ORIGINAL ARTICLE

Detection of epistatic interactions in association mapping populations: an example from tetraploid potato

B Stich and C Gebhardt

Max Planck Institute for Plant Breeding Research, Köln, Germany

Epistatic interactions among loci are expected to contribute substantially to variation of quantitative traits. The objectives of our research were to (i) compare a classical mixed-model approach with a combined mixed-model and analysis of variance approach for detecting epistatic interactions; (ii) examine using computer simulations the statistical power to detect additive–additive, additive–dominance and dominance–dominance epistatic interactions and (iii) detect epistatic interactions between candidate genes for resistance to leaf blight in a set of tetraploid potato clones. Our study was based on the genotypic and phenotypic data of 184 tetraploid potato cultivars as well as computer simulations. The number of significant ($\alpha^* = 1 \times 10^{-6}$)

epistatic interactions ranged for the three examined traits from 3 to 32. Our findings suggested that the combined mixed-model and analysis of variance approach leads in comparison with the classical mixed-model approach not to an increased rate of false-positives. The results of the computer simulations suggested that, if molecular markers are available that are in high LD ($D' > 0.9$) with the trait-coding loci, the statistical power to detect epistatic interactions, which explain 5–10% of the phenotypic variance, was of a size that seems promising for their detection.

Heredity (2011) **107**, 537–547; doi:10.1038/hdy.2011.40; published online 15 June 2011

Keywords: epistasis; association mapping; potato

Introduction

The contribution of epistatic interactions to variation of quantitative traits is supported by classical quantitative genetic studies such as generation mean analyses (for example, Mather and Jinks, 1982). Furthermore, epistatic interactions have also been identified in model species as well as various crops using populations of near-isogenic lines (for example, Eshed and Zamir, 1996; Kroymann and Mitchell-Olds, 2005) or linkage mapping populations (for example, Blanc *et al.*, 2006), where differing conclusions about the importance of epistasis have been made in different studies. It has been suggested that this observation might be because of the differing experimental materials used (Eshed and Zamir, 1996). Another explanation might be the difference in the reproduction type of the examined species, as studies in autogamous species have more often provided evidence for the importance of epistasis than studies in allogamous species (Holland, 2001).

Association mapping methods, which use the population-wide linkage disequilibrium (LD), are increasingly being used in order to identify quantitative trait loci (QTL) (for example, Thornsberry *et al.*, 2001; Yan *et al.*, 2010). However, with very few exceptions (for example, Manicacci *et al.*, 2009; Stracke *et al.*, 2009; Li *et al.*, 2010), all earlier association mapping studies have neglected

the presence of epistasis. This might be because of the high computational burden of mixed-model association mapping approaches, which to our knowledge are the state of the art for performing association analyses in structured populations (Yu *et al.*, 2006). However, no earlier study examined alternative approaches for detection of epistatic interactions (cf. Aulchenko *et al.*, 2007).

The power to detect marker–phenotype associations is a function of the allele frequency, the sample size, the importance of population structure as well as the extent of LD. Therefore, earlier studies examined, based on empirical data sets, the power of various association mapping approaches with respect to the detection of QTL with additive effects (Yu *et al.*, 2006; Stich and Melchinger, 2009). To our knowledge, however, no earlier study examined the power to detect pairwise epistatic interactions using association mapping methods.

Association mapping approaches are particularly useful for species for which either the development of segregating populations is extremely laborious or where linkage analysis is complicated by polysomic inheritance (Luo *et al.*, 2001). The latter is the case for cultivated potato (*Solanum tuberosum*). Potato cultivation is strongly affected by the oomycete *Phytophthora infestans*, which causes the late blight disease. Without chemical control, epidemics of *P. infestans* can lead to complete yield loss. Extensive use of pesticides in crop protection increases production costs, is damaging to the environment and contributes to the emergence of resistant *P. infestans* isolates (Duncan, 1999; Kamoun and Smart, 2005). Therefore, we used the resistance to late blight and the correlated trait plant maturity (PM) as model traits for

Correspondence: B Stich, Max Planck Institute for Plant Breeding Research, Carl-von-Linné Weg 10, Köln 50829, Germany.

E-mail: stich@mpipz.mpg.de

Received 22 November 2010; revised 8 April 2011; accepted 12 April 2011; published online 15 June 2011

our study in order to contribute to a genetic improvement of potato. To reach this goal, 22 candidate loci, which were selected on the basis of (i) colocalization with known QTL for resistance to *P. infestans* or other pathogens; (ii) close linkage with R genes or R gene-like sequences and/or (iii) coding for a gene with known function in pathogen resistance, were examined.

The objectives of our research were to (i) compare a classical mixed-model approach with a combined mixed-model and analysis of variance approach for detecting epistatic interactions; (ii) examine with computer simulations the statistical power to detect additive-additive, additive-dominance and dominance-dominance epistatic interactions and (iii) detect epistatic interactions between candidate loci for resistance to leaf blight in a set of tetraploid potato clones.

Materials and methods

Plant materials

Our study was based on 184 tetraploid potato cultivars from the breeding programs of Böhm-Nordkartoffel-Agrarproduktion OHG (Ebendorf, Germany) and SAKA-Pflanzenzucht GbR (Windeby, Germany). The clones were selected to represent a broad range of breeding materials with respect to field resistance to late blight after excluding very late-maturing genotypes. In addition, no full sibs were included in the set. Further details of the germplasm set are described elsewhere (Pajeroska-Mukhtar *et al.*, 2009).

Phenotypic evaluation

The clones of our study were evaluated for field resistance to late blight during the years 2004 to 2006, with two replications per year at one to two locations as described by Pajeroska-Mukhtar *et al.* (2009). The plots were inoculated with a mixture of two complex field isolates of *P. infestans*. Disease progress was monitored every 3–4 days, using a 1-to-9 scale (1, 0% infection; 9, >90%), and the area under the disease-progress curve (AUDPC) (Fry, 1978) was calculated for each year–location combination. Furthermore, PM was evaluated in uninfected plants grown under standard phytosanitary regimes in comparison with standard potato varieties of known maturity type, using a 1-to-9 scale (1, very late; 9, very early maturing). From AUDPC and PM, maturity-corrected resistance (MCR) was calculated (Pajeroska-Mukhtar *et al.*, 2009).

Collection of genotypic data

The potato clones of our study were fingerprinted using 31 genome-wide-distributed simple sequence-repeat markers. For 21 markers, the allele dosage was scored based on relative band intensities. Furthermore, 22 candidate loci (Table 1) selected on the basis of (i) colocalization with known QTL for resistance to *P. infestans* or other pathogens, (ii) close linkage with R genes or R gene-like sequences and/or (iii) coding for a gene with known function in pathogen resistance were custom-sequenced as described previously by Pajeroska-Mukhtar *et al.* (2009). Single-nucleotide polymorphisms and insertion/deletion

Table 1 Locus name, description, chromosome location and number of polymorphisms (Polym.) of the candidate loci examined in the current study

Locus	Locus description	Chr. no.	No. Polym.
StPen1	Defense signaling gene homologous to <i>Arabidopsis</i> PEN1	1	23
StPAD4-1	Defense signaling gene homologous to <i>Arabidopsis</i> PAD4, linked to the late blight quantitative trait locus (QTL) Pin2A	2	8
GP321	Sequence-tagged site (STS) linked to late blight QTL Pin2B	2	5
Stp23	Gene encoding L-type starch phosphorylase, linked to resistance hot spot including late blight QTL Pin3A	3	8
StSGT1-1	Defense signaling gene homologous to <i>Arabidopsis</i> SGT1-1, linked to late blight QTL Pin3B	3	10
S1b3	Expressed sequence tag (EST) encoding a candidate gene (defense response, strong homology to respiratory burst oxidase), linked to late blight QTL Pin3C	3	7
St4cl	Defense response gene encoding 4-coumarate: CoA ligase, linked to late blight QTL Pin3C	3	11
BA106c14t7	End of a BAC insertion containing a member of the Gro1 resistance gene family, linked to a resistance hot spot including late blight QTL Pin4A	4	10
StAOS1	Defense signaling gene homologous to <i>Arabidopsis</i> AOS	4	8
P4h6	EST encoding a gene with nucleotide binding and leucine-rich repeat domains linked to a resistance hot spot including late blight QTL Pin5A	5	13
BA213c14t7	End of a BAC insertion containing the R1 resistance gene, linked to a resistance hot spot including late blight QTL Pin5A	5	31
239E4left	STS marker linked to the H1 nematode resistance gene and late blight QTL PinVB	5	11
StSGT1-2	Defense signaling gene homologous to <i>Arabidopsis</i> SGT1-2, linked to a resistance hot spot including late blight QTL PinVIA	6	11
StEDS1	Defense signaling gene homologous to <i>Arabidopsis</i> EDS1	6	18
BA228g19t7	End of a BAC insertion containing resistance gene-like sequences	7	9
BA229c2t3	End of a BAC insertion physically linked to the Gro1 resistance gene family	7	7
BA73e8t3	End of a BAC insertion containing resistance gene-like sequences	8	9
CP105	EST marker linked to the <i>R_{per}</i> gene for resistance to late blight	10	11
BA811t5t3	End of a BAC insertion containing resistance gene-like sequences	10	9
StAOS2	Defense signaling gene, functional homolog of <i>Arabidopsis</i> AOS, linked to a resistance hot spot including late blight QTL Pin11B	11	14
OSM8e	Defense response gene, osmotin like	11	10
GP122	STS marker, linked to the <i>Ry-f_{sto}</i> resistance gene	12	15

For references see Pajeroska-Mukhtar *et al.* (2009).

polymorphisms (InDels) detected from sequence alignments are designated in the following sections as molecular markers. The single-nucleotide polymorphism allele dosage in heterozygous individuals (1:3, 2:2 or 3:1) was estimated from the height ratio of the overlapping base-calling peaks manually as well as using the data acquisition and analysis software DAX (Van Mierlo Software Consultancy, Eindhoven, the Netherlands). For our study, only such markers were used, which occurred in at least two potato clones.

Statistical analysis

As described by Pajeroska-Mukhtar *et al.* (2009), an adjusted entry mean M was calculated for each of the 184 potato clones for each of the three examined traits. Furthermore, for each trait, heritability on an entry mean basis was calculated according to the procedure described by Holland *et al.* (2003).

Population structure analysis: A population structure matrix $\mathbf{Q}$ was calculated on the basis of the 31 simple sequence-repeat markers using the software STRUCTURE (Pritchard *et al.*, 2000), which gives for each clone under consideration the probability of membership in each of K subpopulations. In our investigations, the set of 184 clones was analyzed by setting K from 1 to 20 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, following the suggestion of Whitt and Buckler (2003). To determine the most probable value of K , the *ad hoc* criterion described by Evanno *et al.* (2005) was used. The K columns of the $\mathbf{Q}$ matrix add up to 1 and, thus, only the first $K-1$ columns were used as an $\mathbf{S}$ matrix.

Association analysis: We used a mixed-model approach, which is similar to the QK_{Topt} method described by Stich and Melchinger (2009), for identifying markers in the candidate genes whose additive effects are significantly associated with the trait under consideration:

$$M_i = \mu + a_q x_{iqa} + \sum_{u=1}^{K-1} v_u S_{iu} + \check{g}_i + e_i, \quad (1)$$

where M_i is the adjusted entry mean of the i th clone, μ the intercept term, a_q the allele substitution effect of the q th molecular marker, x_{iqa} the genotype indicator of the i th clone at the q th molecular marker with respect to the additive effect (x_{iqa} takes one of five values {0, 1, 2, 3, 4} depending on the dosage of one of the two present alleles), v_u the effect of the u th column of the population structure matrix $\mathbf{S}$, $\check{g}_i$ the total genetic effect of the i th clone except for the fixed effects of the model, and e_i the residual. Only the variation modeled by $\check{g}_i$ and e_i was regarded as random, where it was assumed that $\check{g}_i \sim N(0, 2\mathbf{K}\sigma_g^2)$ and $e_i \sim N(0, \mathbf{R}\sigma_e^2)$. $\mathbf{K}$ was a 184×184 matrix of kinship coefficients that define the degree of genetic covariance between all pairs of clones. We calculated the kinship coefficient K_{yz} between clone y and z on the basis of marker data according to

$$K_{yz} = \frac{U_{yz} - 1}{1 - T} + 1, \quad (2)$$

where U_{yz} is the proportion of marker loci with shared variants between clone y and z , and T is the conditional probability that marker alleles are alike in state, given

that they are not identical by descent. For the examined values $T = 0, 0.025, \dots, 0.975$, the corresponding matrices $\mathbf{K}_T$ were obtained and negative kinship values were set to 0. The optimum T value was estimated as described by Stich *et al.* (2008), and the corresponding $\mathbf{K}_T$ matrix was used as a $\mathbf{K}$ matrix. $\mathbf{R}$ was a 184×184 matrix in which the off-diagonal elements were 0 and the diagonal elements were calculated as the square of the standard errors of the adjusted entry means. Potato clones are highly heterozygous and, thus, σ_g^2 was the total genetic variance except the variance explained by the fixed effects included in the model, and σ_e^2 was the residual variance, both estimated by REML. In order to identify the n molecular markers with significant additive effects, which were designated in the current study as additive cofactors, the above-described mixed model was used in several runs across all molecular markers similar to a multiple forward regression. For this analysis, a P -to-enter criterion of 0.005 was chosen for the applied Wald- F -test (Searle, 1971), with denominator degrees of freedom approximated by the method of Kenward and Roger (1997).

Similarly, we used subsequent runs of the model

$$M_i = \mu + \sum_{j=1}^n a_j x_{ija} + h_{q1} x_{ih_{q1}} + h_{q2} x_{ih_{q2}} + h_{q3} x_{ih_{q3}} + \sum_{u=1}^{K-1} v_u S_{iu} + \check{g}_i + e_i, \quad (3)$$

where n is the number of additive cofactors; h_{q1} , h_{q2} and h_{q3} are the bi-, tri- and tetragenic effects of the q th molecular marker, respectively (Gallais, 2003); and $x_{ih_{q1}}$, $x_{ih_{q2}}$ and $x_{ih_{q3}}$ are the genotype indicators of the i th clone at the q th molecular marker with respect to the bi-, tri- and tetragenic effects; $x_{ih_{q1}}$, $x_{ih_{q2}}$ and $x_{ih_{q3}}$ take one of two values {0, 1} depending on whether the clone under consideration belongs to the corresponding heterozygote class (1) or not (0), to identify using a P -to-enter criterion of 0.005 in the Wald- F -test the m molecular markers with significant dominance effects. These m molecular markers were designated in the current study as dominance cofactors. As in the previous model, only variation modeled by $\check{g}_i$ and e_i was regarded as random.

Two different approaches were applied to identify epistatic interactions. In the mixed-model approach, we performed for all r pairs of molecular markers a Wald- F -test to examine the significance of additive–additive (AA: aa), additive–dominance (AD: $ah_1, ah_2, ah_3, h_1a, h_2a, h_3a$) or dominance–dominance (DD: $h_1h_1, h_1h_2, h_1h_3, h_2h_1, h_2h_2, h_2h_3, h_3h_1, h_3h_2, h_3h_3$) epistatic interaction effects using the following model:

$$M_i = \mu + AA_r \mathbf{x}_{iAA_r} + AD_r \mathbf{x}_{iAD_r} + DD_r \mathbf{x}_{iDD_r} + \sum_{j=1}^n a_j x_{ija} + \sum_{k=1}^m (h_{k1} x_{ih_{k1}} + h_{k2} x_{ih_{k2}} + h_{k3} x_{ih_{k3}}) + \sum_{u=1}^{K-1} v_u S_{iu} + \check{g}_i + e_i, \quad (4)$$

where n and m are the number of additive and dominance cofactors, respectively, and $\mathbf{x}_{iAA_r}$, $\mathbf{x}_{iAD_r}$ and $\mathbf{x}_{iDD_r}$ the incidence vectors relating the genotypic information of the i th clone at the r th pair of molecular markers to AA, AD and DD epistatic interaction effects, respectively. $\mathbf{X}_{AA_r}$, $\mathbf{X}_{AD_r}$ and $\mathbf{X}_{DD_r}$ were calculated as direct

products of the corresponding combinations of $\mathbf{x}_a$, $\mathbf{x}_{h_1}$, $\mathbf{x}_{h_2}$ and $\mathbf{x}_{h_3}$ of the marker pair under consideration. Only variation modeled by $\check{g}_i$ and e_i was regarded as random.

For the residual approach, which was based on the method proposed by Aulchenko *et al.* (2007), we used the following model:

$$M_i = \mu + \sum_{j=1}^n a_j x_{ija} + \sum_{k=1}^m \times (h_{k1} x_{ih_{k1}} + h_{k2} x_{ih_{k2}} + h_{k3} x_{ih_{k3}}) + \sum_{u=1}^{K-1} v_u S_{iu} + \check{g}_i + e_i, \quad (5)$$

to calculate e_i , where only variation modeled by $\check{g}_i$ and e_i was regarded as random. These residuals were then used to examine for all r pairs of molecular markers the significance of *AA*, *AD* and *DD* epistatic interaction effects using the following linear model:

$$e_i = \mu + AA_r \mathbf{x}_{iAA_r} + AD_r \mathbf{x}_{iAD_r} + DD_r \mathbf{x}_{iDD_r} + s_i, \quad (6)$$

where s_i is the residual. For those pairs of molecular markers, for which significant epistatic interactions have been observed using the residual approach, the effects of epistatic interactions as well as the proportion of the explained genotypic variance h_g^2 were calculated using the above-described mixed-model approach (cf. Aulchenko *et al.*, 2007). h_g^2 was computed in our study based on the relative reduction in σ_g^2 when the corresponding terms were added to the model.

In order to compare the distribution of the P -values of the two above-described approaches for identifying epistatic interactions, the mean of squared differences (MSD) between observed and expected P -values was calculated as described by Stich and Melchinger (2009). Multiple testing was considered in our study by using a significance threshold that was derived from computer simulations as described below.

Power simulations: In order to examine the statistical power to detect *AA*, *AD* and *DD* epistatic interaction effects using our data set, the power simulations were based on the genotypic data of the 184 potato clones from which phenotypic data were simulated. We considered one pair of molecular markers for each replication. For each of these two markers, one additional marker was simulated, which has the same allele frequency as the original marker, but shows a predefined level of LD (D' ; Lewontin, 1964) to the original marker. The two simulated markers were then used to calculate the genotypic as well as phenotypic values of the 184 potato clones, and the two original markers were used for the test of association. Both procedures will be outlined in detail below.

In each of the three examined scenarios, we assigned the effects of one of the three types of epistatic interactions *AA*, *AD* or *DD* the value 1 (Appendix). Therewith and using the information about the allele configuration present at each of the two simulated loci, we calculated the genotypic value of each of the 184 potato clones. With the variance of the simulated 184 genotypic values σ_g^2 as well as the predefined proportion of the phenotypic variance explained by the epistatic interaction under consideration h_p^2 (0.01, 0.05, 0.1, 0.15, 0.25, 0.5, 0.75, 1), we

calculated for each of the 184 clones a phenotypic value by adding a realization from a normally distributed variable $N(0, (\frac{1-h_p^2}{h_p^2} \sigma_g^2))$ to the genotypic value. Using this procedure, we scale the simulated noise in such a way that the epistatic interaction under consideration explains a predefined level of the phenotypic variance h_p^2 . Owing to the high computational burden of the above-described mixed-model approach, we used only the residual approach for the power simulations. Based on this method we calculated a P -value for the association of the original marker pair and the simulated phenotypic values. This approach was then replicated for all possible molecular marker pairs of the original genotypic data.

In order to calculate the significance level that was used for the power simulations but also for the above-described association analysis of the empirical data, we simulated a random distributed variable and examined for all possible marker pairs the association of their *AA*, *AD* and *DD* effects with the simulated data. This was replicated 100 times. The nominal type-I error rate α was calculated separately for *AA*, *AD* and *DD* as the average of the 1×10^{-6} quantile of the P -values across the 100 replicates. Thereby the empirical type-I error rate α^* was set to 1×10^{-6} . The statistical power to detect *AA*, *AD* or *DD* epistatic interactions, $1-\beta^*$, was calculated as the proportion of marker pairs for which the P -value was lower than α .

All mixed-model calculations were performed using ASReml release 2.0 (Gilmour *et al.*, 2006), whereas all other analyses were performed using the statistical software R (R Development Core Team, 2011).

Results

The heritabilities of the three phenotypic traits assessed for the 184 potato clones ranged from 0.69 for MCR to 0.94 for PM, and was 0.77 for AUDPC. Pearson's correlation coefficient among the adjusted entry means of the three traits varied between -0.03 for PM versus MCR and 0.84 for AUDPC versus MCR, and was 0.50 for AUDPC versus PM.

The slopes of the power curves were flat for all three types of epistatic effects for $h_p^2 < 0.1$ or $h_p^2 > 0.5$ (Figure 1), but were steep for $0.1 < h_p^2 < 0.5$. For epistatic interactions with $h_p^2 < 0.4$, the power to detect *AA* epistatic interactions was higher than that for detection of *AD* or *DD* interactions, whereas the opposite was true for epistatic interactions with $h_p^2 > 0.4$. The power to detect epistatic interactions, $1-\beta^*$, decreased with decreasing extent of LD between molecular markers and the trait-coding loci (Figure 2).

The MSD between the observed and expected P -values of the mixed-model approach ranged for the nine examined trait-epistatic effect combinations from 0.0000 to 0.0165 (Table 2). The MSD values for the residual approach were on average slightly lower and varied between 0.0003 and 0.0023. The Spearman's rank correlation coefficients between the P -values of the residual and the mixed-model approach ranged from 0.30 to 0.87, and the Spearman's rank correlation coefficients between the effects of the epistatic interactions were higher and varied between 0.62 and 0.93.

The number of additive cofactors ($\alpha = 0.005$) ranged from 3 (PM) to 5 (MCR), whereas the number of dominance

Figure 1 Power to detect additive–additive (circles), additive–dominance (triangles) and dominance–dominance (squares) epistatic interactions ($1-\beta^*$) in the examined set of tetraploid potato clones under the assumption of complete LD between the molecular marker loci and the two corresponding trait-coding loci, which interact epistatically. The empirical type-I error rate α^* was 1×10^{-6} . h_p^2 is the proportion of the phenotypic variance explained by the epistatic interaction. LD, linkage disequilibrium.

cofactors was lower and varied between 0 (PM) and 2 (AUDPC) (Table 3). The proportion of the genotypic variance explained by all additive cofactors simultaneously varied between 50.8% (MCR) and 74.1% (AUDPC), and between 3.3% (MCR) and 4.5% (AUDPC) for dominance cofactors. The number of significant ($\alpha^* = 1 \times 10^{-6}$) epistatic interactions was lowest for PM (3) and highest for MCR (32) (Table 4). By contrast, the proportion of the genotypic variance explained by all epistatic interactions simultaneously was lowest for AUDPC (3.6%) and highest for PM (16.5%). For none of the three traits were AA epistatic interactions identified, whereas the number of DD epistatic interactions was highest, with 35 interactions identified. The significant epistatic interactions were not equally distributed across the examined candidate genes (Figure 3). The candidate locus BA81115t3 showed the highest number of significant ($\alpha^* = 1 \times 10^{-6}$) epistatic interactions, whereas the candidate loci GP321, St4cl and BA213c14t7 were involved in no significant epistatic interactions.

Discussion

Comparison of methods for detecting epistatic interactions

One of the objectives of this research was to detect epistatic interactions in a set of tetraploid potato clones. Owing to the fact that the applied genotyping strategy allows the inference of the allele dosage of the examined molecular markers, a two-locus model requires eight main effects as well as 16 epistatic interaction effects. Regarding these markers as fixed in the statistical

analyses leads to a reduction of the error degrees of freedom and, thus, a reduction of the power to detect epistatic interactions, $1-\beta^*$.

Alternatively, the allele effects and their interactions could be regarded as random and their significance could be assessed using a likelihood ratio test. Association mapping experiments will then examine whether the explained variance of the molecular markers under consideration is 0. The use of likelihood ratio methods for constructing tests for non-zero variance components is a non-standard problem because the null hypothesis that such a variance component is 0 places the true value of the variance parameters on the boundary of the parameter space defined by the alternative hypothesis (Stram and Lee, 1994). Therefore, the likelihood ratio test is in this situation no valid test (Stram and Lee, 1994). This is in accordance with our observation that for an approach with random allele effects the P -values strongly deviated from a uniform distribution (data not shown), which is in accordance with earlier results (Crainiceanu and Ruppert, 2004). Therefore, the allele effects were regarded as fixed in our study.

Mixed-model approaches are currently the method of choice for analysing association mapping experiments (Malosetti *et al.*, 2007). This is because such approaches closely adhere to the nominal type-I error rate in addition to having a high power to detect marker–phenotype associations (for example, Yu *et al.*, 2006). Therefore, in the current study we used a mixed-model approach for detecting epistatic interactions in a set of tetraploid potato clones.

Based on the assumption that only a very small subset of the examined two-way epistatic interactions between the 22 candidate loci are significant, the MSD between observed and expected P -values can be used to evaluate the adherence to the nominal α level (Stich and Melchinger, 2009). The low MSD values observed in our analyses suggested that the mixed-model approach will adhere to the nominal α level (Table 2). However, the drawback of mixed-model approaches is their prolonged computing time (Kang *et al.*, 2008). Nevertheless, any suggestion to reduce the computational burden by limiting statistical testing to only those marker–phenotype associations with significant main effects are probably ill-advised (Phillips, 2008) and, thus, were not followed in our study. This is because epistatic interactions with the largest relative effect size will be those with small main effects. In addition, epistasis is frequently detected in the absence of main effects (Xu and Jia, 2007; Li *et al.*, 2010).

Another way of reducing the computational burden is to use the residual approach instead of a mixed-model approach, which were compared in our study. The mixed-model approach takes about 12 h to examine all two-locus epistatic interactions for one phenotypic trait using an Intel 2.4-GHz Dual Core CPU, whereas the residual approach was considerably faster, taking approximately 20 min. The Spearman's rank correlation coefficient between the P -values of the residual and the mixed-model approach ranged from 0.30 to 0.87 (Table 2). The correlation coefficient between the effects of the epistatic interactions were higher and varied between 0.62 and 0.93. These correlation coefficients were in the range of earlier empirical comparisons of different models for association mapping of QTL with additive

Figure 2 Power to detect additive–additive (top), additive–dominance (left) and dominance–dominance (right) epistatic interactions ($1-\beta^*$) in the examined set of tetraploid potato clones based on an empirical type-I error rate α^* of 1×10^{-6} , where h_p^2 is the proportion of the phenotypic variance explained by the epistatic interaction, and D' the extent of LD between the molecular marker loci and the two corresponding trait-coding loci, which interact epistatically. LD, linkage disequilibrium.

Table 2 MSD between observed and expected P -values for two approaches to identify additive–additive (AA), additive–dominance (AD) and dominance–dominance (DD) epistatic interactions, as well as Spearman's rank correlation coefficient, ρ , between the P -values, as well as the epistatic effects inferred by these two methods for the traits AUDPC and MCR for late blight infection and PM

Trait	Mixed-model approach			Residual approach			ρ			ρ		
	MSD			MSD			P values			Epistatic effects		
	AA	AD	DD	AA	AD	DD	AA	AD	DD	AA	AD	DD
AUDPC	0.0066	0.0132	0.0003	0.0004	0.0023	0.0004	0.35	0.47	0.58	0.70	0.73	0.81
PM	0.0034	0.0042	0.0000	0.0011	0.0119	0.0003	0.67	0.83	0.87	0.87	0.89	0.93
MCR	0.0082	0.0165	0.0026	0.0010	0.0009	0.0021	0.30	0.41	0.47	0.62	0.67	0.75

Abbreviations: AUDPC, area under the disease-progress curve; MCR, maturity-corrected resistance; MSD, mean of squared differences; PM, plant maturity.

effects (Stich and Melchinger, 2009). Furthermore, we observed for six of the examined nine trait–epistatic effect combinations lower MSD values for the residual approach than for the mixed-model approach (Table 2). This finding suggested that the residual approach, in comparison with the mixed-model approach, does not lead to an increased rate of false-positives. Another important criterion to compare different association mapping approaches is the statistical power. Owing to the high computational burden of the mixed-model approach, we were not able to compare both methods with respect to their power to detect epistatic interac-

tions. However, the results of Aulchenko *et al.* (2007) suggested that the power to detect the marker–phenotype associations of the residual approach was in the worst scenario, with approximately 96% of the power of the mixed-model approach only slightly inferior than the latter. In consideration of these facts, we used for all further analyses the residual approach.

Power simulations for the set of tetraploid potato clones
Information on the power to detect epistatic interactions in an association mapping context is required, on the one

Table 3 Polymorphisms in candidate loci selected as additive (A) or dominance (D) cofactors ($\alpha = 0.005$) for the three traits examined in our study: AUDPC and MCR for late blight infection and PM

A				D					
Polymorphism	Alleles	$a \pm s.e.$	h_g^2	Polymorphism	Alleles	$h_1 \pm s.e.$	$h_2 \pm s.e.$	$h_3 \pm s.e.$	h_g^2
AUDPC									
StAOS2 SNP692	G/C	14.3 ± 2.0	58.2	BA213c14t7 SNP218	A/C	17.1 ± 4.0	8.1 ± 5.0	6.1 ± 14.0	2.4
StAOS2 SNP691	A/G	-7.3 ± 2.0	49.1	GP122 SNP5	C/T		24.0 ± 13.3	20.5 ± 6.2	0.2
S1b3 SNP197	T/C	6.5 ± 1.6	6.9						
St4Cl SNP3	C/T	-6.8 ± 1.7	5.9						
Simultaneous fit			74.1						4.5
PM									
StSGT1-1 SNP169	C/T	0.22 ± 0.04	27.2						
239E4left SNP189	G/T	-0.40 ± 0.11	24.1						
StAOS2 SNP691	A/G	-0.35 ± 0.07	20.4						
Simultaneous fit			60.1						
MCR									
StAOS2 SNP692	G/C	3.33 ± 0.35	36.2	StSGT1-1 SNP414	A/G	-1.40 ± 1.5	-3.10 ± 1.0	-2.06 ± 1.5	2.2
StAOS1 SNP2	G/A	2.17 ± 0.57	10.1						
StAOS2 SNP900	A/T	-1.17 ± 0.38	7.3						
GP122 SNP5	C/T	-3.20 ± 0.97	4.8						
S1b3 SNP145	C/T	0.95 ± 0.29	2.9						
Simultaneous fit			50.8						3.3

Abbreviations: AUDPC, area under the disease-progress curve; MCR, maturity-corrected resistance; PM, plant maturity. a is the additive effect of allele substitution; h_1 , h_2 and h_3 are the bi-, tri- and tetragenic (dominance) effects; s.e. indicates the standard error; and h_g^2 is the percentage of the explained genotypic variance. The effects of MCR were multiplied by 100.

hand, for proper interpretation of the findings of association mapping experiments, but also for the design of new experiments. As with the exception of the power to detect QTL by genetic background interactions (Jannink, 2007) this information was lacking, we performed such power simulations in our study. Theoretical considerations suggested that the power to detect epistatic interactions might be influenced by (i) the proportion of the phenotypic variance explained by the epistatic interaction; (ii) the type of epistatic interactions; (iii) the extent of LD between the molecular marker locus and the trait coding locus; and (iv) the population size.

Proportion of the explained phenotypic variance: A higher power $1-\beta^*$ was observed for scenarios in which the epistatic interactions explained a higher proportion of the phenotypic variance (Figure 1). The relationship between $1-\beta^*$ and h_p^2 , however, was not linear but rather S-shaped. This observation is in accordance with results from studies on the power to detect QTL with additive (for example, Charcosset and Gallais, 1996) or epistatic effects (for example, Baierl *et al.*, 2006) in linkage mapping populations.

Type of epistatic interactions: The analysis of heterozygous potato clones allows the discrimination between AA, AD and DD epistatic interactions (Cockerham, 1954). The results from our computer simulations suggested that the $1-\beta^*$ values for epistatic interactions with $h_p^2 < 0.4$ is highest for AA effects and lower for AD and DD effects (Figure 1). This finding is in good accordance with the results of Baierl *et al.* (2006), who examined h_p^2 values < 0.2 and observed in inter-cross populations that the $1-\beta^*$ values for AA epistatic interactions were slightly higher than those for the AD or DD epistatic interactions. Nevertheless, no earlier study observed that, for scenarios with $h_p^2 > 0.4$ the $1-\beta^*$ values were higher for AD

and DD than for AA epistatic interactions. This finding requires further research.

Extent of LD between the molecular marker locus and the trait-coding locus: A higher power $1-\beta^*$ was observed for scenarios with a high extent of LD between the two trait-coding loci and the corresponding molecular markers (Figure 2). This observation is in accordance with results of Kruglyak (1999) on the power to detect markers with main effects in association mapping experiments. Furthermore, we observed that the reduction of the power $1-\beta^*$ observed for a decreasing level of LD between molecular markers and trait-coding loci was not identical for the different effects of epistatic interactions (Figure 2). This finding might be explained by the factor also responsible for the above-outlined observation that in scenarios with $h_p^2 > 0.4$ the $1-\beta^*$ values were higher for AD and DD than for AA epistatic interactions. However, the reason for this observation is unclear at the moment.

Population size: Theoretical considerations as well as results from empirical studies on joint linkage and association mapping experiments (Stich *et al.*, 2007) suggest that the power $1-\beta^*$ can be increased by increasing the sample size. This is because in this case the allele as well as the interaction effects can be estimated more precisely.

However, we also want to advert that using for the detection of epistatic interactions a sample that is considerably smaller than that used in our study will considerably reduce the power $1-\beta^*$. In this case, only epistatic interactions explaining a high proportion of the genotypic variance can be detected, which, however, might not exist for the trait under consideration. Therefore, based on our results from computer simulations we recommend to use at least 150 genotypes for the detection of epistatic interactions.

Table 4 Polymorphisms in candidate loci significantly ($\alpha^* = 1 \times 10^{-6}$) associated with their additive-additive (AA), additive-dominance (AD) and dominance-dominance (DD) epistatic interactions with the three traits examined in our study using the residual approach: AUDPC and MCR for late blight infection and PM

Polymorphism 1	Alleles 1/2	Polymorphism 2	Alleles 3/4	Effects												h ² _g							
				AA				AD				DD				AA	AD	DD					
				aa	ah ₁	ah ₂	ah ₃	h ₁ a	h ₂ a	h ₃ a	h ₁ h ₁	h ₁ h ₂	h ₁ h ₃	h ₂ h ₁	h ₂ h ₂	h ₂ h ₃	h ₃ h ₁	h ₃ h ₂	h ₃ h ₃				
AUDPC	T/C	SlSGT1-2 SNP2	G/A											-40.7		2.3							1.0
	A/G	SlPEN1 SNP18	G/A											-34.1	-43.3	6.6	-20.9	-9.7	19.0				5.6
	A/G	SlPEN1 SNP19	A/T											-25.6	-34.9	3.1	-17.5	-31.0					4.3
	A/T	SlPEN1 SNP18	C/A											-34.0	-43.0	7.5	-20.9	-10.0	21.9				6.6
	A/T	SlPEN1 SNP19	A/T											-26.0	-36.2	2.3	-18.6	-33.8			Simultaneous fit	-	3.6
PM	C/G	P4h6 SNPs	G/C					2.28						-1.82	1.21	-1.03	1.61	-0.67	1.06				18.3
	A/G	P4h6 SNPs	G/C					-0.37	-1.45	0.94				-3.35									15.7
	A/T	P4h6 SNPs	G/C					0.46	-1.23	1.41				-3.25									16.9
																				Simultaneous fit	-	16.5	
MCR	G/A	BA229c213 SNP3	C/T			1.93	0.16		-0.06	0.79													4.5
	C/T	BA811153 SNP183	C/T			0.79	-1.46		2.00	0.19													3.8
	C/T	SlAOS1 SNP1	G/A			1.09	0.04	-0.18	1.20	-3.07													3.2
	C/T	SlSGT1-2 SNP1	A/G			1.40	0.38	0.29	1.05	-2.28													3.4
	C/T	SlPEN1 SNP3	G/A			0.16	0.42	0.33	0.87	-3.26													2.5
	C/T	SlSGT1-1 SNP309	C/T												0.32			7.96	-4.28				4.2
	G/A	BA811153 SNP192	G/C												-13.27							4.9	
	G/A	BA811153 SNP183	C/T																			3.2	
	G/A	BA811153 SNP192	G/C																				3.4
	G/A	SlEDS1 SNP202	G/C																				2.9
	G/A	P4h6 SNP13	A/G																			4.0	
	A/G	Slb3 SNP175	A/G																			3.6	
	G/C	GP122 SNP13	G/A																			3.5	
	G/C	SlPEN1 SNP18	G/A																			4.4	
	A/G	Slp23 SNP1	C/A																				1.0
	A/G	Slp23 SNP7	G/A																				0.0
	C/A	Slp23 SNP1	C/A																				0.7
	C/A	SlSGT1-2 SNP2	G/A																				0.6
	A/G	SlSGT1-1 SNP309	C/T																				3.4
	A/G	SlSGT1-2 SNP2	G/A																				1.9
C/A	SlEDS1 SNP409	G/A																				1.5	
C/A	SlPAD4-1 SNP157	C/T																				3.0	
C/A	P4h6 SNP10	T/A																				1.7	
C/A	SlPEN1 SNP12	A/G																				3.2	
C/T	SlSGT1-2 SNP2	G/A																				3.5	
A/T	SlSGT1-2 SNP2	G/A																				1.7	
G/C	SlSGT1-1 SNP305	T/C																				3.1	
G/C	SlSGT1-1 SNP309	C/T																				2.1	

Table 4 (Continued)

Polymorphism 1	Alleles 1/2	Polymorphism 2	Alleles 3/4	Effects															h ² _g																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																		
				AA						AD						DD			AA	AD	DD																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																
				aa	ah ₁	ah ₂	ah ₃	h ₁ a	h ₂ a	h ₃ a	h ₁ h ₁	h ₁ h ₂	h ₁ h ₃	h ₂ h ₁	h ₂ h ₂	h ₂ h ₃	h ₃ h ₁	h ₃ h ₂				h ₃ h ₃																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																															
SlSGT1-2 SNP2	G/A	P4h6 SNP2	G/A																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																		

Abbreviations: AUDPC, area under the disease-progress curve; MCR, maturity-corrected resistance; PM, plant maturity. aa , ah_1 or ah_2 , as well as h_1a , are the interaction effects of the additive effects of allele substitution a , and the bi-, tri- and tetragenic (dominance) effects h_1 , h_2 and h_3 . h_g^2 is the percentage of the explained genotypic variance. The effects of MCR were multiplied by 100.

Figure 3 Map of pairwise epistatic interactions for the three traits: AUDPC (red) and MCR (blue) for late blight infection and PM (green). The colored boxes around the names of candidate loci indicate that polymorphisms in the respective candidate genes were selected as additive (solid) or dominance (dashed) cofactors. The arrows between candidate loci indicate that polymorphisms in the genes show significant ($\alpha^* = 1 \times 10^{-6}$) additive-additive (solid), additive-dominance (dashed) or dominance-dominance (dotted) epistatic interactions using the residual approach. Details of the candidate loci can be found in Table 1. AUDPC, area under the disease-progress curve; MCR, maturity-corrected resistance; PM, plant maturity.

In summary, the results of our computer simulations suggested that, if molecular markers are available, which are in high LD ($D' > 0.9$) with the two trait-coding loci, even epistatic interactions, which explain only 5–10% of the phenotypic variance, can be detected with a power $1 - \beta^*$ of up to 0.2, which seems to be promising for their detection. A high extent of LD between molecular markers and the two trait-coding loci can be expected in association mapping experiments either when using a very high marker density in genome-wide association mapping studies or as in our study by resequencing functional and/or positional candidate loci. Therefore, it seemed promising to detect epistatic interactions in our set of tetraploid potato clones.

Epistatic interactions in the set of tetraploid potato clones
In the current study, we identified for all three traits significant epistatic interactions using the residual approach. The applied significance levels, which were derived from resampling simulations, are expected to result in an empirical type-I error rate of 1×10^{-6} . As this significance level is slightly lower than that suggested based on Bonferroni-Holm correction (Holm, 1979), we expect that the proportion of biologically interpretable interactions among the identified epistatic interactions is high.

The number of epistatic interactions identified for the three traits differed considerably and ranged from 3 (PM) to 32 (MCR) (Table 4). The high number of identified epistatic interactions might be explained by the fact that

MCR is, in contrast to AUDPC and PM, a meta trait, for which trait values were calculated from the trait values of the other two traits. Therefore, it is expected that MCR has a higher genetic complexity than AUDPC and PM. Another reason for the low number of identified epistatic interactions for PM might be that the candidate loci of this study were selected with respect to their involvement in late blight resistance, but they are no explicit candidates for PM. Therefore, our finding suggested that the candidate loci selection in our study was successful.

Despite our identification of altogether 39 epistatic interactions for the three traits, none of the identified interactions were of the AA type (Table 4). This observation cannot be explained by differences in the significance threshold, as for all three types of epistatic interactions the nominal α level was chosen based on a resampling procedure in such a way that the empirical α^* level was 1×10^{-6} . The results of the power simulations suggested that, for those h_p^2 values expected in our experiment ($h_p^2 < 0.4$) the $1 - \beta^*$ was higher for AA than for AD or DD epistatic interactions (Figure 1). These simulations, however, are based on the assumption that the different types of epistatic interactions explain the same proportion of the variance. Owing to the fact that potato is a highly heterozygous species, the frequency of homozygous genotypes is low (cf. Supplementary Material S1) and, thus, the AA epistatic interactions are expected to explain a lower proportion of the variance than the DD epistatic interactions (cf. Falconer and Mackay, 1996). Therefore, the low proportion of the variance explained by the AA epistatic interactions explains the absence of AA epistatic interactions in our study. The same argumentation as for the absence of AA epistatic interactions can be used to explain the lower number of AD than DD epistatic interactions identified (Table 4).

We observed that the explained proportion of the epistatic interactions was considerably lower than that of the main effects (Tables 3 and 4). This finding is in accordance with earlier studies examining QTL mapping experiments (Xu and Jia, 2007) and can be explained by quantitative genetic theory (for example, Falconer and Mackay, 1996). Furthermore, we think that in association mapping studies, another reason might be the selection of candidate loci. In most cases, these genes are selected from molecular biology studies or studies involving mutant screening approaches, which, in turn, are expected to be more powerful in identifying genes with main rather than epistatic effects.

We observed that the epistatic interactions were not uniformly distributed among the 22 loci. The markers in the candidate locus BA81115t3 showed with markers in eight other candidate loci significant ($\alpha^* = 1 \times 10^{-6}$) epistatic interaction for MCR (Figure 3). This finding might be explained by the fact that this candidate locus is closely linked to a cluster of resistance gene-like sequences (Rickert *et al.*, 2003).

Despite the fact that the epistatic interactions between the molecular markers identified in our study explain only a relatively low proportion of the genotypic variance as compared with the molecular markers with significant main effects, we expect that this information is of value for most potato breeding programs in temperate climates. The use of this information on the epistatic interactions in marker-assisted breeding programs is expected to facilitate the early identification of indi-

viduals carrying specific allele combinations for the three traits examined in our study.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

This work was funded by Grant 0313114 (GABI-CONQUEST-2) from the German Federal Ministry for Education and Research under the program Genome Analysis in the Biological System of Plants (GABI) and the Max Planck Society. We thank the breeding companies Böhm-Nordkartoffel-Agrarproduktion OHG (Ebster, Germany) and SAKA-Pflanzenzucht GbR (Windeby, Germany) for providing phenotypic data. We wish to thank HP Piepho for his insight into mixed-model theory. We thank the associate editor and two anonymous reviewers for valuable suggestions.

References

- Aulchenko YS, de Koning DJ, Haley C (2007). Genome-wide 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.
- Baierl A, Bogdan M, Frommlet F, Futschik A (2006). On locating multiple interacting quantitative trait loci in intercross designs. *Genetics* **173**: 1693–1703.
- Blanc G, Charcosset A, Mangin B, Gallais A, Moreau L (2006). Connected populations for detecting quantitative trait loci and testing for epistasis: an application in maize. *Theor Appl Genet* **113**: 206–224.
- Charcosset A, Gallais A (1996). Estimation of the contribution of quantitative trait loci (QTL) to the variance of a quantitative trait by means of genetic markers. *Theor Appl Genet* **93**: 1193–1201.
- Cockerham CC (1954). An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. *Genetics* **39**: 859–882.
- Crainiceanu CM, Ruppert D (2004). Likelihood ratio tests in linear mixed models with one variance component. *J Roy Stat Soc B Met* **66**: 165–185.
- Duncan JM (1999). *Phytophthora*: an abiding threat to our crops. *Microbiol Today* **26**: 114–116.
- Eshed Y, Zamir D (1996). Less-than-additive epistatic interactions of quantitative trait loci in tomato. *Genetics* **143**: 1807–1817.
- Evanno G, Regnaut S, Goudet J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* **14**: 2611–2620.
- Falconer DS, Mackay TFC (1996). *Introduction to Quantitative Genetics*, 4th edn. Longman Group: London.
- Fry W (1978). Quantification of general resistance of potato cultivars and fungicide effects for integrated control of potato late blight. *Phytopathology* **68**: 1650–1655.
- Gallais A (2003). *Quantitative Genetics and Breeding Methods in Autopolyploid Plants*. INRA: Paris.
- Gilmour A, Gogel B, Cullis B, Thompson R (2006). *ASReml User Guide Release 2.0*. VSN International Ltd: Hemel Hempstead UK.
- Holland J (2001). Epistasis and plant breeding. *Plant Breed Rev* **21**: 27–92.
- Holland JB, Nyquist WE, Cervantes-Martinez CT (2003). Estimating and interpreting heritability for plant breeding: an update. *Plant Breed Rev* **22**: 9–112.
- Holm S (1979). A simple sequentially rejective multiple test procedure. *Scand J Stat* **6**: 65–70.

- Jannink JL (2007). Identifying quantitative trait locus by genetic background interactions in association studies. *Genetics* **176**: 553–561.
- Kamoun S, Smart C (2005). Late blight of potato and tomato in the genomics era. *Plant Dis* **89**: 692–699.
- Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ *et al.* (2008). Efficient control of population structure in model organism association mapping. *Genetics* **178**: 1709–1723.
- Kenward M, Roger J (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* **53**: 983–997.
- Kroymann J, Mitchell-Olds T (2005). Epistasis and balanced polymorphism influencing complex trait variation. *Nature* **435**: 95–98.
- Kruglyak L (1999). Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat Genet* **22**: 139–144.
- Lewontin RC (1964). The interaction of selection and linkage. I. general considerations; heterotic models. *Genetics* **49**: 49–67.
- Li L, Paulo MJ, van Eeuwijk FA, Gebhardt C (2010). Statistical epistasis between candidate gene alleles for complex tuber traits in an association mapping population of tetraploid potato. *Theor Appl Genet* **121**: 1303–1310.
- Luo Z, Hackett C, Bradshaw J, McNicol J, Milbourne D (2001). Construction of a genetic linkage map in tetraploid species using molecular markers. *Genetics* **157**: 1369–1385.
- Malosetti M, van der Linden CG, Vosman B, van Eeuwijk FA (2007). A mixed-model approach to association mapping using pedigree information with an illustration of resistance to *Phytophthora infestans* in potato. *Genetics* **175**: 879–889.
- Manicacci D, Camus-Kulandaivelu L, Fourmann M, Arar C, Barrault S, Rousselet A *et al.* (2009). Epistatic interactions between *Opaque2* transcriptional activator and its target gene *CyPPDK1* control kernel trait variation in maize. *Plant Physiol* **150**: 506–520.
- Mather D, Jinks JL (1982). *Biometrical Genetics. The Study of Continuous Variation*, 3rd edn. Chapman & Hall: New York.
- Pajeroska-Mukhtar K, Stich B, Achenbach U, Ballvora A, Lübeck J, Strahwald J *et al.* (2009). Single nucleotide polymorphisms in the *allene oxide synthase 2* gene are associated with field resistance to late blight in populations of tetraploid potato cultivars. *Genetics* **181**: 1115–1127.
- Phillips PC (2008). Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet* **9**: 855–867.
- Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.
- R Development Core Team (2011). *R: a Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria. ISBN 3-900051-07-0. <http://www.R-project.org>.
- Rickert A, Kim J, Meyer S, Nagel A, Ballvora A, Oefner P *et al.* (2003). First-generation SNP/InDel markers tagging loci for pathogen resistance in the potato genome. *Plant Biotechnol J* **1**: 399–410.
- Searle S (1971). *Linear Models*. Wiley: Chichester, Sussex, UK.
- 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, Möhring J, Piepho HP, Heckenberger M, Buckler ES, Melchinger AE (2008). Comparison of mixed-model approaches for association mapping. *Genetics* **178**: 1745–1754.
- Stich B, Yu J, Melchinger AE, Piepho HP, Utz HF, Maurer HP *et al.* (2007). Power to detect higher-order epistatic interactions in a metabolic pathway using a new mapping strategy. *Genetics* **176**: 563–570.
- Stracke S, Haseneyer G, Veyrieras JB, Geiger HH, Sauer S, Graner A *et al.* (2009). Association mapping reveals gene action and interactions in the determination of flowering time in barley. *Theor Appl Genet* **118**: 259–273.
- Stram DO, Lee JW (1994). Variance-components testing in the longitudinal mixed effects model. *Biometrics* **50**: 1171–1177.
- Thornberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES (2001). *Dwarf8* polymorphisms associate with variation in flowering time. *Nat Genet* **28**: 286–289.
- Whitt SR, Buckler ES (2003). *Plant Functional Genomics: Methods and Protocols*, Vol. 236, chapter Using natural allelic diversity to evaluate gene function. Humana Press Inc.: Totowa, NJ, pp 123–140.
- Xu S, Jia Z (2007). Genomewide analysis of epistatic effects for quantitative traits in barley. *Genetics* **175**: 1955–1963.
- Yan J, Kandianis CB, Harjes CE, Bai L, Kim EH, Yang X *et al.* (2010). Rare genetic variation at *Zea mays crtRB1* increases beta-carotene in maize grain. *Nat Genet* **42**: 322–327.
- 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.

Appendix

Expected genotypic values for a two-locus scenario of a tetraploid species, where alleles 1 and 2 occur at locus 1 (cf. Table 4), and alleles 3 and 4 at locus 2. μ is the general mean; a the additive effect; and h_1 , h_2 and h_3 are the bi-, tri- and tetragenic effects. The indices $l1$ or $l2$ indicate the main effects of locus 1 and 2, respectively

Allele configuration at locus 1	Allele configuration at locus 2				
	3333	3334	3344	3444	4444
1111	$\mu + (0a)_{l1} + (0a)_{l2} + aa$	$\mu + (0a)_{l1} + (1a+h_1)_{l2} - ah_1$	$\mu + (0a)_{l1} + (2a+h_2)_{l2} - ah_2$	$\mu + (0a)_{l1} + (3a+h_3)_{l2} - ah_3$	$\mu + (0a)_{l1} + (4a)_{l2} - aa$
1112	$\mu + (1a+h_1)_{l1} + (0a)_{l2} - h_1a$	$\mu + (1a+h_1)_{l1} + (1a+h_1)_{l2} + h_1h_1$	$\mu + (1a+h_1)_{l1} + (2a+h_2)_{l2} + h_1h_2$	$\mu + (1a+h_1)_{l1} + (3a+h_3)_{l2} + h_1h_3$	$\mu + (1a+h_1)_{l1} + (4a)_{l2} + h_1a$
1122	$\mu + (2a+h_2)_{l1} + (0a)_{l2} - h_2a$	$\mu + (2a+h_2)_{l1} + (1a+h_1)_{l2} + h_2h_1$	$\mu + (2a+h_2)_{l1} + (2a+h_2)_{l2} + h_2h_2$	$\mu + (2a+h_2)_{l1} + (3a+h_3)_{l2} + h_2h_3$	$\mu + (2a+h_2)_{l1} + (4a)_{l2} + h_2a$
1222	$\mu + (3a+h_3)_{l1} + (0a)_{l2} - h_3a$	$\mu + (3a+h_3)_{l1} + (1a+h_1)_{l2} + h_3h_1$	$\mu + (3a+h_3)_{l1} + (2a+h_2)_{l2} + h_3h_2$	$\mu + (3a+h_3)_{l1} + (3a+h_3)_{l2} + h_3h_3$	$\mu + (3a+h_3)_{l1} + (4a)_{l2} + h_3a$
2222	$\mu + (4a)_{l1} + (0a)_{l2} - aa$	$\mu + (4a)_{l1} + (1a+h_1)_{l2} + ah_1$	$\mu + (4a)_{l1} + (2a+h_2)_{l2} + ah_2$	$\mu + (4a)_{l1} + (3a+h_3)_{l2} + ah_3$	$\mu + (4a)_{l1} + (4a)_{l2} + aa$

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