Euphytica (2009) 166:131–144
DOI 10.1007/s10681-008-9849-4

Searching for interacting QTL in related populations of an outbreeding species

Johannes Jansen · Martin P. Boer · Marco C. A. M. Bink · W. Eric van de Weg

Received: 24 April 2008 / Accepted: 17 November 2008 / Published online: 4 December 2008

Abstract Many important crop species are outbreeding. In outbreeding species the search for genes affecting traits is complicated by the fact that in a single cross up to four alleles may be present at each locus. This paper is concerned with the search for interacting quantitative trait loci (QTL) in populations which have been obtained by crossing a number of parents. It will be assumed that the parents are unrelated, but the methods can be extended easily to allow a pedigree structure. The approach has two goals: (1) finding QTL that are interacting with other loci and also loci which behave additively; (2) finding parents which segregate at two or more interacting QTL. Large populations obtained by crossing these parents can be used to study interactions in detail. QTL analysis is carried out by means of regression on predictions of QTL genotypes.

Keywords Epistasis · Related populations · Outbreeding species · QTL analysis · Regression

J. Jansen (✉) · M. P. Boer · M. C. A. M. Bink
Biometris, Wageningen University and Research Centre, P.O. Box 100, 6700 AC Wageningen, The Netherlands
e-mail: johannes.jansen@wur.nl

W. E. van de Weg
Plant Breeding, Wageningen University and Research Centre, P.O. Box 16, 6700 AA Wageningen, The Netherlands

Introduction

Genetic linkage maps based upon molecular genetic markers in combination with the development of powerful statistical techniques have greatly enhanced the search for genes underlying traits that show continuous variation. Commonly, such traits are affected by environmental conditions and by genes acting singly or in combination with other genes. The effect of interaction between alleles at different loci is known as epistasis. Currently, the development of statistical methods for detecting interacting loci receives much attention in the statistical genetic literature. For example, Carlborg et al. (2000, 2004) use a genetic algorithm, Du and Hoeschele (2000) use Gibbs sampling, Jannink and Jansen (2001) use maximum likelihood in a one-dimensional search involving a diallel cross of pure lines, Boer et al. (2002) use penalized regression and Yi et al. (2003) use Bayesian methods. Much of the literature is concerned with populations derived from pure lines; see e.g. Cockerham and Zeng (1996), Kao and Zeng (2002); Zeng et al. (2005) and Melchinger et al. (2007).

Cheverud and Routman (1995) distinguish between a physiological-genetic and a statistical-genetic definition of epistasis. In physiological genetics, epistasis means that genotypic values of individuals with different genotypes at one locus depend functionally on their genotypes at other loci. In statistical genetics, epistasis simply means that

deviations from additivity are present in the population under study. As a consequence, even if physiological-genetic or functional epistasis is present, the actual appearance of statistical-genetic epistasis in a population depends on the segregation types of the parents of that population with regard to the interacting loci. Alvarez-Castro and Carlborg (2007) provide a formal relationship between statistical genetic models for epistasis and functional models for epistasis in different populations in the case of two alleles.

In this paper, we consider a number of populations of an outbreeding species, which have been obtained by crossing a number of parents. At this stage, it will be assumed that the parents are unrelated individuals. The mating design may be considered as an incomplete diallel cross. At the time molecular genetic markers were still unavailable, a population mean was written as a grand mean plus the sum of the average performances of the parents involved plus a deviation term. The average performances were called general combining abilities (GCA). The deviation term was called specific combining ability (SCA) (cf. Bulmer 1985). The presence of statistically significant specific combining abilities is an indication of the presence of non-additivity. However, this form of analysis provides no further information about the origin of the non-additivity: is it caused by dominance (also known as intra-locus interaction) or by interacting genes? Of course, no indication of the positions of the genes could be given.

Jannink and Jansen (2001) considered a situation in which the parents were pure lines, and individuals of the populations also consisted of pure lines (double haploids or RIL). In such a set-up, dominance is absent, and specific combining ability can be attributed solely to the presence of interacting genes. By using molecular markers, they devised a statistical procedure to test whether additive QTL effects of parents behaved consistently over crosses. This approach only requires a linear search of the genome. Due to the absence of a dominance component, a significant result would indicate the presence of epistatic QTL. Jannink (2007, 2008) discuss an approach in which statistical interactions of marker effects with genetic background are related to additive-by-additive epistatic effects in association studies using a mixed model.

In the case we deal with outbreeding species a significant result might still be due to dominance rather than epistasis. This requires that loci which show significant non-additivity in a linear search should be further examined. As a consequence, the linear search has to be followed by fitting models involving dominance and epistatic terms for a limited number of loci. In this study we limit the search for interacting genes to the additive-by-additive component. In populations of limited size it will be very difficult to trace more complex forms of non-additivity. However, it is possible to test whether the within-parent part of additive $\times$ additive component of non-additivity is consistent over crosses. A significant result of this test would imply the existence of more complicated forms of interaction.

The search for interacting genes does not only consist of looking for statistically significant components of non-additivity. The results of the statistical analysis may assist the genetist in elucidating combinations of parents that will produce populations that can be used to estimate the parameters of the physiological genetic model for epistasis. It will be clear at this stage that such parents should be heterozygous at interacting loci. Moreover, the number of individuals in the population should be very large.

Following Haley et al. (1994), all statistical methods developed in this paper are based on regression analysis. The computations are carried out in two-steps: (1) prediction of QTL genotypes from marker data, (2) regression analysis using the predictions of QTL genotypes as regressors. See Knott (2005) for an overview of regression based methods of QTL analysis. A two-step approach is also used by George et al. (2000) in a mixed-model analysis.

The models described in this paper have been developed especially in the context of apple breeding (Gianfranceschi and Soglio 2004; http://www.hidras.unimi.it/index.html). Apple is a diploid outbreeding species with a long juvenile period (5–7 years). Apple varieties are propagated vegetatively by grafting on rootstocks. As a consequence, a selected genotype with favourable combinations of alleles at several (interacting) loci can be maintained forever, and be used as parent in new crosses. New populations, obtained by crossing are added year after year. Many old varieties are still kept in orchards, and used as parent in breeding programs.

In the statistical approach described in this paper each parent is considered as a separate entity having two alleles at each locus. These alleles are considered to be specific to parents, i.e. the total number of alleles is equal to twice the number of parents. Usually, statistical approaches described in the literature use bi-allelic models. The markers that are used are essentially multi-allelic (microsatellites). Up to four marker alleles may be segregating at each locus; in some cases markers carry so-called null alleles which may lead to loss of information.

In "Model and methods" a physiological genetic model and a statistical genetic model will be developed. The physiological genetic model that will be considered is the complementary effects model (Allard 1960). In this paper, it will be assumed that for each gene affecting the trait two different alleles are present; however, the limitation to two alleles is not essential. In "Application", the methods will be applied to simulated data which involve two interacting genes on two different linkage groups as well as one singly 'operating' gene, that is located on the same linkage group as one the interacting genes. Simulated data are used to show how the method works if genes, interacting as well as non-interacting genes, affecting the trait are really present. The set of crosses concerned is similar to one that is used in the EC funded Hidras project on apple (*Malus domestica* L.) (Gianfranceschi and Soglio 2004; http://www.hidras.unimi.it/index.html). The advantages of using pedigreed plants populations have been described by Jannink et al. (2001) and Bink et al. (2002).

Model and methods

Preliminaries

Cheverud and Routman (1995) distinguish between physiological genetics and statistical genetics. Physiological genetics investigates relationships between genotypic means and corresponding genotypes. Statistical genetics investigates properties of populations of individuals. Cheverud and Routman note that physiological genetic interaction between genes may not only contribute to statistical genetic interaction effects but also to statistical genetic additive and dominance effects. This will be illustrated in the following sections.

Physiological genetics

In this section, we consider a situation in which two loci affect a trait. It will be assumed that two alleles may occur at each of the two loci in the group of individuals under study; the alleles will be indicated by the numbers 1 and 2. It will be assumed that in the heterozygotes the order of the alleles (i.e. whether allele 1 has been obtained from the mother and allele 2 from the father, or vice versa) does not affect the genotypic mean. As a consequence, nine different genotypes may occur. The genotypic means are shown in Table 1. For example, $\mu_{12;22}$ denotes the genotypic mean if at locus 1 single copies of allele 1 and 2 are present, and at locus 2 two copies of allele 2.

The nine genotypic means can be arranged (row by row) in a 9×1 vector $\boldsymbol{\mu}$. This vector of genotypic means can be written as a linear function of a grand mean and eight linear contrasts, $\boldsymbol{\mu} = \mathbf{X}\theta$, in which

$$\mathbf{X} = \begin{bmatrix} 1 & -1 & -1 & 1 & 1 & 1 & -1 & -1 & 1 \\ 1 & -1 & 0 & 1 & -2 & 0 & 2 & 0 & -2 \\ 1 & -1 & 1 & 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & 0 & -1 & -2 & 1 & 0 & 0 & 2 & -2 \\ 1 & 0 & 0 & -2 & -2 & 0 & 0 & 0 & 4 \\ 1 & 0 & 1 & -2 & 1 & 0 & 0 & -2 & -2 \\ 1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 & 1 \\ 1 & 1 & 0 & 1 & -2 & 0 & -2 & 0 & -2 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

and $\boldsymbol{\theta} = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \delta_1 \\ \delta_2 \\ \alpha\alpha_{12} \\ \alpha\delta_{12} \\ \alpha\delta_{21} \\ \delta\delta_{12} \end{pmatrix}$.

The linear contrasts in $\boldsymbol{\theta}$ are the additive effects at locus 1 (α_1) and locus 2 (α_2), the dominance effects at locus 1 (δ_1) and locus 2 (δ_2), the additive–additive interaction component ($\alpha\alpha_{12}$), the additive–dominance interaction components ($\alpha\delta_{12}$ and $\alpha\delta_{21}$) and the dominance–dominance interaction component ($\delta\delta_{12}$). The elements of $\boldsymbol{\theta}$ can be obtained from $\boldsymbol{\theta} = \mathbf{L}^{-1}\mathbf{X}^{t}\boldsymbol{\mu}$, in which $\mathbf{L} = \mathbf{X}^{t}\mathbf{X} = \text{diag}(9,6,6,18,18,4,12,12,36)$. In the next section a typical physiological genetic model will be considered.

Table 1 Genotypic means associated with the genotypes that occur in the case two loci with each two alleles affect a trait

Locus 1	Locus 2		
	11	12	22
11	$\mu_{11;11}$	$\mu_{11;12}$	$\mu_{11;22}$
12	$\mu_{12;11}$	$\mu_{12;12}$	$\mu_{12;22}$
22	$\mu_{22;11}$	$\mu_{22;12}$	$\mu_{22;22}$

Complementary gene effects

Genotypic means for the physiological genetic model called complementary gene effects are given in Table 2. At each locus, one of the alleles is completely dominant over the other allele; the combinations of alleles in which the dominant allele is present at both loci are completely dominant over the other combinations. This model could reflect the situation of two critical successive steps in a biochemical pathway. The model was named "complementary gene effects" by Allard (1960, p. 100). In this example, the genotypic mean is equal to zero if at one locus or at both loci the genotype is 22 and the genotypic mean is equal to 9λ if this is not the case; λ is some positive constant, the multiplier 9 is only used to keep values of contrasts reasonably simple. In this case, the value of $\boldsymbol{\theta}$ is equal to $\boldsymbol{\theta} = \lambda(4, -3, -3, -1, -1, \frac{9}{4}, \frac{3}{4}, \frac{3}{4}, \frac{1}{4})^{t}$.

Statistical genetics

In this case, a population will be defined as the offspring of two parents, i and j, say. For two loci, each offspring individual can be characterized by one of 16 possible inheritance states. Information about the inheritance state of offspring individuals will be obtained from marker data. The mean values associated with the inheritance states are given in Table 3. It should be noticed that no information is available about the genotypes of the parents at the two loci. For example, $m_{12;22}$ represents the mean value for an offspring individual that at locus 1 obtained the allele on homolog 1 of parent i and the allele on homolog 2 of parent j, and at locus 2 obtained the allele on homolog 2 of parent i and the allele on homolog 2 of parent j.

The elements of Table 3 can be arranged (row after row) into a 16×1 vector $\mathbf{m}_{ij}$. The vector $\mathbf{m}_{ij}$ can be written as a linear function of a grand mean and 15 linear contrasts, $\boldsymbol{m}_{ij} = \boldsymbol{B}\boldsymbol{c}_{ij}$, in which

$$\boldsymbol{B} = \begin{bmatrix}
1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 \\
1 & -1 & -1 & -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 & 1 & 1 & -1 & 1 & -1 \\
1 & -1 & -1 & 1 & -1 & 1 & -1 & -1 & 1 & 1 & -1 & 1 & 1 & 1 & -1 & -1 \\
1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 \\
1 & -1 & 1 & -1 & -1 & -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 & 1 & 1 & -1 \\
1 & -1 & 1 & -1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 \\
1 & -1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 & -1 & 1 & 1 \\
1 & -1 & 1 & 1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 & -1 \\
1 & 1 & -1 & -1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 & 1 & -1 & 1 & 1 & -1 \\
1 & 1 & -1 & -1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & -1 & 1 & 1 & -1 & 1 \\
1 & 1 & -1 & 1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & -1 & 1 & -1 & 1 & 1 \\
1 & 1 & -1 & 1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 & -1 & -1 \\
1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\
1 & 1 & 1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & -1 \\
1 & 1 & 1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1
\end{bmatrix} \quad \boldsymbol{c}_{ij} = \begin{pmatrix}
m_{ij} \\
a_{i(j);*} \\
a_{j(i);*} \\
a_{*;i(j)} \\
a_{*;j(i)} \\
d_{ij;**} \\
d_{**;ij} \\
aa_{i(j);i(j)} \\
aa_{j(i);j(i)} \\
aa_{i;j} \\
aa_{j;i} \\
ad_{i;ij} \\
ad_{j;ij} \\
da_{ij;i} \\
da_{ij;j} \\
dd_{ij;ij}
\end{pmatrix}$$

Table 2 Genotypic means for the physiological genetic model called complementary gene effects

Locus 1	Locus 2			
	11	12	22	Mean
11	9λ	9λ	0	6λ
12	9λ	9λ	0	6λ
22	0	0	0	0
Mean	6λ	6λ	0	4λ

in which $a_{i(j);*}$ ($a_{j(i);*}$, $a_{*,i(j)}$, $a_{*;j(i)}$) represents the additive effect of parent i (j, i, j) at locus 1 (1, 2, 2) in the cross with parent j (i, j, i); $d_{ij;**}$ ($d_{**;ij}$) represents the dominance effect at locus 1 (2) for the combination of parents i and j; $aa_{i(j);i(j)}$ ($aa_{j(i);j(i)}$) represents the parallel additive-by-additive effect of parent i (j) in combination with parent j (i), $aa_{i;j}$ and $aa_{j;i}$ represent cross-parent additive-by-additive effects, $ad_{i;ij}$, $ad_{j;ij}$, $ad_{ij;i}$ and $ad_{ij;j}$ represent additive-by-dominance effects and $dd_{ij;ij}$ represents the dominance-by-dominance effect. The value of $\boldsymbol{c}$ can be obtained from $\boldsymbol{c}_{ij} = \frac{1}{16}\boldsymbol{B}^{\mathrm{t}}\boldsymbol{m}_{ij}$.

We may write

$$\boldsymbol{c}_{ij} = \begin{pmatrix} m_{ij} \\ a_{i(j);*} \\ a_{j(i);*} \\ a_{*;i(j)} \\ a_{*;j(i)} \\ d_{ij;**} \\ d_{**;ij} \\ aa_{i(j);i(j)} \\ aa_{j(i);j(i)} \\ aa_{i;j} \\ aa_{j;i} \\ ad_{i;ij} \\ ad_{j;ij} \\ da_{ij;i} \\ da_{ij;j} \\ dd_{ij;ij} \end{pmatrix} = \begin{pmatrix} m+g_i & +g_j & +s_{ij} & & \\ a_{i;*} & & +t_{i(j);*} & & \\ & a_{j;*} & +t_{j(i);*} & & \\ a_{*;i} & & +t_{*;i(j)} & & \\ & a_{*;j} & +t_{*;j(i)} & & \\ & & & d_{ij;**} & \\ & & & d_{**;ij} & \\ aa_{i;i} & & +u_{i(j);i(j)} & & \\ & aa_{j;j} & +u_{j(i);j(i)} & & \\ & & & aa_{i;j} & \\ & & & aa_{j;i} & \\ & & & ad_{i;ij} & \\ & & & ad_{j;ij} & \\ & & & da_{ij;i} & \\ & & & da_{ij;j} & \\ & & & dd_{ij;ij} & \end{pmatrix}$$

in which m represents a grand mean, g_i (g_j) represent the general combining ability of parent i (j) and s_{ij} represents the specific combining ability of the combination of parents i and j. Furthermore, $a_{i;*}$ ($a_{j;*}$, $a_{*,i}$, $a_{*;j}$) represents the general part of additive effect of parent i (j, i, j) at locus 1 (1, 2, 2) and $t_{i(j);*}$ ($t_{j(i);*}$, $t_{*,i(j)}$, $t_{*;j(i)}$) represents the specific part of the additive effect of parent i (j, i, j) at locus 1 (1, 2, 2) in the cross with parent j (i, j, i). Finally, $aa_{i;i}$ ($aa_{j;j}$) represents the general part of the parallel additive-by-additive effect of parent i (j) and $u_{i(j);i(j)}$ ($u_{j(i);j(i)}$) represents the specific part of the parallel additive-by-additive effect of parent i (j) in combination with parent j (i).

Table 3 Mean values associated with the 16 inheritance states

Inheritance state				
Locus 1	Locus 2			
	11	12	21	22
11	$m_{11;11}$	$m_{11;12}$	$m_{11;21}$	$m_{11;22}$
12	$m_{12;11}$	$m_{12;12}$	$m_{12;21}$	$m_{12;22}$
21	$m_{21;11}$	$m_{21;12}$	$m_{21;21}$	$m_{21;22}$
22	$m_{22;11}$	$m_{22;12}$	$m_{22;21}$	$m_{22;22}$

Segregation types

For the parents, the segregation types at two loci (separated by a '/') are written as $Q_{11}Q_{12} \times Q_{21}Q_{22}$/$Q_{11}Q_{12} \times Q_{21}Q_{22}$, in which Q_{ph} is the QTL allele of parent p (= 1, 2) on homolog h (= 1, 2). As a consequence, 256 different segregation types may be encountered. As an example, the parameters of the statistical genetic model will be calculated for a number of segregation types in the case of the complementary effects model. Results are shown in Table 4.

The following can be observed from Table 4. For segregation type 11 × 22/11 × 22, the first parent has genotypic mean 9λ, the second parent has genotypic mean 0 and all offspring have genotypic mean 0. For segregation type 11 × 11/22 × 22, both parents have genotypic mean 0, whereas all offspring have genotype 9λ. Both outcomes are a clear indication of non-additivity.

The presence of non-zero additive effects not only requires segregation, but also variation in the genotypic values in the offspring population. This is not the case for the first two segregation types. However, in the case of interaction, additive effects of parents may vary from one cross to the other (Jannink and Jansen 2001). Segregation types 11 × 12/12 × 22

Table 4 Values of the parameters of the statistical genetic model (shown as multipliers of λ) for a number of segregation types in the case of the complementary effects model

Parameter	Segregation type				
	11 × 22/ 11 × 22	11 × 11/ 22 × 22	11 × 12/ 12 × 22	12 × 12/ 12 × 22	12 × 12/ 12 × 12
m_{ij}	9	0	$\frac{9}{2}$	$\frac{27}{8}$	$\frac{81}{16}$
$a_{i(j);*}$	0	0	0	$-\frac{9}{8}$	$-\frac{27}{16}$
$a_{j(i);*}$	0	0	0	$-\frac{9}{8}$	$-\frac{27}{16}$
$a_{*;i(j)}$	0	0	$-\frac{9}{2}$	$-\frac{27}{8}$	$-\frac{27}{16}$
$a_{*;j(i)}$	0	0	0	0	$-\frac{27}{16}$
$d_{ij;**}$	0	0	0	$-\frac{9}{8}$	$-\frac{27}{16}$
$d_{**;ij}$	0	0	0	0	$-\frac{27}{16}$
$aa_{i(j);i(j)}$	0	0	0	$\frac{9}{8}$	$\frac{9}{16}$
$aa_{j(i);j(i)}$	0	0	0	0	$\frac{9}{16}$
$aa_{i;j}$	0	0	0	0	$\frac{9}{16}$
$aa_{j;\mathrm{i}}$	0	0	0	$\frac{9}{8}$	$\frac{9}{16}$
$ad_{i;ij}$	0	0	0	0	$\frac{9}{16}$
$ad_{j;ij}$	0	0	0	0	$\frac{9}{16}$
$da_{ij;i}$	0	0	0	$\frac{9}{8}$	$\frac{9}{16}$
$da_{ij;j}$	0	0	0	0	$\frac{9}{16}$
$dd_{ij;ij}$	0	0	0	0	$\frac{9}{16}$

and 12 × 12/12 × 22 may be considered as belonging to crosses with the second parent j as common parent with genotypic mean 0. In this case, at locus 1 the additive effect of the second parent changes from 0 in the first cross to $-\frac{9}{8}\lambda$ in the second cross. Segregation types 12 × 12/12 × 22 and 12 × 12/12 × 12 may be considered as belonging to crosses with the first parent i as common parent with genotypic mean 9λ. Now, at the first locus the additive effect of the first parent changes from $-\frac{9}{8}\lambda$ in the first cross to $-\frac{27}{16}\lambda$ in the second cross. At the second locus the additive effect of the first parent changes from $-\frac{27}{8}\lambda$ in the first cross to $-\frac{27}{16}\lambda$ in the second cross. Differences in additive effects of parents between crosses are an indication of interaction. Dominance can only be found in crosses in which both parents are segregating at one or both loci.

The additive × additive part of non-additivity can be split into two parts: $aa_{i(j);i(j)}$ ($aa_{j(i);j(i)}$) measures the difference in additive effects between locus 1 and 2 for parent i (j) in combination with parent j (i); $aa_{i;\,j}$ ($aa_{j;\,i}$) measures the difference between the additive effect of parent i (j) at locus 1 and the additive effect of parent j (i) at locus 2. The first component can be split into a general component for parent i (j) over all crosses involving parent i (j) and into a specific component. The latter component is always specific to the combination of parents i and j. The first term can be used in the search for evidence of even more complex forms of non-additivity than the additive-by-additive component. For example, for the segregation types 12 × 12/12 × 22 and 12 × 12/12 × 12 (in which the first parent i may be considered as the common parent), the values of $aa_{i;i}$ are $\frac{9}{8}\lambda$ and $\frac{9}{16}\lambda$, respectively. This is an indication that the interaction has a more complicated form than additive × additive.

It is clear from Table 4 that the information that we will obtain from the data about the complexity of the physiological genetic model depends very much on the combinations of parents that have been used for the making crosses.

Inheritance states

With regard to two loci, the expectation of the observation on offspring individual k from the cross between parents i and j may be written as

$$\omega_{ijk} = \boldsymbol{p}_{ijk}{}^{\mathrm{t}}\boldsymbol{m}_{ij} = \boldsymbol{p}_{ijk}{}^{\mathrm{t}}\boldsymbol{B}\boldsymbol{c}_{ij},$$

in which $\mathbf{p}_{ijk}$ is a 16 × 1 vector containing predictions of inheritance states arranged in the same way as the elements of $\mathbf{m}_{ij}$. For example, $p^{ijk}_{12;21}*$ is the probability that given all marker data, at locus 1

individual k obtained a copy of the allele on homolog 1 of parent i and a copy of the allele on the homolog 2 of parent j; on locus 2 it obtained a copy of the allele on homolog 2 of parent i and a copy of the allele on the homolog 1 of parent j.

At this stage, we will use a vector $\tilde{\mathbf{p}}$ rather than $\mathbf{p}$. If it is assumed that inheritance states are conditionally independent given all available marker data, values of the elements of $\tilde{\mathbf{p}}$ can be obtained using the multiplication rule. For example,

$$\tilde{p}^{ijk}_{12;21} = p^{k}_{i\cdot;\cdot\cdot}\left(1 - p^{k}_{j;\cdot\cdot}\right)\left(1 - p^{k}_{\cdot\cdot;i\cdot}\right)p^{k}_{\cdot\cdot;j},$$

in which $p^{k}_{i\cdot;\cdot\cdot}$ ($p^{k}_{j;\cdot\cdot}$,$p^{k}_{\cdot\cdot;i\cdot}$ $p^{k}_{\cdot\cdot;j}$) denotes the conditional marginal probability that given all available marker data that on locus 1 (1, 2, 2) individual k obtained a copy of the allele on homolog 1 of parent i (j, i, j). For the calculation of predictions of inheritance states use was made of FlexQTL (Bink 2005; http://www.biometris.nl/uk/Software/FlexQTL/). Predictions can also be obtained using Loki (Heath 1997; http://loki.homeunix.net/). In the above approach variances and covariances due to errors in prediction are assumed to be absent (see e.g. Xu 1995). The variances and covariances are small if the density of informative markers is fairly high ($\sim$10 cM). Using the above it can be derived that

$$\begin{aligned}\omega_{ijk} = {} & m_{ij} + q^{k}_{i\cdot;\cdot\cdot}a_{i(j);*} + q^{k}_{\cdot j;\cdot\cdot}a_{j(i);*} + q^{k}_{\cdot\cdot;i\cdot}a_{*;i(j)} \\ & + q^{k}_{\cdot\cdot;\cdot j}a_{*;j(i)} + q^{k}_{i\cdot;\cdot\cdot}q^{k}_{\cdot j;\cdot\cdot}d_{ij;**} + q^{k}_{\cdot\cdot;i\cdot}q^{k}_{\cdot\cdot;\cdot j}d_{**;ij} \\ & + q^{k}_{i\cdot;\cdot\cdot}q^{k}_{\cdot\cdot;i\cdot}aa_{i(j);i(j)} + q^{k}_{\cdot j;\cdot\cdot}q^{k}_{\cdot\cdot;\cdot j}aa_{j(i);j(i)} \\ & + q^{k}_{i\cdot;\cdot\cdot}q^{k}_{\cdot\cdot;\cdot j}aa_{i;j} + q^{k}_{\cdot j;\cdot\cdot}q^{k}_{\cdot\cdot;i\cdot}aa_{j;i} \\ & + q^{k}_{i\cdot;\cdot\cdot}q^{k}_{\cdot\cdot;i\cdot}q^{k}_{\cdot\cdot;\cdot j}ad_{i;ij} + q^{k}_{\cdot j;\cdot\cdot}q^{k}_{\cdot\cdot;i\cdot}q^{k}_{\cdot\cdot;\cdot j}ad_{j;ij} \\ & + q^{k}_{i\cdot;\cdot\cdot}q^{k}_{\cdot j;\cdot\cdot}q^{k}_{\cdot\cdot;i\cdot}da_{ij;i} + q^{k}_{i\cdot;\cdot\cdot}q^{k}_{\cdot j;\cdot\cdot}q^{k}_{\cdot\cdot;\cdot j}da_{ij;j} \cdot \\ & + q^{k}_{i\cdot;\cdot\cdot}q^{k}_{\cdot j;\cdot\cdot}q^{k}_{\cdot\cdot;i\cdot}q^{k}_{\cdot\cdot;\cdot j}dd_{ij;ij}\end{aligned}$$

in which $q^{k}_{i\cdot;\cdot\cdot}$ $(q^{k}_{j;\cdot\cdot},q^{k}_{\cdot\cdot;i\cdot},q^{k}_{\cdot\cdot;j}) = 1 - 2p^{k}_{i\cdot;\cdot\cdot}$ $(p^{k}_{j;\cdot\cdot},p^{k}_{\cdot\cdot;i\cdot},p^{k}_{\cdot\cdot;j})$. As a consequence, estimates of the parameters of the statistical genetic model can be obtained using linear regression on predictions of inheritance states or products of these predictions (cf. Haley et al. 1994).

Approach

The first part of the approach follows the following steps:

1. Decomposition of the observations into general combining abilities (GCA) of parent individuals, specific combining abilities (SCA) of combinations of parent individuals and residual effects. Statistical significance of specific combining abilities indicates the presence of non-additivity (dominance or epistatic effects).
2. Genome-wide linear search of positions that are acting additively.
   In this paper a simple, forward selection approach is used. Other approaches, such as backward elimination or penalised regression (Boer et al. 2002), will not be discussed at this stage. Additive effects are decomposed into general additive effects of parents and specific effects for combinations of parents. Positions for which specific effects are statistically significant are subject to intra-locus or inter-locus non-additivity. Only, these positions will be investigated further in Steps 3 and 4.
3. Analysis of intra-locus non-additivity of positions showing non-additivity in Step 2.
4. Analysis of inter-locus non-additivity of positions showing non-additivity in Step 2. This step will be limited to the additive $\times$ additive component of inter-locus non-additivity. For each cross, the four degrees of freedom of the additive $\times$ additive component of non-additivity can be divided into a part related to the parameters $aa_{i(j);i(j)}$ and $aa_{j(i);j(i)}$, and a part related to the parameters $aa_{i;j}$ and $aa_{j;i}$. Over all crosses, parts (a) and (b) of the additive $\times$ additive component of non-additivity will be decomposed into (1) a general component that can be attributed to parents (with number of degrees of freedom equal to the number of parents), and (2) a specific component that can be attributed to combinations of parents (with number of degrees of freedom equal to twice the number of crosses minus the number of parents). Statistical significance of component (2) would indicate the presence of more complicated forms of non-additivity than the additive $\times$ additive form. Part (c) is always specific to parent combinations.

The second part of the approach starts with identifying parents that are segregating at interacting loci (and preferably non-segregating at other loci). If such parents are present a large population of several

hundred individuals can be made by crossing these parents. This population will be used to validate the findings of the first part of the approach and for fitting/estimating a physiological-genetic model.

Calculations

The statistical analysis can be carried in the form of multiple linear regression and analysis of variance. Calculations were carried out using Genstat (Genstat Committee, 2006). A part of the results will be provided in the form of *P*-values or minus their logarithms to the base 10. In this paper results of statistical tests will be declared significant if the *P*-value is smaller than 0.001 ($-^{10}\log(0.001) = 3$).

Application

Data simulation

The data concern 24 crosses involving 15 parents according to the scheme presented in Table 5. Although in practice the parents may be related by pedigree, it is assumed that the parents are unrelated. In this application, we assume that the individuals of the species concerned are hermaphrodite, so that parents can be used as mother and as father. Two linkage groups of each 100 cM will be considered. On linkage group I two QTL are present: QTL A at 25 cM and QTL B at 75 cM; on linkage group II one QTL is present: QTL C at 50 cM. QTL A and C are interacting in the way described in "Model and methods"; QTL B is additive.

It is assumed that the parents are randomly sampled from a base population in which for each of the three QTL, two alleles (denoted by 1 and 2) are present. For each of the QTL, the two alleles occur with a frequency of 0.5 in the base population. It is assumed that the base population is in Hardy–Weinberg and in linkage equilibrium. For locus A, B and C the genotypes at the QTL are shown in Table 6.

For each cross, genotypes of 50 offspring were obtained. Depending on the QTL genotypes of the parents, the offspring individuals may have QTL genotypes 11, 12, 21 or 22. Trait observations *Y* were obtained using the formula

$$\begin{aligned} Y = {} & 0 * (Q_A \equiv 22 \,\text{OR}\, Q_C \equiv 22) \\ & + 5 * (Q_A \in (11, 12, 21) \,\text{AND}\, Q_C \in (11, 12, 21)) \\ & + 1 * ((Q_B \equiv 22) - (Q_B \equiv 11)) + \sigma E \end{aligned}$$

in which Q_A, Q_B and Q_C denote the QTL genotype at QTL A, B and C, respectively; σ denotes the square root of the error variance and E denotes a standard normal random variate. In this application, σ has been set equal 1. The genotypic means for the parents are also shown in Table 6.

Table 5 Crossing schema

Mother	Father														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1		+							+		+				
2						+									
3				+							+			+	
4															
5		+									+				
6					+										
7				+						+					
8									+					+	+
9										+	+				
10															
11												+			
12													+		
13							+				+				
14			+												
15	+								+						

"+" indicates the presence of the cross

Table 6 Genotypes at the three QTL

Parent	Genotype			Genotypic mean
	A	B	C	
1	22	11	12	−1
2	11	21	22	0
3	11	22	12	6
4	11	11	11	4
5	12	21	12	5
6	21	21	11	5
7	12	22	21	6
8	11	12	22	0
9	22	12	11	0
10	22	12	11	0
11	12	12	21	5
12	22	12	12	0
13	22	12	12	0
14	11	12	12	5
15	22	11	22	−1

On both linkage groups markers were positioned at 0, 10, 20… 100 cM. For the parents, marker data were obtained using the assumptions of Hardy–Weinberg and linkage equilibrium. For all markers the number of alleles was set equal to five. The phases of markers in the parents have been assumed to be known.

Analysis

Step 1 Table 7 shows the analysis of variance corresponding with the classical decomposition into general and specific combining abilities. From Table 7 it may be concluded that only 7.1% of the differences between population means can be attributed to specific combining abilities, whereas 92.9% are attributed to general combining abilities. The fact that specific combining abilities are present ($P < 0.001$) indicates the presence of physiological genetic interaction, i.e. dominance or epistatic effects. In Fig. 1 the population variances have been plotted against the corresponding means. Figure 1 shows that the variance is not constant. The smallest variances are obtained in populations with either a small mean or a large mean. In general, populations with 'average' means have much larger variances. Parent 9 is a parent of populations 2, 16 and 24 (small mean/small variance), 17 (average mean/large variance) and 13 (large mean/small variance).

Table 7 Summary analysis of variance

Step	Source of variation	Df	Sum of squares	Mean square	*F* value	*P* value
1	GCA	14	4,155	296.8	61.69	<0.001
	SCA	9	316	35.1	7.30	<0.001
	Residual 1	1,176	5,658	4.8		
2	Lg I-25 cM common	15	838.7	55.9	24.43	<0.001
	Lg I-25 cM specific	33	401	12.2	5.31	<0.001
	Lg I-75 cM common	15	308	20.5	8.96	<0.001
	Lg I-75 cM specific	33	66	2.0	0.88	0.666
	Lg II-50 cM common	15	806	53.7	23.49	<0.001
	Lg II-50 cM specific	33	618	18.7	8.18	<0.001
	Residual 2	1,032	2,362	2.3		
3	Lg I-25 cM dominance	23	64	2.8	1.38	0.111
	Lg II-50 cM dominance	23	304	13.2	6.54	<0.001
	Residual 3	986	1,989	2.0		
4	Lg I-25 cM × Lg II 50 cM within-common	15	92	6.1	3.26	<0.001
	Lg I-25 cM × Lg II 50 cM within-specific	33	64	1.9	1.04	0.411
	Lg I-25 cM × Lg II 50 cM across	48	164	3.4	1.83	<0.001
	Residual 4	890	1668	1.9		
	Total	1,199	10,129	8.4		

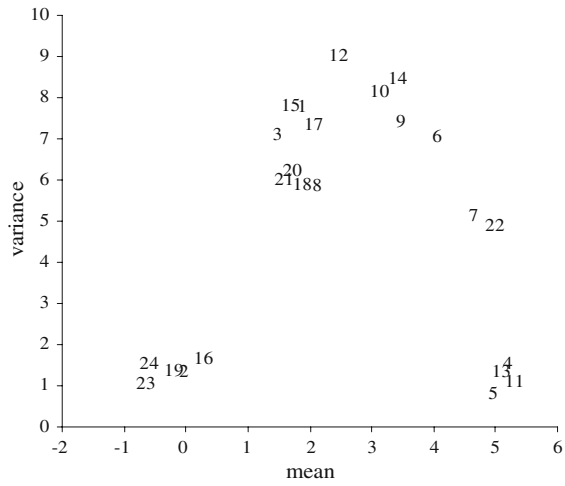

Fig. 1 Population variances plotted against population means; numbers refer to populations

Conclusion 1 From the first step of the analysis (decomposition of variation in GCA, SCA and residual) it may be concluded that non-additive effects play a role in the populations under study.

Step 2 For all 2,400 individuals predictions of inheritance states were obtained for positions 0, 5, 10… 100 cM on both linkage groups. These predictions will be used as regressors in the QTL analysis.

First, the analysis will concentrate on additive effects. A QTL analysis can be carried out in each of the populations separately. Thereby it is ignored that parents are used more than once. Figure 2a, b shows the *P*-values (shown as $-^{10}\log(P\text{-value})$) for testing the absence of additive effects versus the position on the linkage groups. At each position 48 (= 24 × 2) additive effects are estimated. The highest peaks occur on linkage group I around 25 cM and on linkage group II around 50 cM. The corresponding *P*-values are extremely small indicating very significant additive effects. On linkage group I, much smaller peaks occur in some populations around 75 cM. However, in most of the populations no sign of a significant additive effect is found. If it supposed that QTL are present on linkage group I at 25 cM and on linkage group II at 50 cM, the peaks near 75 cM on linkage group I remain present (Fig. 2c, d).

Conclusion 2 The data indicate the presence of three QTL.

Figure 3a, b show the *P*-values associated with the null hypothesis that additive QTL effects of a parent are constant and do not depend on the other parent used in a cross. Now, at each position 15 additive effects are estimated, which leaves 33 degrees of freedom for testing the interaction of additive effects with genetic background. Figure 3a, b shows that the QTL at position 25 cM on linkage group I and the QTL at position 50 cM on linkage group II are subject to interaction of additive effects with genetic background. This may either be intra-locus interaction or inter-locus interaction. Figure 3c, d shows that for the QTL at 75 cM on linkage group I not a trace of interaction of additive effect with genetic background can be found. On the basis of this information, this QTL will be considered as additive.

Conclusion 3 The QTL at position 25 cM on linkage group I and the QTL at position 50 cM on linkage group II are involved in interaction; the QTL at position 75 cM on linkage group I must be considered as additive.

Step 3 Starting from a model which accounts for the conclusions of Step 2. the presence of intra-locus non-additivity was tested for the QTL at position 25 cM on linkage group I and the QTL at position 50 cM on linkage group II (Table 7, Step 3).

Conclusion 4 The QTL at 50 cM on linkage group 2 is subject to intra-locus non-additivity.

Step 4 Starting from a model which accounts for the conclusions of Step 3 the presence of inter-locus non-additivity will be investigated (Table 7, Step 4). The results shown in Table 7 indicate that the cross-specific component of the within-individual part of the additive-by-additive component of non-additivity is by far not significant.

Conclusion 5 The QTL at 25 cM on linkage group I and the QTL at 50 cM on linkage group II show inter-locus non-additivity of the additive × additive form. The results do not indicate the presence of more complex forms of additivity.

Parents with segregating QTL

Figure 4 shows the *t*-values for the cross-specific additive effects of parents for position 25 cM on

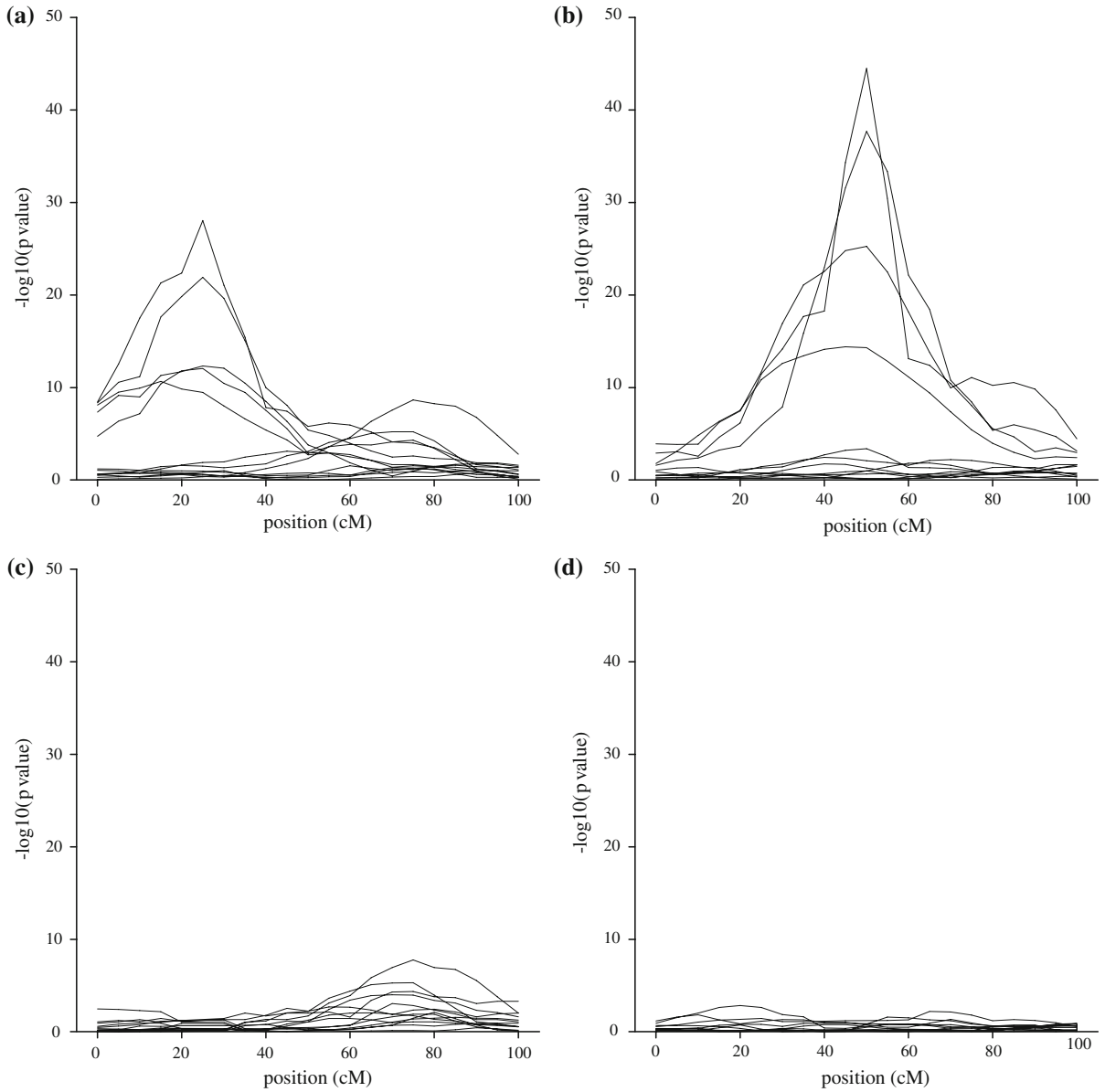

Fig. 2 **a**, **b** *P*-values (shown as $-^{10}\log(P\text{-value})$) for testing the absence of additive effects versus the position on linkage groups I and II, respectively. **c**, **d** *P*-values (shown as $-^{10}\log(P\text{-value})$) for testing the absence of additive effects versus the position on linkage group I and II, respectively, if QTL are assumed at 25 cM on linkage group I and at 50 cM on linkage group II. Different lines refer to different populations

linkage group I and position 50 cM on linkage group II. The solid lines represent significance thresholds for the *t*-test (here taken as ± 2). It follows from Fig. 4, that parents 5 and 11 show significant additive effects in at least one cross. This figure also shows in a clear way the interaction of additive effects and genetic background. A majority of estimated additive effects is significant at neither of the two loci considered.

Conclusion 6 A cross between parents 5 and 11 should be made in order to confirm the results of the current analysis and to investigate the interlocus non-additivity at the two QTL in more detail.

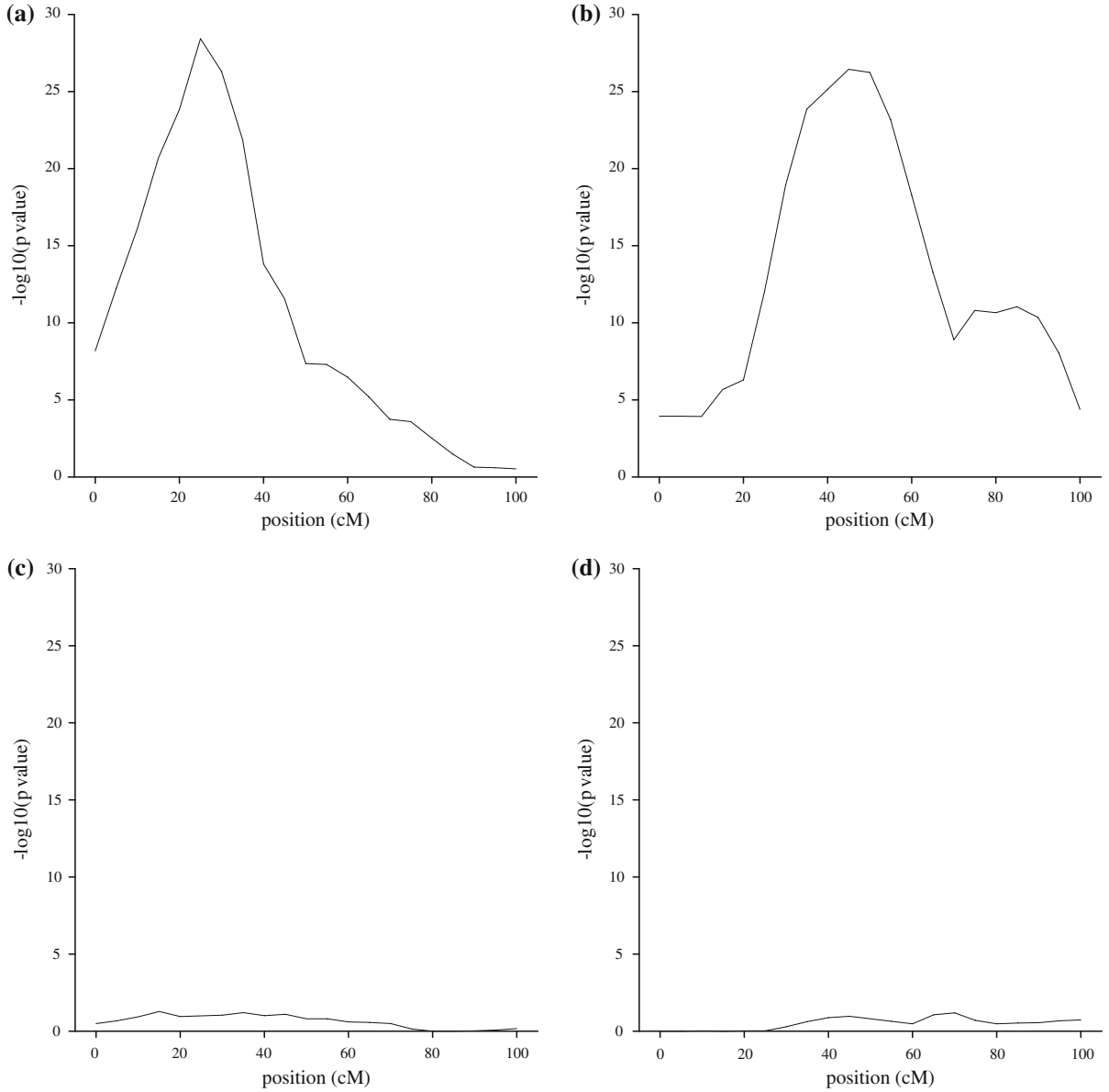

Fig. 3 **a**, **b** *P*-values (shown as $-^{10}\log(P\text{-value})$) for testing the absence of interaction between additive effects and genetic background versus the position on linkage groups I and II, respectively. **c** *P*-values (shown as $-^{10}\log(P\text{-value})$) for testing the absence of interaction between additive effects and genetic background versus the position on linkage groups I and II, respectively, if QTL are assumed at 25 cM on linkage group I and at 50 cM on linkage group II

Discussion

The methods described in this paper enable geneticists and plant breeders working with outbreeding species to identify loci involved in epistatic interactions. In order to identify the presence of interacting loci a simple "top-down" approach is followed to identify the presence of epistatic QTL. As expected, going from "top" to "bottom" the law of diminishing returns holds, i.e. the statistical significance (as expressed by the *P* value) diminishes rapidly with the increasing complexity of effects. In a genetical study like the one discussed in this paper and using the available data, it is not possible to model all aspects of physiological genetic epistasis. The initial aim must be limited to identification of interacting loci and to

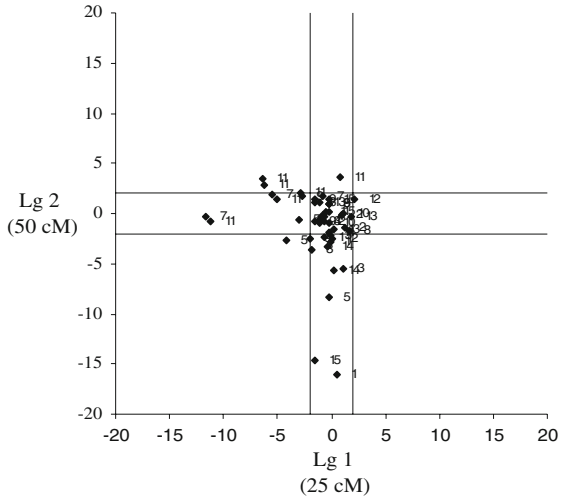

Fig. 4 *t*-Values for cross-specific additive effects; numbers indicate parents; lines represent significance thresholds (±2)

identification of combinations of parent individuals that can provide information about all aspects of the underlying physiological epistatic model.

In the current paper it is assumed that the parents are genetically independent. However, the approach can be extended to allow a pedigree on top of the parents. This would require a matrix of transition probabilities from alleles of the founders of the pedigree to parental alleles. These transition probabilities can be obtained by using FlexQTL (Bink 2005; http://www.biometris.nl/uk/Software/FlexQTL/) or Loki (Heath 1997; http://loki.homeunix.net/). If the number of founder alleles is greater than the number of parental alleles the pedigree will work as a sieve: some alleles pass through, other alleles stay behind. The problems of overparameterization are easily dealt with in a statistical computer package like Genstat (Genstat Committee, 2006).

In the application a biallelic QTL model is used for generating the data. However, the model that is used for analyzing the data does not require assumptions about the number of QTL alleles. In every cross the number of QTL alleles is equal to the total number of inheritance states, which is equal to four in diploid species. The model allows effects of parental alleles to be cross-specific. Allelic effects may also be restricted to be identical over crosses. This allows the testing of hypotheses concerning the presence of QTL × genetic background interaction in a simple way.

In comparatively simple QTL studies involving a single cross between two parents significance thresholds have been obtained in various ways. Churchill and Doerge (1994) used a permutation test and Van Ooijen (1999) used simulation to obtain significance thresholds. In this paper results of statistical tests have been declared significant if the *P*-value is smaller than 0.001 ($-^{10}\log(0.001) = 3$). The *P*-values presented in Table 7 are obtained for the final multiple linear regression model, and ignore the fact that this model is obtained after many steps of selection. Further studies are needed to determine the true frequency of false positives.

Yi et al. (2003) used 300 individuals in their Bayesian analysis of interaction effects in back-cross populations derived from inbred lines. For a pair of loci a back-cross involves four (= 2 × 2) possible inheritance states. If the predictions of the inheritance states can be made without error, this leads to an average of 75 individuals per inheritance state (standard error of a mean = 0.12σ). In the case of one full-sib family of an outbreeding species the analysis of interaction for two loci involves 16 possible inheritance states. This leads to some 18 individuals per inheritance state (standard error of a mean = 0.24σ). As a consequence, analysis of interaction effects in a full-sib family of an outbreeding species would require twice as many individuals in order to obtain the same level of precision as analysis of an inbreeding species. Also, interpretating a 2 × 2 table (back-cross) is much simpler than interpreting a 4 × 4 table (full-sib family of outbreeding species), even if the number of QTL alleles is limited to two.

The percentage of the total genetic variance attributed to interactions will usually be small. This is due to the fact that in the statistical analysis a large part of the contributions of functional epistasis to genetic variation will be attributed to additive effects. Therefore, these small percentages may misleading with regard to the biological importance of the effects of functional epistasis on phenotypic values. In vegetatively propagated crops, it is important to detect epistatic genes, because favourable combinations of alleles that are combined in one genotype can be maintained forever and can be exploited in the development of new varieties.

A great advantage of the method described in this paper is that it can be applied to data obtained from

populations of ongoing breeding programs (cf. Jannink et al. 2001). Initially, no special populations have to be created. A full analysis of interaction can be deferred until real evidence of interaction has been found in the form of QTL × genetic background interaction, and also parents that are heterozygous at epistatic loci have been identified. A further advantage of the current method is that it only requires linear searches of the genome. The power of the method may be further increased by using more sophisticated techniques like penalized regression (Boer et al. 2002).

Acknowledgements Thanks are due to Chris Maliepaard and Fred van Eeuwijk for their comments on an earlier draft of this manuscript. This project has been carried out with the financial support from the Commission of the European Communities (Contract No. LK5-CT-2002-01492), Directorate-General Research Quality of Life and Management of Living Resources Program. This manuscript does not necessarily reflect the Commission's views and in no way anticipates its future policy in this area; its content is the sole responsibility of the authors.

References

Allard RW (1960) Principles of plant breeding. Wiley, New York

Alvarez-Castro JM, Carlborg O (2007) A unified model for functional and statistical epistasis and its application in quantitative trait loci analysis. Genetics 176:1151–1167. doi:10.1534/genetics.106.067348

Bink MCAM (2005) FlexQTL Software: efficient estimation of identity by descent probabilities and QTL mapping in pedigreed populations. Plant and animal genomes XIII conference, 15–19 January, San Diego, CA

Bink MCAM, Uimari P, Sillanpää MJ, Janss LLG, Jansen RC (2002) Multiple QTL mapping in related plant populations via a pedigree-analysis approach. Theor Appl Genet 104:751–762. doi:10.1007/s00122-001-0796-x

Boer MP, Ter Braak CJF, Jansen RC (2002) A penalized likelihood method for mapping epistatic quantitative trait loci with one-dimensional genome searches. Genetics 162:951–960

Bulmer M (1985) The mathematical theory of quantitative genetics. Clarendon Press, Oxford

Carlborg Ö, Andersson L, Kinghorn B (2000) The use of a genetic algorithm for simultaneous mapping of multiple interacting quantitative trait loci. Genetics 155:2003–2010

Carlborg Ö, Hocking PM, Burt DW, Haley CS (2004) Simultaneous mapping of epistatic QTL in chickens reveals clusters of QTL pairs with similar genetic effects on *growth*. Genet Res Camb 83:197–209

Cheverud JM, Routman EJ (1995) Epistasis and its contribution to genetic variance components. Genetics 139:1455–1461

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

Cockerham CC, Zeng Z-B (1996) Design III with marker loci. Genetics 143:1437–1456

Du F-X, Hoeschele I (2000) Estimation of additive, dominance and epistatic variance components using finite locus models implemented with a single-site Gibbs and a descent graph sampler. Genet Res Camb 76:187–198

George AW, Visscher PM, Haley CS (2000) Mapping quantitative trait loci in complex pedigrees: a two-step variance component approach. Genetics 156:2081–2092

Gianfranceschi L, Soglio V (2004) The European Project HIDRAS: innovative multidisciplinary approaches to breeding high quality disease resistant apples. Acta Hortic 663:327–330

Haley CS, Knott SA, Elsen J-M (1994) Mapping quantitative trait loci in crosses between outbred lines using least squares. Genetics 136:1195–1207

Heath SC (1997) Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. Am J Hum Genet 61:748–760. doi:10.1086/515506

Jannink J-L (2007) Identifying quantitative trait locus by genetic background interactions in association studies. Genetics 176:553–561. doi:10.1534/genetics.106.062992

Jannink J-L (2008) QTL × genetic background interaction: predicting inbred progeny value. Euphytica 161:61–69. doi:10.1007/s10681-007-9509-0

Jannink J-L, Jansen RC (2001) Mapping epistatic quantitative trait loci with one-dimensional genome searches. Genetics 157:445–454

Jannink J-L, Bink MCAM, Jansen RC (2001) Using complex plant pedigrees to map valuable genes. Trends Plant Sci 6:337–342. doi:10.1016/S1360-1385(01)02017-9

Kao C-H, Zeng Z-B (2002) Modeling epistasis of quantitative trait loci using Cockerham's model. Genetics 160:1243–1261

Knott SA (2005) Regression-based quantitative trait loci mapping: robust, efficient and effective. Philos Trans R Soc B 360:1435–1442

Melchinger AE, Utz HF, Piepho H-P, Zeng Z-B, Schön CC (2007) The role of epistasis in the manifestation of heterosis: a systems-oriented approach. Genetics 177:1815–1825. doi:10.1534/genetics.107.077537

Van Ooijen JW (1999) LOD significance thresholds for QTL analysis in experimental populations of diploid species. Heredity 83:613–624. doi:10.1038/sj.hdy.6886230

Xu S (1995) A comment on the simple regression method for interval mapping. Genetics 141:1657–1659

Yi N, Xu S, Allison DB (2003) Bayesian model choice and search strategies for mapping interacting quantitative trait loci. Genetics 165:867–883

Zeng Z-B, Wang T, Zou W (2005) Modeling quantitative trait loci and interpretation of models. Genetics 169:1711–1725. doi:10.1534/genetics.104.035857