

ORIGINAL ARTICLE

A general method for controlling the genome-wide type I error rate in linkage and association mapping experiments in plants

BU Müller¹, B Stich² and H-P Piepho¹

¹Institute for Crop Science, Bioinformatic Unit, Universität Hohenheim, Stuttgart, Germany and ²Max Planck Institute for Plant Breeding Research, Quantitative Crop Genetics, Köln, Germany

Control of the genome-wide type I error rate (GWER) is an important issue in association mapping and linkage mapping experiments. For the latter, different approaches, such as permutation procedures or Bonferroni correction, were proposed. The permutation test, however, cannot account for population structure present in most association mapping populations. This can lead to false positive associations. The Bonferroni correction is applicable, but usually on the conservative side, because correlation of tests cannot be exploited. Therefore,

a new approach is proposed, which controls the genome-wide error rate, while accounting for population structure. This approach is based on a simulation procedure that is equally applicable in a linkage and an association-mapping context. Using the parameter settings of three real data sets, it is shown that the procedure provides control of the GWER and the generalized genome-wide type I error rate ($GWER_g$). *Heredity* (2011) **106**, 825–831; doi:10.1038/hdy.2010.125; published online 20 October 2010

Keywords: association mapping; genome-wide type I error rate; linkage mapping; mixed model; Monte Carlo simulation; parametric bootstrap

Introduction

Of central importance for marker-assisted selection is the estimation of positions and effects of quantitative trait loci (QTL). Two of the most commonly used tools for estimating the position of QTL are classical linkage mapping (Lander and Botstein, 1989) and association mapping (Bodmer, 1986; Thornsberry *et al.*, 2001; Yu *et al.*, 2006; Sun *et al.*, 2010). The difference between both methods is that in linkage mapping, there are only a few opportunities for recombination to occur within families and pedigrees with known ancestry. This results in a relatively low mapping resolution (Flint-Garcia *et al.*, 2003). By contrast, for association mapping, historical recombination and natural genetic diversity of the different populations lead to a higher mapping resolution (Ersoz *et al.*, 2008; Zhu *et al.*, 2008).

The resolution of association mapping depends on the structure and degree of linkage disequilibrium across the genome. Linkage disequilibrium caused by population structure and familial relatedness lead to false positive results if not controlled correctly in the statistical analysis (Pritchard *et al.*, 2000; Yu *et al.*, 2006). Genetic and non-genetic factors, like recombination, drift and selection, affect the structure of linkage disequilibrium (Stich *et al.*, 2005). To overcome these problems and to reduce the effect of the population structure, several procedures

have been proposed, including the logistic regression ratio test (*Q* model) (Pritchard *et al.*, 2000; Thornsberry *et al.*, 2001), linear mixed models with effects for subpopulations (Brescghello and Sorrels, 2006) and a unified mixed model approach (*QK* model) (Yu *et al.*, 2006). In the *QK* mixed model, Bayesian clustering (Pritchard *et al.*, 2000) is used to estimate probabilities for subpopulation membership (matrix *Q*), which are used to fit fixed effects, whereas random effects are fitted with covariance proportional to the relative kinship matrix *K* (Hardy and Vekemans, 2002). Both *Q* and *K* account for population structure when scanning for marker trait association (Yu *et al.*, 2006).

One major concern in the context of both linkage and association mapping studies is the statistical power and the control of false positive associations (type I error rate). A false positive association occurs when a significant QTL is declared where none really exists. A genome-wide type I error occurs if at least one false QTL is declared. In both linkage and association mapping, multiple testing needs to be accounted for to control the genome-wide type I error rate (GWER).

Different methods were proposed for linkage mapping to control the GWER. Traditionally, the type I error rate has been controlled by a Bonferroni correction. This correction is conservative and sacrifices statistical power because it cannot exploit the correlation structure among the multiple tests. Several alternative analytical methods have been proposed (Davies, 1977; Lander and Botstein, 1989; Feingold *et al.*, 1993; Rebai *et al.*, 1994; Dupuis and Siegmund, 1999; Piepho, 2001; Li and Ji, 2005) that exploit the correlation structure of multiple tests on the same chromosome.

Correspondence: Professor H-P Piepho, Institute for Crop Science, University of Hohenheim, Fruwirthstrasse 23, Stuttgart 70593, Germany.

E-mail: piepho@uni-hohenheim.de

Received 18 January 2010; revised 4 August 2010; accepted 8 August 2010; published online 20 October 2010

A further approach to control the GWER commonly used in linkage mapping is the permutation test of Churchill and Doerge (1994) and Doerge and Churchill (1996). This approach depends on no distributional assumption and is characterized by simplicity and applicability to different experimental populations. In this approach, the trait values are permuted relative to the genotypic data. A disadvantage of the permutation test procedure is the computational workload. To compute a critical threshold for a GWER of 0.01, 10 000 permutations of the trait values are necessary, in which for a GWER of 0.05, 1000 permutations are recommended (Churchill and Doerge, 1994).

Although permutation testing is the standard method in linkage mapping, it is not applicable in an association mapping context because permutation would destroy any correlations between trait and population structure (Aulchencko *et al.*, 2007). This would be inappropriate because a valid test must control for any such structure. Furthermore, analytical methods as proposed for linkage mapping are not available for association mapping.

Another error rate that has been used for linkage mapping and association mapping is the false discovery rate (FDR). Loosely speaking, FDR is the ratio of false positives among detections. This approach was proposed by Benjamini and Hochberg (1995) and for genome-wide studies by Storey and Tibshirani (2003). The popularity of the FDR stems from the fact that it leads to more liberal thresholds than the GWER. Chen and Storey (2006), however, have shown that it is difficult to interpret the FDR when applied to genome-wide linkage scans, because the FDR counts multiple true discoveries as being distinct even though they are from the same underlying gene (De Silva and Ball, 2007). As the marker density applied for association mapping studies will dramatically increase in the near future (Donnelly, 2008), the FDR does not seem to be an appropriate error rate concept for association studies. For this reason, we will not consider it further.

Use of the GWER can lead to conservative tests, if there are numerous QTL. Control of the GWER requires that not a single false positive result occurs among all tests, and it may be argued that this requirement is too stringent in the presence of many QTL. Therefore, Chen and Storey (2006) proposed to relax the definition of GWER by allowing a small number $k > 0$ of false positives, the so-called generalized genome-wide k -error rate (GWER _{k}). The usual GWER corresponds to $k = 0$.

In this study a new approach for controlling both GWER and GWER _{k} is proposed. This method, which is based on simulation, is equally applicable in linkage and association mapping. In the simulation procedure, S random samples from the same multivariate normal distribution are generated under the null hypothesis. For each sample, the test statistic is calculated for each QTL. The critical value, which is used as threshold for controlling GWER _{k} , is given by the α -quantile of the simulated distribution of S values of the $(k+1)$ th smallest P -value. The simulation reflects both the population structure and the correlation of tests. The performance of the method is assessed for three different real data sets regarding different GWER _{k} ($k = 0, 1, 2$ and 5).

Materials and methods

Plant materials, phenotypic data and molecular markers

To assess the performance of our method, we used three empirical data sets that were described in detail by Stich *et al.* (2008) (winter wheat) and by Stich and Melchinger (2009) (sugar beet and rapeseed).

Winter wheat: A total of 303 winter wheat genotypes (*Triticum aestivum*) developed by KWS Lochow GmbH (Bergen-Wohlde, Germany) was used for this study. The entries were evaluated for grain yield in a series of five breeding trials at four to six locations, with the number of entries per trial ranging from 36 to 110. All 303 inbreds were fingerprinted by KWS Lochow GmbH following standard protocols with 36 simple sequence repeat markers and one single nucleotide polymorphism marker. The 37 marker loci were randomly distributed across 19 of the 21 wheat chromosomes.

Sugar beet: A total of 178 sugar beet inbreds (*Beta vulgaris*) of the pollen parent heterotic pool of the KWS SAAT AG (Einbeck, Germany) were analyzed. 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 among others for beet yield. All entries were fingerprinted with 59 simple sequence repeat markers and 41 single nucleotide polymorphism marker, both randomly distributed across the sugar beet genome. The fingerprinting was done by the KWS SAAT AG following standard protocols.

Rapeseed: A total of 136 rapeseed (*Brassica napus*) inbreds of the Norddeutsche Pflanzenzucht Hans-Georg Lembke KG (Holtsee, Germany) were studied. All entries were evaluated in a series of field trials, in which data were collected for thousand-kernel weight. Furthermore, all entries were fingerprinted with 59 genome-wide distributed simple sequence repeat markers by Saaten-Union Resistenzlabor GmbH (Hovedissen, Germany) following standard protocols.

Statistical analyses

Phenotypic data analyses: In the study of Stich *et al.* (2008) the empirical type I error rates of association mapping approaches, which were based on adjusted entry means calculated by a two-step analysis, were only slightly higher than that of approaches in which the phenotypic data analysis and the association analysis were performed in one step (one-step analysis) (also see Möhring and Piepho, 2009). We therefore calculated adjusted entry means (winter wheat and rapeseed) or entry means (sugar beet) in the first step (for more details, see Stich *et al.*, 2008; Stich and Melchinger, 2009) for each entry under consideration. These estimates were then used in a second step for the association analyses.

Population structure analyses: For each of the three above mentioned data sets, the kinship matrix K was calculated based on the available marker data using the software package SPAGeDi (Hardy and Vekemans, 2002), in which negative kinship values between entries were set to 0. We used the first p principal components of an allele frequency matrix (PC-matrix) instead of the Q

matrix of STRUCTURE (Pritchard *et al.*, 2000), as previous studies suggested that both methods are comparable with respect to adherence to the nominal α level, but the former requires much less computational effort (Yu *et al.*, 2006; Zhao *et al.*, 2007). The explained variance of the first p principal components was about 25% (Stich and Melchinger, 2009).

Method for controlling GWER

To scan the genome for QTL in linkage mapping or association mapping, we use a mixed linear model to represent the phenotypic data. At each putative QTL position/marker, we test the null hypothesis of no QTL effect. Under this hypothesis, the null model for genotype means can be written as

$$\mathbf{y} = \mathbf{X}_0\boldsymbol{\beta}_0 + \mathbf{e} \quad (1)$$

where $\mathbf{y}' = (y_1, y_2, \dots, y_G)$, y_i is the mean of the i -th genotype ($i = 1, \dots, G$), $\boldsymbol{\beta}_0$ is a vector of fixed effects, \mathbf{X}_0 is the corresponding design matrix and \mathbf{e} is a random residual. In association mapping, \mathbf{X}_0 might represent the probabilities of subpopulation membership (\mathbf{Q} matrix) or PC-matrix of allele frequencies and, possibly, cofactors accounting for major background QTL, whereas \mathbf{e} models genetic correlation due to coancestry and identically distributed noise, that is, $\text{var}(\mathbf{e}) = \mathbf{V} = 2A\sigma_\lambda^2 + \mathbf{I}\sigma^2$, where \mathbf{A} is the numerator relationship matrix. Alternatively, \mathbf{A} could be replaced by the kinship matrix \mathbf{K} (Yu *et al.*, 2006), which was done in this study. For the rapeseed data set, \mathbf{e} models $\text{var}(\mathbf{e}) = \mathbf{V} = \mathbf{I}\sigma^2$, because \mathbf{A} was similar to \mathbf{I} and no changes were visible in the log likelihoods when fitting the full model including \mathbf{A} .

To test the null hypothesis at the q th putative position ($q = 1, 2, \dots, Z$), we augment the null model by

$$\mathbf{W}_q\boldsymbol{\alpha}_q \quad (2)$$

where $\boldsymbol{\alpha}_q$ is the vector of fixed genetic effects at the q th putative position and \mathbf{W}_q is the associated design matrix. Notably, the dimension of $\boldsymbol{\alpha}_q$ may vary among markers, depending on the genetic model and the number of alleles per marker. Furthermore, we need to cater for the possibility that marker information may be missing, especially in association mapping, in which imputation is not straightforward. The approach taken in this study is to simply discard records of individuals with missing information at the q th marker when testing the q th marker, meaning that different subsets of the data will be used for different markers. We therefore add a subscript q also to the data vector \mathbf{y} , writing \mathbf{y}_q . Thus, \mathbf{y}_q contains all records with complete data at the q th marker. Consequently, the design matrix \mathbf{W}_q will have rows only for observations in \mathbf{y}_q . The marker-specific data vector may be formally defined as follows. Let \mathbf{B} be a $G \times Z$ indicator matrix of zeros and ones, with rows corresponding to genotypes and columns to markers, reflecting the missing data pattern and let \mathbf{D}_q be computed by $\text{diag}(\mathbf{b}_q)$, deleting all rows that have zeros only where \mathbf{b}_q is the q th column of \mathbf{B} . We then have

$$\mathbf{y}_q = \mathbf{D}_q\mathbf{y}.$$

\mathbf{D}_q selects from \mathbf{y} all observations that have complete data for the q th marker. The reduced data vector \mathbf{y}_q has variance

$$\text{var}(\mathbf{y}_q) = \mathbf{V}_{qq} = \mathbf{D}_q\mathbf{V}\mathbf{D}_q'.$$

The full model can be written compactly as

$$\mathbf{y}_q = \mathbf{X}_q\boldsymbol{\beta}_q + \mathbf{e}_q, \quad (3)$$

where $\mathbf{X}_q = (\mathbf{D}_q\mathbf{X}_0, \mathbf{W}_q)$ and $\boldsymbol{\beta}_q' = (\boldsymbol{\beta}_0', \boldsymbol{\alpha}_q')$. The null hypothesis at the q th position can be stated as

$$\mathbf{H}_0 : \mathbf{H}_q\boldsymbol{\beta}_q = \mathbf{0}, \quad (4)$$

where \mathbf{H}_q is a suitable matrix of known constants. The size and form of \mathbf{H}_q depend on the putative position q , for example, on the number of marker alleles. Furthermore, the null hypothesis pertains to $\boldsymbol{\alpha}_q$ only, that is, $\mathbf{H}_q = (\mathbf{0}_q \tilde{\mathbf{H}}_q)$, where $\mathbf{0}_q$ is a null matrix with number of columns corresponding to those of $\mathbf{D}_q\mathbf{X}_0$ and $\tilde{\mathbf{H}}_q$ states the null hypothesis pertaining to $\boldsymbol{\alpha}_q$. For example, when \mathbf{H}_0 states equality of all additive allele effects at a locus, then $\tilde{\mathbf{H}}_q = (\mathbf{I}_{n(q)}, -\mathbf{1}_{n(q)})$, where $n(q)$ equals the number of marker alleles minus one. Thus,

$$\mathbf{H}_0 : \mathbf{H}_q\boldsymbol{\beta}_q = \tilde{\mathbf{H}}_q\boldsymbol{\alpha}_q = \mathbf{0}$$

When \mathbf{V} is known, the Wald statistic

$$T_q = \hat{\boldsymbol{\beta}}_q' \mathbf{H}_q' (\mathbf{H}_q (\mathbf{X}_q' \mathbf{V}_{qq}^{-1} \mathbf{X}_q)^{-1} \mathbf{H}_q')^{-1} \mathbf{H}_q \hat{\boldsymbol{\beta}}_q, \quad (5)$$

where $\hat{\boldsymbol{\beta}}_q = (\mathbf{X}_q' \mathbf{V}_{qq}^{-1} \mathbf{X}_q)^{-1} \mathbf{X}_q' \mathbf{V}_{qq}^{-1} \mathbf{y}_q$, has an exact central χ^2 -distribution with $\text{rank}(\mathbf{H}_q)$ degrees of freedom. In practice, \mathbf{V} needs to be estimated from the data based on the null model (1). In this case, one may use the Kenward–Roger method to approximate the distribution of T_q . Provided the number of genotypes G is not small, Equation (5) will have an approximate χ^2 -distribution. We expect the approximation to be very accurate in most practical cases, so long as the number of genotypes is not very small (for example, < 50).

Simulation of the joint distribution of T_1, T_2, \dots, T_Z

It is convenient to re-write T_q as

$$T_q = \hat{\boldsymbol{\gamma}}_q' \mathbf{M}_{qq}^{-1} \hat{\boldsymbol{\gamma}}_q, \quad (6)$$

where $\hat{\boldsymbol{\gamma}}_q = \mathbf{H}_q \hat{\boldsymbol{\beta}}_q$ and

$$\mathbf{M}_{qq} = \text{var}(\hat{\boldsymbol{\gamma}}_q) = \mathbf{H}_q (\mathbf{X}_q' \mathbf{V}_{qq}^{-1} \mathbf{X}_q)^{-1} \mathbf{H}_q' \quad (7)$$

Under the global null hypothesis the joint distribution of

$$\hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\gamma}}_1', \hat{\boldsymbol{\gamma}}_2', \dots, \hat{\boldsymbol{\gamma}}_Z') \quad (8)$$

is multivariate normal with zero mean and variance-covariance matrix

$$\text{var}(\hat{\boldsymbol{\gamma}}) = \begin{pmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} & \dots & \mathbf{M}_{1Z} \\ \mathbf{M}_{21} & \mathbf{M}_{22} & \dots & \mathbf{M}_{2Z} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{M}_{Z1} & \mathbf{M}_{Z2} & \dots & \mathbf{M}_{ZZ} \end{pmatrix}, \quad (9)$$

where $\mathbf{M}_{qq'} = \text{cov}(\hat{\boldsymbol{\gamma}}_q, \hat{\boldsymbol{\gamma}}_{q'}) = \mathbf{H}_q (\mathbf{X}_q' \mathbf{V}_{qq}^{-1} \mathbf{X}_q)^{-1} \mathbf{X}_q' \mathbf{V}_{qq'}^{-1} \mathbf{X}_{q'} (\mathbf{X}_{q'}' \mathbf{V}_{q'q'}^{-1} \mathbf{X}_{q'})^{-1} \mathbf{H}_{q'}'$ ($q, q' = 1, \dots, Z$). This result is explained in more detail in the Appendix. Notably, when $q = q'$, then Equation (9) simplifies to Equation (7).

For simulating $\hat{\boldsymbol{\gamma}}_{sim}$, it is convenient to compute Equation (9), obtain a decomposition $\text{var}(\hat{\boldsymbol{\gamma}}) = \mathbf{P}\mathbf{P}'$, where the number of columns in \mathbf{P} equals the rank r of $\text{var}(\hat{\boldsymbol{\gamma}})$, store \mathbf{P} in memory during iterations, and at each iteration simulate $\hat{\boldsymbol{\gamma}}$ as $\hat{\boldsymbol{\gamma}}_{sim} = \mathbf{P}\mathbf{u}_{sim}$, where \mathbf{u}_{sim} is a vector of r -independent standard normal deviates. We can use the singular value decomposition

$$\text{var}(\hat{\boldsymbol{\gamma}}) = \mathbf{U}\mathbf{F}\mathbf{U}'$$

where F is a diagonal matrix, first diagonal elements of which are the r non-zero singular values of $\text{var}(\hat{y})$, whereas the remaining ones are zero. We can then choose

$P = (U\sqrt{F})_r$, where $(M)_r$ is given by the first r columns of M .

To compute a critical threshold for the Wald tests controlling the GWER at level α , we may generate S random samples \hat{y}_{sim} from this same multivariate normal distribution. For each sample, we compute the corresponding test statistics T_q ($q=1, \dots, Z$). As test statistics T_q may involve hypotheses with differing degrees of freedom for different q , we convert each T_q to the point-wise P -value p_q based on a χ^2 distribution with degrees of freedom equal to $\text{rank}(H_q)$. Conversion to P -values allows us to use the same rejection region for all QTL (Storey, 2002). Subsequently we determine the minimum of p_q across positions ($p_{q(\min)}$). The critical value is given by the α -quantile of the simulated distribution of S values of p_{\min} .

The approach can be extended further using the GWER $_k$ approach of Chen and Storey (2006), which defines a genome-wide error to occur when more than k point-wise tests are falsely declared significant. In this more general case, the $(k+1)$ th lowest p_q across positions is determined in each simulation run. Notably the ordinary GWER corresponds to $k=0$.

Simulation study

The performance of the above method is verified by simulation. As the method for determining the threshold is also based on simulation, there are two levels of simulation: (1) an inner simulation that generates the thresholds for a given data set, and (2) an outer simulation that generates data to be analyzed by a mixed model.

The simulation scheme can be described as follows:

- Do $i = 1$ to n ($n = \text{number of outer loops}$)
- Generate a data set y_{sim} from a multivariate normal distribution with zero mean, using restricted maximum likelihood estimates of V of a real data set.
 - Determine threshold based on simulation with S runs of the inner loops, using y_{sim} and X_0 and W_q ($q = 1, \dots, Z$) from real data set.
 - Evaluate significance tests for scan of i th simulated data set y_{sim} and determine the $(k+1)$ th ordered P -value across the positions.

End

Determine the threshold P -value for $\text{GWER}_k = \alpha$ as the α -quantile of the $n(k+1)$ th ordered P -values.

To start a simulation, we analyze a real data set under the global H_0 based on model (1), obtain an estimate of V and then compute its Cholesky decomposition according to $V = LL'$. In each run of the outer loop, we then simulate data under the global H_0 as

$$y_{sim} = Lv,$$

where v is a vector of independent standard normal deviates. The same L is used in all iterations of the outer loop, so L needs to be stored throughout the whole simulation.

Results

The proposed method for controlling the GWER_k (Chen and Storey, 2006) was tested on three empirical data sets of commercial plant-breeding programs.

The threshold computation and the analysis of the PC-K mixed model were repeated 1000 times, meaning there were 1000 inner simulations and 1000 outer simulations. Notably, for a test to be declared significant, the P -value had to remain below the threshold P -value. At a nominal error rate of 5%, a 95% prediction interval for the observed error rate has lower limit of 3.65% and upper limit of 6.35% when 1000 runs converged. Thus, an observed error rate should not exceed 64 cases or fall below 36 cases of the 1000 simulations if tests control α exactly. For the sugar beet data set only 978 outer simulations converged. The 95% prediction interval therefore has a lower limit of 3.63% and an upper limit of 6.37%. We also computed Bonferroni-adjusted prediction intervals based on the 12 cases studied (Table 1). For 1000 runs and for the 978 runs of the sugar beet data set, the Bonferroni adjusted limits are 30 and 70, respectively. The empirical error rates for the GWER_k are given in Table 1.

The nominal GWER could be maintained for the winter wheat data set of KWS Lochow. For $\text{GWER}_{k=0}$ in 6.0% of simulations, the critical threshold was higher than the P -values of the PC-K mixed models. The threshold for $\text{GWER}_{k=0}$ was 0.00139, which is higher than the Bonferroni-corrected threshold (0.00135). The extension of Chen and Storey (2006) led to further reduction of times the critical threshold was higher than the P -values of the PC-K mixed models. The critical

Table 1 Empirical levels in counts and percent of converged cases for the GWER_k ($k=0, 1, 2$, and 5) at nominal level of 5% for the three different data sets and the prediction interval of upper and lower limit

Data set	GWER_k	Empirical level (absolute count)	Empirical level (in %)	Prediction interval (95%) for absolute rejection frequency (in %)	Bonferroni-adjusted prediction interval (95%) for absolute rejection frequency (in %)
Wheat	0	60	6.0	3.649–6.351	3.027–6.973
Wheat	1	49	4.9	3.649–6.351	3.027–6.973
Wheat	2	36	3.6	3.649–6.351	3.027–6.973
Wheat	5	29	2.9	3.649–6.351	3.027–6.973
Sugar	0	60	6.1	3.634–6.366	3.006–6.993
Sugar	1	53	5.4	3.634–6.366	3.006–6.993
Sugar	2	46	4.7	3.634–6.366	3.006–6.993
Sugar	5	23	2.4	3.634–6.366	3.006–6.993
Rapeseed	0	63	6.3	3.649–6.351	3.027–6.973
Rapeseed	1	50	5.0	3.649–6.351	3.027–6.973
Rapeseed	2	33	3.3	3.649–6.351	3.027–6.973
Rapeseed	5	27	2.7	3.649–6.351	3.027–6.973

threshold was passed for $\text{GWER}_{k=1}$ in 4.9% of the simulations, for $\text{GWER}_{k=2}$ in 3.6%, and for the $\text{GWER}_{k=5}$ in 2.2% of the simulations (Table 1).

For the sugar beet data set of KWS, the nominal GWER could be kept. In 6.1% of the simulations for $\text{GWER}_{k=0}$, the threshold was higher than the P -values of the PC-K mixed model. The threshold for $\text{GWER}_{k=0}$ was 0.00052 and therefore higher than the threshold corrected by the Bonferroni method (0.00050). Furthermore, for the modified GWER_k with $k=1, 2$ and 5 the nominal rate of 5% could be maintained. In 5.4% of the simulations, a type I error occurred for $\text{GWER}_{k=1}$, in 4.7% of the simulations for the $\text{GWER}_{k=2}$ and in 2.4% of the simulations for the $\text{GWER}_{k=5}$ (Table 1).

Our method could also satisfactorily control the nominal GWER for the third data set. For the rapeseed data set of Norddeutsche Pflanzenzucht, the threshold for the $\text{GWER}_{k=0}$ was higher than the P -values of the PC-K mixed model in 6.3% of the simulations. The threshold for $\text{GWER}_{k=0}$ was 0.000937 and therefore also higher than the Bonferroni-corrected threshold that had the value 0.000847. For the $\text{GWER}_{k=1}$, the empirical error rate was 5.0%; for $\text{GWER}_{k=2}$ it was 3.3%. The empirical $\text{GWER}_{k=5}$ was 2.7% (Table 1).

Discussion

Error rates for controlling the multiple testing in linkage and association mapping experiments include the FDR, which was proposed by Benjamini and Hochberg (1995) and Storey and Tibshirani (2003), and the GWER and its extension GWER_k , which was proposed by Chen and Storey (2006). For linkage mapping, different approaches were proposed, which control the GWER, like the Bonferroni correction, the permutation procedure (Churchill and Doerge, 1994; Doerge and Churchill, 1996) and several analytical methods for specific population structures (Davies, 1977; Lander and Botstein, 1989; Feingold *et al.*, 1993; Rebai *et al.*, 1994; Dupuis and Siegmund, 1999; Piepho, 2001). Thus, at present there do not seem to be tailor-made methods for controlling GWER for association mapping experiments. This study has proposed a simulation-based approach for controlling the type I error rate, which includes the information of the population structure. The approach is akin to that proposed by Edwards and Berry (1987) in the context of multiple mean comparisons in linear models, and it is also similar in spirit to the method of Zou *et al.* (2004) in the context of linkage mapping. The simulation approach can also be regarded as a parametric bootstrap procedure (Efron and Tibshirani, 1993). The simulations of the proposed method based on the three commercial plant breeding data sets have shown that the calculated thresholds provide reasonable, slightly conservative control of the genome-wide type I error rate.

An advantage of our proposed method over the permutation procedure of Churchill and Doerge (1994) is that the information of the population structure is accounted for in our threshold computation. The associations between trait and population structure are not destroyed like for the permutations procedure. Aulchenko *et al.*, 2007 proposed an approach, in which residuals from a mixed model fit ignoring markers, but corrected for family effects are used for the permutations

test. The method was developed in an animal breeding context for genetically homogeneous populations, but its principles could be applied to the more general setting considered here. Residuals from a mixed model fit will typically display correlation and heteroscedasticity arising from the estimation of model effects, which may affect the performance of the method. Our procedure does not have these limitations, because the null distribution is simulated rather than computed from permutations.

Li and Ji (2005), Seaman and Müller-Myhsok (2005) and Conneely and Boehnke (2007) suggested methods to adjust the P -value regarding the correlation structure of the markers. These approaches are therefore similar to our approach; but they do not account for population structure. Moreover, the approaches of Seaman and Müller-Myhsok (2005) and Conneely and Boehnke (2007) need imputation, if there are missing values in the marker data. The occurrence of missing values can be handled without imputation by our proposed method.

For the three data sets used in this study, the computation time for one approximate threshold was 1 min and 23 s for the rapeseed data set up to 9 min and 20 s for the sugar beet data set (Intel Pentium Dual central processing unit, 2.20 GHz, 1.95 GB random access memory). The computational time depends on the number of markers and on the number of genotypes. The computational time increases mainly due to the generation of the matrix M , if there are more markers. Furthermore, the computational time is increased by the number of genotypes because mixed model analysis takes longer time. The computational time could be reduced, if necessary, by performing threshold computation separately for each chromosome and using a Bonferroni correction across chromosomes (Piepho, 2001). Moreover, when the number of markers by far exceeds the number of genotypes, it will be computationally more efficient to simulate data y instead of test statistics T_q (Supplementary Information).

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

We thank the breeding companies KWS, KWS Lochow and Norddeutsche Pflanzenzucht for providing the data sets within the GABI BRAIN project. This study was supported by the GABI GAIN project (Grant no FKZ0315072C).

References

- Aulchenko YS, de Koning DJ, Haley C (2007). Genomewide rapid association using mixed model and regression: a fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. *Genetics* **177**: 577–585.
- Benjamini Y, Hochberg Y (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B* **85**: 289–300.
- Bodmer WF (1986). Human genetics: the molecular challenge. Cold spring harbour symp. *Quant Biol* **51**: 1–13.

- Breseghele F, Sorrells ME (2006). Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics* **172**: 1165–1177.
- Chen L, Storey JD (2006). Relaxed significance criteria for linkage analysis. *Genetics* **173**: 2371–2381.
- Churchill GA, Doerge RW (1994). Empirical threshold values for quantitative trait mapping. *Genetics* **138**: 963–971.
- Conneely KN, Boehnke M (2007). So many correlated tests, so little time! Rapid adjustment of *P*-values for multiple correlated tests. *Am J Hum Genet* **81**: 1158–1168.
- Davies RB (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **64**: 247–254.
- De Silva HN, Ball RD (2007). Linkage disequilibrium mapping concepts. In: Oraguzie NC, Rikkerink EHA, Gardiner SE, De Silva HN (eds). *Association Mapping in Plants*. Springer: New York, NY, USA.
- Doerge RW, Churchill GA (1996). Permutation tests for multiple loci affecting a quantitative character. *Genetics* **142**: 285–294.
- Donnelly P (2008). Progress and challenges in genome-wide association studies in humans. *Nature* **456**: 728–731.
- Dupuis J, Siegmund D (1999). Statistical methods for mapping quantitative trait loci from a dense set of markers. *Genetics* **151**: 373–386.
- Edwards D, Berry J (1987). The efficiency of simulation-based multiple comparisons. *Biometrics* **43**: 913–928.
- Efron B, Tibshirani RJ (1993). *An introduction to the bootstrap*. Chapman & Hall, London.
- Ersöz ES, Yu J, Buckler ES (2008). Applications of linkage disequilibrium and association mapping in maize. In: Kriz A, Larkins B (eds). *Molecular Genetic Approaches to Maize Improvement*. Springer: Dordrecht, The Netherlands.
- Feingold EP, Brown PO, Siegmund D (1993). Gaussian models for genetic linkage analysis using complete high-resolution maps of identity by descent. *Am J Hum Genet* **53**: 234–251.
- Flint-Garcia SA, Thornsberry JM, Buckler ES (2003). Structure of linkage disequilibrium in plants. *Ann Rev Plant Biol* **54**: 357–374.
- Hardy OJ, Vekemans X (2002). SPAGeDI: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol Ecol Notes* **2**: 618–620.
- Lander ES, Botstein D (1989). Mapping Mendelian factors underlying quantitative traits using RFLP markers. *Genetics* **121**: 185–199.
- Li J, Ji L (2005). Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity* **95**: 221–227.
- Möhrling J, Piepho HP (2009). Comparison of weighting in two-stage analysis of plant breeding trials. *Crop Sci* **49**: 1977–1988.
- Piepho HP (2001). A quick method for computing approximate thresholds for quantitative trait loci detection. *Genetics* **157**: 425–432.
- Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000). Association mapping in structured populations. *Am J Hum Genet* **67**: 170–181.
- Rebai A, Goffinet B, Mangin B (1994). Approximate thresholds of interval mapping tests for QTL detection. *Genetics* **138**: 235–240.
- Seaman SR, Müller-Myhsok B (2005). Rapid simulation of *P*-values for product methods and multiple-testing adjustment in association studies. *Am J Hum Genet* **76**: 399–408.
- Stich B, Melchinger AE (2009). Comparison of mixed-model approaches for association mapping in rapeseed, potato, sugar beet, maize, and *Arabidopsis*. *BMC Genomics* **10**: 94.
- Stich B, Melchinger AE, Frisch M, Maurer HP, Heckenberger M, Reif JC (2005). Linkage disequilibrium in European elite maize germplasm investigated with SSRs. *Theor Appl Genet* **111**: 723–730.
- Stich B, Möhring J, Piepho HP, Heckenberger M, Buckler ES, Melchinger AE (2008). Comparison of mixed-model approaches for association mapping. *Genetics* **178**: 1745–1754.
- Storey JD (2002). A direct approach to false discovery rates. *J R Stat Soc Ser B Stat Methodol* **64**: 479–498.
- Storey JD, Tibshirani R (2003). Statistical significance for genome-wide studies. *Proc Natl Acad Sci USA* **100**: 9440–9445.
- Sun G, Zhu C, Kramer MH, Yang SS, Song W, Piepho HP et al. (2010). Comparing different R² statistics for mixed model association mapping. *Heredity* **105**: 333–340.
- Thornsberry JM, Goodmann MM, Doebley J, Kresovich S, Nielsen D, Buckler IV ES (2001). Dwarf8 polymorphisms associate with variation in flowering time. *Nat Genet* **28**: 286–289.
- Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF et al. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet* **38**: 203–208.
- Zhao J, Paulo MJ, Jamar D, Lou P, Van Eeuwijk F, Bonnema G et al. (2007). Association mapping of leaf traits, flowering time, and phytate content in *Brassica rapa*. *Genome* **50**: 963–973.
- Zhu C, Gore M, Buckler ES, Yu J (2008). Status and prospects of association mapping in plants. *Plant Genome* **1**: 5–19.
- Zou F, Fine JP, Hu J, Lin DY (2004). An efficient resampling method for assessing genome-wide statistical significance in mapping quantitative trait loci. *Genetics* **168**: 2307–2316.

Appendix

Let $\mathbf{y}_q = \mathbf{D}_q \mathbf{y}$ and $\mathbf{y}_{q'} = \mathbf{D}_{q'} \mathbf{y}$. Then

$$\begin{aligned} \text{var} \left[\begin{pmatrix} \mathbf{D}_q \\ \mathbf{D}_{q'} \end{pmatrix} \mathbf{y} \right] &= \begin{pmatrix} \mathbf{D}_q \\ \mathbf{D}_{q'} \end{pmatrix} \mathbf{V}(\mathbf{D}_q' \ \mathbf{D}_{q'}') \\ &= \begin{pmatrix} \mathbf{V}_{qq} & \mathbf{V}_{qq'} \\ \mathbf{V}_{q'q} & \mathbf{V}_{q'q'} \end{pmatrix}, \end{aligned}$$

where

$$\mathbf{V}_{qq'} = \text{cov}(\mathbf{y}_q, \mathbf{y}_{q'}) = \mathbf{D}_q \mathbf{V} \mathbf{D}_{q'}'.$$

Similarly, noting that $\hat{\mathbf{y}}_q = \mathbf{C}_q \mathbf{y}_q$ with $\mathbf{C}_q = \mathbf{H}_q (\mathbf{X}_q' \mathbf{V}_{qq}^{-1} \mathbf{X}_q) \mathbf{X}_q' \mathbf{V}_{qq}^{-1}$, we have

$$\begin{pmatrix} \hat{\mathbf{y}}_q \\ \hat{\mathbf{y}}_{q'} \end{pmatrix} = \begin{pmatrix} \mathbf{C}_q & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{q'} \end{pmatrix} \begin{pmatrix} \mathbf{y}_q \\ \mathbf{y}_{q'} \end{pmatrix} = \begin{pmatrix} \mathbf{C}_q & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{q'} \end{pmatrix} \begin{pmatrix} \mathbf{D}_q \\ \mathbf{D}_{q'} \end{pmatrix} \mathbf{y}$$

and

$$\begin{aligned} \text{var} \begin{pmatrix} \hat{\mathbf{y}}_q \\ \hat{\mathbf{y}}_{q'} \end{pmatrix} &= \\ &= \begin{pmatrix} \mathbf{C}_q & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{q'} \end{pmatrix} \begin{pmatrix} \mathbf{D}_q \\ \mathbf{D}_{q'} \end{pmatrix} \mathbf{V}(\mathbf{D}_q' \ \mathbf{D}_{q'}') \begin{pmatrix} \mathbf{C}_q' & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{q'}' \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{M}_{qq} & \mathbf{M}_{qq'} \\ \mathbf{M}_{q'q} & \mathbf{M}_{q'q'} \end{pmatrix}, \end{aligned}$$

where $M_{qq'} = C_q V_{qq'} C_q'$. Inserting the expression for C_q we find

$$\begin{aligned} M_{qq} &= C_q V_{qq} C_q' \\ &= H_q (X_q' V_{qq}^{-1} X_q)^{-1} X_q' V_{qq}^{-1} V_{qq} V_{qq}^{-1} X_q (X_q' V_{qq}^{-1} X_q)^{-1} H_q' \\ &= H_q (X_q' V_{qq}^{-1} X_q)^{-1} H_q' \end{aligned}$$

and

$$\begin{aligned} M_{qq'} &= C_q V_{qq'} C_q' \\ &= H_q (X_q' V_{qq}^{-1} X_q)^{-1} X_q' V_{qq}^{-1} V_{qq'} V_{q'q}^{-1} X_{q'} \\ &\quad \times (X_{q'}' V_{q'q}^{-1} X_{q'})^{-1} H_{q'}'. \end{aligned}$$

Supplementary Information accompanies the paper on Heredity website (<http://www.nature.com/hdy>)