

The Use of a Genetic Algorithm for Simultaneous Mapping of Multiple Interacting Quantitative Trait Loci

Örjan Carlborg,* Leif Andersson* and Brian Kinghorn†

*Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala Biomedical Center, S-751 24 Uppsala, Sweden and †Department of Animal Science, University of New England, Armidale, New South Wales 2351, Australia

Manuscript received February 23, 2000

Accepted for publication April 28, 2000

ABSTRACT

Here we describe a general method for improving computational efficiency in simultaneous mapping of multiple interacting quantitative trait loci (QTL). The method uses a genetic algorithm to search for QTL in the genome instead of an exhaustive enumerative ("step-by-step") search. It can be used together with any method of QTL mapping based on a genomic search, since it only provides a more efficient way to search the genome for QTL. The computational demand decreases by a factor of ~ 130 when using genetic algorithm-based mapping instead of an exhaustive enumerative search for two QTL in a genome size of 2000 cM using a resolution of 1 cM. The advantage of using a genetic algorithm increases further for larger genomes, higher resolutions, and searches for more QTL. We show that a genetic algorithm-based search has efficiency higher than or equal to a search method conditioned on previously identified QTL for all epistatic models tested and that this efficiency is comparable to that of an exhaustive search for multiple QTL. The genetic algorithm is thus a powerful and computationally tractable alternative to the exhaustive enumerative search for simultaneous mapping of multiple interacting QTL. The use of genetic algorithms for simultaneous mapping of more than two QTL and for determining empirical significance thresholds using permutation tests is also discussed.

RAPID development in molecular genetics has led to the development of dense genetic maps, which are a powerful tool for studying the molecular basis for quantitative genetic variation. Several methods for mapping quantitative trait loci have been proposed in the literature. These can be classified into four major groups: least-squares-based methods (*e.g.*, Haley and Knott 1992; Martinez and Curnow 1992), maximum-likelihood-based methods (*e.g.*, Lander and Botstein 1989), nonparametric methods (*e.g.*, Kruglyak and Lander 1995), or Bayesian methods (*e.g.*, Thomas and Cortessis 1992; Hoeschele and VanRaden 1993a,b).

Haley and Knott (1992) and Jansen (1992) have indicated how a simultaneous search for multiple interacting quantitative trait loci (QTL) can be performed by a stepwise search of the genome. These methods have not been used in practice for genome-wide searches for QTL due to the high computational demand. Jansen (1992, 1993), Jansen and Stam (1994), and Zeng (1993a,b) have developed methods for mapping multiple noninteracting QTL. These methods repeatedly search the genome for QTL in one dimension. The QTL identified in previous searches are used as cofactors in successive searches to improve the power of QTL

mapping. Recently, Kao *et al.* (1999) described a method for simultaneous mapping of multiple interacting QTL. Due to computational constraints, the search is restricted to regions where there are prior indications of QTL. The method is therefore, as implemented today, only a quasi-simultaneous QTL mapping method.

In a simultaneous genome-wide search for multiple QTL, methods based on enumerative search methods (*e.g.*, stepwise searches at fixed centimorgan intervals throughout the genome) rapidly become computationally intractable as the number of QTL in the model increases. Searching for one QTL requires n evaluations, where n is the number of positions evaluated; a search for multiple QTL requires $n!/(k!(n-k)!)$ evaluations, where k is the number of fitted QTL. A more efficient strategy to search the genome is needed to decrease the computational demand. Since the surface of the QTL mapping test statistic contains many local maxima, the method must be robust (*i.e.*, not get stuck on local extremes). It should at the same time be computationally feasible. In this study, we have investigated the properties of a genetic algorithm as a tool for a simultaneous genome-wide search for two interacting QTL.

Genetic algorithms are search algorithms, based on the mechanisms of genetics and natural selection, whose basic principles were first designed by John Holland (Holland 1975). The central theme during their continuing development has been robustness (Goldberg

Corresponding author: Ö. Carlborg, Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala Biomedical Ctr., Box 597, S-751 24 Uppsala, Sweden.
E-mail: orjan.carlborg@hgen.slu.se

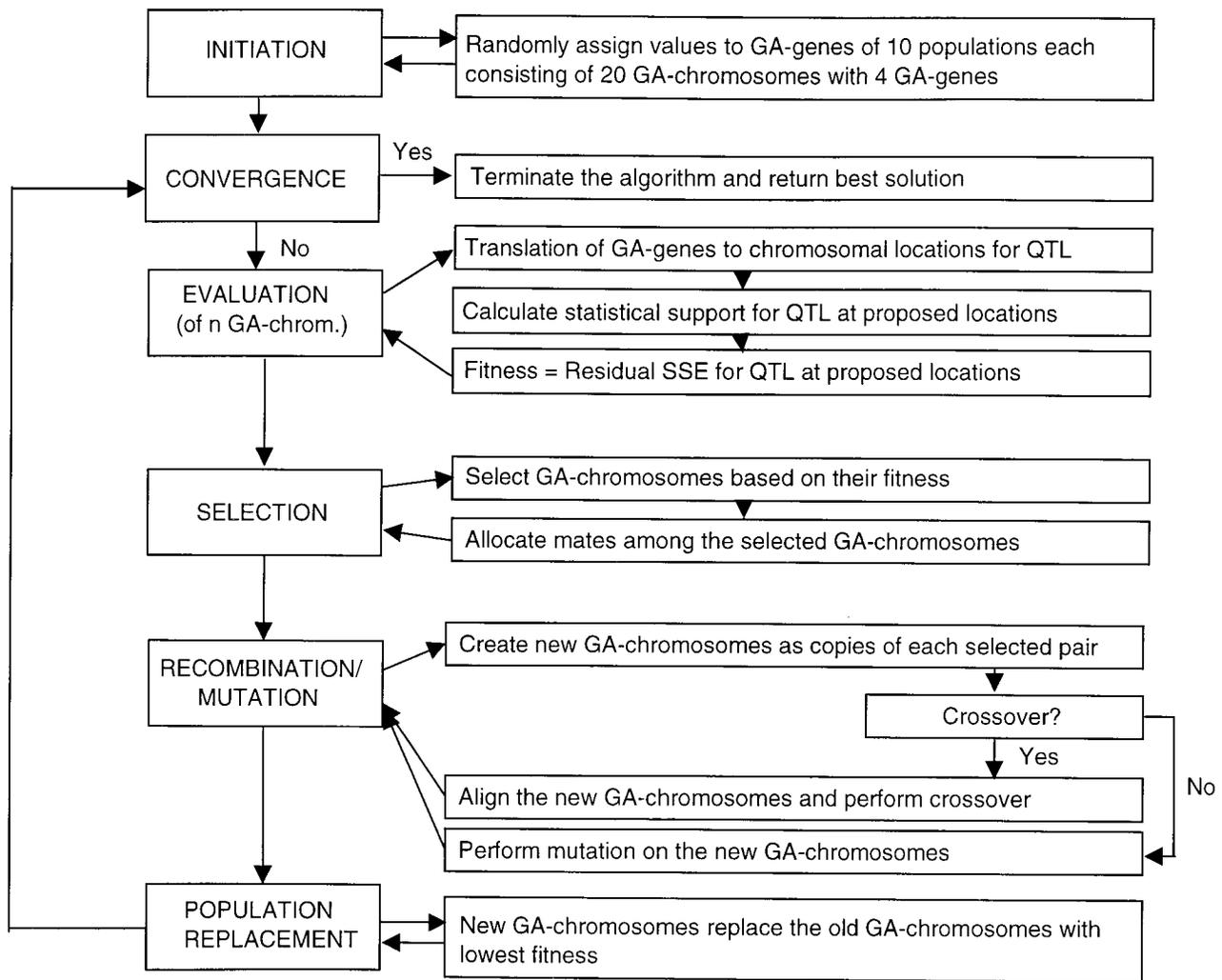


Figure 1.—Outline for the implementation of a genetic algorithm for simultaneous mapping of two interacting QTL.

1989), which make them interesting as a tool for QTL mapping. We have used a genetic algorithm library named PGAPack (Levine 1996). This library can be called from both C and FORTRAN programs and is freely available from http://www-fp.mcs.anl.gov/CCST/research/reports_pre1998/comp_bio/stalk/pgapack.html and works on several computer platforms.

For simplicity, we have investigated the search properties of a genetic algorithm using a weighted least-squares method to search for two interacting QTL in a cross between inbred lines. However, the method is applicable for any QTL mapping method, including methods for outbred crosses and general pedigrees, and for searches in any number of dimensions, that is for any number of (interacting) QTL.

METHOD

The principle of a genetic algorithm: The implementation of a genetic algorithm for QTL mapping is outlined in Figure 1. The genetic algorithm (ga) is initiated

by randomly generating a population, each member of which constitutes a vector of coded parameter values. The parameter values in the vectors are linked together like genes in a linkage group. Each vector is named a ga-chromosome and each cell in the vector is named a ga-gene. We have used ga-chromosomes containing four ga-genes with real values between 0 and 1. The ga-genes are coded representatives for the chromosomal positions of the two QTL we are fitting (two ga-genes are used to represent the chromosome numbers and two are used to represent the respective locations on these chromosomes).

After initiation, the population member that constitutes the best solution is sought. This is done by selection of good individual ga-chromosomes on the basis of their evaluation values from an objective function (in our case the QTL mapping method). During the evaluation process, the real values of the ga-genes on each ga-chromosome are translated to putative QTL locations. The statistical support for QTL at the locations given by the ga-chromosome is evaluated and each ga-chromo-

some becomes associated with an evaluation value, named as its fitness. Ga-chromosomes can then be selected and used further in the algorithm on the basis of their fitness. The standard selection operator in PGAPack is tournament selection, which operates like a tennis tournament, where the fitnesses of the ga-chromosomes are compared in pairs (matches) in several consecutive rounds. The winners of each match (the most fit ga-chromosomes) advance to the next round. The tournament continues until the desired number of ga-chromosomes are left. This scheme gives more variance in solutions as less fit ga-chromosomes can become selected, while the most fit solution is always selected.

The selected ga-chromosomes are used to generate new ga-chromosomes for the next generation by a process resembling genetic recombination. The ga-chromosomes are paired and the linkage between ga-genes is broken at one or several places along the ga-chromosome. New ga-chromosomes are generated by combining stretches of ga-genes (parameter values) from the old ga-chromosomes. The standard recombination operator in PGAPack is two-point crossover, which is defined as the process where the linkage is broken at two locations on the ga-chromosome before exchanging stretches of ga-genes. The standard probability of a crossover event occurring during generation of a new ga-gene is 0.85. A small amount of new variation is also introduced by mutation of the ga-genes, *i.e.*, the value of the ga-gene is changed from its present value either using a mathematical function or assigning it to a new random value from the range of parameter values. The standard mutation frequency, defined as the probability of applying the mutation operator on a ga-gene, is 0.001. Propagation of the algorithm through sufficient generations is expected to lead to convergence to the optimal solution according to the objective function.

The implementation of the genetic algorithm: The genetic algorithm was used to find the location in the genome with the strongest support for a QTL effect and high robustness was our major objective when implementing the ga. The computational performance was compared to other search methods when robust ga settings had been found. The robustness of the method was controlled by evaluating the ability of the ga to repeatedly find the global extremes in the simulated data set with 18 QTL described below and real data from a wild boar \times Large White intercross consisting of 191 individuals (Andersson *et al.* 1994).

The ga was parameterized with the chromosome number and the chromosomal location as separate parameters to obtain maximal blending of parameters during the search. A real-valued genetic algorithm was chosen due to simple translation of ga-gene values to QTL locations. The real-valued ga using the standard settings in PGAPack (Levine 1996) had problems with premature convergence when mapping in populations where several pairs of QTL had effects of similar magnitude. To

increase robustness, the ga settings were changed to maintain variation in the ga-population until termination. The mutation frequency was increased from 0.001 to 0.4 (0.4 was chosen, but any value from 0.2 to 0.5 works well) to introduce more new variation and a mutation-operator that chooses new ga-gene values in the whole parameter range was chosen to maximize the new variation generated by mutation. The ga was also set to not allow creation of new ga-chromosomes, that are duplicates of already existing ga-chromosomes. With these settings, the ga was reasonably robust, but convergence at local extremes was sometimes observed when the data contained several extremes of similar magnitude. The PGAPack manual indicated that practitioners often prefer to have recombination and mutation as exclusive operators. This means that if recombination has occurred during generation of a new ga-chromosome, mutation is not to occur, and if no recombination has occurred, mutation is to be performed with a certain probability for each ga-gene. Implementation of this procedure led to a slightly better performance.

Splitting of the ga-population into several smaller ga-populations has been suggested as a means to increase robustness without excessive increase in computational demand. During optimization, each ga-population is in turn allowed to converge at an extreme that is not identified by any previous populations. The detected extremes are then compared to identify the global extreme. The use of multiple populations is implemented by repetitive calls of PGAPack in a DO-loop structure. Three populations of size 50, 10 populations of size 20, and 20 populations of size 10 were compared and the best results were for 10 populations of size 20. With these settings, the ga-based search gave good performance in all our examined populations. In most cases, the ga found the global extreme in <5 populations, but occasionally all 10 populations were needed. A decrease in the number of populations should therefore be avoided unless the major limitation is computational performance.

Genetic algorithms are good at finding regions where the extremes are located but can have difficulty finding the precise location. One way to overcome this is to perform a local exhaustive search in the vicinity (± 5 cM) of the extreme found by each ga-population. We noted a significant improvement in the location estimates after implementing this.

The objective function used was the residual sum of squared errors from a weighted least-squares approach to QTL mapping. The method is the extension of the method of Jansen (1992) to the two-loci linear model $G = m + A_1 + A_2 + D_1 + D_2 + AA_{12} + AD_{12} + AD_{21} + DD_{22}$ as indicated by the author. The parameters of the model are explained below. Markers have not been used as cofactors and successive iterations in the EM algorithm have been removed to increase the computational efficiency during the evaluation procedure. The modi-

fications needed for the single QTL mapping procedure described by Jansen and Stam (1994) when implementing the two-QTL model included duplication of each individual nine times (instead of three times, *i.e.*, once for every possible two-QTL genotype) and the use of an expanded design matrix (X). The design matrix for the two-locus linear model has been described by Jana (1971). The weight for each observation was taken to be the product of the conditional probabilities of the single-QTL genotypes given the markers (Haley and Knott 1992) at each of the two fitted QTL. The estimates of the model parameters can be found as

$$\hat{\beta} = (X^T W X)^{-1} X^T W Y$$

$$\sigma^2 = (1/N) (Y - X\hat{\beta})^T W (Y - X\hat{\beta}),$$

where Y is the complete data vector, X is the design matrix for the complete data, W is the diagonal matrix of weights, β is the vector of the regression parameters, σ^2 is the normal variance, and N is the number of individuals (Jansen and Stam 1994).

The residual sums of squared errors can then be calculated as

$$\text{SSE} = (Y - X\hat{\beta})^T W (Y - X\hat{\beta}).$$

The method described above can easily be extended to take account of background QTL in the analysis. Two extra *ga*-genes are added to the genetic algorithm and two extra columns are added to the X matrix for each background QTL. The extra *ga*-genes represent the chromosomal location for the QTL and the columns in the design matrix are to contain the QTL indicator variables a and d (Haley and Knott 1992) for a QTL at the location given by the *ga*-genes. The rest of the evaluation procedure is the same as before. We have evaluated the increase in computational demand for a simultaneous search for more than two QTL using this method, but have not investigated any other properties.

Simulations: To evaluate the robustness of the genetic algorithm and to show the advantages of mapping multiple QTL simultaneously, data were simulated for five genetic models (described below) with 100 replicates of 520 F_2 individuals from a cross between two inbred lines. The simulated genome consisted of 20 chromosomes, each 100 cM in length, carrying fully informative marker loci at the ends and at 10-cM intervals. Cross-overs were generated using Haldane's mapping function without interference (Haldane 1919).

Four different combinations of interacting and additive QTL were evaluated. All contained two epistatically interacting QTL and a varying number of fully additive QTL, which were all fixed in the parental lines and simulated at random locations in the genome. Only one QTL was allowed on each chromosome. The number of additive QTL were in turn 16, 8, 4, and 0. The additive effects for the combination with 16 fully additive QTL were 0.4 for QTL 1–4, 0.3 for QTL 5–8, 0.2 for QTL

9–12, and 0.1 for QTL 13–16. These effects correspond to the genetic variances 0.08, 0.045, 0.020, and 0.005, respectively. The QTL effects for the models with 8 and 4 additive QTL were similar, but with 2 or 1 QTL in each of the four QTL effect classes. The genetic variance explained by the pair of interacting QTL is given in Table 2. The narrow-sense heritability was set to 0.5. The simulated effects of the QTL were chosen such that mapping of the QTL using an exhaustive enumerative search would detect the pair of interacting QTL in >80% of the simulated populations.

To evaluate the sensitivity of the search methods to differences in detection power of the statistical mapping methods, lower QTL effects were simulated for the recessive epistatic model by using three different heritabilities, 0.5, 0.1, and 0.01. The different heritabilities were obtained by increasing the residual variance relative to the QTL effects. The detection powers for these heritabilities were 80, 70, and 50%, using the exhaustive enumerative search.

Each simulated population was analyzed using two different methods. Method 1 was simultaneous mapping of two interacting QTL using a genetic algorithm. Method 2 was mapping of two interacting QTL by first performing a one-dimensional search, using the method described by Haley and Knott (1992), to identify the most significant single QTL and then subsequently searching for a second QTL showing epistatic interactions with the previously detected QTL. This latter procedure is from here on called a conditional QTL search. The second search is in one dimension and uses the two-QTL epistatic model, where the location of the first QTL is fixed at the location of the most significant single QTL. Both methods used the previously described weighted least-squares method. Detection of the epistatic QTL was declared upon identification of two QTL on the chromosomes where the simulated major QTL were located, *i.e.*, no significance thresholds have been used at this stage. The comparison between the *ga*-based and the conditional search is not fair, since the conditional search is model dependent in the sense that it relies on the contribution of epistasis to additive and dominance variances for detection of the first QTL. Thus, it does not guarantee giving the right answer when searching for two QTL, while the *ga* does. This comparison has been included anyway to show how the more theoretically correct genetic algorithm compares to the conditional search that is frequently used in practice.

Epistatic model: The epistatic model used in the regression was the two-locus linear model $G = m + A_1 + A_2 + D_1 + D_2 + AA_{12} + AD_{12} + AD_{21} + DD_{22}$, where A_1 is the additive effect at locus 1, A_2 is the additive effect at locus 2, D_1 is the dominance effect at locus 1, D_2 is the dominance effect at locus 2, AA_{12} is the interaction between A_1 and A_2 , AD_{12} is the interaction between A_1 and D_2 , AD_{21} is the interaction between A_2 and D_1 , and

TABLE 1
The relationships among the eight genetic parameters producing digenic segregation ratios in the F₂ generation characteristic of classical epistasis (from Jana 1971)

| Nature of epistasis | Relationship among parameters ^a | F ₂ ratio |
|---------------------|---|----------------------|
| Complementary | $A_1 = A_2 = D_1 = D_2 = AA_{12} = AD_{12} = AD_{21} = DD_{12}$ | 9:7 |
| Duplicate | $A_1 = A_2 = D_1 = D_2 = -AA_{12} = -AD_{12} = -AD_{21} = -DD_{12}$ | 15:1 |
| Dominant | $A_1 = D_1 \neq A_2, A_2 = D_2 = -AA_{12} = -AD_{12} = -AD_{21} = -DD_{12}$ | 12:3:1 |
| Recessive | $A_1 = D_1 \neq A_2, A_2 = D_2 = AA_{12} = AD_{12} = AD_{21} = DD_{12}$ | 9:3:4 |
| Inhibitory | $A_1 = -A_2 = D_1 = -D_2 = -AA_{12} = -AD_{12} = -AD_{21} = -DD_{12}$ | 13:3 |

^a A_1 is the additive effect at locus 1; A_2 is the additive effect at locus 2; D_1 is the dominance effect at locus 1; D_2 is the dominance effect at locus 2; AA_{12} is the interaction between A_1 and A_2 ; AD_{12} is the interaction between A_1 and D_2 ; AD_{21} is the interaction between A_2 and D_1 ; and DD_{12} is the interaction between D_1 and D_2 .

DD_{12} is the interaction between D_1 and D_2 (Jana 1971). The epistatic interactions used for the simulation study were the classical models of epistasis: complementary, dominant, duplicate, recessive, and inhibitory epistasis (Jana 1971). Complementary epistasis is observed when a defect in either of two genes gives the same mutant phenotype, giving an expected Mendelian segregation ratio of 9:7 (Table 1). In this case, functional copies of both genes must be present to produce the dominant phenotype. Duplicate epistasis is observed when a defect in two genes gives a mutant phenotype and the expected segregation ratio is 15:1. In this case, a functional copy of only one of the two genes must be present to produce the dominant phenotype. Dominant, recessive, and inhibitory epistasis occurs when one gene blocks the phenotypic expression of a second gene. For dominant epistasis, the dominant allele at the first locus is also dominant over the alleles at the second locus. The phenotypic effects of the second locus are therefore expressed only when the individual is recessive homozygote at the first locus. This gives an expected segregation ratio of 12:3:1. Recessive epistasis occurs when the recessive homozygote at one locus is dominant over the alleles at the other locus. The phenotypic effects for the second locus are therefore expressed only when the individual is dominant homozygous or heterozygous at the first locus. The expected segregation ratio is in this case 9:3:4. Inhibitory epistasis works in the same way as

dominant epistasis and is the special case when the two genes have equal-sized effects with opposite signs. The expected segregation ratio is here 13:3. The relationships among the genetic parameters for these five genetic models are given in Table 1. The translation of the genetic parameters to the genotypic effects of the two interacting QTL are given in Figure 2. The genotypic effects and genotypic variances for the two interacting QTL for the five epistatic models are given in Table 2.

Computational efficiency: The computational efficiency was compared by calculating the number of statistical evaluations performed during QTL mapping. Three search methods were compared: an exhaustive enumerative search, a genetic algorithm-based search, and a conditional search. Comparisons were made for mapping two, three, and four QTL in a genome size of 2000 cM using 1-cM resolution.

RESULTS

The simulation study was performed on data where 18, 10, 6, and 2 QTL together explained all genetic variation. Depending on the genetic model simulated, the amount of additive and nonadditive genetic variation explained by the two major QTL varies. We have not used any significance levels at this stage, but instead defined the power of the exhaustive search-based

$$\underbrace{\begin{pmatrix} QQQQ & QQQq & QQqq \\ QqQQ & QqQq & Qqqq \\ qqQQ & qqQq & qqqq \end{pmatrix}}_{QTL\ genotypes} = \underbrace{\begin{pmatrix} Y + A_1 + A_2 + AA_{12} & Y + A_1 + D_2 + AD_{12} & Y + A_1 - A_2 - AA_{12} \\ Y + D_1 + A_2 + AD_{21} & Y + D_1 + D_2 + DD_{12} & Y + D_1 - A_2 - AD_{21} \\ Y - A_1 + A_2 - AA_{12} & Y - A_1 + D_2 - AD_{12} & Y - A_1 - A_2 + AA_{12} \end{pmatrix}}_{Phenotypic\ descriptions}$$

Figure 2.—The complete descriptions of the phenotypes associated with the nine QTL genotypes in a common genetic background according to the model in Crow and Kimura (1970). A_1 is the additive effect at locus 1, A_2 is the additive effect at locus 2, D_1 is the dominance effect at locus 1, D_2 is the dominance effect at locus 2, AA_{12} is the interaction between A_1 and A_2 , AD_{12} is the interaction between A_1 and D_2 , AD_{21} is the interaction between A_2 and D_1 , and DD_{12} is the interaction between D_1 and D_2 . The quantity Y is the midparent value, represented as the residual phenotype when the i th and j th loci are not considered.

TABLE 2

The genetic effects of the simulated QTL for the five epistatic models used in the simulation study

| Nature of epistasis | QTL genotype effect | | | | | | | | | V_g^a |
|---------------------|---------------------|------|------|------|------|------|------|------|------|---------|
| | QQQQ | QQQq | QQqq | QqQQ | QqQq | Qqqq | qqQQ | qqQq | qqqq | |
| Complementary | 1.2 | 1.2 | -0.4 | 1.2 | 1.2 | -0.4 | -0.4 | -0.4 | -0.4 | 0.63 |
| Duplicate | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | -1.8 | 0.34 |
| Dominant | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | -1.4 | -1.4 | 0.6 | 0.50 |
| Recessive | -0.6 | -0.6 | 1.4 | -0.6 | -0.6 | 1.4 | -0.4 | -0.4 | -0.4 | 0.58 |
| Inhibitory | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | -1.5 | -1.5 | 0.5 | 0.61 |

^a V_g is the genetic variance explained by the two epistatic QTL.

method as the percentage of the cases where the lowest residual sum of squares is at the simulated positions that explain the highest amount of genetic variance. We have chosen to describe the efficiency of the QTL mapping method as its relative efficiency of detection as compared to the method using the exhaustive enumerative search, which is the limit of the power of the statistical QTL mapping method. The results from the simulation study with 18 QTL (Figure 3) showed that the genetic algorithm-based search had higher relative efficiency to detect the simulated pair of epistatically interacting QTL than the conditional search for all epistatic models tested, except the complementary, where the methods were both fully efficient. The genetic algorithm had a relative efficiency of 100% for all epistatic models but the duplicate. The conditional search had between 86 and 96% relative efficiency for the dominant, recessive, and inhibitory epistatic models and 100% relative efficiency for the complementary model.

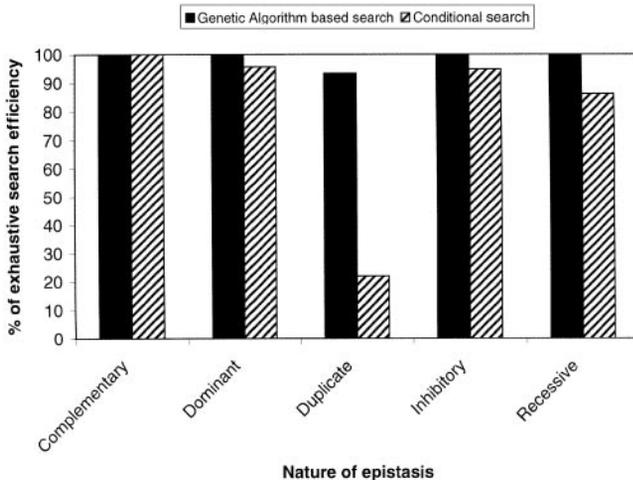


Figure 3.—Efficiency of genetic algorithm and conditional QTL searches for two QTL with simulated complementary, dominant, duplicate, recessive, and inhibitory epistatic interactions. The efficiency is measured as the relative efficiency of QTL mapping using the search algorithm compared to the efficiency of the exhaustive enumerative search. The results are based on 100 replicates per model and method.

The difference in relative efficiency for the search methods was very large for the duplicate epistatic model, where the conditional search had a relative efficiency of only 21%, while the genetic algorithm-based search had a relative efficiency of 93%. The relative efficiency of the genetic algorithm-based and the conditional QTL search methods remained constant when the power of the statistical QTL mapping method was decreased from 80 to 50% by lowering the heritability from 0.5 to 0.01. Decreasing the number of background QTL from 16 to 8 and from 8 to 4 only affected the relative efficiencies of the search methods marginally. In the simulation where two interacting QTL explained all genetic variation, both methods had a relative efficiency of 100%.

The genetic algorithm increases the computational demand by a factor of 3–5 when compared to the conditional search (Table 3). The improvement in computational efficiency of the ga, as compared to the exhaustive enumerative search, was by a factor 133 for two QTL. An expansion of the search to additional dimensions by also searching for background QTL simultaneously leads to further computational advantages for the ga-based search. Improvements are in the order of 65,000 for three QTL and 1.7×10^7 for four simultaneous QTL.

DISCUSSION

This study has shown that the genetic algorithm has large potential for improving the computational efficiency for simultaneous mapping of multiple interacting QTL. It can be used together with any QTL mapping method based on a multidimensional search in the genome. Implementation is simplified by the use of existing software libraries and the results from our simulation study show that the method is robust in finding the pair of QTL in the genome with the highest statistical support. The computational advantage of using a genetic algorithm increases with the complexity of the QTL mapping method and the number of QTL simultaneously searched for. The greatest potential benefit of using this method is therefore when using multidimensional searches based on computationally demanding likelihood QTL mapping methods.

TABLE 3

Number of statistical evaluations when searching for two, three, or four QTL using a conditioned search for additional interacting QTL, or a simultaneous search for two or more interacting QTL using a genetic algorithm, or an exhaustive enumerative search with a stepsize of 1 cM

| Search method | No. of evaluations | | |
|--------------------------------|--------------------|-------------------|----------------------|
| | Two QTL | Three QTL | Four QTL |
| Conditional search | 4×10^3 | 6×10^3 | 8×10^3 |
| Genetic algorithm-based search | 1.7×10^4 | 2×10^4 | 4×10^4 |
| Exhaustive enumerative search | 2.0×10^6 | 1.3×10^8 | 6.6×10^{11} |

In our simulation study, the efficiency, defined as the relative efficiency to that of an exhaustive enumerative search, of the genetic algorithm-based search was 100% for all epistatic models but the duplicate, where it was 93%. The genetic algorithm had thus, regardless of the model of epistatic interaction, almost as high relative efficiency as the exhaustive search, using only a fraction of the CPU time. The conditional QTL search had reasonable relative efficiency (86–100%) for all genetic models but the duplicate, where it was only 21%. There was no indication that the relative efficiencies of the genetic algorithm or the conditional QTL searches were reduced at lower QTL effects or varied number of background QTL. The conditional search was thus more sensitive to the type of epistatic interaction and the size of the QTL effects than the genetic algorithm. There was also a tendency toward a generally lower relative efficiency for the conditional search method for the other epistatic interaction models. The conclusion from the simulation study is thus that the genetic algorithm is a good alternative to the exhaustive enumerative search and that it has potential to be of great use in practice, where the time for analysis is limited. It should be noted that the detection power of the genetic algorithm was achieved using a general genetic algorithm software package for ease of implementation. A custom-written genetic algorithm optimized for the problem will most likely achieve higher power of detection in a more computationally efficient way.

The simultaneous mapping methods used for this study were based on least squares. We do not expect that the use of a maximum-likelihood method instead of least-squares regression would alter our conclusions as the two methods generally give very similar results (Haley and Knott 1992; Martinez and Curnow 1992). The least-squares method was used because of its speed of computation.

The method proposed here to detect interacting QTL would in practice be used as a complement to existing QTL mapping methods. The genome should first be searched in one dimension to identify QTL with high individual effects, after which a search would be performed to identify additional QTL, both interacting and noninteracting, using the genetic algorithm-based si-

multaneous search. The statistical support for the additional QTL can then be evaluated for the QTL positions indicated by the ga. The genetic algorithm-based search can then be used repetitively for successive searches for additional QTL in higher dimensions.

One major issue when using multiple QTL models in practice is to draw statistical inferences about the support for additional QTL. Randomization testing (Churchill and Doerge 1994) is an intuitively attractive way to derive appropriate threshold values for declaring significant QTL effects. This method is, however, computationally demanding and has until now not been feasible to use for QTL mapping methods with high computational demands. The use of a genetic algorithm as described here would make randomization testing feasible for more complex mapping methods and multiple QTL models and thus make these methods more useful for analysis of real data. The computational improvement in randomization testing will be considerable when using the genetic algorithm-based search.

The genetic algorithm is a general tool to search large parameter spaces and could be of use in many other areas in QTL mapping. In this study, we have used a genetic algorithm in the search for two interacting QTL in a cross between inbred lines, but the method can also be used for analyses of crosses between outbred lines and in searches for more than two QTL. For analyses of outbred lines, the genetic algorithm could also be used when testing for QTL segregation within the founder lines. This would be implemented by using a genetic algorithm to group the haplotypes from the founders in allelic groups and in this way obtaining the most likely allelic constitution for the founders and other individuals in the pedigree. This results in greater detection power because of more extreme probabilities of identity by descent of chromosomal regions between phenotyped individuals and each founder.

We thank Julius van der Werf and Susan Mezaros for many useful discussions about genetic algorithms and statistical methods for QTL mapping and Lena Andersson-Eklund for valuable comments on the manuscript. This work was supported by the National Graduate School in Scientific Computing, the Food21 project (MISTRA), the Swedish Medical Research Council (MFR), and a travel grant from the Swedish University of Agricultural Sciences.

LITERATURE CITED

- Andersson, L., C. S. Haley, H. Ellegren, S. A. Knott, M. Johanson *et al.*, 1994 Genetic mapping of quantitative trait loci for growth and fatness in pigs. *Science* **263**: 1771–1774.
- Churchill, G. A., and R. W. Doerge, 1994 Empirical threshold values for quantitative trait mapping. *Genetics* **138**: 963–971.
- Crow, J. F., and M. Kimura, 1970 *An Introduction to Population Genetics Theory*. Harper and Row, New York.
- Goldberg, D. E., 1989 *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison and Wesley, Reading, MA.
- Haldane, J. B. S., 1919 The combination of linkage values and the calculation of distances between loci of linked factors. *J. Genet.* **8**: 299–309.
- Haley, C. S., and S. A. Knott, 1992 A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69**: 315–324.
- Hoeschele, I., and P. VanRaden, 1993a Bayesian analysis of linkage between genetic markers and quantitative trait loci. I. Prior knowledge. *Theor. Appl. Genet.* **85**: 953–960.
- Hoeschele, I., and P. VanRaden, 1993b Bayesian analysis of linkage between genetic markers and quantitative trait loci. II. Combining prior knowledge with experimental evidence. *Theor. Appl. Genet.* **85**: 946–952.
- Holland, J., 1975 *Adaption in Natural and Artificial Systems*. The University of Michigan Press, Ann Arbor, MI.
- Jana, S., 1971 Simulation of quantitative characters from qualitatively acting genes. I. Nonallelic gene interactions involving two or three loci. *Theor. Appl. Genet.* **41**: 216–226.
- Jansen, R. C., 1992 A general mixture model for mapping quantitative trait loci by using molecular markers. *Theor. Appl. Genet.* **85**: 252–260.
- Jansen, R. C., 1993 Interval mapping of multiple quantitative trait loci. *Genetics* **135**: 205–211.
- Jansen, R. C., and P. Stam, 1994 High resolution of quantitative traits into multiple loci via interval mapping. *Genetics* **136**: 1447–1455.
- Kao, C-H., Z-B. Zeng and R. D. Teasdale, 1999 Multiple interval mapping for quantitative trait loci. *Genetics* **152**: 1203–1216.
- Kruglyak, L., and E. S. Lander, 1995 A nonparametric approach for mapping quantitative trait loci. *Genetics* **139**: 1421–1428.
- Lander, E. S., and D. Botstein, 1989 Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**: 185–199.
- Levine, D., 1996 *Users Guide to the PGAPack Parallel Genetic Algorithm Library*. Argonne National Laboratory, Mathematics and Computer Science Division, Argonne, IL.
- Martinez, O., and R. N. Curnow, 1992 Estimating the locations and sizes of effects of quantitative trait loci using flanking markers. *Theor. Appl. Genet.* **85**: 480–488.
- Thomas, D., and V. Cortessis, 1992 A Gibbs sampling approach to linkage analysis. *Hum. Hered.* **42**: 63–76.
- Zeng, Z-B., 1993a Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. *Proc. Natl. Acad. Sci. USA* **90**: 10972–10976.
- Zeng, Z-B., 1993b Precision mapping of quantitative trait loci. *Genetics* **136**: 1456–1468.

Communicating editor: C. Haley