Mapping interspecific genetic architecture in host-parasite interaction system

Jian Yang, Weiren Wu, and Jun Zhu

Institute of Bioinformatics, College of Agriculture & Biotechnology, Zhejiang University, Hangzhou, P. R. China, 310029.

Corresponding Author:

Jun Zhu

Institute of Bioinformatics, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou, P. R. China, 310029.

Phone: (86)571-86971731.
Fax: (86)571-86971498.
Email: jzhu@zju.edu.cn.

Key words: Host-parasite interaction; Interspecific genetic architecture; Between-species interaction; QTL mapping; Monte Carlo simulation.
ABSTRACT

Under a hypothesis that the host-parasite interaction system is governed by genome-for-genome interaction, we propose a genetic model that integrates genetic information from both of the host and parasite genomes. The model can be used for mapping quantitative trait loci (QTLs) conferring the interaction between host and parasite and detecting interactions among these QTLs. A one-dimensional genome scan strategy is used to map QTLs in both of the host and parasite genomes simultaneously conditioned on selected pairs of markers controlling the background genetic variation; a two-dimensional genome scan procedure is conducted to search for epistasis within the host and parasite genomes and interspecific QTL by QTL interactions between the host and parasite genomes. Permutation test is adopted to calculate empirical threshold to control the experimental-wise false positive rate of detected QTLs and QTL interactions. Monte Carlo simulations were conducted to examine the reliability and efficiency of the proposed models and methods. Simulation results illustrated that our methods could provide reasonable estimates of the parameters and adequate powers for detecting QTLs and QTL by QTL interactions.
INTRODUCTION

The interaction between a host and a parasite is a ubiquitous biological phenomenon, which has significant impact on ecology, evolution and agriculture. Host-parasite interaction drives the coevolution of both the interacting species (Hamilton 1980; Thompson 1994). It is believed that host-parasite coevolution has generated much of the biological diversity on the Earth (Rauscher 2001). Processes analogous to the host-parasite coevolution also occur in agricultural systems. When breeders release disease-resistant crop varieties, new virulent pathogen strains will rapidly emerge to overcome the resistance of the crop.

Host-parasite interaction is genetically controlled by both of the host and parasite genomes. Most studies on the genetic mechanisms of host-parasite interaction and coevolution have been focused on related major genes of both the host (termed resistance genes) and the parasite (termed infection or virulence genes). Several models have been proposed to describe the relationship between the host resistance gene and the parasite infection gene. In plants, the most prevalent model is the gene-for-gene (GFG) model, in which the disease resistance of a host occurs only when its resistance ($R$) gene is matched by the parasite avirulence ($AVR$) gene (Flor 1971; Thompson and Burdon 1992). In animals, the most prevalent model is the matching alleles model, in which a host can defend against any parasite whose genotype does not match the host’s genotype (Agrawal and Lively 2001; Peters and Lively 1999). These two models represent two ends of a continuum (Agrawal and Lively 2001).

The GFG model has become the theoretical basis for plant disease-resistance breeding and has substantial molecular evidences. It has been known that $R$ and $AVR$ genes are responsible for the mutual recognition between plants and pathogens.
During the last decade, more than 40 plant $R$ genes have been cloned (Martin et al. 2003), all of which follow the GFG model. Besides, two pairs of $R$ and $AVR$ genes ($Pto/AVR-Pto$ and $Pita/AVR-Pita$) have been proved to interact directly via their encoded proteins (Jia et al. 2000; Scofield et al. 1996; Tang et al. 1996). Although the resistance and infection genes play the most important roles in the host-parasite interaction, they are not the only genes involved. In fact, host-parasite interaction is a complicated biological process, in which many genes apart from the resistance/infection genes in either the host or the parasite genome are involved. In addition, many host-parasite interactions do not rely on the resistance/infection genes, but are presumably controlled by many genes with minor effects. Therefore, host-parasite interaction is actually a complex system controlled by many genes of both species. To completely understand the genetic basis of host-parasite interaction, it is necessary to study not only the resistance/infection genes, but also many related minor genes.

With the advent of molecular marker technology, mapping of quantitative trait loci (QTLs) have become a powerful tool for dissecting the genetic architectures of complex traits. Systematic statistical methods for QTL mapping have been well established (Haley and Knott 1992; Kao et al. 1999; Lander and Botstein 1989; Ljungberg et al. 2004; Zeng 1994). However, all the genetic models for QTL mapping proposed up to date are limited to single species. Consequently, QTL mapping for host-parasite interaction can only be performed either in the host or in the parasite. In fact, most QTL mapping studies on the plant-pathogen interaction reported have been focused on the plant (host) side. Since the host-parasite interaction involves the genetic control of both of the host and parasite genomes, a comprehensive understanding of the genetic basis of host-parasite interaction requires...
knowing not only the genes involved in host and parasite, but also the interaction between host genes and parasite genes. Hence, current genetic models of QTL mapping are not adequate for the full dissection of the genetic architecture of host-parasite interaction.

By considering host and parasite as an integrated biological system and assuming a genome by genome interaction mechanism in the system, we propose a general quantitative genetic model for the host-parasite interaction. The genetic model fits our current knowledge on the genetic basis of host-parasite interaction and enables us to simultaneously detect QTLs and their interactions related to host and parasite. Monte Carlo simulations are conducted for demonstrating the application of the host-parasite model to mapping QTLs under a special experimental design.

**GENETIC MODELS**

In quantitative genetics, there is a basic formula for describing a quantitative trait:

\[ P = \mu + G + e \]  

where \( P \) is the phenotypic value, \( \mu \) is the mean of population, \( G \) is the genotypic effect, and \( e \) is the environmental deviation. Although the original formula is limited to a single species, it can be extended to describe the interaction between a host and a parasite. Suppose \( P_{HP} \) is the phenotypic value of a trait (e.g. disease index) reflecting the host-parasite interaction. Obviously, the trait is genetically controlled by both of the host genome and the parasite genome. Therefore, the genotypic effect \( G \) can be partitioned into components due to host genotypic effect \( G_H \), parasite genotypic effect \( G_P \), and genotypic interaction between host and parasite \( G_{HP} \), so that the basic formula can be rewritten as
This extended formula provides us with a general genetic model for the analysis of
host-parasite interaction.

A detailed genetic model of host-parasite interaction depends on both the genetic
structures of the host and the parasite in specific experimental design. Without loss of
generality, we can use the plant-fungus interaction as a model to address the basic
principle of our methodology. An obvious feature of this system is that the sporophyte
generation is generally dominant in the host (plant) life cycle, while the gametophyte
generation is generally dominant in the parasite (fungus) life cycle. Therefore, the
host is generally diploid and the parasite is generally haploid. For simplicity, we
consider the situation of using a population of doubled haploid lines (DHLs) of the
host derived from a cross between two pure lines \( H_1^P \) and \( H_2^P \) and a population of
haploid strains (HSs) of the parasite derived from a cross between two strains \( P_1^P \)
and \( P_2^P \). Suppose the number of DHLs is \( n_H \) and that of HSs is \( n_P \). Both the host
population and the parasite population have been genotyped with molecular markers,
and their corresponding genetic maps have been constructed. The host DHLs are
randomly inoculated by the parasite HSs and the symptom of each inoculation
combination is measured as an index of interaction between the host DHL and the
parasite HS. There is a constraint that each host line or parasite strain is involved in at
least one inoculation.

Based on the above experimental design, we can set up a genetic model of the
plant-fungus interaction. Suppose that there are \( s_H \) QTLs in the host genome \( Q^H \),
\( s_p \) QTLs in the parasite genome \( Q^P \), \( t_H \) epistatic QTL pairs in the host genome
\( (E Q^H) \), \( t_p \) epistatic QTL pairs in the parasite genome \( (E Q^P) \), and \( t_{HP} \) pairs of
interspecific QTL interactions between the host genome and the parasite genome \((E_Q^{HP})\). When tri-locus or higher order interactions are ignored, the host-parasite (plant-fungus) interaction can be expressed with the following genetic model

\[
y_{ij} = \mu + \sum_{k=1}^{4} \alpha_k^H \xi_{ki} + \sum_{l=1}^{4} \alpha_l^P \xi_{lj} + \sum_{m=n}^{4} \alpha_{mn}^{HH} \xi_m \xi_n + \sum_{q=r}^{4} \alpha_{qr}^{PP} \xi_{qj} + \sum_{u,v}^{4} \alpha_{uv}^{HP} \xi_u \xi_v + \varepsilon_{ij}
\]

where \(y_{ij}\) is the trait value (symptom) of the \(i\)-th host (plant) DHL and the \(j\)-th parasite (fungus) HS; \(\mu\) is the model mean; \(\alpha_k^H\) is the additive effect of the \(k\)-th \(Q^H\); \(\alpha_l^P\) is the additive effect of the \(l\)-th \(Q^P\); \(\alpha_{mn}^{HH}\) is the \(E_Q^{HH}\) effect between the \(m\)-th and the \(n\)-th \(Q^H\); \(\alpha_{qr}^{PP}\) is the \(E_Q^{PP}\) effect between the \(q\)-th and the \(r\)-th \(Q^P\); \(\alpha_{uv}^{HP}\) is the \(E_Q^{HP}\) effect between the \(u\)-th \(Q^H\) and the \(v\)-th \(Q^P\); \(\xi_{ki}\) (\(\xi_{mi}\), \(\xi_{ni}\) or \(\xi_{ui}\)) is a dummy variable taking a value of 1 or \(-1\) when the genotype of the \(k\)-th (\(m\)-th, \(n\)-th or \(u\)-th) \(Q^H\) in the \(i\)-th DHL is like that of \(P_i^H\) or \(P_2^H\); \(\xi_{lj}\) (\(\xi_{mj}\), \(\xi_{nj}\) or \(\xi_{uj}\)) is also a dummy variable taking a value of 1 or \(-1\) when the genotype of the \(l\)-th (\(q\)-th, \(r\)-th or \(v\)-th) \(Q^P\) in the \(j\)-th HS is like that of \(P_i^P\) or \(P_2^P\); and \(\varepsilon_{ij}\) is the residual error, which is generally assumed to follow a normal distribution with zero mean.

**METHODOLOGY OF QTL MAPPING**

Model (3) provides a theoretical basis for simultaneously mapping host (plant) and parasite (fungus) QTLs underlying the host-parasite interaction, and for estimating the intra-genomic and inter-genomic QTL interactions. When fitting (3), it involves a problem of multi-dimensional search of multiple loci in two genomes. We adopt a two-step mapping strategy, by which QTLs are first searched for individual effects using one-dimensional (1D) scan throughout the host and parasite genomes,
and then for intra- and inter-genomic QTL interactions using two-dimensional (2D) scan with the previously mapped individual QTLs fixed in the model. After the mapping procedure is finished, all the detected QTLs and QTL interactions are integrated into (3) to estimate their effects.

**Scanning for QTLs in the host and parasite genomes:** The 1D scan is performed following the principle of composite interval mapping (Zeng 1994). But we select marker intervals (pairs of adjacent markers) instead of individual markers as cofactors in the model. This would enable more efficient control of the effects of background QTLs presumably located in the selected marker intervals, because the effect of a QTL could be completely absorbed by its flanking markers (Zeng 1993).

Suppose \(v_H\) marker intervals in the host and \(v_P\) marker intervals in the parasite are selected as cofactors, respectively. Define two dummy variables, \(\phi_H\) and \(\phi_P\), which take values of 1 and 0 when the test is performed on a putative QTL in the host genome or 0 and 1 when in the parasite genome. Thus, according to (3), the model for testing a putative QTL in either the host genome or the parasite genome can be written as

\[
y_{ij} = \mu + a_H^i x_{ij} + a_P^j z_{ij} + \sum_{k} (\alpha_k^i \xi_{ki}^- + \alpha_k^i \xi_{ki}^+) + \sum_{l} (\beta_l^j \zeta_{lj}^- + \beta_l^j \zeta_{lj}^+) + \epsilon_{ij}
\]

\[
= \mu + a_H^i x_{ij} + a_P^j z_{ij} + \xi_{ki}^- \alpha_k^i + \xi_{ki}^+ \alpha_k^i + \zeta_{lj}^- \beta_l^j + \zeta_{lj}^+ \beta_l^j + \epsilon_{ij}
\]  

(4)

where \(a_H^i\) (or \(a_P^j\)) is the additive effect of the putative QTL in the host (or parasite) genome; \(x_{ij}\) (or \(z_{ij}\)) is the expected value of the indicator variable of the putative QTL in the \(i\)-th host DHL (or \(j\)-th parasite HS), conditioned on the genotypes of its flanking markers (Jiang and Zeng 1997); \(\alpha_k^i\) and \(\alpha_k^i\) (or \(\beta_l^j\) and \(\beta_l^j\)) are the additive effects of the left and the right markers of the \(k\)-th (or \(l\)-th) marker interval in the host (or parasite) genome; \(\xi_{ki}^-\) and \(\xi_{ki}^+\) (or \(\zeta_{lj}^-\) and \(\zeta_{lj}^+\)) are indicator variables.
for the genotypes of the left and the right markers of the \( k \)-th (or \( l \)-th) marker interval in the \( i \)-th host DHL (or \( j \)-th parasite HS) following the similar definitions of \( \xi_\beta \) (or \( \zeta_\phi \)) in (3).

Model (4) can be expressed in matrix form as

\[
\mathbf{y} = \mathbf{W}_1 \mathbf{b}_1 + \mathbf{W}_2 \mathbf{b}_2 + \mathbf{\varepsilon}
\]

where \( \mathbf{y} \) is an \( n \times 1 \) vector of trait values; \( \mathbf{b}_1 = [\mu, \alpha', \beta']' \) and \( \mathbf{b}_2 = [a^R \phi_i, a^P \phi_p]' \); \( \mathbf{W}_1 \) and \( \mathbf{W}_2 \) are the design matrices corresponding to \( \mathbf{b}_1 \) and \( \mathbf{b}_2 \), respectively; \( \mathbf{\varepsilon} \) is an \( n \times 1 \) vector of residual errors. A general equation for the expected reduction sum of squares of \( \mathbf{b}_2 \) can be obtained,

\[
E[\text{SSR}(\mathbf{b}_2 | \mathbf{b}_1)] = E(\mathbf{y}' \mathbf{W} W^* \mathbf{y} - \mathbf{y}' \mathbf{W}_1 \mathbf{y}) = \text{tr}[\mathbf{W}_2^2 \mathbf{M}_1 \mathbf{W}_2 E(\mathbf{b}_2 \mathbf{b}_2')] + \sigma_e^2 (r_w - r_w_r)
\]

where \( \mathbf{W} = (\mathbf{W}_1 : \mathbf{W}_2) \); \( \mathbf{M}_1 = \mathbf{I} - \mathbf{W}_1 (\mathbf{W}_1' \mathbf{W}_1)^{-1} \mathbf{W}_1' \); \( r_w \) is the rank of \( \mathbf{W} \) and \( r_{w_r} \) is the rank of \( \mathbf{W}_1 \). Accordingly, under the null hypothesis \( H_0: \mathbf{b}_2 = 0 \), \( F \)-test can be performed based on Henderson method III (SEARLE et al. 1992)

\[
F = \frac{\text{SSR}(\mathbf{b}_2 | \mathbf{b}_1) / (r_w - r_w_r)}{\text{SSE} / (n - r_w)}
\]

The \( F \)-test can be performed across the host genome (\( \phi_i = 1 \) and \( \phi_p = 0 \), for \( H_0: a^H = 0 \)) and the parasite genome (\( \phi_i = 0 \) and \( \phi_p = 1 \), for \( H_0: a^P = 0 \)) with a step length of 1 cM. When the \( F \)-values in a region exceed a pre-defined critical threshold, a QTL is indicated at the position of the maximum \( F \) value.

In regard to the selection of cofactor marker intervals, we adopt a method similar to that of marker pair selection (MPS) (PIEPOHO and GAUCH 2001) to select cofactor marker intervals. The MPS approach has the advantage that markers enter the model in adjacent pairs, which reduces the number of models to be considered, thus
alleviating the problem of over-fitting and increasing the chances of detecting QTLs.

For the \( k \)-th marker interval of the host \((M_k^{-H}, M_k^{+H})\) or the \( l \)-th marker interval of the parasite \((M_l^{-P}, M_l^{+P})\), a single marker interval model can be written as

\[
y_{ij} = \mu + \alpha_i \xi_{il} \phi_{il} + \alpha_i \xi_{il} \phi_{il} + \beta_i \xi_{il} \phi_{il} + \beta_i \xi_{il} \phi_{il} + \epsilon_{ij}
\]

where all the parameters and variables have the same definitions as in (4). With equation (8), \( F \)-tests are used for testing all marker intervals based on Henderson method III. Marker intervals showing \( F \)-values greater than the given critical threshold are taken as candidate marker intervals.

**Detecting intra- and inter-genomic interactions between QTLs:** Suppose \( s_H \)

\( Q^H \)s and \( s_P \) \( Q^P \)s have been mapped in the 1D scan. Obviously, interaction between these QTLs can be tested easily by model selection technique analogous to the strategy of multiple interval mapping (Kao et al. 1999). However, there may be interactions involving some QTLs not detected by the 1D scan due to small or even no individual effects. In order to find out all possible QTL interactions, we adopt a 2D scan procedure, which tests any possible pair of loci within and between the host and parasite genomes, conditioned on the effects of the \( Q^H \)s and \( Q^P \)s mapped by the 1D scan. In addition, we select marker interval pairs that show significant interactions between them and put them into the model as cofactors to control background variance. Suppose we select \( f_H \) and \( f_P \) pairs of marker intervals within the host and parasite genomes and \( f_{HP} \) pairs of marker intervals between the host and parasite genomes. For simplicity, the interaction between two marker intervals (\( A \) and \( B \)) is approximately represented by the interactions between the left markers (\( A^- \) and \( B^- \)) and between the right markers (\( A^+ \) and \( B^+ \)) of the two intervals. Thus, the significance of \( EQ^H \), \( EQ^P \) or \( EQ^{HP} \) of any pair of loci (denoted by \( f \) and \( h \)) can be
tested using with $F$-statistic based on Henderson method III using the following model

$$y_{ij} = \mu + a_{HI} x_{ij} x_i \phi_h + a_{PP} z_{ij} z_j \phi_p + a_{HP} x_{ij} z_{ij} (1 - \phi_h) (1 - \phi_p) + \sum_{k=1}^{s_h} a_k^{HI} z_{kj} + \sum_{l=1}^{s_p} a_l^{PP} z_{lj} + \sum_{f=1}^{f_m} \left[ \alpha \alpha^{A^+ B^-} z_{ij} + \alpha \alpha^{A^- B^+} z_{ij} \right]$$

$$+ \sum_{m} \left[ \beta \beta^{A^+ B^-} z_{im} z_{im} + \beta \beta^{A^- B^+} z_{im} z_{im} \right]$$

$$+ \sum_{m} \left[ \alpha \beta^{A^+ B^-} z_{im} z_{im} + \alpha \beta^{A^- B^+} z_{im} z_{im} \right] + \epsilon_{ij} \quad (9)$$

where $a_{HI}$ and $a_{PP}$ are QTL interactions within the host and parasite genomes, respectively; $a_{HP}$ is QTL interaction between the host and parasite genomes; $\alpha \alpha^{A^+ B^-}$ and $\alpha \alpha^{A^- B^+}$ are interactions between the left and between the right markers of the $l$-th ($m$-th) pair of marker intervals in the host (parasite) genome; $\alpha \beta^{A^+ B^-}$ and $\alpha \beta^{A^- B^+}$ are interactions between the left and between the right markers of the $n$-th pair of marker intervals between the host and parasite genomes; the remaining parameters and variables have the same definitions as above. The 2D genome scan can be performed within the host ($\phi_h = 1$ and $\phi_p = 0$), for $H_0 : a^{HI} = 0$) or the parasite ($\phi_h = 0$ and $\phi_p = 1$, for $H_0 : a^{PP} = 0$) genome for intra-genomic QTL interactions, and between these two genomes ($\phi_h = 0$ and $\phi_p = 0$, for $H_0 : a^{HP} = 0$) for inter-genomic QTL interactions.

Selection of significant interactive marker interval pairs used as cofactors in (9) is performed with the following model using $F$-statistic based on Henderson method III.
\[ y_j = \mu + \alpha \alpha^{X \beta} \xi_i \xi_j \phi_i + \alpha \alpha^{X \beta} \xi_i \xi_j \phi_i \\
+ \beta \beta^{X \beta} \xi_i \xi_j \phi_i + \beta \beta^{X \beta} \xi_i \xi_j \phi_i \\
+ \alpha \beta^{X \beta} \xi_i \xi_j (1-\phi_i)(1-\phi_j) + \alpha \beta^{X \beta} \xi_i \xi_j (1-\phi_i)(1-\phi_j) \\
+ \sum_i (\alpha_i \xi_i + \alpha_i \xi_i) + \sum_j (\beta_j \xi_j + \beta_j \xi_j) + \epsilon_j \tag{10} \]

All the parameters and variables in (10) have been defined above. The cofactor marker intervals used in (4) are also fixed in (10) to reduce residual variance and increase the power of detecting marker interval interactions. We can perform this test on all possible pairs of marker intervals in the integrated host-parasite genetic system and consequently get a 2D $F$-statistic profile, according to which marker interval pairs corresponding to the peaks of $F$-values that exceed the pre-defined significance threshold value can be selected.

**Threshold determination and model selection:** Every step of the analysis discussed above (marker interval selection, 1D scanning for QTLs, marker interval pair selection, or 2D scanning for QTL pairs) involves the problem of testing multiple null hypotheses. Therefore, multiple test correction is needed. We can adopt the method of permutation tests (CHURCHILL and DOERGE 1994; DOERGE and CHURCHILL 1996) to control the experimental-wise false positive rate. For each of the steps, we only shuffle the relationship between the trait and the effect(s) to be tested but leave the relationships between the trait and the other effects unchanged because if the relationship between the trait and the non-tested effects are destroyed, an artificially low empirical threshold could be obtained from the permutation tests and thus an enormously high false positive rate would be obtained (CHURCHILL and DOERGE 1994). This is done by shuffling the row order of design matrix of the testing effects (e.g. $W_2$ in equation 5).

Furthermore, in each of the steps, the $F$-statistic profile is plotted and the peaks
that exceed the critical $F$-value calculated by permutation tests are identified as significant marker intervals, QTLs, marker interval pairs, or QTL pairs. However, some of the peaks could be ghost peaks due to the correlation of closely linked markers and random noise etc. So, we perform a stepwise model selection with also the aforementioned $F$-statistic as criteria to remove possible ghost peaks (marker intervals, QTLs, marker interval pairs, or QTL pairs). The threshold of $F$-value at a significance level of $\alpha/K$ is used to control the type I error rate in the model selection, where $K$ is the number of tests in each forward or backward selection process in the stepwise selection procedure.

**MONTE CARLO SIMULATIONS**

Real datasets necessary to validate the concept and method presented in the present study are not yet available in the public domain. To examine the feasibility, reliability and robustness of the proposed models and analysis methods, we performed simulation studies. The genetic maps of host and parasite were assumed to have 5 and 3 chromosomes, respectively. Eleven markers evenly spread over each of the host and parasite chromosomes with spacing of 10 cM. Assumed that there were 5 $Q^h$'s and 3 $Q^p$'s (Table 1), and 1 $EQ^h$, 1 $EQ^p$ and 2 $EQ^{hp}$'s (Table 2); 60% of the trait variation was attributed to the assumed interspecific genetic architecture; the relative contribution (RC) of individual QTLs (additive effects) and QTL pairs (interaction effects) ranged 3.22%~9.46% and 3.45%~6.57%, respectively; the population size of host DHLs and parasite HSs was 200, respectively. The number of host-parasite inoculation combinations was set as three cases, namely 400 (Case I), 600 (Case II) and 800 (Case III). Every case was simulated for 200 replicates. For every analysis
step in every case, an empirical threshold of $F$ statistic at the experimental-wise significance level of 0.05 was estimated with 1000 permutation tests.

Simulation results are shown in Tables 1 and 2. In general, our method could provide reasonably accurate estimates of the parameters (positions, additive effects and interaction effects). The precisions of parameter estimates and the statistical powers of QTL detection increased as the increase of sample size. The increase was considerable for those QTLs or QTL pairs with small RCs, indicating that a large sample is desirable for detecting QTLs and QTL interactions with minor RCs. The powers of detecting QTLs with RCs larger than 5.0% were all larger than 90.0% in all the three cases. The simulation results demonstrated the feasibility of our method for mapping QTLs underlying the host-parasite interaction.

(insert Tables 1 and 2 here)

DISCUSSION

Host-parasite interaction is a complex system controlled by many genes from the host and parasite genomes. For understanding the genetic basis of host-parasite interaction, it is necessary to study the underlying gene effects and their interactions from both of the host genome and the parasite genome. The method proposed in this paper integrates the host genome and the parasite genome into a genetic model so that host QTLs and parasite QTLs as well as their interactions can be simultaneously detected.

In the past, because of no suitable QTL mapping method, studies of QTL mapping for host-parasite interaction were limited to single species especially of the host side (KOVER and CAICEDO 2001). Such single-species-analysis approach can not comprehensively dissect the interspecific genetic architecture underlying the
host-parasite interaction system, because it can not detect the interactions between
host and parasite QTLs. In addition, the separated analysis approach would possibly
decrease the statistical powers of detecting QTLs and epistasis but increase the
standard errors of the estimates of QTL positions and effects in comparison with the
method proposed in this paper. This was demonstrated by analyzing the
aforementioned simulated data based on the host genome and the parasite genome
separately (results not presented). Take Case I for example. The powers of detecting
$Q_1^H$, $Q_3^H$, $Q_4^H$ and $Q_5^H$ decreased to 73.0%, 82.0%, 97.0% and 97.0%, and the
powers of detecting $Q_1^P$, $Q_2^P$ and $Q_3^P$ decreased to 86.5%, 55.5% and 97.%
Sample size is an important factor determining the statistical power in QTL
mapping. With the objective of mapping QTLs in both of the host and the parasite
genomes and detecting their QTL interactions, relatively large population sizes are
required for both the host and the parasite. We conducted a simulation study with a
sample size of 400 generated from 200 host DHLs and 20 parasite HSs. In that case,
each host DHL was inoculated by 2 different parasite HLs and each parasite HS was
inoculated to 20 different host DHLs. As compared with the result of Case I, the
powers of detecting parasite QTLs decreased by about half and the precision of effect
estimation of these QTLs remarkably decreased. It was not only because of the small
size of the parasite population, but also because that many data points were dependent
resulting from the multiple inoculations of each parasite HS. In our opinion, an ideal
design for our method is using two large populations (at least larger than 200) of host
and parasite with same size, and inoculating each host line by each different parasite
strain for only one time. However, in practices, it might be quite difficult to conduct
such a balance design. Our simulation study demonstrated that if one host line (or
parasite strain) was involved in inoculation with no more than 4 times, it would not
generate much independent data but still result in reasonable mapping results. In addition, the accuracy of estimated QTL effects is another important issue in QTL mapping. It was showed in the simulation study that the biases of all of the QTL effect estimates were less than 5%. However, the genome-wide scan approach tended to over-estimate the QTL effects (Goring et al. 2001). Any conclusion based on QTL effect estimates should take consideration of the estimation bias.

In principle, the general genetic model proposed in this paper could be applicable to all sorts of host-parasite interactions and the corresponding QTL mapping methodology would be suitable for all sorts of interactions between eukaryotic hosts and parasites. The detailed QTL mapping method presented could be directly applied to researching the genetic architecture of plant diseases caused by fungal pathogens. A suitable example would be rice blast, the most devastating fungal disease of rice, because both the genomes of rice and the fungal pathogen (Magnaporthe grisea) have been sequenced.

Every species interacts with other species in the ecosystem. There are various interactions between species such as competition and symbiosis. The host-parasite interaction is only one case of them. The method of QTL mapping for host-parasite interaction could be applied to other types of between-species interaction in principle, as long as a suitable trait could be found to quantitatively reveal the between-species interaction. Sometimes there might be several traits reflecting the between-species interaction from either or both of these species. For example, in an interaction system of food competition between two species, the growth rate of either species could be taken as a trait representing the interaction between the two species. In this case, it would be recommended to analyze the traits jointly for QTL mapping. Several methods of joint analysis of multiple traits for QTL mapping have been proposed.
Following similar principles, the genetic model of between-species interaction could also be extended for QTL mapping based on multiple traits. This would probably increase the statistical power of QTL mapping.

Genotype by environment interaction could also be considered in the between-species interaction analysis. The interaction between two species might be different under various environments. It has been revealed that the geographical separation may cause unique ecological and evolutionary dynamics of host and parasites at different locations (Benkman 1999; Burdon and Thrall 1999). The reciprocal selection between plants and parasite usually occurs in a subset of locations or coevolutionary hotspots (Thompson 1999). Therefore, as a prospect, a more general genetic model of between-species interaction should not only integrate the genetic information of both species, but also consider the influence of the macro- and micro-environmental factors on the expression of the genes and gene interaction to fit the interacted traits from both species.
Acknowledgments

We thank Prof. William Hill at Edinburg University for insightful comments. We also thank two anonymous reviewers for useful comments and suggestions on the earlier version of the manuscript. This research was supported in part by the National Basic Research Program (973 Program) of China (2004CB117306) and the National Natural Science Foundation of China (30671123).


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


Table 1 Estimation of individual QTLs for positions, additive effects, and statistical powers.

<table>
<thead>
<tr>
<th>QTL</th>
<th>True value</th>
<th>Estimate of position</th>
<th>Estimate of effect</th>
<th>Power (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Chrom.</td>
<td>Pos. (cM)</td>
<td>Effect</td>
<td>RC (%)</td>
</tr>
<tr>
<td>$Q_1^H$</td>
<td>1</td>
<td>41.0</td>
<td>3.0</td>
<td>3.69</td>
</tr>
<tr>
<td>$Q_2^H$</td>
<td>2</td>
<td>18.0</td>
<td>0.0</td>
<td>0.00</td>
</tr>
<tr>
<td>$Q_3^H$</td>
<td>3</td>
<td>87.0</td>
<td>–3.4</td>
<td>4.75</td>
</tr>
<tr>
<td>$Q_4^H$</td>
<td>4</td>
<td>20.0</td>
<td>4.0</td>
<td>6.57</td>
</tr>
<tr>
<td>$Q_5^H$</td>
<td>5</td>
<td>48.0</td>
<td>–4.8</td>
<td>9.46</td>
</tr>
<tr>
<td>$Q_1^P$</td>
<td>1</td>
<td>24.0</td>
<td>3.7</td>
<td>5.62</td>
</tr>
<tr>
<td>$Q_2^P$</td>
<td>2</td>
<td>65.0</td>
<td>–2.8</td>
<td>3.22</td>
</tr>
<tr>
<td>$Q_3^P$</td>
<td>3</td>
<td>73.0</td>
<td>–4.4</td>
<td>7.95</td>
</tr>
</tbody>
</table>

* Mean and standard error (in parentheses) of 200 simulation replicates.
Table 2

Estimation of pair-wise QTLs for epistasis effects, and statistical powers.

<table>
<thead>
<tr>
<th>QTL&lt;sub&gt;i&lt;/sub&gt;</th>
<th>QTL&lt;sub&gt;j&lt;/sub&gt;</th>
<th>True value</th>
<th>Effect</th>
<th>RC (%)</th>
<th>Estimate of effect *</th>
<th>Power (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Case I Case II Case III</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Case I Case II Case III</td>
<td></td>
</tr>
<tr>
<td>H&lt;sub&gt;1&lt;/sub&gt;</td>
<td>Q&lt;sub&gt;5&lt;/sub&gt;</td>
<td>–2.9</td>
<td>3.45</td>
<td></td>
<td>–2.85 (0.53) –2.74 (0.58) –2.73 (0.54)</td>
<td>60.5 92.0 99.5</td>
</tr>
<tr>
<td>P&lt;sub&gt;1&lt;/sub&gt;</td>
<td>Q&lt;sub&gt;3&lt;/sub&gt;</td>
<td>3.0</td>
<td>3.69</td>
<td></td>
<td>2.93 (0.55) 2.79 (0.59) 2.68 (0.62)</td>
<td>61.0 94.5 97.5</td>
</tr>
<tr>
<td>H&lt;sub&gt;2&lt;/sub&gt;</td>
<td>Q&lt;sub&gt;1&lt;/sub&gt;</td>
<td>4.0</td>
<td>6.57</td>
<td></td>
<td>3.83 (0.59) 3.81 (0.57) 3.77 (0.53)</td>
<td>74.5 96.0 98.5</td>
</tr>
<tr>
<td>H&lt;sub&gt;4&lt;/sub&gt;</td>
<td>Q&lt;sub&gt;3&lt;/sub&gt;</td>
<td>–3.5</td>
<td>5.03</td>
<td></td>
<td>–3.36 (0.68) –3.42 (0.50) –3.43 (0.48)</td>
<td>95.0 98.5 100.0</td>
</tr>
</tbody>
</table>

* Mean and standard error (in parentheses) of 200 simulation replicates.