Theor Appl Genet (2005) 110: 561–566
DOI 10.1007/s00122-004-1872-9

ORIGINAL PAPER

H. P. Piepho

Statistical tests for QTL and QTL-by-environment effects in segregating populations derived from line crosses

Received: 13 May 2004 / Accepted: 23 October 2004 / Published online: 18 January 2005

Abstract Quantitative trait locus (QTL) studies in plants frequently employ phenotypic data on a population of lines (doubled haploid lines, recombinant inbred lines, etc.) tested in multiple environments. An important feature of such data is the genetic correlation among observations on the same genotype in different environments. Detection of QTL-by-environment interaction requires tests which take this correlation into account. In this article, a comparison was made of the properties of several such tests by means of simulation. The results indicate that a split-plot analysis of variance (ANOVA), being an approximate method, tends to be too liberal under departures from the Huynh-Feldt condition. A standard two-way ANOVA, which ignores genetic correlation, yields inappropriate tests and should be avoided. In contrast, mixed model approaches as well as univariate and multivariate repeated-measures ANOVA yield valid results. This supports the use of a flexible mixed model framework in more complex settings, which are difficult to tackle by repeated-measures ANOVA.

Introduction

Quantitative trait locus (QTL) studies frequently employ phenotypic data obtained on a population of lines [doubled-haploid (DH) lines, recombinant inbred lines (RILs), etc.] tested in multiple environments. The detection of QTL main effects and of QTL-by-environment interaction in such studies requires appropriate statistical tests. A simple method is to set up a linear model with QTL effects and the corresponding QTL-by-environment interaction in the linear predictor, assuming independent residual errors (Tinker and Mather 1995; Pillen et al. 2003). This approach does not account for the fact that the same genotypes are grown in each environment, which introduces genetic correlation among phenotypic observations on the same genotype. While part of this correlation will be explicitly modelled by QTL and QTL-by-environment effects, some unexplained genetic correlation will remain and affect the residual variation (Piepho 2000; Piepho and Pillen 2004). The present paper explores the effect of ignoring residual genetic correlation by means of simulation. In addition, a comparison is made of the different approaches accounting for genetic correlation, with a focus on the least squares approach to QTL mapping (Haley and Knott 1992), which constitutes an approximation of the full-likelihood approach (Lynch and Walsh 1998).

One option is to use a split-plot model, in which each genotype is considered to be a main plot, while observations from different environments play the role of split plots (Utz and Melchinger 1996). The resulting analysis will provide exact tests, providing the residual variance-covariance matrix meets the Huynh-Feldt (sphericity, circularity) condition, which requires that the variance of a difference is the same for each pair of observations from a common genotype (Winer et al. 1991). The Huynh-Feldt condition is met, for example, when the genetic variances are the same in each environment and all pairwise genetic covariances are identical. This structure is also known as the compound symmetry model (Wolfinger 1996). When the genetic variance-covariance structure departs from the Huynh-Feldt condition, the split-plot analysis is at best approximately valid.

In the case of departure from the Huynh-Feldt condition, one may resort to repeated-measures analysis of variance (ANOVA) procedures, in which observations from multiple environments are regarded as different traits (Jiang and Zeng 1995; Falconer and Mackay 2001). There are basically two options. One is to compute a standard ANOVA table and adjust the degrees of

This paper is dedicated to Prof. Dr. H. F. Utz on the occasion of his 65th birthday.

Communicated by H. C. Becker

H. P. Piepho
Bioinformatics Unit, Universität Hohenheim,
Fruwirthstrasse 23, 70599 Stuttgart, Germany
E-mail: piepho@uni-hohenheim.de

freedom using Box's method (Geisser and Greenhouse 1958; Huynh and Feldt 1976; Winer et al. 1991). The other is to employ multivariate ANOVA (MANOVA) tests (Winer et al. 1991). The repeated-measures ANOVA approach leaves the variance-covariance matrix of the data unstructured. This may be contrasted to a mixed model approach, in which a specific model is assumed for the variance-covariance structure, possibly allowing some gain in efficiency.

The present paper compares all of these methods by means of simulation. For simplicity, simulations are based on a single-point analysis. Results are equally relevant to interval mapping approaches, to more complex models including effects for gene-by-gene interaction (epistasis, etc.) and to other types of population analysed either by maximum likelihood or a least squares approximation (Lynch and Walsh 1998). The objective is to investigate whether the mixed model framework, which offers great flexibility to accommodate complex designs, is competitive to other methods restricted to simpler settings.

Theory

For illustration, we consider a doubled-haploid (DH) population and a single-point analysis, which may be seen as a special case of interval mapping when a QTL is located exactly at a marker. The design is assumed to be laid out in randomised complete blocks. Environments are taken to be a fixed factor. Analysis will be based on genotype means per environment. The linear model, ignoring residual genetic effects of undetected QTLs, is

$$y_{ij} = \mu + x_i\alpha + u_j + x_i a_j + e_{ij}, \tag{1}$$

where y_{ij} is the phenotypic mean of the i-th DH line in the j-th environment, μ is the general mean, x_i is the -marker covariate ($=0$ for the one parent, $=1$ for the other parent), α is the QTL main effect (fixed), u_j is the environmental main effect (fixed), a_j is the QTL-by-environment interaction effect in the j-th environment (fixed) and e_{ij} is the random residual error associated with a mean y_{ij}.

With minor modifications, the model may be extended to other types of populations, to (composite) interval mapping and to more complex forms of gene action (epistasis, etc.). All effects except that of the residual e_{ij} are fixed. In order to account for the undetected QTLs, which induce a genetic correlation, we assume that the residual e_{ij} comprises both environmental errors as well as residual environment-specific genetic effects (Piepho 2000; Piepho and Pillen 2004). The crucial point here is that the residual genetic effects of the same genotype in different environments are not necessarily independent, for example, when the same genes are expressed across environments. As a result, observations y_{ij} of the same genotype i in different environments ($j=1, \ldots, J$) are potentially correlated, i.e., the variance-covariance matrix

$$\mathbf{V} = \mathrm{var}\,(y_{i1},\ y_{i2},\ \ldots,\ y_{iJ}) = \mathrm{var}\,(e_{i1},\ e_{i2},\ \ldots,\ e_{iJ}) \tag{2}$$

may have non-zero covariances on the off-diagonal. Eq. 1 in conjunction with the variance-covariance structure in Eq. 2 is taken to be the correct model, and distributional properties of different test statistics are considered under this assumed model. The following four methods are used for analysis:

1. Simple two-way ANOVA, assuming that $\mathbf{V} = \mathbf{I}_J\sigma^2$, where $\mathbf{I}_J$ is the identity matrix of dimension J. For example, when $J=3$, the model is
$$\mathbf{V} = \begin{pmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{pmatrix}, \tag{3}$$
i.e., all covariances are equal to zero.
2. Split-plot type ANOVA, assuming that $\mathbf{V}$ has the compound symmetry structure, i.e., $\mathbf{V}=\mathbf{R}\sigma^2$ with correlation matrix $\mathbf{R}=\rho\mathbf{J}_J + (1-\rho)\mathbf{I}_J$, where ρ is a common correlation, and $\mathbf{J}_J$ is a J-by-J matrix of ones everywhere. For example, when $J=3$, the model is
$$\mathbf{V} = \sigma^2\begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix} = \begin{pmatrix} \sigma^2 & \rho\sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \rho\sigma^2 & \sigma^2 \end{pmatrix}. \tag{4}$$
This analysis corresponds to a decomposition of the random term e_{ij} as
$$e_{ij} = g_i + h_{ij}, \tag{5}$$
where g_i is a random genetic main effect, which induces a correlation among errors e_{ij} of the same genotype, and h_{ij} is an independent homoscedastic residual. In this analysis, a genotype is treated as a main plot, and g_i corresponds to a main-plot error, while the performance in the j-th environment constitutes a split plot and h_{ij} is a split-plot error term. The ANOVA is depicted in Table 1.
3. Repeated-measures ANOVA, which regards genotypes as subjects and assumes that $\mathbf{V}$ is unstructured. This analysis will yield the same test for QTL main effects as the split-plot analysis depicted in Table 1, and the test will be exact (Geisser and Greenhouse 1958). There are two options for testing the QTL-by-environment interaction: (1) The ANOVA F-statistic F_3 in Table 1 is referred to an F-distribution with $(J-1)\epsilon$ and $(J-1)(N-2)\epsilon$ degrees of freedom (df), where
$$\varepsilon = (J-1)^{-1}\frac{\left(\sum_{j=1}^{J}\lambda_j\right)^2}{\sum_{j=1}^{J}\lambda_j^2} \quad \left[(J-1)^{-1} \leqslant \varepsilon \leqslant 1\right] \tag{6}$$
and λ_j ($j = 1,.., J$) are the latent roots of $\mathbf{V}-J^{-1}1_J1'_J\mathbf{V}$. The Box's correction ϵ may be estimated from the data, as suggested by Geisser and Greenhouse (1958) and Huynh and Feldt (1976). The test is an approximate one. (2) The interaction is tested by a multivariate ANOVA for repeated measures (Winer et al. 1991). When there is only one QTL-by-environment

Table 1 ANOVA table for split-plot analysis of the multi-environment test of a QTL mapping population (adapted from Geisser and Greenhouse 1958)

Source	*df*[a]	Sum of squares	*F*
Environments	$(J-1)$	Q_1	$F_1 = (N-2)\ Q_1/Q_5$
Marker (class)	1	Q_2	$F_2 = (N-2)\ Q_2/Q_3$
Genotypes within marker[b]	$(N-2)$	Q_3	
Environment × marker	$(J-1)$	Q_4	$F_3=(N-2)\ Q_4/Q_5$
Environment × genotypes within marker (residual)[c]	$(J-1)(N-2)$	Q_5	

[a] J, Number of environments; N, total number of genotypes (DH lines, RILs or others)
[b] Corresponds to main-plot error
[c] Corresponds to split-plot error

interaction effect per environment, as in the example considered in simulations, the test is Hotelling's T^2, which is exact. Incidentally, Box's correction factor ϵ is a measure of the departure from the Huynh-Feldt condition, which requires that $v_{jj} + v_{j'j'} - 2\,v_{jj'} = 2\lambda$ for all pairs of environments, where $v_{jj'}$ is the jj'-th element of $\mathbf{V}$ (Winer et al. 1991), in which case $\epsilon = 1$. The smaller the ϵ, the larger the departure from the Huynh-Feldt condition, which is required for the split-plot F-test of the QTL-by-environment interaction to be valid.

4. Mixed model analysis, imposing a model on $\mathbf{V}$ (factor-analytic, heterogeneous compound symmetry, compound symmetry or other) (see Table 2; Wolfinger 1996). The special case of an unstructured model for $\mathbf{V}$ corresponds to a repeated-measures MANOVA. By way of example, for $J=3$ the factor-analytic model with homoscedastic residual is

$$\mathbf{V} = \phi\phi' + \mathbf{I}_3\sigma^2 = \begin{pmatrix} \phi_1^2+\sigma^2 & \phi_1\phi_2 & \phi_1\phi_3 \\ \phi_2\phi_1 & \phi_2^2+\sigma^2 & \phi_2\phi_3 \\ \phi_3\phi_1 & \phi_3\phi_2 & \phi_3^2+\sigma^2 \end{pmatrix}, \tag{7}$$

where $\phi = (\phi_1, \phi_2, \phi_3)$, while the heterogeneous compound symmetry model can be expressed as

$$\mathbf{V} = \mathbf{S}^{-1}\mathbf{R}\mathbf{S}^{-1} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \rho\sigma_1\sigma_3 \\ \rho\sigma_2\sigma_1 & \sigma_2^2 & \rho\sigma_2\sigma_3 \\ \rho\sigma_3\sigma_1 & \rho\sigma_3\sigma_2 & \sigma_3^2 \end{pmatrix}, \tag{8}$$

where $\mathbf{S}$ is diag(σ_1, σ_2, σ_3) and $\mathbf{R}$ is the same correlation matrix as that for the compound symmetry model of the split-plot ANOVA. The Kenward-Roger (1997) method is used to approximate the denominator degrees of freedom.

Method 1 ignores genetic correlation altogether. Method 2 imposes the rather restrictive compound symmetry structure for the variance-covariance matrix $\mathbf{V}$, while methods 3 and 4 allow more flexibility. Method 3 is completely general, while method 4 allows models intermediate in complexity between unstructured and compound symmetry and includes these two extremes as special cases.

Table 2 Empirical type-I error rates and power of tests for QTL main effects and for QTL-by-environment interaction; single-point analysis; three environments and 250 DH per marker class, 10,000 simulation runs; tests for a nominal significance level of 5%, conditional on the putative QTL position. The simulated type-I error rate is in bold type when above +0.02 of the nominal level of 0.05, and the corresponding power results are italicised. This was done to indicate that the tests are invalid in these cases

Hypothesis[a]	Rejection rate								
	QTL main effect			QTL interaction					
	ANOVA[b]	Split-plot[c, d]	Mixed[e]	ANOVA[b]	Split-plot[c]	G–G[d]	H–F[d]	MANOVA[d]	Mixed[e]
True model: factor-analytic, $\mathbf{V} = \phi\phi' + \mathbf{I}_3\sigma^2$ with $\phi = (\phi_1, \phi_2, \phi_3)$ and $\phi_1=1$, $\phi_2=3$, $\phi_3=5$, $\sigma^2=1$ ($\epsilon=0.610$):									
H_0	**0.181**	0.047	0.047	0.004	**0.073**	0.049	0.049	0.051	0.051
H_{A1}	*0.659*	0.403	0.403	0.226	*0.636*	0.567	0.567	0.531	0.533
H_{A2}	*0.650*	0.394	0.395	0.011	*0.142*	0.101	0.101	0.372	0.374
True model: heterogeneous compound symmetry, $\mathbf{S}=$diag(σ_1, σ_2, σ_3), $\mathbf{R}=\rho\ \mathbf{J}_3 + (1-\rho)\mathbf{I}_3$, $\mathbf{V}=\mathbf{S}^{-1}\ \mathbf{R}\mathbf{S}^{-1}$, $\sigma_1=1$, $\sigma_2=3$, $\sigma_3=5$, $\rho=0.9$ ($\epsilon=0.779$):									
H_0	**0.184**	0.054	0.054	0.005	**0.074**	0.051	0.051	0.052	0.051
H_{A1}	*0.673*	0.429	0.429	0.253	*0.647*	0.583	0.583	0.488	0.488
H_{A2}	*0.677*	0.429	0.429	0.148	*0.149*	0.108	0.108	0.346	0.345
True model: compound symmetry, $\mathbf{R} = \rho\ \mathbf{J}_3 + (1-\rho)\mathbf{I}_3$, $\mathbf{V} = \mathbf{R}\sigma^2$, $\rho=0.9$, $\sigma^2=16$ ($\epsilon=1.000$):									
H_0	**0.241**	0.053	0.053	0.000	0.050	0.050	0.050	0.050	0.050
H_{A1}	*0.367*	0.122	0.122	0.000	0.741	0.741	0.740	0.740	0.741
H_{A2}	*0.376*	0.123	0.123	0.000	0.155	0.155	0.155	0.154	0.154

[a] $\eta_j = \alpha + a_j$, where α is the QTL main effect and a_j is the QTL-by-environment interaction in the j-th environment; H_0: $\eta_1=\eta_2=\eta_3=0$; H_{A1}: $\eta_1=0.1$, $\eta_2=0.4$, $\eta_3=0.9$; H_{A2}: $\eta_1=0.5$, $\eta_2=0.6$, $\eta_3=0.3$
[b] Two-way ANOVA assuming independent errors
[c] Split-plot ANOVA assuming the Huynh-Feld condition is met
[d] G-G and H-F, Repeated-measures ANOVA using the Greenhouse-Geisser and Huynh-Feldt correction of the degrees of freedom; MANOVA, multivariate ANOVA test of interaction (exact); F-test for QTL main effect is identical to that obtained by a split-plot analysis
[e] Mixed model analysis assuming the same model for $\mathbf{V}$ that was used to generate the data

Simulation study

The assumption has been made that the number of genotypes is the same in each of the two marker classes, i.e., $n_1 = n_2 = n = N/2$, where N is the total number of DH lines, n_1 is the number of DH lines with $X = 0$ and n_2 is the number of DH lines with $X = 1$. n was set at 250, which is a realistic sample size for sufficiently powered QTL mapping studies (Utz et al. 2000). The cases of $n = 10$ and 25 lines per marker class were also simulated to study the small-sample behaviour of mixed model analysis in comparison to the other methods. This was done with the understanding that in QTL projects the number of lines will usually be much larger. Simulations were performed for $J = 3$ and 9 environments. Data were simulated assuming different models for **V** (Table 2). For each setting, the number of simulation runs was 10,000. All methods were implemented using the GLM and MIXED procedures of the SAS system (SAS Institute, Cary, N.C.). We assumed that for the mixed model analysis the type of model for **V** is known. In practice, the model for **V** needs to be selected based on the data. The model selection step adds uncertainty, which is not considered in the simulation study.

Results are presented in Tables 2 and 3. The simulated Type-I error rate is boldfaced when above +0.02 of the nominal level of 0.05, and the corresponding power results are italicised. This was done to indicate that the tests are invalid in these cases. It is seen that the Type-I error rate cannot generally be controlled if the genetic correlation across environments is ignored (standard two-way ANOVA). The split-plot ANOVA yields correct F-tests for QTL main effects, and it is on the liberal side for the tests of QTL-by-environment interaction. The fact that Box's correction ϵ is ≤ 1 shows that violation of the Huynh-Feldt condition ($\epsilon < 1$) will generally yield a liberal test for interaction. The other methods generally provide satisfactory control of the Type-I error rate for $n = 25$ and 250. With $n = 10$, the mixed model test for interaction yields too liberal rejection rates (Table 3), but control is satisfactory for $n = 25$.

In terms of power, mixed model analysis fares well compared to the repeated ANOVA approaches (Box's correction and MANOVA), when departure from the Huynh-Feldt condition is marked ($\epsilon << 1$). For $J = 3$ environments and $n = 250$ lines per marker class (Table 2), the mixed model analysis is rather similar to MANOVA, which is mainly due to the small number of environments. With $J = 9$ environments, gain in parsimony by fitting a specific structure to **V** tends to be more marked, in which case a small edge in favour of the

Table 3 Empirical Type-I error rates and power of tests for QTL main effects and for QTL-by-environment interaction, assuming the factor-analytic model, $\mathbf{V} = \phi\phi' + \mathbf{I_J}\sigma^2$ (all parameter values for **V** taken from Table 2; for $J = 9$, each of the three environments was replicated three times), single-point analysis, 10,000 simulation runs and tests for a nominal significance level of 5%, conditional on the putative QTL position. The simulated Type-I error rate is in bold type when outside +0.02 of the nominal level of 0.05, and the corresponding power results are italicised. This was done to indicate that the tests are invalid in these cases

Hypothesis[a]	Rejection rate								
	QTL main effect			QTL interaction					
	ANOVA[b]	Split-plot[c, d]	Mixed[e]	ANOVA[b]	Split-plot[c]	G-G[d]	H-F[d]	MANOVA[d]	Mixed[e]
$n = 10$, $J = 3$ ($\epsilon = 0.610$):									
H_0	**0.197**	0.047	0.051	0.008	**0.074**	0.053	0.049	0.048	0.053
$H_{A1'}$	*0.639*	0.358	0.368	0.236	*0.596*	0.534	0.518	0.457	0.489
$H_{A2'}$	*0.641*	0.396	0.381	0.019	*0.143*	0.106	0.097	0.317	0.349
$n = 25$, $J = 3$ ($\epsilon = 0.610$):									
H_0	**0.184**	0.054	0.055	0.006	**0.077**	0.053	0.051	0.050	0.053
$H_{A1''}$	*0.655*	0.393	0.397	0.225	*0.625*	0.559	0.555	0.495	0.508
$H_{A2''}$	*0.649*	0.385	0.390	0.015	*0.140*	0.102	0.099	0.356	0.370
$n = 10$, $J = 9$ ($\epsilon = 0.203$):									
H_0	**0.447**	0.052	0.056	0.014	**0.145**	0.062	0.055	0.052	0.065
$H_{A1''}$	*0.838*	0.374	0.383	0.308	*0.761*	0.584	0.562	0.253	0.480
$H_{A2''}$	*0.843*	0.376	0.388	0.277	*0.251*	0.116	0.103	0.349	0.564
$n = 25$, $J = 9$ ($\epsilon = 0.203$):									
H_0	**0.438**	0.048	0.050	0.008	**0.136**	0.055	0.053	0.053	0.057
$H_{A1''}$	*0.845*	0.392	0.397	0.292	*0.775*	0.595	0.587	0.362	0.450
$H_{A2''}$	*0.836*	0.395	0.398	0.018	*0.250*	0.108	0.105	0.508	0.588
$n = 250$, $J = 9$ ($\epsilon = 0.203$):									
H_0	**0.449**	0.046	0.046	0.006	**0.139**	0.048	0.048	0.053	0.054
H_{A1}	*0.839*	0.413	0.413	0.300	*0.771*	0.606	0.605	0.432	0.444
H_{A2}	*0.837*	0.405	0.406	0.019	*0.244*	0.103	0.103	0.597	0.603

[a] H_0, H_{A1} and H_{A2} as specified in Table 2; $H_{A1'}$, $H_{A2'}$, parameter values of H_{A1} and H_{A2} for (η_1, η_2, η_3) multiplied by $\sqrt{10}$; $H_{A1''}$, $H_{A2''}$, parameter values of H_{A1} and H_{A2} for (η_1, η_2, η_3) multiplied by 5; for $J = 9$, each of the three environments from Table 3 was repeated three times; $\eta_j = \alpha + a_j$, where α is the QTL main effect and a_j is the QTL-by-environment interaction in the j-th environment

[b, c, d, e] See Table 2

mixed model analysis is expected, and this was confirmed in the simulation study (Table 3). The advantage in power is largest when the number of lines is small.

The approximate analysis based on Box's correction may be more powerful than both MANOVA and mixed model analysis, depending on the type of departure from the null hypothesis. Simulation evidence by other authors suggests that, broadly speaking, Box's corrected *F*-test and MANOVA *F*-test for repeated measures designs tend to yield comparable power, there being no clear-cut advantage of one procedure over the other (Huynh 1978; Rogan et al. 1979).

The standard split-plot ANOVA *F*-test for QTL-by-environment interaction as well as the standard two-way ANOVA *F*-test for the QTL main effect tend to be too liberal under departures from the Huynh-Feldt condition, while the two-way ANOVA *F*-test is too conservative with regard to the QTL-by-environment interaction effect. In contrast, when this condition is met, a split-plot analysis is the best choice due to the smaller number of variance parameters that need to be estimated compared to MANOVA and the associated gain in power for detecting interaction. In fact, the ANOVA *F*-test is the most powerful test in this situation (Huynh 1978), and it will then be asymptotically equivalent to both the mixed model analysis and repeated MANOVA.

Discussion

The simulation shown here has demonstrated that both the mixed model approaches and the univariate and multivariate repeated-measures ANOVA can be generally recommended for QTL-by-environment interaction analysis. The mixed model analysis does fairly well for all three variance-covariance structures studied, and it will be uniformly more powerful than MANOVA, when the variance-covariance structure is correctly specified. The split-plot ANOVA is an approximate method that tends to be too liberal under departures from the Huynh-Feldt condition. A standard two-way ANOVA, which ignores genetic correlation, should be avoided.

Jiang and Zeng (1995) have proposed the MANOVA approach for QTL-by-environment interaction within a full likelihood-based interval mapping framework accounting for the mixture of normals induced by the QTL model. Alternatively, one may use a least squares approximation and regress phenotypes on the expected QTL genotype, given the flanking markers, as proposed by Haley and Knott (1992). In this case, one of the methods proposed in this paper could be used, i.e., repeated-measures ANOVA or a mixed model analysis. An important practical consideration is that both of these methods yield test statistics that either exactly or asymptotically have χ^2-distributions conditionally on the putative QTL. Thus, the approximate method proposed by Piepho (2001) and other analytic approximations reviewed therein can be used to quickly determine critical thresholds for scanning the whole genome. When a least squares approach is used, the method of Davies (2002) is particularly useful for determining critical thresholds. A competitive, though computationally more demanding alternative is to use permutation thresholds (Churchill and Doerge 1994). Tinker and Mather (1995) argued that a simple test statistic may be used for detecting QTL-by-environment interaction, which ignores genetic correlation, since a permutation approach will fix the problem when the genome-wise Type-I error rate is controlled. While it is true that permutation of multi-environment phenotypic records among genotypes can always be used to obtain a valid null distribution, ignoring the genetic correlation in the construction of the test statistic may sacrifice some power.

When the number of genotypes is large, as is necessary in QTL mapping studies designed to achieve reasonable power (Utz et al. 2000), there are plenty of degrees of freedom for estimating **V**. This is the main reason why repeated MANOVA and mixed model analysis yielded comparable results for all simulated settings with $n = 250$. Clearly, both methods are asymptotically equivalent, providing the structure for **V** is correctly specified in the mixed model analysis. Discrepancies in power are also small when the difference in the number of parameters in **V** is rather pronounced as, for example, for the unstructured model and the factor-analytic model with one factor and a homoscedastic residual in the case of a large number of environments. For $J = 9$ environments, as studied by this simulation, these models have 45 and 10 parameters, respectively. Notable power differences are found only when the number of degrees of freedom for estimating **V** is small (Table 3). From this it may be concluded that in QTL studies it is usually affordable to fit an unstructured model for **V** in a mixed model analysis. When the number of lines is limited and the number of environments is large, however, it is sometimes difficult to fit this model in a mixed model setting. Also, the parsimony principle suggests that simpler models may be preferable in terms of predictive power. The factor-analytic model, which imposes a certain structure on **V**, seems to be a very useful model for approximating an unstructured model (Piepho 1998). According to one referee, experience from simulations suggests that factor-analytic models often provide stable and parsimonious model fits, even when the true model is unstructured.

This paper has considered environments to be a fixed factor, which is appropriate when environments have been purposefully selected and when the analysis focusses on the detection of QTL-by-environment interaction. Depending on context and sampling design, it may be more appropriate to take environments as random; for example, when the number of environments is large, trial sites have been randomly selected and the researcher is interested in assessing mean performance as well as the stability of QTL effects across environments in a target region. For a full treatment of this case, see Piepho (2000).

The main purpose of this paper has been to demonstrate that a mixed model analysis is competitive to repeated measures MANOVA, where both are applicable. The major advantage of the mixed model approach is that it is also applicable when repeated measures MANOVA or Box's correction are difficult to implement or simply will not work. For example, in a mixed model framework it is straightforward to account for spatial within-trial variation and for other random effects (incomplete blocks, rows, columns, years, locations, repeated measures in space and time, etc.) (Piepho 2000; Eckermann et al. 2001; Verbyla et al. 2003; Piepho et al. 2003, 2004). Also, accommodating a three-way analysis of QTL-by-year-by-location analysis with random years and/or locations is easy by mixed modelling, but rather cumbersome within a repeated MANOVA framework. Moreover, missing data on some genotypes in some of the environments can by accommodated efficiently by a mixed model, while repeated MANOVA requires complete data for each genotype to be included in the analysis. Thus, mixed modelling has much to commend it relative to other methods. It is hoped that the present simulation study increases confidence in the suitability of mixed model procedures among the QTL mapping community.

Acknowledgements I thank Andreas Büchse for helpful comments on earlier versions of this manuscript.

References

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. Genetics 138:963–971

Davies RB (2002) Hypothesis testing when a nuisance parameter is present only under the alternative: linear model case. Biometrika 89:484–489

Eckermann PJ, Verbyla AP, Cullis BR, Thompson R (2001) The analysis of quantitative traits in wheat mapping populations. Aust J Agric Res 52:1195–1206

Falconer DS, Mackay TFC (2001) Introduction to quantitative genetics, 4th edn. Pearson Education, Harlow

Geisser S, Greenhouse SW (1958) An extension of Box's results on the use of the *F*-distribution in multivariate analysis. Ann Math Stat 29:885–891

Haley CS, Knott SA (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69:315–324

Huynh H (1978) Some approximate tests for repeated measurement designs. Psychometrika 43:161–175

Huynh H, Feldt LS (1976) Estimation of the Box correction for degrees of freedom from sample data in randomized block and split plot designs. J Educ Stat 1:69–82

Jiang CJ, Zeng ZB (1995) Multiple trait analysis of genetic mapping for quantitative trait loci. Genetics 140:1111–1127

Kenward MG, Roger JH (1997) Small sample inference for fixed effects from restricted maximum likelihood. Biometrics 53:983–997

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

Piepho HP (1998) Empirical best linear unbiased prediction in cultivar trials using factor analytic variance-covariance structures. Theor Appl Genet 97:195–201

Piepho HP (2000) A mixed model approach to mapping quantitative trait loci in barley on the basis of multiple environment data. Genetics 156:253–260

Piepho HP (2001) A quick method to compute approximate LOD thresholds for QTL detection. Genetics 157:425–432

Piepho HP, Pillen K (2004) Mixed modelling for QTL × environment interaction analysis. Euphytica 137:147–153

Piepho HP, Büchse A, Emrich K (2003) A hitchhiker's guide to the mixed model analysis of randomized experiments. J Agron Crop Sci 189:310–322

Piepho HP, Büchse A, Richter C (2004) A mixed modelling approach to randomized experiments with repeated measures. J Agron Crop Sci 190:230–247

Pillen K, Zacharias A, Léon J (2003) Advanced backcross QTL analysis in barley (*Hordeum vulgare* L.). Theor Appl Genet 107:340–352

Rogan JC, Keselman HJ, Mendoza JL (1979) Analysis of repeated measures. Br J Math Stat Psychol 32:269–286

Tinker NA, Mather DE (1995) Methods for QTL analysis with progeny replicated in multiple environments. J Agric Genomics 1 (http://www.cabi-publishing.org/gateways/jag/index.html)

Utz HF, Melchinger AE (1996) PLABQTL: a program for composite interval mapping of QTL. J Agric Genomics 1 (http://www.cabi-publishing.org/gateways/jag/index.html)

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

Verbyla AP, Eckermann PJ, Thompson R, Cullis BR (2003) The analysis of quantitative trait loci in multi-environment trials using a multiplicative mixed model. Aust J Agric Res 54:1395–1408

Winer BJ, Brown DR, Michels KM (1991) Statistical principles in experimental design, 3rd edn. McGraw-Hill, New York

Wolfinger RD (1996) Heterogeneous variance-covariance structures for repeated measures. J Agric Biol Environ Stat 1:205–230