM, Li Y, Scott LJ, Scheet PA, Sundvall J, Watanabe RM, Nagaraja R, Ebrahim S, Lawlor DA, Ben-Shlomo Y, Davey-Smith G, Shuldiner AR, Collins R, Bergman RN, Uda M, Tuomilehto J, Cao A, Collins FS, Lakatta E, Lathrop GM, Boehnke M, Schlessinger D, Mohlke KL, Abecasis GR: **Newly identified loci that influence lipid concentrations and risk of coronary artery disease.** *Nat Genet* 2008, **40:**161-169.
7. Kraakman ATW, Niks RE, Berg PMMM Van den, Stam P, Van Eeuwijk FA: **Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars.** *Genetics* 2004, **168:**435-446.
8. Breseghello F, Sorrells ME: **Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars.** *Genetics* 2006, **172:**1165-1177.
9. Malosetti M, Linden CG van der, Vosman B, van Eeuwijk FA: **A mixed-model approach to association mapping using pedigree information with an illustration of resistance to *Phytophthora infestans* in potato.** *Genetics* 2007, **175:**879-889.
10. Olsen KO, Halldorsdottir SS, Stinchcomb JR, Weinig C, Schmitt J, Purugganan MD: **Linkage disequilibrium mapping of Arabidopsis *CRY2* flowering time alleles.** *Genetics* 2004, **167:**1361-1369.
11. Stich B, Melchinger AE, Frisch M, Maurer HP, Heckenberger M, Reif JC: **Linkage disequilibrium in European elite maize germplasm investigated with SSRs.** *Theor Appl Genet* 2005, **111:**723-730.
12. Pritchard JK, Stephens M, Rosenberg NA, Donnelly P: **Association mapping in structured populations.** *Am J Hum Genet* 2000, **67:**170-181.
13. Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES: **Dwarf8 polymorphisms associate with variation in flowering time.** *Nat Genet* 2001, **28:**286-289.
14. Yu J, Buckler ES: **Genetic association mapping and genome organization of maize.** *Curr Opin Biotech* 2006, **17:**155-160.
15. Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DN, Holland JB, Kresovich S, Buckler ES: **A unified mixed-model method for association mapping that accounts for multiple levels of relatedness.** *Nat Genet* 2006, **2:**203-208.
16. Bernardo R, Murigneux A, Karaman Z: **Marker-based estimates of identity by descent and alikeness in state among maize inbreds.** *Theor Appl Genet* 1996, **93:**262-267.
17. Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C, Tang C, Toomajin C, Zheng H, Dean C, Marjoram P, Nordborg M: **An Arabidopsis example of association mapping in structured samples.** *PLoS Genet* 2007, **3(1):**e4.
18. Bernardo R: **Estimation of coefficient of coancestry using molecular markers in maize.** *Theor Appl Genet* 1993, **85:**1055-1062.
19. Lynch M: **Estimation of relatedness by DNA fingerprinting.** *Mol Biol Evol* 1988, **5:**584-599.
20. Stich B, Möhring J, Piepho H-P, Heckenberger M, Buckler ES, Melchinger AE: **Comparison of mixed-model approaches for association mapping.** *Genetics* 2008, **178:**1745-1754.
21. Pajerowska-Mukhtar K, Stich B, Achenbach U, Ballvora A, Lübeck J, Strahwald J, Tacke E, Hofferbert H-R, Ilarionova E, Bellin D, Walkemeier B, Basekow R, Kersten B, Gebhardt C: **Single nucleotide polymorphisms in the *allene oxide synthase 2* gene of potato (*Solanum tuberosum*) are associated with maturity-corrected resistance to late blight in tetraploid breeding populations.** *Genetics* 2009 in press.

22. Satterzadeh A, Achenbach U, Lübeck J, Strahwald J, Tacke E, Hofferbert H-R, Rothsteyn T, Gebhardt C: **Single nucleotide polymorphism (SNP) genotyping as basis for developing a PCR-based marker highly diagnostic for potato varieties with high resistance to *Globodera pallida* pathotype Pa2/3.** *Mol Breeding* 2006, **18:**301-312.
23. Fry WE: **Quantification of general resistance of potato cultivars and fungicide effects for integrated control of potato late blight.** *Phytopathology* 1978, **68:**1650-1655.
24. Burba M, Puscz W: **Über die Verwendung von Aluminiumsalzen an Stelle von basischen Bleiacetaten zur Klärung von kalten wäßrigen Breiextrakten der Rübe.** *Z Zuckerindustrie* 1976, **26:**249-251.
25. Schneider K, Schäfer-Pregel R, Borchardt DC, Salamini F: **Mapping QTLs for sucrose content, yield and quality in a sugar beet population fingerprinted by EST-related markers.** *Theor Appl Genet* 2002, **104:**1107-1113.
26. Nordborg N, Hu TT, Ishino Y, Jhaveri J, Toomajian C, Zheng H, Bakker E, Calabrese P, Gladstone J, Goyal R, Jakobsson M, Kim S, Morozov Y, Padhukasahasram B, Plagnol V, Rosenberg NA, Shah C, Wall JD, Wang J, Zhao K, Kalbfleisch T, Schulz V, Kreitman M, Bergelson J: **The pattern of polymorphism in Arabidopsis thaliana.** *PLoS Biology* 2005, **3:**e196.
27. Gallais A: **Quantitative genetics and breeding methods in autopolyploid plants.** *Paris: INRA* 2003.
28. Hardy OJ, Vekemans X: **SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population level.** *Mol Ecol Notes* 2002, **2:**618-620.
29. Pritchard JK, Stephens M, Donelly P: **Inference of population structure using multilocus genotype data.** *Genetics* 2000, **155:**945-959.
30. Whitt SR, Buckler ES: **Using natural allelic diversity to evaluate gene function.** In *Plant functional genomics: methods and protocols* Edited by: Grotewald E. Clifton: Humana Press; 2003:123-139.
31. Evanno G, Regnaut S, Goudet J: **Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study.** *Mol Ecol* 2005, **14:**2611-2620.
32. Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ES: **Structure of linkage disequilibrium and phenotypic associations in the maize genome.** *Proc Natl Acad Sci U S A* 2001, **98(20):**11479-11484.
33. Magnussen S: **An algorithm for generating positively correlated Beta-distributed random variables with known marginal distributions and a specified correlation.** *Comput Stat Data An* 2004, **46:**397-406.
34. Gilmour AR, Gogel BJ, Cullis BR, Thompson R: **ASReml User Guide Release 2.0.** Hermel Hempstead UK: VSN International Ltd; 2006.
35. Liu K, Goodman M, Muse S, Smith JS, Buckler E, Doebley J: **Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites.** *Genetics* 2003, **165:**2117-2128.
36. Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S: **Genetic structure and diversity in *Oryza sativa* L.** *Genetics* 2005, **169:**1631-1638.
37. Long AD, Langley CH: **The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits.** *Genome Res* 1999, **9:**720-731.
38. Piepho H-P, Williams ER, Fleck M: **A note on the analysis of designed experiments with complex treatment structure.** *Hort Science* 2006, **41:**446-452.
39. Holland JB, Nyquist WE, Cervantes-Martinez CT: **Estimating and interpreting heritability for plant breeding: an update.** *Plant Breed Rev* 2003, **22:**9-112.