Genet. Res., Camb. (2006), **88**, *pp.* 119–131. © 2006 Cambridge University Press
doi:10.1017/S0016672306008391 Printed in the United Kingdom

Gene–environment interactions in complex diseases: genetic models and methods for QTL mapping in multiple half-sib populations

HAJA N. KADARMIDEEN*, YONGJUN LI AND LUC L. G. JANSS
Statistical Animal Genetics Group, Institute of Animal Science, Swiss Federal Institute of Technology, ETH Zentrum (UNS D7), Universitaetstrasse 65, Zurich 8092, Switzerland

(*Submitted 9 January 2006 and in revised form 28 May and 17 July 2006*)

Summary

An interval quantitative trait locus (QTL) mapping method for complex polygenic diseases (as binary traits) showing QTL by environment interactions (QEI) was developed for outbred populations on a within-family basis. The main objectives, within the above context, were to investigate selection of genetic models and to compare liability or generalized interval mapping (GIM) and linear regression interval mapping (RIM) methods. Two different genetic models were used: one with main QTL and QEI effects (QEI model) and the other with only a main QTL effect (QTL model). Over 30 types of binary disease data as well as six types of continuous data were simulated and analysed by RIM and GIM. Using table values for significance testing, results show that RIM had an increased false detection rate (FDR) for testing interactions which was attributable to scale effects on the binary scale. GIM did not suffer from a high FDR for testing interactions. The use of empirical thresholds, which effectively means higher thresholds for RIM for testing interactions, could repair this increased FDR for RIM, but such empirical thresholds would have to be derived for each case because the amount of FDR depends on the incidence on the binary scale. RIM still suffered from higher biases (15–100% over- or under-estimation of true values) and high standard errors in QTL variance and location estimates than GIM for QEI models. Hence GIM is recommended for disease QTL mapping with QEI. In the presence of QEI, the model including QEI has more power (20–80% increase) to detect the QTL when the average QTL effect is small (in a situation where the model with a main QTL only is not too powerful). Top-down model selection is proposed in which a full test for QEI is conducted first and then the model is subsequently simplified. Methods and results will be applicable to human, plant and animal QTL mapping experiments.

1. Introduction

Most biological traits and diseases in animals, plants and humans have a complex or multifactorial inheritance (controlled by numerous interacting genes with small and large effects) and are affected as well as by environmental factors such as nutrition, hygiene, lifestyle and management. A chromosomal region that contains one or more genes that influence a multifactorial trait is known as a quantitative trait locus or QTL. Standard segregation analysis in informative families or experimental crosses to map QTLs is well established using either regression or maximum likelihood based approaches (Haley & Knott, 1992), generalized linear (mixed) models (Kadarmideen *et al.*, 2000; Kadarmideen & Dekkers, 2001; Thomson, 2003) or Bayesian methods (Xu & Yi, 2000). The power of such analyses to detect and map QTLs (under different genetic models) depends, among other factors, on the proportion of the phenotypic variance explained by QTLs, the design and size of the segregating population, and the type and number of genetic markers.

It is well known that gene expression differs depending on the environmental conditions, which could be due to gene by environment correlations or interactions. A QTL by environment interaction (QEI) may exist when a QTL is detected at a specific

* Corresponding author. e-mail: haja.kadarmideen@inw.agrl.ethz.ch

map location in one environment but not in another and/or when the effects of a QTL between environments are significantly different. Mostly, QTL mapping methods consider gene and environmental effects independently of each other and do not take into account interaction among such effects. That is, the effects of QTLs are assumed be the same across all environmental conditions. Common environmental factors include, for example, soil types, day length and general temperature regimes, etc. affecting plant growth; disease pathogenesis and fertility; farm hygiene and nutrition for animal growth, disease development and fertility; and cancer or diabetes episodes due to lifestyle changes or exposure to a poor environment in humans. Genes or QTLs with an effect on these phenotypes may in fact be interacting with these environmental factors, contributing to the manifestation of a given phenotype.

In the genetic epidemiology of human diseases, QEI models would be of help when counselling on the relative risks for contracting a given disease depending on the environment. For example, while exposure to ultraviolet light increases the risk of developing skin cancer in non-carriers of xeroderma pigmentosum (XP) mutations, the combination of these mutations and exposure to ultraviolet light vastly increases the risk of skin cancer (Hunter, 2005). In companion animals, many congenital diseases or cancer risks are not well studied for QEI interactions, but these may exist, as they do in humans. In farm animal/plant/tree breeding terms, QEI models would be of help for predicting the genetic merit of individuals (at the QTL level) depending on the environment in which they are reared, paving a way for environment-specific gene-assisted selection.

While including QEI in genetic models is useful, ignoring it when present would lead to a poor accuracy/confidence interval for the genomic position of QTLs and low power to detect them, as shown by Wang *et al*. (1999). The QEI effects can be expected to be present especially in the prevalence of certain diseases among individuals raised in different environmental conditions: for example, when a QTL confers resistance to a pathogen specific to particular environments.

Some studies have investigated gene–gene interaction (epistasis) as well as QEI in QTL mapping for normally distributed traits in crosses from inbred lines or breeds (e.g. Jansen *et al*., 1995; Wang *et al*., 1999; Juenger *et al*., 2005). Most often diseases are observed in binary form, the organism being either 'healthy' or 'diseased'. No study has thus far developed a QTL mapping method for a QTL with QEI in an outbred population on a within-family basis where (disease) data are observed as binary traits or compared different genetic models (with and without QEI) and statistical methods (linear versus liability methods) for binary traits.

It is expected that linear models would show spurious interactions for binary traits (such as diseases) due to scale effects while threshold models would not in principle, suffer from such artefacts. No study has so far investigated this phenomenon or proved this expectation in the context of QTL mapping for binary traits (regardless of which designs were used). The objective of this study was to investigate whether linear models are indeed inappropriate for mapping QTL with interactions for binary traits. Hence we compare generalized interval mapping (GIM) with linear regression interval mapping (RIM) methods when QEI are indeed present or absent for (binary) disease episodes in within-family analyses for an outbred population.

2. Materials and methods

(i) *QTL mapping experimental design and liability models*

Multiple half-sib families (with the ith family having n_i progeny) distributed across j different environments (for $j=1$ to r) are considered, with the common parent in each family being called the 'sire' (sires $i=1$ to s) and its mates 'dams'. The expression of the sire QTL alleles is assumed to differ between environments such that the progeny of the same sire deviate from the main QTL effect by an amount equal to QEI effects, depending on which environment they are in. It is assumed here that these environments are quite distinct and limited in number (e.g. tropical versus temperate climate). Let N be the total number of progeny across all sires and environments. Progeny of sires are affected by a multifactorial disease that is observed in one of two categories as healthy or diseased (binary scores, 0 or 1). Genotype data for multiple marker brackets are available on the offspring and on their sires, but dams may or may not have that information. All sires are assumed to be fully informative (heterozygous) for the marker bracket under consideration. Sires can be QTL informative or uninformative, with alleles Q_1 and Q_2 (with frequencies f and $1-f$, respectively).

(ii) *Genetic threshold models*

Threshold or liability models for genetic analysis of traits measured in categorical or binary scales (e.g. diseases) in pedigree populations have been described previously for QTL mapping (e.g. Kadarmideen *et al*., 2000; Kadarmideen & Dekkers, 2001) and segregation analyses for detecting major genes (e.g. Kadarmideen & Janss, 2005). Briefly, let y_{ijk} be a disease (binary) response variable observed (as $y_{ijk}=0$

https://doi.org/10.1017/S0016672306008391 Published online by Cambridge University Press

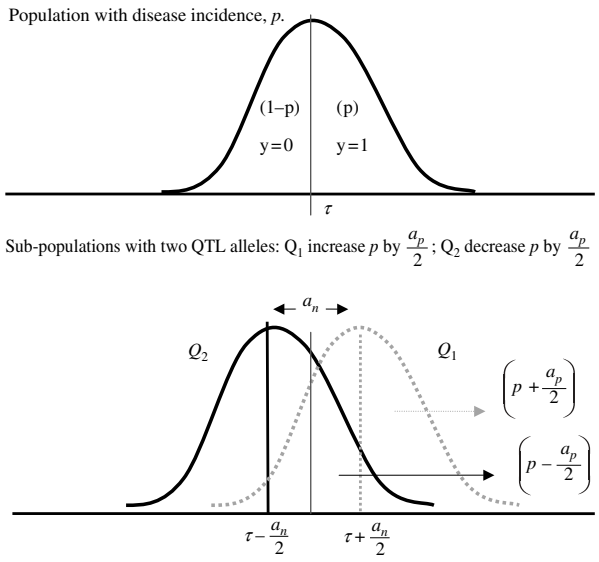

Fig. 1. Illustration of linkage between QTL effects on liability versus observed scales for binary disease traits. Top: General population with incidence p and threshold τ. Bottom: The subpopulation shown with an unbroken line has the Q_2 allele and that shown with a dotted line has the Q_1 allele.

for a 'healthy' and $y_{ijk}=1$ for a 'disease' phenotype) in offspring k in sire family i and environment j. Threshold theory assumes that y_{ijk} results from an underlying linear predictor z_{ijk}, called liability, that is normally distributed and that models the probability of $y_{ijk}=1$ using the Normal cumulative distribution function (CDF) as the *link* function.

Application of the liability concept to QTL mapping for binary traits is illustrated in Fig. 1, where the population threshold, τ, for the disease results in incidence p in the general population which is stratified into two QTL subpopulations depending on inheritance of either a 'healthy' Q_2 or 'disease' Q_1 allele from the common parent. The liability of these two subpopulations in Fig. 1 differs from the liability of the general population (due to inheritance of the healthy or disease allele) and hence in their incidences; their differences correspond to substitution effects on the liability scale, a_n, or on the probability scale, a_p, as shown below. In the case of QEI, these substitution effects themselves differ between environments. Based on Fig. 1, the QTL substitution effect as a difference of two subpopulations (Q_1-Q_2) on the probability scale is

$$a_p=\left(p+\frac{a_p}{2}\right)-\left(p-\frac{a_p}{2}\right)$$

and on the liability scale is

$$a_n=\left(\tau+\frac{a_n}{2}\right)-\left(\tau-\frac{a_n}{2}\right).$$

(iii) *The null model*

The null model is a linear statistical model for liability which does not fit genetic effects at QTL to disease data but only environmental effects and polygenic effects of the common parent i, as:

$$z_{ijk}=\mu+u_i+\varepsilon_j+e_{ijk} \quad (1)$$

with $i=1, 2, \ldots, s$, $j=1, 2, \ldots, r$ and $k=1, 2, \ldots, m$, where z_{ijk} is the liability of the kth offspring from the ith sire family, in the jth environment, μ is the overall mean, u_i is the polygenic mean for the ith sire family as fixed effect, ε_j is the fixed effect of environment j, and e_{ijk} is a residual, with $e_{ijk}\sim N(0, \sigma_e^2)$.

(iv) *The QTL model*

The QTL model fits a QTL effect to disease data in addition to those in the null model, as:

$$z_{ijk}=\mu+u_i+\varepsilon_j+c_{ijk}\beta_i+e_{ijk} \quad (2)$$

where β_i is the QTL substitution effect for the ith sire and c_{ijk} is the conditional probability of transmission of allele Q_1 from the ith sire to the kth offspring. The conditional probabilities for QTL allele transmission from sire to offspring (c_{ijk}) were assigned as in Kadarmideen *et al.* (2000). This is a function of the marker genotypes of offspring and parents, marker–QTL linkage phase in 'sires', marker allele frequencies and genomic position at which the model is fitted.

(v) *The QTL–environment interaction (QEI) model*

The QEI model fits the QEI term in the genetic model in addition to those in the QTL model (2) assuming the presence of QEI effects in the data. For simplicity, it is assumed that background polygenes are not interacting with environments, but only the QTLs. The QEI model is written as

$$z_{ijk}=\mu+u_i+\varepsilon_j+c_{ijk}\beta_{ij}+e_{ijk}. \quad (3)$$

Note that the QEI part, $c_{ijk}\beta_{ij}$, essentially results in r different expressions (substitution effects, a_n or a_p) of the QTL from the same sire across environments ($s\times r$ effects) due to nesting sire QTL effects within environments. The estimated parameters from this model actually represent both QTL as well as QEI effects, as seen by writing the above model equivalently as: $z_{ijk}=\mu+u_i+\varepsilon_j+c_{ijk}\beta_i+c_{ijk}\beta_{ij}+e_{ijk}$. However, there is a linear dependency between QTL versus QEI effects in this model. There are two ways to estimate QTL and QEI effects: first, one of the r environments within all s QTL effects would be set to zero, resulting in ($sr-s$) QEI effects and s main QTL effects, with a total of sr estimable effects; second,

https://doi.org/10.1017/S0016672306008391 Published online by Cambridge University Press

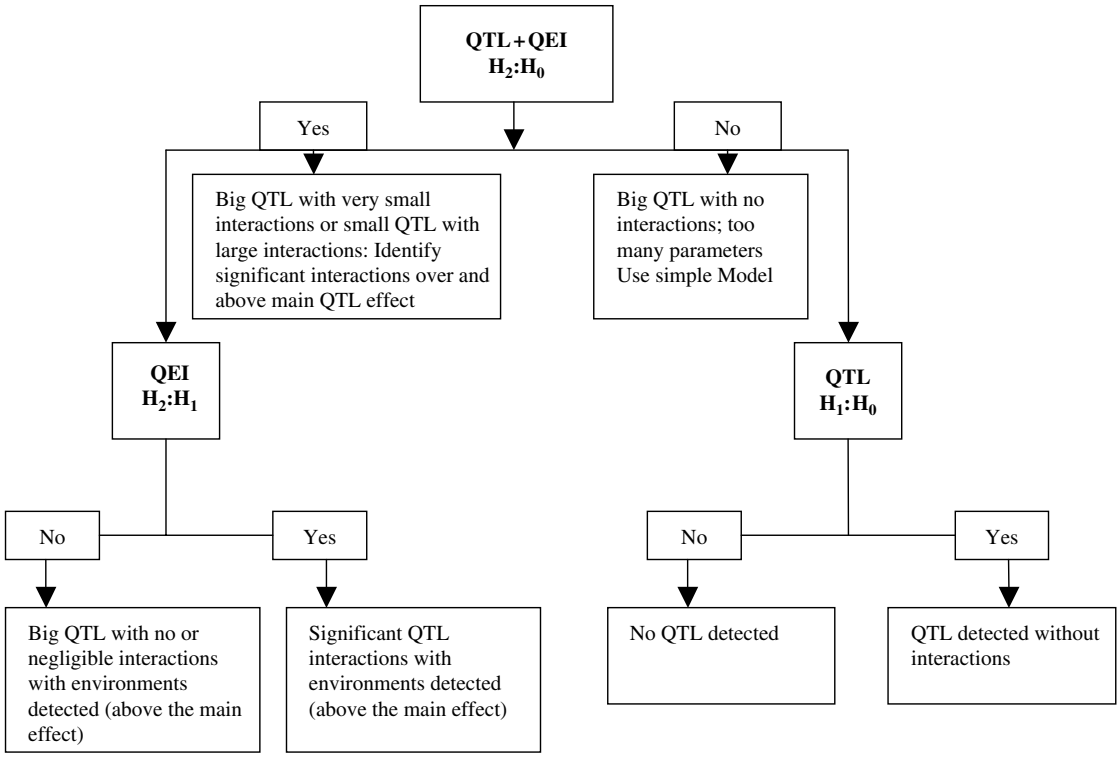

Fig. 2. Top-down selection scheme of genetic models to detect and map QTLs with or without environmental interactions for binary disease traits.

s main QTL effects are to be set to zero with a total of *sr* estimable effects. The two approaches are equivalent, but tests derived from each have different interpretations. The QEI model (3) was set up where *s* main QTL effects were set to zero to be able to estimate *sr* QEI effects.

(vi) *Statistical methods*

(a) *Generalized interval mapping* (*GIM*)

The general form of log-likelihood for the binomial distribution of disease episodes from *s* unrelated sire families across *r* environments, *L*, can be written, based on principles in Kadarmideen *et al.* (2000), as the sum of the log-likelihood for each offspring over all families, environments and interaction of family-specific QTLs with the environment:

$$L=\sum_{i=1}^{s}\sum_{j=1}^{r}\sum_{k=1}^{m}[y_{ijk}\ln(\pi_{ijk})+(1-y_{ijk})\ln(1-\pi_{ijk})]. \quad (4)$$

Parameters were estimated as the joint maximum likelihood of u_is, ε_js, β_is, β_{ij} and recombination rate between QTL and markers, depending on the genetic model (1, 2 or 3) using a Newton–Raphson algorithm, and the corresponding likelihood is computed as shown in the Appendix of Kadarmideen *et al.* (2000). With GIM, QTL and QEI variances were estimated as shown in Appendix A. Computation of the log-likelihood under different genetic models, L_{null}, L_{QTL} and $L_{QTL+QEI}$, for GIM is shown in Appendix B.

To test for the presence of a potentially interacting QTL we adopt here a top-down model selection procedure (Fig. 2). This top-down scheme was used with the rationale that bottom-up testing may fail to find interacting QTLs when they have a small main effect only; also top-down model selection here is readily feasible because the number of interaction models to be tested remains small. This is unlike the case of testing for QTL × QTL interactions (epistasis), where the number of potential interaction models is large and a top-down selection scheme may be computationally prohibitive (see also our Section 4). Our top-down model selection starts with testing for joint main (QTL) and interaction effects (QEI) using model (3) and the null hypothesis of $\beta_{ij}=0$ for all *i* and *j*, with the LR test statistic $\mathrm{LR}_{\mathrm{QTL+QEI}}=2[L_{QTL+QEI}-L_{null}]$. This test may be positive also when there is a main effect only; therefore, if positive, model selection proceeds to test for the presence of interaction above the main effect, using $\mathrm{LR}_{\mathrm{QEI}}=2[L_{QTL+QEI}-L_{QTL}]$. If the QTL + QEI test is negative, this does not yet preclude the presence of a QTL with main effect only, and therefore, if the QTL + QEI test fails, the top-down approach proceeds with a test for a QTL with main effect only based on model (2) and using $\mathrm{LR}_{\mathrm{QTL}}=2[L_{QTL}-L_{null}]$.

https://doi.org/10.1017/S0016672306008391 Published online by Cambridge University Press

Table 1. *Parameters used in simulation of QTL and QTL–environment interactions effects and variances for binary disease data with different QTL substitution effects across environments*

	QTL substitution effect (α_j)						$\sigma^2_{QTL_r}$	
σ^2_{QEI}	E1	E2	E3	E4	E5	Average	liability scale	Binary scale
$\bar{q}=0{\cdot}05$								
0	0·05	0·05	0·05	0·05	0·05	0·05	1·3	0·2
10 %	−0·29	−0·25	0·13	0·21	0·45	0·05	41·0	5·7
$\bar{q}=0{\cdot}15$								
0	0·15	0·15	0·15	0·15	0·15	0·15	11·3	1·7
10 %	−0·19	−0·14	0·23	0·31	0·55	0·15	50·7	6·5
$\bar{q}=0{\cdot}30$								
0	0·30	0·30	0·30	0·30	0·30	0·30	45·0	6·5
10 %	−0·04	0·01	0·38	0·46	0·70	0·30	84·8	9·3

(b) *Regression interval mapping* (*RIM*)

The RIM based on three different genetic models was exactly the same as described for GIM, except that y_{ijk} replaces liability z_{ijk} and likelihoods are computed based on the residual sums of squares (RSS) as shown below. Estimates of σ^2_{QTL}, $\sigma^2_{QTL+QEI}$ and σ^2_{QEI} were computed as for GIM.

As per our top-down approach, we first test for significance of the presence of a QTL as well as the QEI and compute the LR test statistic as: $\mathrm{LR}_{\mathrm{QTL+QEI}} = N \ln(RSS_{null}/RSS_{QTL+QEI})$, where $RSS_{QTL+QEI}$ is the RSS obtained from fitting the QEI model (3) under the alternative hypothesis ($\beta_{ij} \neq 0$ for at least one *ij*). If positive test results are found for joint QTL and QEI effects, the next test is for significance of the presence of only the QEI where the LR test statistic is computed as $\mathrm{LR}_{\mathrm{QEI}} = N \ln(RSS_{QTL}/RSS_{QTL+QEI})$. If the test results for joint QTL and QEI effects are negative, the test is for significance of the presence of a QTL in the marker bracket. The LR test statistic for this test was computed as $\mathrm{LR}_{\mathrm{QTL}} = N \ln(RSS_{null}/RSS_{QTL})$ where RSS_{null} is the RSS obtained from fitting a null model under the null hypothesis (null model 1), $\beta_i = 0$ for all *i*, and RSS_{QTL} is the RSS obtained from fitting the QTL model (QTL model 2) under the alternative hypothesis, $\beta_i \neq 0$ for at least one *i*.

(vii) *Simulation*

(a) *Phenotype and QEI data*

Phenotypic values of offspring were first generated on the liability scale as $z_{ijk} = \mu + u_i + \varepsilon_j + \alpha_j + e_{ijk}$, where u_i is the polygenic effect of the *i*th sire, ε_j is the environmental effect of the *j*th environment, α_j is the substitution effect for sire QTL allele (Q_1 or Q_2) in the *j*th environment and e_{ijk} are residuals. This bi-allelic QTL was located at 5 cM from the left marker in a 20 cM marker bracket and was assumed to be interacting with the environment, which meant that it had different QTL substitution effects among environments. Table 1 gives the environment-specific QTL effects, specific for assumed QEI variances, 0 or 10 %, as well as the actual QTL variances calculated based on the method described in Appendix C.

A total heritability, including QTL and QEI effects, of 0·25 on the liability scale was used in all cases. The polygenic effects, u_i, were sampled from $N[0, 0{\cdot}25\sigma^2_u]$, where σ^2_u is the additive genetic variance ($0{\cdot}25\sigma^2_u$ is a sire variance). One environmental effect with five levels was simulated; ε_j was sampled from $N[0, \sigma^2_{ev}]$ where σ^2_{ev} is the environmental variance, which was set equal to 0 or 10 % of the total phenotypic variance on the liability scale. The distribution of progeny across the five environments was random for 20 sires. A progeny of the *i*th sire had a QTL allele effect of $\frac{1}{2}\alpha_j$ when the progeny inherited the Q_1 allele and $-\frac{1}{2}\alpha_j$ when the progeny inherited the Q_2 allele, based on binomial sampling, with equal frequency. Residuals, *e*, were sampled from $N[0, \sigma^2_p - (0{\cdot}25\sigma^2_u + \sigma^2_{ev} + 0{\cdot}5\sigma^2_{QTL_r})]$, where σ^2_p is the phenotypic variance.

(b) *Parameters for simulations*

Binary datasets were simulated for population incidences *p*, equal to 0·25 and 0·50, corresponding to standardized thresholds (*t*) equal to 0·67 and 0·0 and for the number of offspring per sire (n_i) equal to 200 and 500. The mean QTL effects on the underlying liability scale ($\bar{q}$) were equal to 0·05, 0·15 and 0·30 phenotypic standard deviation units ($\sigma_p = 1$) each with a QEI variance (σ^2_{QEI}) of 0 and 10 % of phenotypic variance. Incidences of 25 % and 50 % were chosen to represent binary data with large scale effects (around 25 % incidence) and with very minimal scale effects (around 50 % incidence). Scale effects are related to the binomial variance, $p(1-p)$, which is fairly flat around 50 % but decreases sharply from about 35 % down to 15 % incidence. The above combinations of

https://doi.org/10.1017/S0016672306008391 Published online by Cambridge University Press

Table 2. *Empirical significance threshold values for QTL interval mapping for a binary trait in 20 half-sib families based on linear regression (RIM) and threshold (GIM) models for testing different effects (ST_1, overall QTL effect; ST_2, overall QTL effect and QEI effect; and ST_3, QEI effect)*

Incidence	*n*	Model	Level	ST_1	ST_2	ST_3
25%	200	RIM	5%	33·85	135·56	110·87
			1%	40·53	147·80	122·87
		GIM	5%	33·45	127·44	103·58
			1%	39·50	137·37	114·14
	500	RIM	5%	33·90	140·44	116·11
			1%	40·45	153·45	128·39
		GIM	5%	33·67	129·60	105·44
			1%	40·27	141·68	115·88
50%	200	RIM	5%	33·80	130·55	105·83
			1%	39·88	141·86	116·28
		GIM	5%	33·23	132·04	107·45
			1%	35·47	137·25	111·68
	500	RIM	5%	33·73	129·87	105·28
			1%	39·94	142·89	116·03
		GIM	5%	33·73	129·17	104·70
			1%	39·78	141·89	115·52

Results are based on 10 000 replicates of binary data for different combinations of progeny group sizes (*n*) and disease incidences (*p*). The expected significance thresholds from a chi-square distribution are: for ST_1 (19 d.f.) 30·14 (5%) and 36·19 (1%); for ST_2 (95 d.f.) 118·75 (5%) and 129·97 (1%); and for ST_3 (76 d.f.) 97·35 (5%) and 107·58 (1%).

parameters resulted in 24 types of disease data to be analysed by RIM and GIM. Since the QEI model being developed can also be suitable for normally distributed traits, we kept a few representative simulations on the underlying liability scale to be analysed by RIM. They correspond to six datasets created by three main QTL effects (0·05, 0·15 and 0·3 phenotypic standard deviation units) and two G × E variances (0 and 10%) for n_i, equal to 500. Further, a few more combinations of parameters were used in simulations to test the effect of the number of different levels of environments (=2) and low/high (5% or 20%) G × E variances. Table 1 gives QTL, QEI effects and the different variances simulated.

3. Results

(i) *Empirical significance thresholds and simulations without QEI*

Table 2 shows the empirical significance threshold (ST) values for testing of main effects and interactions for different disease incidences and population sizes using the data simulated under the null model. Overall, these empirical thresholds lie about 8–12% above the expected χ^2 table values. In one particular case, however, RIM deviates from GIM, which is for testing interactions at 25% incidence: here empirical thresholds for RIM are elevated upwards by 14–19% compared with χ^2 table values, whereas empirical thresholds for GIM in this case actually deviate least from the table value (+6 to +9%). This difference between RIM and GIM is not seen at 50% incidence or for testing of main effects, pointing to scale effects on the binary scale being the cause of these elevated ST values for RIM. A subtle effect of the progeny group size on significance thresholds can be seen in that threshold levels are higher for the larger progeny group size, but only for the testing of interactions at 25% disease incidence, and not for other situations.

(ii) *Power to detect QTL and QEI*

Table 3 shows the power for detection of the QTL main effect and QEI from simulated data with a main QTL effect but without QEI interaction using nominal values for significance thresholds or empirical significance thresholds. As no QEI is present, results from the third test (H2 : H1) testing for interaction above the main effect indicate the false detection rate (FDR) for testing of QEI by RIM and GIM. As expected from the results in Table 2, use of nominal significance thresholds leads to (extra) increased FDR by RIM when testing for interactions at 25% disease incidence; GIM here has a FDR of 12–16%, while RIM shows a FDR around 23%. In other cases, such as 50% disease incidence or when testing for main effects, no differences between RIM and GIM are discernible. The test combining main and interaction effects (H2 : H0), which is not supposed to give more significant results than the test for main effect only (H1 : H0) in these data without QEI, can also be seen to be elevated for RIM compared with GIM for 25% disease incidence and for small QTL effects when using nominal significance thresholds. When using empirical threshold levels, RIM has somewhat smaller power for the H2 : H0 test at 25% disease incidence when QEI is not present, but only marginally. In practice, this (somewhat) reduced power for RIM has no implications, because after rejection of the H2 : H0 test one would still proceed to test for a main QTL effect only using the H1 : H0 test (see Fig. 2). When using empirical thresholds, Table 2 shows that the pure interactions test (H2 : H1) is corrected to yield the desired 5% FDR. In the absence of QEI, small QTL effects remain largely undetected with power <10%.

Data on the power to detect QTLs and their interactions with environments (QEI) based on RIM and GIM when QEI interaction is present (10% QEI variance) are given in Table 4. Here, empirical thresholds are used which imply elevated thresholds for RIM when testing interaction effects at 25%

https://doi.org/10.1017/S0016672306008391 Published online by Cambridge University Press

Table 3. *Power for detection of QTL main effect and QTL–environment interactions (QEI) for a binary trait at 5% nominal significance level and 5% empirical significance level when no QEI is present, shown for two incidence levels (p), different average main effect ($\bar{q}$) of the QTL, and for linear regression (RIM) and threshold models (GIM)*

			At nominal 5% level			At empirical 5% level		
Incidence	Model	$\bar{q}$	$H_1 : H_0$	$H_2 : H_0$	$H_2 : H_1$	$H_1 : H_0$	$H_2 : H_0$	$H_2 : H_1$
25%	RIM	0·05	13·2	27·5	23·5	6·6	6·1	5·6
		0·15	25·9	34·8	22·4	16·0	7·6	5·4
		0·3	71·9	59·7	23·1	58·5	25·8	5·8
	GIM	0·05	12·8	14·0	13·6	7·7	6·4	6·0
		0·15	27·4	24·8	12·4	16·7	12·8	6·0
		0·3	71·9	48·4	9·7	58·1	28·6	3·7
50%	RIM	0·05	12·0	18·7	14·8	5·0	5·0	4·7
		0·15	28·0	25·4	14·0	15·1	9·5	4·8
		0·3	79·7	57·8	15·0	69·1	31·6	5·8
	GIM	0·05	12·3	21·5	16·7	6·1	5·6	4·5
		0·15	28·4	29·0	16·7	16·9	9·3	5·0
		0·3	80·4	60·2	15·9	71·7	30·9	5·4

$H_1 : H_0$ tests for the presence of a QTL main effect, $H_2 : H_0$ tests for the presence of a QTL main effect and QEI jointly and $H_2 : H_1$ tests only QEI over the main QTL effect. Results are based on 1000 replicates of binary data on 20 half-sib families with 200 progeny per family. The empirical significance levels used are in Table 2.

disease incidence, so it might be expected that RIM loses some power in these situations compared with GIM. In fact GIM consistently has somewhat better power at 50% incidence; at 25% incidence GIM is only better for the small main effect, while RIM has more power for the medium and large main effect. Overall, differences in power between RIM and GIM remain small, and these results indicate that the use of empirical thresholds, which elevate thresholds for RIM to safeguard RIM against detection of spurious interactions, does not negatively affect the power of RIM compared with GIM when QEI is present.

An interesting effect can be observed when the power of a test for a QTL main effect (H1 : H0) when QEI is present (Table 4) is compared with that when QEI is absent (Table 3) for the same QTL main effect. Although QEI is not modelled, power to detect a QTL by its main effect increases when the QTL also has QEI effects. For instance, a QTL with main effect of 0·3 at a disease incidence of 25% with empirical thresholds (in Table 3) can be detected with 58% power using the test for main effect only, whereas Table 4 (compare at $n = 200$) shows that this same size of QTL can be detected using a test for main effect only with 75–77% power. This phenomenon is further referred to as the 'absorption of interaction effects in the main effect' (see also Section 4). In this respect, however, there is no difference between RIM and GIM. As expected, power increases for larger progeny group size, but power is not clearly affected by disease incidence. The effects of incidence are that power is consistently lower for detecting a main effect at 50% than at 25% incidence, while the situation is variable for detecting an interacting QTL with the H2 : H0 test.

As can be seen further, the presence of QEI when modelled and used in testing can improve the power to detect a QTL. This is especially critical when the segregating QTL has a small main effect ($\bar{q} = 0{\cdot}05$) but interacts with the environment as evidenced by the increase in power for QEI tests ranging from approximately 20% to 80% in Table 4. This confirms the ability of the full QTL + QEI model to detect QTLs with small main effect but interacting with the environment that would otherwise have been missed by the simple QTL model.

Further simulations with low (5%) and high (20%) G × E variances (results not shown), in addition to the 0 versus 10% G × E variances reported here, showed that the power to detect interacting QTLs by LR_{QEI} tests was 5–15% higher when $\sigma^2_{G\times E}$ was 20% than when it was 5%. This confirmed that the power to detect interacting QTLs proportionately increases as G × E variances increased from 0 to 5% and 10% to 20%, as expected. Like other tests, LR_{QEI} tests tend to have less power to detect interactions at lower incidence (here 25%) than at intermediate incidence (here 50%), especially when the overall QTL substitution effect is small (0·05).

(iii) *QTL and QEI variances*

Estimated QTL variances (σ^2_{QTL}) and QTL + QEI variances ($\sigma^2_{QTL+QEI}$) are given in Table 5 for RIM

https://doi.org/10.1017/S0016672306008391 Published online by Cambridge University Press

Table 4. *Power for detection of a QTL main effect and QTL–environment interaction (QEI) for a binary trait at a 5% empirical significance level when QEI is present, shown for different progeny group size (n), different disease incidence (p), different main effects ($\bar{q}$), and for linear regression (RIM) and threshold models (GIM)*

n	Incidence	Model	$\bar{q}$	$H_1 : H_0$	$H_2 : H_0$	$H_2 : H_1$
200	25%	RIM	0·05	10·0	39·5	41·2
			0·15	28·5	56·5	46·3
			0·3	77·0	85·2	55·3
		GIM	0·05	8·6	41·4	36·2
			0·15	31·8	55·9	39·0
			0·3	74·7	78·4	36·8
	50%	RIM	0·05	6·2	44·0	47·2
			0·15	17·7	50·7	46·1
			0·3	69·6	78·4	47·0
		GIM	0·05	7·2	44·9	49·4
			0·15	18·9	53·0	46·7
			0·3	71·0	79·2	47·2
500	25%	RIM	0·05	17·6	84·7	83·4
			0·15	65·7	96·6	89·3
			0·3	99·4	100·0	95·3
		GIM	0·05	16·0	88·8	88·3
			0·15	59·9	95·7	88·0
			0·3	99·6	100·0	89·0
	50%	RIM	0·05	8·0	91·8	94·3
			0·15	42·7	96·7	94·0
			0·3	98·5	100·0	94·2
		GIM	0·05	7·9	93·7	94·8
			0·15	43·2	97·2	94·9
			0·3	98·4	100·0	94·5

QTL–environment interaction (σ^2_{QEI}) was 10% of phenotypic variance. $H_1 : H_0$ tests for the presence of a QTL main effect, $H_2 : H_0$ tests for the presence of a QTL main effect and QEI jointly and $H_2 : H_1$ tests only QEI over the main QTL effect. Results are based on 1000 replicates of binary data on 20 half-sib families. Empirical threshold levels imply elevated thresholds for RIM for 25% incidence in order to safeguard RIM against spurious detection of interactions.

and in Table 6 for GIM. In these tables, an estimate of 'pure' interaction variance (σ^2_{QEI}) is also constructed from the difference between these two variance components. Estimated variances are on the binary scale for RIM and on the underlying liability for GIM.

For both RIM and GIM, QTL variances were also estimated as given in Appendix A, which accounted for covariances; those estimates were very similar to the QTL variances estimated directly from the QTL model (2) and hence are not shown in Tables 5 or 6.

RIM and GIM deviate considerably in performance with the model for QTL + QEI, with RIM showing a gross overestimation. In the absence of QEI, for instance with $p = 25\%$, $n = 200$ and for the large main effect ($q = 0{\cdot}3$), RIM overestimates total QTL variance with the QTL + QEI model by 177%. GIM in this same situation overestimates total QTL variance by 'only' 58%. Note that this is a situation where the model is actually inappropriate because QEI is not present, and that we might therefore expect some overestimation, but the relatively poorer performance of RIM points to relatively larger problems for RIM in assessing interaction effects in binary data. When QEI is present RIM continues to perform relatively badly, for instance for the same situation with $p = 25\%$, $n = 200$ and for the large main effect ($q = 0{\cdot}3$), RIM overestimates total QTL variance by 147%, while GIM in this case overestimates total QTL variance by 38%. For the large progeny group sizes, overestimation is reducing, and at 50% and with QEI present GIM even starts to slightly underestimate total QTL variance; RIM in this case still overestimates the total QTL variance by around 100%. The main-effects model posed fewer problems for both RIM and GIM, with RIM doing fairly well while GIM appears to overestimate the QTL effect in the absence of QEI. Obviously, when QEI is present, use of a main-effects model will not allow the capture of all QTL variance present, and both RIM and GIM underestimate total QTL variance. Finally, the phenomenon of 'absorption' of interaction effects into the main effect, when using a main-effects model, can also be seen in the estimation of variances. This phenomenon is present at 25% incidence but not at 50% incidence for both RIM and GIM. For instance, for a main QTL effect of $\bar{q} = 0{\cdot}3$, progeny group size of $n = 200$ and with RIM analysis (Table 5), estimated QTL variance with a main-effects model increases from 5·1 (without QEI effects) to 6·9 (with QEI effects) at 25% incidence, i.e. although the main effect is the same, the estimated main effect variance increases by 35% when QEI is present, apparently picking up part of the interaction effect. The picking up of extra variance in the main effect is not seen at 50% incidence. This phenomenon, and absence of it at 50% incidence, is systematically seen for the other QTL main effects, progeny group sizes and for GIM analyses.

4. Discussion

Gene or QTL by environment interaction (QEI) is an important mode of gene action from an evolutionary, medical and quantitative genetics point of view. However, it is often neglected while trying to detect and map genes controlling complex polygenic traits (QTLs) by interval mapping methods, especially in animal populations. This paper has developed and investigated QEI models with a focus on mapping QTLs for binary (disease) traits.

https://doi.org/10.1017/S0016672306008391 Published online by Cambridge University Press

Table 5. *Estimates of QTL variance for interval mapping of a QTL with different average effects* ($\bar{q}$) *and different interaction* (*QEI*) *effects* (σ^2_{QEI}) *based on linear regression* (*RIM*) *models for different disease incidence* (*p*) *and progeny group size* (*n*)

					Estimated		
Incidence	*n*	σ^2_{QEI}	$\bar{q}$	Simulated	σ^2_{QTL}	$\sigma^2_{QTL+QEI}$	σ^2_{QEI}[a]
25%	200	0	0·05	0·2	1·2	13·9	13·1
			0·15	1·7	2·0	14·8	13·0
			0·3	6·5	5·1	18·0	13·0
		10%	0·05	5·7	1·6	16·7	15·5
			0·15	6·5	2·9	18·5	15·7
			0·3	9·3	6·9	23·0	16·0
	500	0	0·05	0·2	0·5	5·6	5·2
			0·15	1·7	1·4	6·6	5·2
			0·3	6·5	4·8	10·0	5·2
		10%	0·05	5·7	0·9	8·5	7·7
			0·15	6·5	2·4	10·3	8·0
			0·3	9·3	6·8	15·1	8·3
50%	200	0	0·05	0·2	1·5	17·5	16·4
			0·15	1·7	2·7	18·9	16·4
			0·3	6·5	7·7	24·0	16·3
		10%	0·05	5·7	1·6	21·2	20·1
			0·15	6·5	2·9	22·6	20·0
			0·3	9·3	7·8	27·7	19·9
	500	0	0·05	0·2	0·7	7·0	6·5
			0·15	1·7	2·1	8·5	6·4
			0·3	6·5	7·5	13·7	6·3
		10%	0·05	5·7	0·7	10·6	10·0
			0·15	6·5	2·1	12·1	10·0
			0·3	9·3	7·6	17·4	9·8

Statistical models with only a main QTL effect (estimating σ^2_{QTL}) or with a main and QEI effect (estimating $\sigma^2_{QTL+QEI}$) were used with results based on 1000 replicates of binary data. All variances are $\times 1000$.

[a] σ^2_{QEI}, estimate of pure interaction variance computed from $\sigma^2_{QTL+QEI} - \sigma^2_{QTL}$.

(i) *False positive rate and significance thresholds*

One of the ideas behind a comparison of the linear (RIM) versus threshold model (GIM) is that a linear model may find spurious interactions for binary traits due to scale effects on the binomial scale. This phenomenon was indeed demonstrated, with RIM showing elevated false positive rates for detection of interacting QTLs; when using empirical thresholds this led to elevated significance thresholds for RIM for testing interactions. This elevated false positive rate for RIM was present only at 25% disease incidence (where scale effects are large), not at 50% incidence (where scale effects are very small). RIM did not show aberrations when testing for main effects and with the use of continuous liability data (results not shown), which confirms our earlier findings (Kadarmideen *et al*., 2000). These false detections of QTLs overwhelmingly indicate that linear models are inappropriate particularly when testing interactions in binary data with incidences deviating from 50%, i.e. where scale effects are important. We expect this to be a general phenomenon for testing any interactions in binary data. The increased false positive rate can be corrected by the use of empirical significance thresholds, which is therefore strongly recommended when using linear models to test interactions in binary data. Although for QTL mapping the use of empirical thresholds is fairly common, this is not the case in many other areas of statistical analysis. Such empirical significance thresholds would have to be computed for each individual case because the empirical levels depend notably on disease incidence, and also somewhat on the size of the dataset and possibly on other factors. This of course makes the use of linear models to analyse interactions in binary data somewhat cumbersome in practice. In all subsequent work performed here we used empirical significance thresholds in order to compare RIM and GIM on an equal footing.

https://doi.org/10.1017/S0016672306008391 Published online by Cambridge University Press

Table 6. *Estimates of QTL variance for interval mapping of a QTL with different average effects* ($\bar{q}$) *and different interaction* (*QEI*) *effects* (σ^2_{QEI}) *based on threshold models* (*GIM*) *for different disease incidences* (*p*) *and progeny group size* (*n*)

					Estimated		
Incidence	*n*	σ^2_{QEI}	$\bar{q}$	Simulated	σ^2_{QTL}	$\sigma^2_{QTL+QEI}$	σ^2_{QEI}[a]
25%	200	0	0·05	1·3	11·1	22·5	11·8
			0·15	11·3	22·0	36·0	13·0
			0·3	45·0	61·5	71·4	6·4
		10%	0·05	41·0	16·4	55·0	39·0
			0·15	50·7	35·6	73·3	40·0
			0·3	84·8	80·0	116·9	39·0
	500	0	0·05	1·3	4·6	8·5	4·5
			0·15	11·3	15·9	19·6	4·0
			0·3	45·0	55·7	58·9	2·0
		10%	0·05	41·0	8·4	39·7	33·0
			0·15	50·7	25·3	54·0	31·0
			0·3	84·8	74·4	98·6	29·0
50%	200	0	0·05	1·3	8·7	20·1	11·1
			0·15	11·3	19·4	31·3	10·0
			0·3	45·0	58·7	70·8	10·0
		10%	0·05	41·0	8·7	51·3	42·0
			0·15	50·7	20·2	62·7	41·0
			0·3	84·8	58·8	101·0	39·0
	500	0	0·05	1·3	4·2	5·5	1·0
			0·15	11·3	15·2	16·1	1·0
			0·3	45·0	55·2	54·5	1·0
		10%	0·05	41·0	4·3	32·6	29·2
			0·15	50·7	15·5	43·5	29·1
			0·3	84·8	55·5	83·4	26·1

Statistical models with only a main QTL effect (estimating σ^2_{QTL}) or with a main and QEI effect (estimating $\sigma^2_{QTL+QEI}$) were used with results based on 1000 replicates of binary data. All variances are ×1000.

[a] σ^2_{QEI}, estimate of pure interaction variance computed from $\sigma^2_{QTL+QEI} - \sigma^2_{QTL}$.

(ii) *Power of detecting QTL and QEI effects*

The $LR_{QTL+QEI}$ test is the first test from our top-down model selection scheme and is a joint test for QTL main effects and QEI. In the absence of QEI, RIM has slightly lower power for this $LR_{QTL+QEI}$ test at 25% disease incidence. This is logically the result of the use of elevated empirical significance thresholds for RIM in this case. In the other tests for the QTL main effect (LR_{QTL}) and for 'pure' interaction only (LR_{QEI}), RIM and GIM have comparable performance based on the empirical significance thresholds. In the absence of QEI it was also seen that the $LR_{QTL+QEI}$ test has lower power than the LR_{QTL} test, which is logically due to the higher number of, here unnecessary, parameters fitted in the QTL + QEI model. This initial overparameterization and low power is not a problem because when the $LR_{QTL+QEI}$ test fails, our model selection scheme proceeds with the LR_{QTL} test.

When QEI was present, differences between RIM and GIM were more complicated. At 50% incidence GIM consistently had (somewhat) better power for the interaction tests ($LR_{QTL+QEI}$ and LR_{QEI}) so that GIM will conclude more often for the correct model with QEI. Applying the correct model can have a dramatic impact on detection of a QTL: when QEI is present but not detected, a subsequent test for the main effect (LR_{QTL}) can be nearly powerless when the main effect is small. At 25% incidence, however, differences between RIM and GIM were variable and were found to depend on the size of the main effect. In conclusion, RIM is better for the medium and larger main effects, while GIM is better for a small main effect.

Juenger *et al.* (2005) studied epistasis and G × E interaction using a bottom-up testing strategy: i.e. detect interesting marker loci with a QTL versus null model (our LR_{QTL}) then test whether a detected QTL has interactions by comparing the QTL + QEI model

https://doi.org/10.1017/S0016672306008391 Published online by Cambridge University Press

with the null model. Our results, however, indicate that under such a scheme QTLs with a small main effect but with important QEI may remain undetected because they fail the initial main-effects test (LR_{QTL}). This was also observed by Wang *et al*. (1999), who found that their simple QTL models failed to detect interacting QTLs with a small main effect.

We confirmed the phenomenon that interaction effects are confounded and partly 'absorbed' in the main effect when fitting a main-effect model only, as also seen by Culverhouse *et al*. (2002) and Purcell & Sham (2004). Additionally, we showed that this effect is not present at 50% incidence, indicating that binomial scale effects would seem to be implicated in this phenomenon. This 'absorption' of interaction effects which happens at 25% but not at 50% incidence, explains certain cases where 50% incidence may give less power than 25% incidence.

(iii) *Estimation of QTL and QEI effects and variances*

We assessed the ability of models to estimate QTL and QEI effects in terms of the estimated variances associated with these QTL and/or QEI effects. It was shown that RIM yields a gross overestimation of QTL variance (up to 177%) when using the QTL + QEI model, both when QEI is present and when it is absent. Although the use of elevated significance thresholds for RIM could correct for its increased false positive rate when modelling interactions, accurate inference on interaction effects therefore remains very problematic for RIM. GIM performs fairly well with the QTL + QEI model, when QEI is either present or absent, usually with some overestimation but, compared with RIM, at a more modest level.

Use of a main-effects model did not allow the capture of all QTL variance present, and both RIM and GIM underestimate total QTL variance. Similar results were reported by Wang *et al*. (1999). The phenomenon already discussed above of interactions effects being 'absorbed' in the main effect by RIM was also seen here, again typically at 25% incidence but not at 50%. The overestimation of QTL + QEI variances by RIM much more frequently and substantially is the evidence that linear models can find spurious interaction effects on the observed scale.

(iv) *QTL location*

The mean and standard deviation of QTL location across 1000 replicates (results not shown) depended on the power, which in turn depended on a number of factors, including the genetic model used. The choice of genetic model was shown to be important when mapping QTLs, with an incorrect model leading to a mean estimated QTL location remote from the true location with high standard deviation. As expected, a powerful design (large segregating QTL, large progeny groups/genotypings, and intermediate incidence) always resulted in mapping a QTL to its true location. Comparing RIM and GIM, the significant differences in power did not translate to significant differences in mapped locations by RIM and GIM when QEI existed at intermediate incidences. The finding that a small interacting QTL is more precisely mapped by QEI models is an important result. Although we simulated a single marker bracket for single QTL mapping, results are valid for QTL mapping with multiple marker brackets, as shown by Kadarmideen & Dekkers (1999) for non-interacting QTLs.

(v) *General*

In this study we used fixed-effect models, but one might think of treating polygenes, QTLs and QEI effects as random in a mixed-model approach, which needs further investigation (e.g. as in Wang *et al*., 1999). In the case of random models, more parameters can be included and all variances can be estimated. So far, however, the fixed models have proved useful to demonstrate our main aim of investigating possible biases with RIM to model and test interactions in binary data. Computational constraint is a common problem in mapping a QTL with interactions; the complexity increases when statistical methods used are non-linear (e.g. threshold models). In such situations, we recommend the use of the marker regression mapping method of Kadarmideen & Dekkers (1999), as this is based on a simple single regression approach. We used large half-sib families but the basic models and methods developed here are applicable to any type of families. The models and methods focused on binary traits (or diseases specifically) but in practice there may also be continuous traits showing QEI (e.g. flowering time in *Arabidopsis*, body condition or milk yield in cows, immune responses in chickens or humans). In this case, the top-down methodology and results would still apply. In fact, results from applying RIM to continuous liability data (results not shown) indicated that comparison of genetic models is valid on both normal (liability) and binary scales. One could map QTLs for each environment but this would result in different QTL locations. Biologically, there is only one QTL located on the genome but acting differently between environments. The method developed in this study has essentially shown how to map a QTL to one location but with many within-family across-environment QTL effects. In addition, in theory, pooling QTL/marker data on offspring across varied environments should result in high power, precision of estimates and lower bias than when mapping QTLs separately in each environment.

5. Conclusions

This is the first study to develop a QTL interval mapping method by liability-threshold models for binary traits (e.g. diseases) that show gene by environment interactions (QEI) with a focus on outbred populations on a within-family across-environments basis. A top-down testing procedure was proposed to help choose appropriate genetic models and obtain maximum power in detecting interacting QTLs. We have shown that a simple QTL model often fails to detect small interacting QTLs; only genetic models that model QEI are able to detect and map such QTLs. The power to detect interacting QTLs increases proportionately as G × E variances increase. It was shown that the use of a linear model to model and test interactions in binary data poses several problems. First, the linear model had an increased false positive rate for testing interactions which was attributable to scale effects on the binary scale. This increased false positive rate can be corrected by the use of empirical significance thresholds, which are therefore mandatory for the correct testing of interactions in binary data using a linear model. Secondly, however, the linear model with interaction terms continued to suffer from very large biases for assessing interaction effects/variances. A phenomenon already reported elsewhere that interaction effects can become 'absorbed' in the main effect also appears associated with scale effect in binary data. A non-linear (threshold) model generally performed well. This is the first study to prove the principle that linear models are not suitable for analysis of binary or categorical data with interactions (at the polygenic and/or QTL level) and that threshold models are better suited for this purpose.

Appendix A. Computation of QTL and QEI variances

From the QTL model (2), an estimate of QTL variance (σ^2_{QTL}) was computed as in Kadarmideen *et al.* (2000) as $\sigma^2_{QTL} = \sigma^2_{\beta} - \bar{\sigma}^2_{pe}$, where σ^2_{β} is the variance of 'best' estimates of QTL substitution effects (β_i) across sires and $\bar{\sigma}^2_{pe}$ is the average prediction error variance of QTL substitution effects for the best-fitting model (obtained from the diagonal of the inverse of the left-hand side, **LHS**, matrix). From the QEI model (3), the variance of environment-specific QEI effects, due to the fitted model, actually represents variance of both the main QTL and QEI effects and hence calculation based on $\sigma^2_{\beta_{ij}}$ could only provide an estimate of the sum of QTL and QEI variance ($\sigma^2_{QTL+QEI}$) as $\sigma^2_{QTL+QEI} = \sigma^2_{\beta_{ij}} - \bar{\sigma}^2_{pe(\beta_{ij})}$, where $\sigma^2_{\beta_{ij}}$ is the variance of 'best' estimates of environment-specific QTL substitution effects (β_{ij}) and $\bar{\sigma}^2_{pe(\beta_{ij})}$ is the average prediction error variance of QTL + QEI effects for the best-fitting model. The average prediction error variance of QTL + QEI effects ($\bar{\sigma}^2_{pe(\beta_{ij})}$) is calculated as

$$\bar{\sigma}^2_{pe(\beta_{ij})} = \frac{1}{sr} \sum_{i=1}^{sr} \mathrm{diag}\{(\mathrm{LHS})^{-1}\} i.$$

An estimate of σ^2_{QEI} was obtained as

$$\sigma^2_{QEI} = \sigma^2_{QTL+QEI} - \sigma^2_{QTL}.$$

However, the most appropriate way is to derive σ^2_{QEI} directly from the QEI model, as follows: First, an overall QTL effect estimate ($\beta_{i.}$) and an overall prediction error variance ($pe(\beta_{i.})$) for each sire need to be calculated; $\beta_{i.}$ can be calculated as $\beta_{i.} = \frac{1}{r}\sum_{j=1}^{r}\beta_{ij}$. The calculation of $pe(\beta_{i.})$ is not straightforward because it is the variance of the average, which needs to include all covariances. Estimates of β_{ij} are calculated from mixed-model equations, which can be written as an $sr \times 1$ vector η with elements $[\beta_{11}, \ldots, \beta_{s1}, \ldots, \beta_{12}, \ldots, \beta_{s2}, \ldots, \beta_{1r}, \ldots, \beta_{sr}]'$. Let $\mathbf{LHS}^{-1}_{\beta}$ be the part in $\mathbf{LHS}^{-1}$ corresponding to η. Let $\boldsymbol{K_i}$ be an $sr \times 1$ vector with values of $1/r$ for the elements corresponding to β_{ij} in the vector η, where $j = 1$ to r, and with a value of zero for the rest of the elements. Then, $pe(\beta_{i.})$ can be calculated as $pe(\beta_{i.}) = \boldsymbol{K}_i' \ \mathbf{LHS}^{-1}_{\beta} \ \boldsymbol{K_i}$. Therefore, $\sigma^2_{QTL} = \sigma^2_{\beta_{i.}} - \frac{1}{s}\sum_{i=1}^{s} pe(\beta_{i.})$. Based on the methodology used here to estimate variances, the estimated QTL or QEI variances could become negative when $\bar{\sigma}^2_{pe}$ or $\bar{\sigma}^2_{pe(\beta_{ij})}$ are larger than the corresponding σ^2_{β} or $\sigma^2_{\beta_{ij}}$; in this case a value of zero for variances was used.

Appendix B. Computation of log-likelihoods

The general form of likelihood for GIM was given in equation (4). To test the significance for the presence of QTL effects with the null hypothesis of $\beta_i = 0$ for all i, the log-likelihood in equation (4) under the null model, L_{null}, is maximized with $\pi_{ijk} = \Phi(\theta_{ijk})$, where $\theta_{ijk} = \mathbf{E}(z_{ijk}) = \mu + u_i + \varepsilon_j$ as defined in the null model. Similarly, to test for the presence of a QTL in the marker bracket, the likelihood (in equation 4) under the alternative hypothesis ($\beta_i \neq 0$ for at least one i), L_{QTL}, is maximized with $\pi_{ijk} = \Phi(\theta_{ijk})$, where $\theta_{ijk} = \mathbf{E}(z_{ijk}) = \mu + u_i + \varepsilon_j + c_{ijk}\beta_i$ as defined in the QTL model. Finally, to test for the presence of a QTL in the marker bracket as well its interaction with the environment, QEI, simultaneously, the log-likelihood in the equation under the assumption of QEI ($\beta_{ij} \neq 0$) for at least one ij, $L_{QTL+QEI}$, is maximized with $\pi_{ijk} = \Phi(\theta_{ijk})$, where $\theta_{ijk} = \mathbf{E}(z_{ijk}) = \mu + u_i + \varepsilon_j + c_{ijk}\beta_{ij}$ as defined in QEI model (3).

Appendix C. Simulation of QTL by environment interaction effects

In simulation, the substitution effect of the QTL in environment j (α_j) included two parts: the overall

https://doi.org/10.1017/S0016672306008391 Published online by Cambridge University Press

QTL substitution effect ($\bar{q}$) and the deviation from the overall effect (qe_j). The $\bar{q}$ was determined by overall QTL variance (σ^2_{QTL}) as $\bar{q}=\sqrt{\frac{\sigma^2_{QTL}}{2f(1-f)}}$. The QTL by environment interaction part, qe_j, was sampled from a normal distribution with a mean of zero and a variance of σ^2_{QEI}. Then the total effect conferred by the QTL in a given environment was $\alpha_j=\bar{q}+qe_j$ ($j=1$ to r). Different QTL effects across environments were fixed in a given replicate but varied between sire families. Although, QEI variances were set to 0 or 10%, the realized QTL variances may differ due to sampling size and number of environments. The realized QTL variance ($\sigma^2_{QTL_r}$) was calculated as the mean of QTL variances for each environment, 1 to r: $\sigma^2_{QTL_r}=\frac{2f(1-f)}{r}\sum_{j=1}^{r}\alpha_j^2$, where r is the number of environments and f is the frequency of the Q_1 allele inherited from the sire. Since data were simulated on liability scales, these realized values were on the Normal scale. For the purpose of comparison of results on a binary or observed scale from RIM, the α_j needs to be transformed from a liability to a probability scale before computing the realized variances on binary scales. This was done based on Kadarmideen *et al.* (2000). The true QTL effect of the *jth* environment on the probability scale (α_{jp}) was computed based on the true environment-specific QTL effect on the liability scale α_j as $\alpha_{jp}=\Phi(\alpha_j)$, where Φ is a normal cumulative density function. Then the QTL variances for each environment and the mean of these variances were computed.

References

Culverhouse, R., Suarez, B. K., Lin, J. & Reich, T. (2002). A perspective on epistasis: limits of models displaying no main effect. *American Journal of Human Genetics* **70**, 461–471.

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

Hunter, D. J. (2005). Gene–environment interactions in human diseases. *Nature Review Genetics* **6**, 287–298.

Jansen, R. C. J., Ooijen, W. Van, Stam, P., Lister, C. & Dean, C. (1995). Genotype-by-environment interaction in genetic mapping of multiple quantitative trait loci. *Theoretical and Applied Genetics* **91**, 33–37.

Juenger, T. E. S., Sen, A., Kirk, S. & Simms, E. L. (2005). Epistasis and genotype–environment interaction for quantitative trait loci affecting flowering time in *Arabidopsis thaliana*. *Genetica* **123**, 87–105.

Kadarmideen, H. N. & Dekkers, J. C. M. (1999). Regression on markers with uncertain marker transmission for QTL mapping in half-sib designs. *Genetics Selection Evolution* **31**, 437–455.

Kadarmideen, H. N. & Dekkers, J. C. M. (2001). Generalized marker regression and interval QTL mapping methods for binary traits in half-sib family designs. *Journal of Animal Breeding and Genetics* **118**, 297–309.

Kadarmideen, H. N. & Janss, L. L. G. (2005). Evidence of major gene from Bayesian segregation analyses of liability to osteochondral diseases in pigs. *Genetics* **171**, 1195–1206.

Kadarmideen, H. N., Janss, L. L. G. & Dekkers, J. C. M. (2000). Power of QTL mapping for polygenic binary traits using generalized and regression interval mapping in multi-family half-sib designs. *Genetical Research* **76**, 305–317.

Purcell, S. & Sham, P. C. (2004). Epistasis in quantitative trait locus linkage analysis: interaction or main effect? *Behavior Genetics* **34**, 143–152.

Thomson, P. C. (2003). A generalized estimating equations approach to quantitative trait locus detection of non-normal traits. *Genetics Selection Evolution* **35**, 257–280.

Wang, D. L., Zhu, J., Li, Z. K. & Paterson, A. H. (1999). Mapping QTLs with epistatic effects and QTL× environment interactions by mixed linear model approaches. *Theoretical and Applied Genetics* **99**, 1255–1264.

Xu, S. Z. & Yi, N. J. (2000). Mixed model analysis of quantitative trait loci. *Proceedings of the National Academy of Sciences of the USA* **97**, 14542–14547.