CHAPTER 10

MODELLING THE GENETIC BASIS OF RESPONSE CURVES UNDERLYING GENOTYPE × ENVIRONMENT INTERACTION

F.A. VAN EEUWIJK, M. MALOSETTI AND M.P. BOER

Laboratory of Plant Breeding, Department of Plant Sciences, Wageningen University, P.O. Box 386, 6700 AJ Wageningen, The Netherlands
Biometris, Wageningen University, P.O. Box 100, 6700 AC Wageningen, The Netherlands
E-mail: fred.vaneeuwijk@wur.nl

Abstract. To increase tolerance to abiotic stresses in breeding programmes, typically families and collections of genotypes are evaluated in series of trials (environments) representing different levels of stress. The statistical analysis of the data from such trials concentrates on modelling the phenotypic behaviour of the genotypes across the set of environments. This phenotypic behaviour can be modelled in the form of genotype-specific linear and non-linear response curves in relation to environmental characterizations. Non-parallelism of the response curves indicates genotype × environment interaction. Identification of the genetic basis of the parameters determining the response curves will help in the development of breeding programmes for improving abiotic stress tolerance and understanding genotype × environment interaction. In this paper we present two strategies for locating quantitative trait loci for response-curve parameters and estimation of their allele effects. The procedures are illustrated by an application to drought stress in maize.

INTRODUCTION

Strategies for improving tolerance to abiotic stress in plant breeding almost invariably test collections of genotypes, whether segregating or not, across a series of trials chosen to represent as well as possible an environmental gradient relevant to the stress of interest. In such experiments, the genotypes will show differential performance across the stress gradient, as tolerant genotypes will do relatively better under stress conditions, whereas this advantage will disappear in the absence of stress. The differential performance of tolerant versus non-tolerant genotypes in relation to the severity of stress produces genotype × environment interaction (GEI), i.e., the phenomenon that differences between genotypes are environment-

J.H.J. Spiertz, P.C. Struik and H.H. van Laar (eds.), Scale and Complexity in Plant Systems Research: Gene-Plant-Crop Relations, 115-126.

dependent. An adequate analysis of GEI is a prerequisite for success in breeding programmes for abiotic-stress tolerance. In plant breeding, GEI is mostly modelled within the context of analysis of variance models with emphasis on test procedures for detecting GEI. When GEI is found significant, the consequence is that genotypes that perform well under non-stress conditions, cannot automatically be recommended for stress environments. Of course, after the establishment of the existence of GEI, one may be interested to identify the genetic and environmental causes for the observed GEI. Various classes of statistical models have been developed that describe GEI in terms of differential genotypic sensitivities to environmental variables, where the models differ with respect to the character of the explanatory variables that are included (for reviews, see Van Eeuwijk et al. 1996; Van Eeuwijk 2006).

In plant breeding there is a tendency to describe GEI in terms of differences in linear responses to environmental variables, probably because restricted environmental gradients allow linear functions to give acceptable fits. In contrast, physiology-based approaches to modelling plant responses in relation to abiotic-stress gradients typically come in the form of non-linear functions. Whatever the chosen specification for the functional relationship between plant phenotype and stress intensity, an attractive option for studying the genetic basis of stress tolerance and GEI is the mapping of quantitative trait loci (QTLs) for the parameters underlying genotype-specific response curves. By treating estimated curve parameters as standard phenotypic traits, QTLs can be identified for the curve parameters and for the genetic basis of GEI. However, although this approach to the identification of the genetic basis of stress tolerance and GEI is straightforward and requires only standard QTL-mapping software, it has some drawbacks. These drawbacks include neglect of estimation error and correlations between parameters, which can lead to faulty inference on QTLs. A solution is given by an integrated modelling approach to GEI and QTL mapping.

In this paper, we present strategies to identify the genetic basis of GEI and response curves. We first give a brief general repository of QTL-mapping methods. Next, we look at a two-step QTL-mapping approach, in which in the first step the parameters are estimated, and in the second step, these estimates are treated as standard phenotypic traits. The following section describes an integrated modelling framework for GEI and QTL mapping. Finally, some of the methods are illustrated by an example on drought stress in maize.

QTL MAPPING FOR SIMPLE TRAITS BY REGRESSION AND MIXED MODELS

In the regression interpretation of QTL mapping, QTLs can be found by regressing the phenotypic response for genotype i, $\underline{P}_i$, on a quantitative or categorical variable, x_i, where x_i represents a function of QTL genotype probabilities (Haley and Knott 1992; Lynch and Walsh 1998). We will use the convention to underline random variables. As the QTL genotypes cannot be observed, but the marker genotypes can, QTL genotype probabilities are estimated from observed marker genotypes. To give

an example, consider a population of doubled haploids. In such a population, at each locus there are only two possible, homozygous genotypes, say A and B, corresponding to the genotypes of the first and second parent, respectively. For a particular marker locus, x_i then takes the value 1 whenever the marker genotype is equal to A and –1 whenever B is the case. We can calculate for each marker a corresponding regressor, or better, genetic predictor, x_i, and then correlate this predictor with the phenotypic response $\underline{P}_i$. Everywhere where a test statistic, like the F-statistic, for the regression of $\underline{P}_i$ on x_i is significant according to some pre-set criterion, we can conclude that a QTL must be close, with the best estimate for the position depending on the maximum of the test statistic over a certain genome region.

We will elaborate these ideas more formally. We start by defining a model for the r-th phenotypic observation on genotype i, again underlining random variables, as

$$\underline{P}_{ir} = \mu + G_i + \underline{\varepsilon}_{ir} \tag{1}$$

with $\underline{P}_{ir}$ the phenotypic observation, μ the general mean, G_i the underlying genotypic contribution as deviation from the mean, and $\underline{\varepsilon}_{ir}$ a random error. For the random variables, we will assume that they are normally distributed. Fitting a QTL model to the response $\underline{P}_{ir}$, merely means partitioning the genotypic effect, G_i, in a part due to regression on x_i, and a random residual $\underline{G}_i^*$:

$$\underline{P}_{ir} = \mu + x_i \alpha + \underline{G}_i^* + \underline{\varepsilon}_{ir} \tag{2}$$

The parameter α represents a QTL effect for a putative QTL locus at the position corresponding to the genetic predictor x_i. Model (2) is a mixed model because it contains two random terms besides the fixed general mean and the QTL effect, α. Most general-purpose statistical packages, like SAS, SPSS, Genstat, S-Plus and R, have facilities to fit mixed models like Model (2) and other mixed models that will be mentioned below. A general test for the significance of fixed parameters, like α, in a mixed model, like (2), cannot be a standard F-test. An appropriate test for such a parameter in a mixed-model framework is the Wald test (Verbeke and Molenberghs 2000), which is produced by the packages just mentioned.

The use of Model (2) for a genome scan with genetic predictors calculated exclusively at marker positions is called marker regression, while Model (2) with genetic predictors calculated at and in between marker positions is called simple interval mapping (Lynch and Walsh 1998). To change Model (2) into a multiple QTL model with n_Q QTLs, we can write

$$\underline{P}_{ir} = \mu + \sum_{q=1}^{n_Q} x_{iq} \alpha_q + \underline{G}_i^* + \underline{\varepsilon}_{ir}. \tag{3}$$

Model (3) shows that the building of a multiple QTL model can be interpreted as a subset-selection problem (Broman and Speed 2002), i.e., we want to find the set of

genetic predictors that best explains the phenotypic response in terms of QTLs. As a strategy to identify such a subset, composite interval mapping has been developed. Analogously to the situation for simple interval mapping, in composite interval mapping, genome scans are performed by evaluating the effect of individual genetic predictors on a one-dimensional grid along the genome. However, to improve power, the effects of possible QTLs elsewhere on the genome are neutralized as much as possible by a set of so-called co-factors, a set of markers close to putative QTLs identified earlier, for example, in a simple interval-mapping genome scan (Lynch and Walsh 1998). A mixed model that can be used for composite interval mapping is

$$\underline{P}_{ir} = \mu + \sum_{c \in C} x_{ic} \alpha_c + x_{iq} \alpha_q + \underline{G}_i^* + \underline{\varepsilon}_{ir}, \quad (4)$$

with C representing the set of co-factors appropriate for use in combination with x_q, the genetic predictor being tested for possible association with a QTL. The set of co-factors varies in relation to the genome position as individual genetic predictors, x_c, are dropped from the co-factor set when their position comes too close to that of x_q.

QTL MAPPING OF EARLIER ESTIMATED CURVE PARAMETERS

In the previous section, we presented mixed-model methodology for QTL mapping of simple phenotypic traits. For identifying the genetic basis of response curves, we can apply the above methodology without modifications when we are willing to interpret estimated curve parameters as if they were simple phenotypic traits. The consequence of the latter assumption is that we need to ignore the precision of the estimated parameters in subsequent analyses and this can lead to incorrect conclusions on QTL existence, location and effects, where the degree of inaccuracy will increase with the imprecision of the estimates. A very simple approach to QTL mapping of response curves, thus, consists of two steps. First, estimate curve parameters for individual genotypes by means of linear or non-linear regression. Second, treat the curve parameters of the first step as a phenotypic trait in its own right, and apply a QTL-mapping approach to this 'trait'.

More important than the statistical-technical difference between linear and non-linear regression models is the type of environmental variables that is used in the set of explanatory environmental variables. We can distinguish two types of environmental variables or characterizations: those that refer to explicitly measured physical and biological variables, like temperature, rainfall and disease pressure, and those that are implicit in the sense that they are calculated from the phenotypic responses in the environment. For example, the mean response of all genotypes in an environment can serve as an implicit, integrated indicator of environmental quality.

Examples of implicit environmental descriptions occur in some well-known statistical models for describing GEI. A popular class of models for GEI describes GEI by differential genotypic sensitivity to environmental characterizations that are themselves linear functions of observed phenotypes. In the regression on the mean

model, probably better known as the Finlay-Wilkinson model (Finlay and Wilkinson 1963), the characterization for a particular environment, j, is just the average phenotypic performance across all genotypes minus the grand mean, denoted by E_j. The model for the phenotypic mean of genotype i in environment j, $\overline{\underline{P}}_{ij}$, reads

$$\overline{\underline{P}}_{ij} = \mu + G_i + \beta_i E_j + \underline{\delta}_{ij} \tag{5}$$

with μ as the general mean, G_i the average performance of genotype i across all environments given as a deviation from the general mean, a measure for wide adaptation, β_i the sensitivity of genotype i to the environmental characterization E_j, with $\overline{\beta}_i$ =1, and $\underline{\delta}_{ij}$ the error attached to the mean for genotype i in environment j. The parameters G_i and β_i can first be estimated by linear regression and the estimates $\hat{G}_i$ and $\hat{\beta}_i$ can subsequently be introduced as ordinary phenotypic variables in a QTL-mapping procedure.

Model (5) can also be written in a form that better emphasizes the connection of the regression on the mean model with models for GEI by replacing the slopes β_i, which are on average 1, by the slopes b_i, which are on average zero,

$$\overline{\underline{P}}_{ij} = \mu + G_i + E_j + b_i E_j + \underline{\delta}_{ij} \tag{6}$$

The regression on the mean model, (5)/(6), has rather restricted versatility for modelling differences in phenotypic responses between genotypes across environments. A more flexible model, following the same philosophy, characterizing the environment on the basis of the phenotypic trait itself, is the 'additive main effects and multiplicative interactions' model (Gollob 1968; Mandel 1969; Gabriel 1978; Gauch Jr 1988):

$$\overline{\underline{P}}_{ij} = \mu + G_i + E_j + \sum_{k=1}^{K} a_{ki} b_{kj} + \underline{\delta}_{ij} \tag{7}$$

with hypothetical environmental characterizations b_{kj} that create maximum discrimination for the corresponding genotypic sensitivities, a_{ki}. The number of multiplicative terms necessary for an adequate description of the data is K. Various test procedures have been developed to assess K (Gollob 1968; Cornelius et al. 1996). Estimates for the genotypic sensitivities, $\hat{a}_{ki}$, and environmental characterizations, $\hat{b}_{kj}$, can be obtained by least-squares estimation procedures. Estimated genotypic sensitivities according to Model (7) can be mapped by ordinary QTL-mapping procedures.

The response Models (5) and (7) are attractive for plant breeders, because they do not require an explicit environmental characterization. However, when physical or biological descriptions of the environment are available, physiologically more attractive models would include explicit references to such descriptions. Factorial

regression models are linear models with multiplicative terms for GEI that can model differential genotypic responses to one or more physical or biological environmental characterizations (Denis 1988; Van Eeuwijk et al. 1996). An example with one environmental variable, z_j, has the form

$$\overline{\underline{P}}_{ij} = \mu + G_i + E_j + b_i z_j + \underline{\delta}_{ij} \quad (8)$$

where z_j could be a function of temperature during a critical growth stage, radiation, nitrogen, water, etc., and b_i then is the corresponding genotypic sensitivity. Equivalently, z_j, can be a stress index obtained from a crop growth simulation. Model (8) can also model non-linear responses, for example by including polynomial terms,

$$\overline{\underline{P}}_{ij} = \mu + G_i + E_j + b_{1i} z_j + b_{2i} z_j^2 + \underline{\delta}_{ij} \quad (9)$$

where b_{1i} and b_{2i} represent the sensitivity of genotype i to the linear and quadratic term in the environmental variable z. Even the inclusion of response surfaces in various dimensions does not present statistical-technical problems, although the number of environments necessary for sufficiently precise estimation of the increasing number of regression parameters will not often be reached in plant-breeding programmes.

When good explicit environmental characterizations are available, it is often preferable to model the genotypic responses by parametric linear and non-linear regression functions based on physiological insights, control equations (Reymond et al. 2003; Tardieu 2003; Tardieu et al. 2005) or meta-mechanisms (Hammer et al. 2005), instead of working with polynomial approximations to these non-linear functions. A general expression for non-linear genotypic responses in one dimension is

$$\overline{\underline{P}}_{ij} = f(\theta_i, z_j) + \underline{\delta}_{ij} \quad (10)$$

with f representing a non-linear function in the parameter vector θ_i for genotype i and z_j, as before, the value for the environmental variable z in environment j. Model (10) is equally applicable in temporal contexts, with z related to time, as in spatial contexts, where z typically is related to soil and management conditions. When z is time related, the error term $\underline{\delta}_{ij}$ demands careful modelling of possible auto-correlations between observations at short intervals. Two illustrative examples of physiological modelling of response curves followed by QTL mapping of the estimated curve parameters are Reymond et al. (2003) for linear parameters and Yin et al. (2005) for non-linear parameters.

For all models discussed in this section, genotype-specific parameters can first be estimated and then subjected to a standard QTL analysis. This practice will work reasonably well as long as the hypothesized curves fit the observed data well across

the full set of genotypes and the standard errors for the parameter estimates are relatively small in comparison to the parameter estimates themselves. Still, a better approach is to model the curve parameters directly as functions of underlying QTLs. The next section describes how to do this.

MODELLING GENOTYPIC RESPONSES AS FUNCTIONS OF QTLS

An integrated approach to the problem of mapping the genetic basis of response curves departs from the development of a multi-environment model for genotypic responses observed across a series of environments. A QTL model for the mean of genotype i in environment j with the possibility of the QTL effect being environment-specific is

$$\underline{P}_{ij} = \mu + E_j + x_i \alpha_j + \underline{\delta}_{ij} \tag{11}$$

with α_j standing for the environment-specific QTL effect in environment j. The generalization of Model (11) to a multi-QTL model would look like

$$\underline{P}_{ij} = \mu + E_j + \sum\nolimits_{q=1}^{n_Q} x_{iq} \alpha_{jq} + \underline{\delta}_{ij} \tag{12}$$

In Model (12) not necessarily each QTL needs to exhibit environment-specific expression. For some QTLs, the expression across environments may be more or less constant, so that a single, main QTL effect would suffice. A more correct QTL model for a multi-environment trial is then

$$\underline{P}_{ij} = \mu + E_j + \sum\nolimits_{q^*=1}^{n_{Q^*}} x_{iq^*} \alpha_{q^*} + \sum\nolimits_{q=1}^{n_Q} x_{iq} \alpha_{jq} + \underline{\delta}_{ij} \tag{13}$$

with the first set of QTLs, Q^*, just having constant expression across environments and the second set, Q, having environment-specific expression. The variance–covariance matrix for the residuals $\underline{\delta}_{ij}$ should be flexible enough to allow for heterogeneity of variance across environments and heterogeneity of correlations between environments due to genetic effects not modelled by the QTL part of the model (Piepho 2000; Verbyla et al. 2003; Malosetti et al. 2004; Piepho and Pillen 2004).

A QTL model in which QTL expression is modelled in direct dependence on environmental variables can be obtained from Model (11) by regressing the QTL effects, α_j, on an environmental variable z_j, $\underline{\alpha}_j = a_0 + a_1 z_j + \underline{\alpha}^*_j$:

$$\underline{P}_{ij} = \mu + E_j + x_i a_0 + x_i a_1 z_j + x_i \underline{\alpha}^*_j + \underline{\delta}_{ij} \tag{14}$$

where a_0 represents the QTL main-effect expression, the part that is constant across environments, while a_1 is a proportionality constant that shows how much the

phenotype will change per unit change in the environmental variable z, this phenotypic change being conditioned on the QTL genotype information contained in x_i. The part of the QTL effect α_j that is not described by the regression on z determines a random residual QTL effect, $\underline{\alpha}_j^*$. Model (14) can be extended in obvious ways by the incorporation of further polynomial terms in z, and by incorporating different environmental variables for different QTLs.

For non-linear response curves as described in Model (10), a QTL model can be constructed by modelling each of the genotypic parameters in the parameter vector θ in terms of underlying linear (multi-)QTL models. The QTL model for the k-th genotypic parameter is then

$$\underline{\theta}_i^k = \mu^k + \sum_{q=1}^{n_{Q^k}} x_{iq} \alpha_q^k + \underline{\varepsilon}_i^k , \tag{15}$$

with the superscript k referring to the parameter within the vector θ, μ^k an intercept term, α_q^k the effect of the q-th QTL for the parameter θ^k, $q = 1...n_Q$, Q^k the set of QTLs underlying θ^k, and $\underline{\varepsilon}_i^k$ a residual term. The set of QTLs underlying a particular genotypic parameter can differ between parameters. Substituting Model (15) for each of the parameters θ^k in Model (10) will convert the latter phenotypic model into a QTL model for non-linear responses.

All QTL models treated so far, except the model for the non-linear responses, are linear mixed models, and parameter estimation and testing follow standard theory for this type of models (Verbeke and Molenberghs 2000). The QTL model for the non-linear responses is a non-linear mixed model (Davidian and Giltinan 2003; Malosetti et al. 2006). Estimation and testing for this class of models is more complex and requires special procedures. Such procedures are present in SAS and S-Plus/R.

None of the QTL models discussed so far is critically dependent on the use of segregating populations of offspring from crosses between inbred parents. The models are equally applicable to the analysis of arbitrary collections of varieties as in marker–trait association analyses. The difference between the former linkage and the latter linkage disequilibrium analyses resides mainly in the incorporation of extra random terms to correct for the varying level of genetic relatedness that characterizes arbitrary collections of varieties (Malosetti 2006; Yu et al. 2006).

EXAMPLE

To illustrate some of the concepts above, we reanalysed data from the CIMMYT maize-breeding programme on drought tolerance, consisting of yield evaluations for 211 F2-derived F3 families across eight trials with varying levels of water and nitrogen stress. Detailed descriptions and more analyses of these data can be found in Malosetti et al. (in press), Van Eeuwijk et al. (2001; 2002) and Vargas et al. (2006). In this chapter, we will present some results related to chromosome 1.

To model GEI in the maize data, we fitted a regression on the mean Model (5), an AMMI model (7) with two terms for interaction, and a number of factorial

regression models (8) trying out a series of environmental variables, among which the minimum temperature during flowering and an environmental contrast between trial performance at the location of Poza Rica versus that at Tlaltizapán. The latter contrast is an example of a qualitative environmental variable. Figure 1 shows LOD profiles of simple interval-mapping scans with MapQTL (Van Ooijen 2004) for a selection of five parameters estimated in the GEI analyses. Figure 1 shows that there is some indication for a QTL main effect (intercept) in the region of 140-180 cM and no proof for QTLs related to the slopes in the regression on the mean model. For the AMMI-2 scores, the slopes for minimum temperature during flowering, and the genotypic contrasts for performance at Poza Rica versus Tlaltizapán, the LOD profiles look very similar, indicating a significant QTL for those parameters at 130-150 cM. All three parameters represent differential genotypic sensitivity to the same environmental contrast. The minimum temperature at flowering was higher at Tlaltizapán than at Poza Rica. The same environmental contrast between those two locations determined the second AMMI axis.

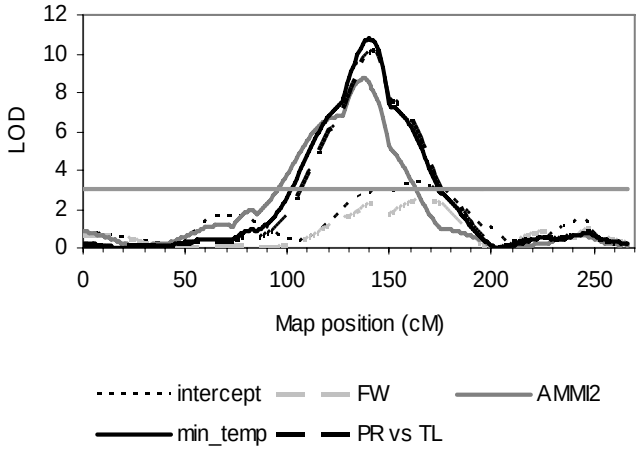

Figure 1. *LOD profiles for simple interval scans for chromosome 1 on different parameters characterizing phenotypic responses across environments for a set of F2-derived F3 lines in the CIMMYT drought-stress programme for maize. Intercept = genotypic mean across environments, FW = genotypic slope from the regression on the mean model, AMMI2 = genotypic score for the second multiplicative term in an AMMI model, min-temp = genotypic slope from factorial regression with minimum temperature during flowering, PR vs TL = genotypic contrasts for the average difference in performance between the locations Poza Rica and Tlaltizapán. Threshold for significance was chosen at an LOD score of 3*

We subsequently modelled the whole of the genotype × environment data by a mixed model with environment-specific QTLs, Model (11) and detected again a QTL in the region from 130 to 150 cM. The next step in a mixed-model QTL analysis of multi-environment data then consists of searching for environmental variables that can explain the differential QTL expression across environments.

Figure 2 shows the regression of the QTL effects on the minimum temperature during flowering. With increasing temperature, the QTL allele coming from the high-yielding, drought-susceptible parent gives less advantage. At minimum temperatures around 10 °C, the temperature at Tlaltizapán, the QTL allele of the drought-susceptible, but high-yielding parent, still confers a yield increase of around 0.6 tons per hectare in comparison to the QTL allele from the drought-tolerant parent. At temperatures around 20 °C, the temperature at Poza Rica, the yield advantage of the QTL allele from the high-yielding parent has disappeared.

In the Sections *QTL mapping of earlier estimated curve parameters* and *Modelling genotype responses as functions of QTLs,* we discussed the step-wise QTL analysis on estimated parameters and the mixed-model QTL analysis of multi-environment data. Both types of QTL analyses for response curves produced similar conclusions for the example data. However, in general the multi-environment mixed-model QTL analysis would have our preference because of its more appropriate representation of variances and correlations and its more transparent statistical properties.

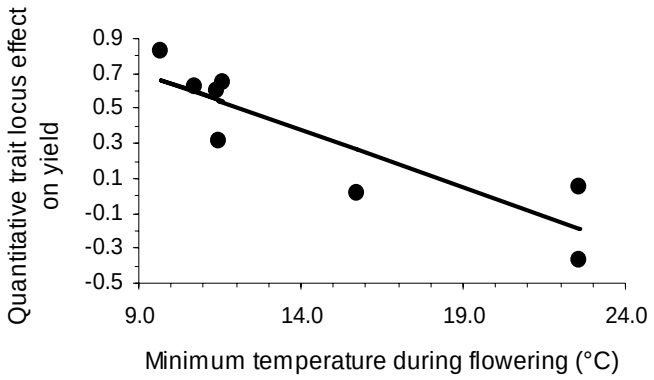

***Figure 2**. QTL effect at chromosome 1 as function of minimum temperature during flowering time*

ACKNOWLEDGEMENTS

This chapter was written as part of competitive project number 4 within the Generation Challenge Program: “An eco-physiological – statistical framework for the analysis of G×E and QTL×E as occurring in abiotic stress trials, with applications to the CIMMYT drought stress programs in tropical maize and bread wheat”. We thank Jean-Marcel Ribaut (GCP and CIMMYT) for permission to use the data in the example, and José Crossa and Mateo Vargas (CIMMYT) for preparation of the data and discussion.

REFERENCES

Broman, K.W. and Speed, T.R., 2002. A model selection approach for the identification of quantitative trait loci in experimental crosses. *Journal of the Royal Statistical Society. Series B. Statistical Methodology,* 64 (4), 641-656.

Cornelius, P.L., Crossa, J. and Seyedsadr, M.S., 1996. Statistical tests and estimators of multiplicative models for genotype-by-environment interaction. *In:* Kang, M.S. and Gauch, H.G. eds. *Genotype-by-environment interaction*. CRC Press, Boca Raton, 199-234.

Davidian, M. and Giltinan, D.M., 2003. Nonlinear models for repeated measurement data: an overview and update. *Journal of Agricultural Biological and Environmental Statistics,* 8 (4), 387-419.

Denis, J.-B., 1988. Two-way analysis using covariates. *Statistics,* 19 (1), 123-132.

Finlay, K.W. and Wilkinson, G.N., 1963. The analysis of adaptation in a plant-breeding programme. *Australian Journal of Agricultural Research,* 14 (6), 742-54.

Gabriel, K.R., 1978. Least squares approximation of matrices by additive and multiplicative models. *Journal of the Royal Statistical Society. Series B. Statistical Methodology,* 40 (2), 186-196.

Gauch Jr, H.G., 1988. Model selection and validation for yield trials with interaction. *Biometrics,* 44 (3), 705-715.

Gollob, H.F., 1968. A statistical model which combines features of factor analytic and analysis of variance techniques. *Psychometrika,* 33 (1), 73-115.

Haley, C.S. and Knott, S.A., 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity,* 69 (4), 315-324.

Hammer, G.L., Chapman, S., Van Oosterom, E., et al., 2005. Trait physiology and crop modelling as a framework to link phenotypic complexity to underlying genetic systems. *Australian Journal of Agricultural Research,* 56 (9), 947-960.

Lynch, M. and Walsh, B., 1998. *Genetics and analysis of quantitative traits*. Sinauer Associates Inc., Sunderland.

Malosetti, M., 2006. *Mixed model methodology for the identification of genetical factors underlying trait variation in plants*. PhD Thesis, Wageningen University, Wageningen.

Malosetti, M., Voltas, J., Romagosa, I., et al., 2004. Mixed models including environmental covariables for studying QTL by environment interaction. *Euphytica,* 137 (1), 139-145.

Malosetti, M., Visser, R.G.F., Celis-Gamboa, C., et al., 2006. QTL methodology for response curves on the basis of non-linear mixed models, with an illustration to senescence in potato. *Theoretical and Applied Genetics,* 113 (2), 288-300.

Malosetti, M., Ribaut, J.-M., Vargas, M., et al., in press. Multi-trait multi-environment QTL modelling for drought stress adaptation in maize. *In:* Spiertz, J.H.J., Struik, P.C. and Van Laar, H.H. eds. *Scale and complexity in plant systems research: gene-plant-crop relations: proceedings of the Frontis workshop on scale and complexity in plant systems research: gene-plant-crop relations, Wageningen, The Netherlands, April 24-26, 2006*. Springer, Dordrecht. Wageningen UR Frontis Series no. 21.

Mandel, J., 1969. The partitioning of interaction in analysis of variance. *Journal of Research of the National Bureau of Standards. Section B. Mathematical Sciences,* 73B, 309-328.

Piepho, H.P., 2000. A mixed-model approach to mapping quantitative trait loci in barley on the basis of multiple environment data. *Genetics,* 156 (4), 2043-2050.

Piepho, H.P. and Pillen, K., 2004. Mixed modelling for QTL x environment interaction analysis. *Euphytica,* 137 (1), 147-153.

Reymond, M., Muller, B., Leonardi, A., et al., 2003. Combining quantitative trait loci analysis and an ecophysiological model to analyze the genetic variability of the responses of maize leaf growth to temperature and water deficit. *Plant Physiology,* 131 (2), 664-675.

Tardieu, F., 2003. Virtual plants: modelling as a tool for the genomics of tolerance to water deficit. *Trends in Plant Science,* 8 (1), 9-14.

Tardieu, F., Reymond, M., Muller, B., et al., 2005. Linking physiological and genetic analyses of the control of leaf growth under changing environmental conditions. *Australian Journal of Agricultural Research,* 56 (9), 937-946.

Van Eeuwijk, F.A., 2006. Genotype by environment interaction: basics and beyond. *In:* Lamkey, K.R. and Lee, M. eds. *Plant breeding: the Arnel R. Hallauer international symposium*. Blackwell Publishing, Oxford, 155-170.

Van Eeuwijk, F.A., Denis, J.-B. and Kang, M.S., 1996. Incorporating additional information on genotypes and environments in models for two-way genotype by environment tables. *In:* Kang, M.S. and Gauch, H.G. eds. *Genotype-by-environment interaction*. CRC Press, Boca Raton, 15-50.

Van Eeuwijk, F.A., Crossa, J., Vargas, M., et al., 2001. Variants of factorial regression for analysing QTL by environment interaction. *In:* Gallais, A., Dillmann, C. and Goldringer, I. eds. *Quantitative genetics and breeding methods: the way ahead: proceedings of the 11th meeting of the EUCARPIA section biometrics in plant breeding, Paris, France, 30/31 August - 1 September, 2000*. Institut National de la Recherche Agronomique, Paris, 107-116.

Van Eeuwijk, F.A., Crossa, J., Vargas, M., et al., 2002. Analysing QTL-environment interaction by factorial regression, with an application to the CIMMYT drought and low-nitrogen stress programme in maize. *In:* Kang, M.S. ed. *Quantitative genetics, genomics and plant breeding*. CAB International, Wallingford, 245-256.

Van Ooijen, J.W., 2004. *MapQTL 5: software for the mapping of quantitative trait loci in experimental populations*. Kyazma, Wageningen.

Vargas, M., Van Eeuwijk, F.A., Crossa, J., et al., 2006. Mapping QTLs and QTL × environment interaction for CIMMYT maize drought stress program using factorial regression and partial least squares methods. *Theoretical and Applied Genetics,* 112 (6), 1009-1023.

Verbeke, G. and Molenberghs, G., 2000. *Linear mixed models for longitudinal data*. Springer Verlag, New York.

Verbyla, A.P., Eckermann, P.J., Thompson, R., et al., 2003. The analysis of quantitative trait loci in multi-environment trials using a multiplicative mixed model. *Australian Journal of Agricultural Research,* 54 (11/12), 1395-1408.

Yin, X., Struik, P.C., Van Eeuwijk, F.A., et al., 2005. QTL analysis and QTL-based prediction of flowering phenology in recombinant inbred lines of barley. *Journal of Experimental Botany,* 56 (413), 967-976.

Yu, J.M., Pressoir, G., Briggs, W.H., et al., 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics,* 38 (2), 203-208.