Quantitative trait loci mapping in plant genetics by α-design experiments and molecular genetic marker systems

by

W. Urfer[1], S. Mejza[2] and F. Hering[1]

[1]University of Dortmund, Department of Statistics, Vogelpothsweg 87, D-44227 Dortmund

[2]Agricultural University, Wojska Polskiego 28, PL-60-637 Poznan, Poland

Summary

Research concerning the quantitative trait loci (QTL) mapping in plant genetics usually consists of two stages. The first stage is concerned with collecting data while the second one, based on the data collected, is concerned with a proper QTL study. The final inferences are strictly connected with the quality of the two approaches applied in both stages. Data to be analyzed come from an experiment dealing with offsprings obtained from a crossing system of several lines. The genotypes then are observed in some natural or quasi natural environment.

The QTL studies are based on so called genotype adjusted means. In α-designs the adjusted means can be calculated in many ways, which will be presented in this paper. We also give an EM-algorithm for the estimation of genetic parameters and comment on recent biometrical research in molecular plant genetics. Finally we mention some activities in the new field of bioinformatics.

1. Introduction

Characters of agronomic importance show quantitative variation in crop species. This variation is the result of multiple segregeting loci, whose gene expression can be modulated by the environment. Genetic improvement of such quantitative traits is difficult because the effects of individual genes controlling these traits cannot be identified. The loci that control quantitative traits are called quantitative trait loci, abbreviated henceforth as QTLs.

An important goal in genetics and plant breeding is to identify and chracterize QTLs. The recent advances in molecular genetics have allowed the construction of genetic linkage maps based on molecular markers. The plant geneticist and statistician can then look for correlations between the mapped markers and the trait of interest in controlled breeding experiments to obtain information about the regions of the genome that control the trait.

The general QTL studies usually consist of two stages. The first stage is connected with collecting data while the second one, based on data collected concerns a proper QTL studies. The final inferences are strictly connected with quality of approaches applied in both stages. Data to be analysed come from an experiment dealing with offsprings obtained from a crossing system of several lines. The genotypes then there are observed in some natural or quasi natural environments.

To obtain as good as possible data concerning genotypes the experimenter uses some optimal experimental design. The design should allow to estimate the genetical characteristics with maximal precision. In QTL experiment we observe only genotypes that are different but which are not additionally grouped or split. It means that most of QTL experiments use one-factorial design such as for example completely randomized design, block design, nested block design, row-column design, block design with nested rows and columns, split plot design, split block designs etc..

All of the above designs are proper with respect to the particular experimental situation. The experiment considered in the paper was carried out in a so called α-design (see Melchinger et al., 1998).

α-designs belong to the class of the nested block designs (NBDs). These designs are applied when experimental material is divided into two nested systems of blocks, (groups) i.e. superblocks and blocks. The blocks are nested within superblocks. Block designs of this type have been investigated by Hering and Mejza (1997). The NBDs allow to eliminate real or potential heterogeneity of experimental units under above nested structure.

2. Estimation

Let n experimental units be divided into r superblocks and additionally, each of them let be divided into b blocks of size k. Then we have $n=rbk$. Let v denote the number of genotypes.

The observation obtained in an experiment carried out in NBD is usually modelled as follows (see Mejza and Mejza 1989, Mejza 1994):

$$y = 1\mu + C'\alpha + D'\beta + \Delta'\gamma + e + \varepsilon \ . \tag{1}$$

Here y denotes an $n\times1$ vector of observation, 1 denotes $n\times1$ vector of ones, μ the general mean. C' and D' are $n\times r$ and $n\times rb$ design matrices for superblocks and blocks, respectively, and α,β correspond to $r\times1$ and $rb\times1$ vectors of superblock and block effects. Furthermore Δ' is the $n\times v$ design matrix for genotypes corresponding to $v\times1$ vector γ of treatment effects, e and ε are $n\times1$ vectors of unit errors and technical errors.

The statistical properties of the model (1) resulting from a three step randomization (i.e. randomization of superblocks, blocks within superblocks and randomization of units within blocks) are as follows:

$$E(y) = 1\mu + \Delta'\gamma$$

$$\begin{aligned}
&\alpha \sim (0, V_\alpha^-),\ V_\alpha = \sigma_\alpha^2\left(I_r - (bk)^{-1}J_r\right),\\
&\beta \sim (0, V_\beta),\quad V_\beta = \sigma_\beta^2\, I_r \otimes \left(I_b - b^{-1}J_b\right),\\
&e \sim (0, V_e),\quad V_e = \sigma_e^2\, I_r \otimes I_b \otimes \left(I_k - k^{-1}J_k\right),\\
&\varepsilon \sim (0, \sigma^2\,\mathrm{I}),
\end{aligned} \tag{2}$$

where σ_α^2, σ_β^2, σ_e^2, and σ^2 are superblock, block, unit and technical error variances, respectively, I_t is identity matrix of degree t, $J_t = 1_t 1_t'$ is the txt matrix of ones, $\otimes$ denotes the Kronecker product of matrices.

The design considered has orthogonal block structure (cf. Nelder, 1965), hence the covariance matrix V=Cov(y) can be expressed as

$$V = G'V_\alpha G + D'V_\beta D + \sigma^2 I_n \tag{3}$$

or

$$V = \tau_0 P_0 + \tau_1 P_1 + \tau_2 P_2 + \tau_3 P_3 \ , \tag{4}$$

where P_i are family of orthogonal projectors, summing to identity matrix, of the form:

$$P_0 = n^{-1}J_n,,\quad P_1 = (rb)^{-1}G'G - n^{-1}J_n,\quad P_2 = k^{-1}D'D - (rb)^{-1}G'G,\quad P_3 = I_n - k^{-1}D'D,$$

and, respectively. τ_i are variance components of the form

$$\tau_0 = \sigma^2, \quad \tau_1 = rb\sigma_\alpha^2 + \sigma^2, \quad \tau_2 = k\sigma_\beta^2 + \sigma^2, \quad \tau_3 = \sigma_e^2 + \sigma^2 \ .$$

The projectors define so called strata, i.e. general area stratum, inter-superblock stratum, inter-block stratum and inter-plot stratum.

Because of $\sum_{i=0}^{3} P_i = I$ we have

$$Iy = (P_0 + P_1 + P_2 + P_3)y = P_0 y + P_1 y + P_2 y + P_3 y = y_0 + y_1 + y_2 + y_3 \,.$$

It means that overall analysis of model (1) can be split into stratum analyses, defined by models

$$\mathrm{y_i} = \mathrm{P_i y}, \quad \mathrm{Cov(y_i)} = \tau_\mathrm{i}\, \mathrm{P_i}, \quad i=0,1,2,3 \,. \tag{5}$$

Because of algebraic properties of P_i (i=0,1,2,3) to the analysis of the strata models (5) we can use ordinary least squares method. The normal equation then may be written as

$$\Delta P_i \Delta' \gamma_i = \Delta P_i y \tag{6}$$

and $\gamma_t = (\Delta P_i \Delta')^- \Delta P_i y$ is a vector of normal equation solutions with

$$Cov(\gamma_i) = \tau_i \Delta P_i \Delta' \,. \tag{7}$$

The matrices $\Delta P_i \Delta' = C_i$ are called stratum information matrices for estimation genotype effects. All stratum statistical properties of the design are connected with the pattern of that matrices.

The normal equation for estimating genotype effects in model (1) under the assumption that τ_i are known can be written as:

$$(\tau_1^{-1} C_1 + \tau_2^{-1} C_2 + \tau_3^{-1} C_3)\gamma_0 = \tau_1^{-1} Q_1 + \tau_2^{-1} Q_2 + \tau_3^{-1} Q_3 \tag{8}$$

where $Q_i = \Delta P_i y$.

The genotype arrangement in NBD can be characterised by two incidence matrices, i.e. by $N_1 = \Delta G'$ and $N = \Delta D'$. The first one characterises the genotype arrangements on superblocks while the second one on blocks.

Let $N_1 1 = N1 = R$ denote the vector of genotype replication.

The above considerations simplify for a particular class of NBDs called α-design. α-designs are special NBDs in which all genotypes are replicated exactly α-times in every superblock, i.e. $R=\alpha \cdot r \cdot 1$.

It means that α-design is equireplicate design. In this class of design we have: $C_1=0$, and $Q_1=0$,

$$C_2 = k^{-1}NN' - (\tfrac{\alpha r}{v})I_v, \quad C_3 = \alpha r I_v - k^{-1}NN' . \tag{9}$$

The QTL techniques are based on data that are a sample from normal population with constant variance and they are independent. The difference lies in expected values only. Hence, the problem that will be considered concerns how to find genotype mean estimates following above requirements.

<u>Method 1 (Williams, 1977)</u>

Let us note that in an α-design there is no information on genotypes in the inter-superblock stratum. Hence, practically we can remodel our data so that

$$y = D'\beta + \Delta'\gamma + e , \tag{10}$$

$$E(y) = \Delta'\gamma, \quad V_1 = Cov(y) = \sigma_\beta^2\, D'D + \sigma_0^2 I . \tag{11}$$

In this approach it is assumed that the sample of units was drawn in two stages from an infinite number of potential blocks and from an infinite number of potential units within blocks. In this case $\sigma_0^2 = \sigma_e^2 + \sigma^2$. Formally, Williams (1977) considers the model with $V_1 = \sigma_0^2(\varphi k^{-1} D'D + I_n)$, where $\varphi = k\sigma_\beta^2 / (\sigma_e^2 + \sigma^2)$. He gives an iterative procedure to estimate the vector of genotype effects and an approximation of V_1.

From that we can get the required estimates of genotype effects used in QTL studies.

<u>Method 2</u>

Let us note that the normal equations for estimating genotype effects are of the form

$$\left(\tau_2^{-1}C_2 + \tau_3^{-1}C_3\right)\gamma_{23} = \tau_2^{-1}Q_2 + \tau_3^{-1}Q_3 . \tag{12}$$

Instead of the true values of τ we can use estimates obtained by stratum ANOVA and solve that equation under known variance components. Then we have

$$\gamma_{23} = \left(\hat{\tau}_2^{-1} C_2 + \hat{\tau}_3^{-1} C_3\right)^{-} \left(\hat{\tau}_2^{-1} Q_2 + \hat{\tau}_3^{-1} Q_3\right). \tag{13}$$

and

$$\mathrm{Var}(\gamma_{23}) = \left(\hat{\tau}_2^{-1} C_2 + \hat{\tau}_3^{-1} C_2\right)^{-} \left(\hat{\tau}_2^{-1} \Delta P_2 + \hat{\tau}_3^{-1} P_3\right) \sum_{i=0}^{3} \hat{\tau}_i P_t \left(\hat{\tau}_2^{-1} P_2 \Delta' + \hat{\tau}_3^{-1} P_3 \Delta'\right) \cdot$$
$$\cdot \left(\hat{\tau}_2^{-1} C_2 + \hat{\tau}_3^{-1} C_2\right)^{-} = \left(\hat{\tau}_2^{-1} C_2 + \hat{\tau}_3^{-1} C_3\right)^{-} . \tag{14}$$

The problem is that we need generalized inverses and therefore the solution is in general not unique.

To obtain a unique solution we may impose linear restrictions upon the parameters. For example we can add the matrix gZZ' to the matrix $\tau_2^{-1} C_1 + \tau_3^{-1} C_3 = C$. Here q is nonzero constant and Z is an any arbitrary vector such that $Z'1 \neq 0$. Then the matrix $C+qZZ'$ is nonsingular and $(C+qZZ')^{-} = (C+qZZ')^{-1}$.

Finally, as estimates of genotype mean (adjusted means) we may take $\tilde{\gamma} = \bar{y}1 + \gamma_{23}$, where $\bar{y}$ is the general mean. Having dispersion matrix nonsingular it is easy to obtain genotype mean estimates uncorrelated.

Method 3

As we see from (12) in α-designs the information is included into two strata i.e. inter-block and inter-plot stratum. Usually, almost all information is included in the inter-plot stratum. Therefore it is sensible to restrict the investigation to that stratum. This is the idea of the third method. It means that we use an approach appropriate to the so called intra-block analysis of the block designs.

Method 4

In fact the α-designed experiment is incomplete with respect to superblocks. Especially when α=1 then with respect to superblocks we have a complete randomized block design. Finally, assuming that the block effects can be omitted in one linear model we can treat the whole experiment as one set up in complete randomized block designs. In this case it is enough to take only usual means as the genotype mean estimates.

Open question

Which method is the most suitable for the QTL study? From the statistical point of view only method 3 guarantees at least unbiasedness of the genotype means. This results from the fact that ordinary least square estimators are unbiased in the mixed linear model case. We cannot say anything about the statistical properties of estimators obtained by other methods. Moreover, by proper design of experiments we can nearly gain the whole information from the third stratum. α-designs are usually highly efficient. This explains why they are used often in agricultural and genetical experiments.

3. Quantitative trait locus mapping in plant genetics using molecular genetic marker systems

The recent advent of molecular markers has created a great potential for the understanding of quantitative inheritance. In parallel to rapid developments in molecular marker technologies biometrical models have been constructed, refined and generalized for detecting, mapping and estimating the effects of quantitative trait loci (QTL). Melchinger et al. (1998) evaluated testcross progenies of 344 F_3 lines in combination with two unrelated testers plus additional testcross progenies from an independent but smaller sample of 107 F_3 lines from the same cross in combination with the same two testers for grain yield and four other important agronomic traits.

For a more detailed statistical analysis of this data set A.E. Melchinger and H.F. Utz from the Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, provided plant height measurements of an F_2 –population of maize, which was genotyped for a total of 89 restriction fragment length polymorphism marker loci. Recombination frequencies between marker loci were estimated by multi-point analyses and transformed into map distances in centi-Morgan by Haldane’s mapping function.

The aim of our statistical approach is to find Maximum-Likelihood-estimates of QTL locations and effects including their estimated standard errors.

We consider experimental populations derived from a cross between two parental inbred lines P_1 and P_2. Two flanking markers for an interval, where a putative QTL is being tested, have alleles M,m and N,n. If the F_1 individuals are selfed or intermated, it produces an F_2 –population with nine observable marker genotypes.

We consider a QTL in the F_2 –population in which the frequencies of genotypes QQ, Qq and qq are ¼, ½ and ¼ . The genetic model for a QTL is given by

$$G=\begin{pmatrix} G_2 \\ G_1 \\ G_0 \end{pmatrix}=\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\mu+\begin{pmatrix} 1 & -1/2 \\ 0 & 1/2 \\ -1 & -1/2 \end{pmatrix}\begin{pmatrix} a \\ d \end{pmatrix}=1_3\mu+DE \qquad (1)$$

where the genetic parameter μ is the mean and a and d denote the additive and dominance effects of QTL in the F_2 –population.

The data consists of two parts

y_j , j=1,2,...,n for the quantitative trait value

and x_j , j=1,2,...,n for the genetic markers and other explanatory variables.

The composite interval mapping model from Kao and Zeng (1997) is proposed as

$$y_j = a\,x_j^* + d\,z_j^* + x_j\beta + \varepsilon_j\,, \quad j=1,2,\ldots,n \qquad (2)$$

where $$x_j^* = \begin{cases} 1 & \text{if the QTL is QQ} \\ 0 & \qquad\qquad\quad \text{Qq} \\ -1 & \qquad\qquad\quad \text{qq} \end{cases}$$

and $$z_j^* = \begin{cases} 1/2 & \text{if the QTL is Qq} \\ -1/2 & \text{otherwise} \end{cases} .$$

y_j is the quantitative trait value of the j-th individual, a and d are the additive effects of the putative QTL, β is the partial regression coefficient vector including the mean μ, and $\varepsilon_j \sim N(0,\sigma^2)$ is a random error.

Let $$g_j\left(x_j^*, z_j^*\right)=\begin{cases} p_{j1} \text{ if } x_j^*=1 \text{ and } z_j^*=-1/2 \\ p_{j2} \text{ if } x_j^*=0 \text{ and } z_j^*=\ 1/2 \\ p_{j3} \text{ if } x_j^*=-1 \text{ and } z_j^*=-1/2 \end{cases}$$

be the distribution of QTL genotype specified by x_j^* and z_j^* .

We treat the unobserved QTL genotypes (x_j^* and z_j^*) as missing data, denoted by $y_{(mis,j)}$ and treat trait y_j, selected markers and explanatory variables (x_j) as observed data, denoted by $y_{(obs,j)}$.

Kao und Zeng (1997) apply the EM algorithm to obtain the Maximum-Likelihood estimates of $\theta=(p,a,d,\beta,\sigma^2)$.

At a given position, p can be determined and the EM algorithm is used for obtaining the ML estimates of a, d, β, and σ^2.

In the E-step we compute the conditional expected complete-data log likelihood with respect to the conditional distribution of Y_{mis} given Y_{obs} and the current estimated parameter value $\theta^{(t)}$, given by

$$Q\left(\theta|\theta^{(t)}\right)=\sum_{j=1}^{n}\sum_{i=1}^{3}\log\left[\varphi\left(\frac{y_j-\mu_{ji}}{\sigma}\right)p_{ji}\right]\cdot\pi_{ji}^{(t)} \quad ,$$

where
$$\pi_{ji}=\frac{p_{ji}\,\varphi\left(\frac{y_i-\mu_{ji}^{(t)}}{\sigma^{(t)}}\right)}{\sum_{i=1}^{3}p_{ji}\,\varphi\left(\frac{y_j-\mu_{ji}^{(t)}}{\sigma^{(t)}}\right)} \ ,$$

p_{ji} are conditional probabilities of QTL genotypes given marker genotypes. They are given by Kao and Zeng (1997) for F_2-populations.

$\varphi(\bullet)$ is a standard normal probability function,

$\mu_{j1}=a-\frac{d}{2}+x_j\beta$, $\mu_{j2}=\frac{d}{2}+x_j\beta$, $\mu_{j3}=-a-\frac{d}{2}+x_j\beta$.

Taking the derivatives of $Q(\theta|\theta^{(t)})$ with respect to each parameter, Kao and Zeng (1997) give the following result:

$$a^{(t+1)}=\frac{\sum_{j=1}^{n}\left[\left(\pi_{j1}^{(t)}-\pi_{j3}^{(t)}\right)\left(y_j-x_j\beta^{(t)}\right)-\frac{1}{2}\left(\pi_{j3}^{(t)}-\pi_{j1}^{(t)}\right)d^{(t)}\right]}{\sum_{j=1}^{n}\left(\pi_{j1}^{(t)}+\pi_{j3}^{(t)}\right)} \ ,$$

$$d^{(t+1)}=\frac{\sum_{j=1}^{n}\frac{1}{2}\left[\left(-\pi_{j1}^{(t)}+\pi_{j2}^{(t)}-\pi_{j3}^{(t)}\right)\left(y_j-x_j\beta^{(t)}\right)-\left(\pi_{j3}^{(t)}-\pi_{j1}^{(t)}\right)a^{(t+1)}\right]}{\frac{1}{4}\sum_{j=1}^{n}\left(\pi_{j1}^{(t)}+\pi_{j2}^{(t)}+\pi_{j3}^{(t)}\right)} \ ,$$

$$\beta^{(t+1)}=(X'X)^{-1}X'\left[Y-\Pi^{(t)}\,DE^{t+1)}\right]$$

and $$\sigma^{2(t+1)} = \frac{1}{n}\Big[\big(Y - X\beta^{(t+1)}\big)'\big(Y - X\beta^{(t+1)}\big) - 2\big(Y - X\beta^{(t+1)}\big)'\Pi^{(t)}\,DE^{(t+1)} + E'^{(t+1)}\,V^{(t)}\,E^{(t+1)}\Big],$$

where $\Pi = \{\pi_{ji}\}_{nx3}$,

$$V = \begin{pmatrix} 1'\Pi(D_1 \# D_2) & 1'\Pi(D_1 \# D_2) \\ 1'\Pi(D_2 \# D_1) & 1'\Pi(D_2 \# D_2) \end{pmatrix},$$

# denotes Hadamard product of the columb vectors of the genetic design matrix D.

Additionally to ML-estimates of a, d, β, and σ^2, Kao and Zeng (1997) give general formulas for the asymptotic variance-covariance matrix of the ML-estimates. In a recent master-thesis Emrich (1999) was able to present explicit formulas for the M-step which where different from the above results. At every position, the position parameter p can be predetermined and only a, d, β, and σ^2 are involved in estimation and testing. If the tests are significant in a chromosome region, the position with the largest LRT statistic is inferred to be the estimate of the QTL position p, and the MLE's at this position are the estimates of a, d, β, and σ^2.

For our plant genetic project it is important to construct confidence intervals for the QTL positions and effects. The asymptotic variances can be used to calculate these confidence intervals. For a large sample, the (1-α)% confidence interval for a position p can be approximated by $\left(\hat{p} - z_{1-\alpha/2}\, S_{\hat{p}} ; \hat{p} + z_{1-\alpha/2}\, S_{\hat{p}}\right)$.

In our approach, the lack of knowledge on the number of locations of the most important QTL's contributing to the trait is a major problem. Stephens and Fisch (1998) utilize reversible jump Markov-chain-Monte-Carlo-methodology to compute posterior densities not only for the parameters, given the number of QTL, but also for the number of QTL itself.

The EM-algorithm is also used by Selinski and Urfer (1998) for the estimation of toxicokinetic parameters which are an essential component in the risk assessment of potential harmful chemicals. This is a first step to analyse the biological processes which are involved in the formation of DNA adducts and might therefore lead to the development of cancer.

More details on this research are given by Urfer and Becka (1996). Jansen (1996) describes a Monte-Carlo expectation-maximization-algorithm for fitting multiple-QTL models to genetic data which are highly incomplete. Such complicated situations occur when dominant and missing markers are used.

Many PCR-(Polymerase Chain Reaction) based genetic markers behave like dominant markers. Co-dominant markers such as restriction fragment length polymorphism (RFLP) show three band patterns in an F_2 population in electrophoretic gels, each representing one genotype of a probe. Dominant markers such as random amplified polymorphic DNA (RAPD) can show only two patterns: presence or absence of a band. A heterozygote can have the same band pattern as one of the homozygotes. Jiang and Zeng (1997) derive a general algorithm to deal with dominant and missing markers in F_2 derived from two inbred lines using hidden Markov chains.

In recent years considerable progress has been made in the development of new statistical methods of QTL analysis. These new methods have been implemented in user-friendly, widely applicable software packages such as PLABQTL, written for routine QTL analysis in Plant Breeding and Biology by Utz and Melchinger (1996). The QTL cartographer from Basten et al. (1999) is another suite of programs for mapping QTLs onto a genetic linkage map. The programs use linear regression, interval mapping or composite interval mapping methods to dissect the underlying genetics of quantitative traits. The mapping program uses a dynamic algorithm that allows several statistical models to be fitted and compared, including various gene actions, QTL-environment interaction and dose linkage.

Scientific progress in molecular genetics leads to new challenges in statistics and computer science using Markov chain Monte-Carlo-algorithms and hidden Markov chains. The new field of bioinformatics is concerned with analyzing genomic data in order to help elucidate biological processes, diagnose diseases and invent new bioactive drugs. The Helmholtz Network on Bioinformatics (HNB) directed by T. Lengauer integrates the bioinformatics software throughout Germany and brings the new potential for genomic analysis to increased use in molecular biology and medicine.

Acknowledgements

The first author would like to thank Professors H.F. Utz and A.E. Melchinger from the Institute of Pant Breeding, Seed Science and Population Genetics, University of Hohenheim, for introducing him to the problems of marker assisted selection in plant breeding. The financial support from the Graduate College and the Collaborative Research Centre at the University of Dortmund is gratefully acknowledged.

References:

Basten, C.J., Weir, B.S. and Zeng, Z.-B., 1999: QTL Cartographer, Version 1.13. *Department of Statistics, North Carolina State University,* Raleigh, NC.

Emrich, K., 1999: Estimation of effects of genes in plant breeding. **Master Thesis.** *Department of Statistics, University of Dortmund.*

Hering, F. and Mejza, S., 1997: Incomplete split-block design. *Biom. J.* **93**, 227-238.

Jiang, C. and Zeng, Z.-B., 1997: Mapping quantitative trait loci with dominant and missing markers in various crosses from two inbred lines. *Genetica* **101**, 47-58.

Jansen, R.C., 1996: A general Monte Carlo method for mapping multiple quantitative trait loci. *Genetics* **142**, 305-311.

Kao, C.-H., and Zeng, Z.-B., 1997: General formulas for obtaining the MLE's and the asymptotic variance-covariance matrix in mapping quantitative trait loci when using the EM algorithm. *Biometrics* **53**. 653-665.

Mejza I., Mejza S., 1989: A note on model building in block designs. *Biuletyn Oceny Odmian*, 187-201.

Mejza S., 1994: On modelling of experiments in natural sciences. *Listy Biom. – Biom. Letters* **31**, 79-100.

Melchinger, A.E., Utz, H.F. and Schön, C.C., 1998: Quantitative trait locus (QTL) mapping using different testers and independent populaton samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. *Genetics* **149**, 383-403.

Nelder J.A., 1965: The analysis of randomized experiments with orthogonal block structure. *Proc. Roy. Soc.* A **283**, 147-178.

Selinski, S. and Urfer, W., 1998: Interindividual and interoccasion variability of toxicokinetic parameters in population models. *Technical Report* **38/1998**, University of Dortmund.*

Stephens, D.A. and Fisch, R.D., 1998: Bayesian analysis of quantitative trait locus data using reversible jump Markov chain Monte Carlo. *Biometrics* **54**, 1334-1347.

Urfer, W. and Becka, M., 1996: Exploratory and model-based inference in toxicokinetics. In: Morgan, B.J.T. (Ed.): *Statistics in Toxicology*, Oxford University Press, pp. 198-216.

Utz, H.F. and Melchinger, A.E., 1996: PLABQTL: A program for composite interval mapping of QTL. *J. Quant. Trait Loci.* **2**. 1-5.

Williams E.R., 1977: Iterative analysis of generalized lattice designs. *Austral. J. Statist.* **19**, 39-42.

*Availabe from the world wide web: http://www.statistik.uni-dortmund.dc/sfb475/sfblit.htm