An Efficient Resampling Method for Assessing Genome-Wide Statistical Significance in Mapping Quantitative Trait Loci

Fei Zou,*,[1] Jason P. Fine,[†] Jianhua Hu* and D. Y. Lin*

*Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina 27599-7420 and
[†]Department of Statistics, University of Wisconsin, Madison, Wisconsin 53706

ABSTRACT

Assessing genome-wide statistical significance is an important and difficult problem in multipoint linkage analysis. Due to multiple tests on the same genome, the usual pointwise significance level based on the chi-square approximation is inappropriate. Permutation is widely used to determine genome-wide significance. Theoretical approximations are available for simple experimental crosses. In this article, we propose a resampling procedure to assess the significance of genome-wide QTL mapping for experimental crosses. The proposed method is computationally much less intensive than the permutation procedure (in the order of 10^2 or higher) and is applicable to complex breeding designs and sophisticated genetic models that cannot be handled by the permutation and theoretical methods. The usefulness of the proposed method is demonstrated through simulation studies and an application to a Drosophila backcross.

A number of statistical methods are available for mapping QTL in experimental populations, such as backcrosses (BCs) and F_2's. The interval-mapping method of LANDER and BOTSTEIN (1989) uses two markers flanking a region where the QTL may fall and evaluates the LOD score at each genome position. This method has been implemented in several freely distributed software packages (LINCOLN *et al.* 1993; BASTEN *et al.* 1997; MANLY and OLSON 1999) and is commonly used in practice. Various extensions, including composite-interval mapping (CIM; ZENG 1993, 1994), the multiple-QTL model (JANSEN and STAM 1994), and multiple-interval mapping (MIM; KAO and ZENG 1997; KAO *et al.* 1999), can be used to map multiple QTL. BROMAN (2001) and DOERGE (2001) provided excellent reviews of QTL-mapping methods.

All the aforementioned methods entail a common problem: how to determine the threshold of the test statistic. This is not a trivial problem. Many factors, such as genome size, genetic map density, informativeness of markers, and proportion of missing data, may affect the distribution of the test statistic. The usual pointwise significance level based on the chi-square approximation is inadequate because the entire genome (or at least several regions) is tested for the presence of QTL and the test statistics are not independent among loci.

Theoretical approximations have been developed to determine threshold and power (LANDER and BOTSTEIN 1989; DUPUIS and SIEGMUND 1999; REBAI *et al.* 1994,

1995) in some standard designs. For backcross populations, LANDER and BOTSTEIN (1989) showed that with an infinitely dense map, the LOD score may be approximated in large samples by an Ornstein-Uhlenbeck diffusion process. DUPUIS and SIEGMUND (1999) derived a similar result for F_2. ZOU *et al.* (2001) extended the results to more general experimental designs. The asymptotic calculations are straightforward, but require relatively dense maps with fairly evenly spaced markers. The parameters needed in the calculations are model specific and are difficult to determine for complicated designs. Furthermore, the calculations are applicable only to single-QTL models and not to multiple-QTL mapping.

REBAI *et al.* (1994, 1995) noted that, for interval mapping, the position of the QTL is a parameter that presents only under alternative hypotheses. On the basis of this observation, they found an explicit formula for the upper bound for the BC and F_2 populations and derived a conservative threshold using the results of DAVIES (1977, 1987). The formula is algebraically involved and an approximation to the upper bound is usually necessary. PIEPHO (2001) proposed an efficient numerical method to compute the thresholds in REBAI *et al.* (1994, 1995) for general designs. His simulation results indicate that the approximation is generally conservative when markers are relatively dense.

To avoid asymptotic approximations, one may use permutation testing (CHURCHILL and DOERGE 1994). The idea is to replicate the original analysis many times on data sets generated by randomly reshuffling the original trait data while leaving the marker data unchanged. This approach accounts for missing marker data, actual marker densities, and nonrandom segregation of marker

[1]*Corresponding author:* Department of Biostatistics, University of North Carolina, 3107D McGavran-Greenberg Hall, CB 7420, Chapel Hill, NC 27599-7420. E-mail: fzou@bios.unc.edu

alleles. However, this method is computationally intensive. For MIM, where model selection is involved, ZENG *et al.* (1999) proposed using a bootstrap resampling method for hypothesis testing. However, the heavy computational burden has limited the use of the bootstrap test (Z-B. ZENG, personal communication). Furthermore, permutation testing is limited to situations in which there is complete exchangeability under the null hypothesis. It is this exchangeability that ensures the validity of inference based on the permutation distribution. It is unclear how to apply the bootstrap method in ZENG *et al.* (1999) to the situation where a nonlinear model, such as logistic regression or a Poisson model, is used to map multiple QTL with MIM, since the bootstrap procedure in ZENG *et al.* (1999) is performed on model-based residuals.

In this article, we propose a resampling method to assess the genome-wide significance for QTL mapping. The method is less computationally demanding than permutation tests, more accurate than theoretical approximations when rigid requirements of theoretical approximations are not satisfied, and applicable to more complicated designs and models than the theoretical and permutation methods are. The performance of the proposed method is assessed through simulation studies. An illustration with data from the *Drosophila* backcross of ZENG *et al.* (2000) is provided.

METHODS

We use single-QTL mapping in an F_2 population as a working example to illustrate the rationale of the proposed method. We assume that the trait is normally distributed. Extensions to nonnormal traits and other mapping populations as well as MIM/CIM are described later.

Resampling methods for single-QTL models: For mapping a quantitative trait, a series of genetic markers are observed over the entire genome or in some specific regions depending on the purpose of the experiment. Specifically, we observe L genetic markers $\{M_{il}; l = 1, \ldots, L\}$ located at positions $\{s_l; l = 1, \ldots, L\}$ along the genome for subject i ($i = 1, \ldots, n$). We also observe the trait value y_i for the ith subject ($i = 1, \ldots, n$). The goal of the QTL mapping is to use the marker information to search for QTL associated with the trait.

At each fixed location d of the genome, the conditional probabilities of the unobserved QTL genotypes can be inferred using flanking markers and the distribution of the quantitative trait given the markers follows a discrete mixture model. Specifically, for a given locus d, the log-likelihood for $\theta = (a, b, \mu, \sigma^2)$ takes the form

$$l(\theta; d) = \sum_{i=1}^{n} l_i(\theta; d), \quad (1)$$

where $l_i(\theta; d) = \log[\pi_i(qq; d)\phi((y_i - \mu + a)/\sigma) + \pi_i(qQ; d)\phi((y_i - \mu - b)/\sigma) + \pi_i(QQ; d)\phi((y_i - \mu - $

$a)/\sigma)]$, $\phi(x)$ is the density of a standard normal random variable, μ is the grand mean, a and b are the additive and dominant effects of the QTL, respectively, and $\pi_i(k; d) = \Pr(\text{QTL genotype at locus } d \text{ of subject } i \text{ is } k|\text{subject } i\text{'s marker genotypes})$ with $k = qq$, qQ, and QQ (three possible QTL genotypes in F_2). Note that the conditional probability $\pi_i(k; d)$ depends on the flanking marker genotypes, the distance between the two flanking markers, as well as the distances between the putative QTL locus d and the right and left markers; see chapter 15 of LYNCH and WALSH (1998) for details. For other mapping populations, such as advanced intercrosses and advanced backcrosses, the likelihood still takes the form of (1), although the conditional probabilities $\pi_i(k; d)$ are calculated differently (LYNCH and WALSH 1998, Chap. 15). With missing markers, $\pi_i(k; d)$ is conditional on the genotypes of the two closest flanking markers if both are available or on the genotype of the single marker if only one flanking marker is available.

The maximum-likelihood estimator (MLE) $\hat{\theta} \equiv (\hat{a}, \hat{b}, \hat{\mu}, \hat{\sigma}^2)$ can be obtained by maximizing $l(\theta; d)$ directly or by using the EM algorithm (DEMPSTER *et al.* 1977) in which the unknown QTL genotypes are treated as missing data. Let the maximum-likelihood estimator of θ under H_0: $a = b = 0$ be denoted by $\tilde{\theta} \equiv (0, 0, \tilde{\mu}, \tilde{\sigma}^2)$. Then the likelihood-ratio test statistic (LRT) for testing H_0: $a = b = 0$ against H_1: $a \neq 0$ and/or $b \neq 0$ at location d takes the form

$$\text{LRT}(d) = 2[l(\hat{\theta}; d) - l(\tilde{\theta}; d)], \quad (2)$$

which is approximately chi-square distributed with 2 d.f. We can replace ϕ in (1) with a nonnormal density function, such as exponential (continuous phenotype) and binomial or Poisson (discrete phenotype). In practice, the true distribution is often unknown and the normal model is used almost exclusively with continuous phenotypes. As is shown later in SIMULATION STUDIES, the proposed method is quite robust to model misspecification. When the normal model is used to fit nonnormal data (chi-square data as in our simulation), the empirical type I error based on our proposed method is well controlled at the targeted level (Table 1).

In multipoint linkage analysis, we maximize $\text{LRT}(d)$ or $\text{LOD}(d)$ over all possible values of d in the genome. Thus, it is necessary to derive the distribution of $\text{LRT}(d)$ as a stochastic process indexed by the genome location d. To this end, it is more convenient to work with the score test statistic for testing the same hypothesis. The equivalence between the likelihood ratio and score test statistics in large samples is shown in mathematical statistics texts, such as COX and HINKLEY [1974, Sect. 9.3 (iii)]. The reason for working with the score test statistic is that it can be approximated by a sum of independent random vectors so that its large-sample distribution, when regarded as a stochastic process in the genome location, can be readily derived. The same large-sample distribution also applies to the likelihood-ratio statistic

TABLE 1

Comparison of the proposed, theoretical, and empirical thresholds in an F$_2$ population and an advanced intercross F$_3$ population

	Setup				Empirical[f]		Proposed[g]		Dense-map[h]		Sparse-map[i]	
Dist[a]	Pop[b]	a[c]	d[d]	Δ[e]	α = 0.05	α = 0.01	α = 0.05	α = 0.01	α = 0.05	α = 0.01	α = 0.05	α = 0.01
Normal	F$_2$	0	0	2	11.84	15.32	11.45	15.07	13.37	17.12	11.80	15.37
				10	10.75	14.28	10.38	13.93	13.37	17.12	10.15	13.53
				20	9.80	13.18	9.76	13.28	13.37	17.12	9.15	12.39
	F$_3$	0	0	2	12.42	16.08	11.97	15.60	14.30	18.02	12.32	15.84
				10	10.87	14.47	10.66	14.20	18.02	12.32	10.30	13.59
				20	10.04	13.48	9.97	13.48	18.02	12.32	9.11	12.24
	F$_2$	0.5	0	2	—	—	11.47	15.10	13.37	17.12	11.80	15.37
				10	—	—	10.40	13.95	13.37	17.12	10.15	13.53
				20	—	—	9.78	13.31	13.37	17.12	9.15	12.39
	F$_2$	0	0.7	2	—	—	11.53	15.15	13.37	17.12	11.80	15.37
				10	—	—	10.44	13.99	13.37	17.12	10.15	13.53
				20	—	—	9.81	13.33	13.37	17.12	9.15	12.39
	F$_2$	0.4	0.4	2	—	—	11.48	15.10	13.37	17.12	11.80	15.37
				10	—	—	10.40	13.95	13.37	17.12	10.15	13.53
				20	—	—	9.78	13.30	13.37	17.12	9.15	12.39
χ^2	F$_2$	0	0	2	11.66	15.40	11.07	14.69	13.37	17.12	11.80	15.37
				10	10.23	14.06	10.09	13.66	13.37	17.12	10.15	13.53
				20	9.66	13.55	9.56	13.09	13.37	17.12	9.15	12.39

[a] Error distribution.
[b] Mapping population (F$_2$ or F$_3$).
[c] Additive effect.
[d] Dominant effect.
[e] Marker distance (in centimorgans).
[f] Percentiles of the test statistic based on the 10,000 simulated data sets under H$_0$.
[g] Average of thresholds from 10,000 simulated data sets.
[h] Theoretical thresholds based on the dense-map assumption.
[i] Theoretical thresholds based on the sparse-map assumption.

since it is equivalent to the score test statistic in large samples.

For the interval mapping of the F$_2$ population, we write $\theta = (a, b, \mu, \sigma^2)$ and we are interested in testing the null hypothesis H$_0$: $\beta \equiv (a, b) = 0$ in the presence of the nuisance parameter $\eta \equiv (\mu, \sigma^2)$. In the sequel, the general notation of θ will be used, where β pertains to the parameter of primary interest and η to the nuisance parameter, so that general QTL models other than the specific F$_2$ model are encompassed.

Let $U_{\beta,i}(\theta; d) = \partial l_i(\beta, \eta; d)/\partial\beta$ and $U_{\eta,i}(\theta; d) = \partial l_i(\beta, \eta; d)/\partial\eta$. These are the contributions of the ith subject to the score functions for β and η. Further, let $U(d) = \Sigma_i U_{\beta,i}(0, \tilde{\eta}; d)$, where $\tilde{\eta}$ is the restricted MLE of η under H$_0$: $\beta = 0$, $i.e.$, the solution of the equation $\Sigma_i U_{\eta,i}(0, \eta; d) = 0$. Note that $U(d)$ is the score function for β evaluated at $\beta = 0$ and $\eta = \tilde{\eta}$. It follows from Taylor series expansions and the law of large numbers that $n^{-1/2}U(d)$ has the same asymptotic distribution as $n^{-1/2}\Sigma_{i=1}^{n}U_i(d)$, where

$$U_i(d) = U_{\beta,i}(0, \eta; d)$$
$$- \Sigma_{\beta\eta}(0, \eta; d)\Sigma^{-1}_{\eta\eta}(0, \eta; d)U_{\eta,i}(0, \eta; d), \quad (3)$$

and $\Sigma_{\beta\eta}(\beta, \eta; d)$ and $\Sigma_{\eta\eta}(\beta, \eta; d)$ are, respectively, the

limits of $n^{-1}\partial^2 l(\beta, \eta; d)/\partial\beta\partial\eta$ and $n^{-1}\partial^2 l(\beta, \eta; d)/\partial\eta^2$ as n goes to infinity [Cox and HINKLEY 1974, Sect. 9.3 (iii)]. Since $U_i(d)$ involves only the information from the ith subject, the $U_i(d)$ ($i = 1, \ldots, n$) are independent zero-mean random variables for any given d. Thus, it follows from the multivariate central limit theorem that the process $n^{-1/2}U(d)$ is asymptotically a zero-mean Gaussian process, where the covariance between $n^{-1/2}U(d_1)$ and $n^{-1/2}U(d_2)$ at any two given positions d_1 and d_2 is $\Xi(d_1, d_2)$, the limit of $n^{-1}\Sigma_i U_i(d_1)U_i^T(d_2)$. The replacement of the unknown parameters in (3) by their sample estimators yields

$$\hat{U}_i(d) = U_{\beta,i}(0, \tilde{\eta}; d) - \{\partial^2 l(0, \tilde{\eta}; d)/\partial\beta\partial\eta\}\{\partial^2 l(0, \tilde{\eta}; d)/\partial\eta^2\}^{-1}$$
$$\times U_{\eta,i}(0, \tilde{\eta}; d). \quad (4)$$

The restricted MLE $\tilde{\eta}$ in (4) may be replaced by the unrestricted MLE $\hat{\eta}$. By the law of large numbers and the consistency of the maximum-likelihood estimators, $\Xi(d_1, d_2)$ can be consistently estimated by $\hat{\Xi}(d_1, d_2) = n^{-1}\Sigma_{i=1}^{n}\hat{U}_i(d_1)\hat{U}_i^T(d_2)$.

The score test statistic for H$_0$: $\beta = 0$ against H$_1$: $\beta \neq 0$ at location d takes the form

$$W(d) = \hat{U}^T(d)\hat{V}^{-1}(d)\hat{U}(d),$$

where $\hat{U}(d) = \Sigma_i \hat{U}_i(d)$ and $\hat{V}(d) = n\hat{\Xi}(d, d)$ [Cox and Hinkley 1974, Sect. 9.3 (iii)]. It can be shown that $W(d)$ is asymptotically equivalent to $\mathrm{LRT}(d)$ [Cox and Hinkley 1974, Sect. 9.3 (iii)]. To assess the genome-wide statistical significance, we need to evaluate the distribution of $\max_d W(d)$. In general, this is not analytically tractable. We propose a resampling method similar to that of Lin *et al.* (1993) to approximate the distribution of $\max_d W(d)$. The idea of simulating thresholds using a score test statistic is mentioned in Rebaï *et al.* (1994).

Define

$$U^*(d) = \sum_{i=1}^{n} \hat{U}_i(d) G_i,$$

where G_i ($i = 1, \ldots, n$) are independent standard normal random variables. Let

$$W^*(d) = U^{*T}(d) \hat{V}^{-1}(d) U^*(d). \qquad (5)$$

In (5), we regard the $\hat{U}_i(d)$ in $U^*(d)$ and $\hat{V}$ as fixed and the G_i in $U^*(d)$ as random. Conditional on the observable data, $U^*(d)$ is normal with mean 0 at each location d and the covariance between $n^{-1/2}U^*(d_1)$ and $n^{-1/2}U^*(d_2)$ equals $\hat{\Xi}(d_1, d_2)$, which converges to $\Xi(d_1, d_2)$. It follows that the conditional distribution of $n^{-1/2}U^*(d)$ given the observed data converges to the same limiting distribution of $n^{-1/2}\hat{U}(d)$. Consequently, the distribution of $W(d)$ can be approximated by that of $W^*(d)$. Our resampling method is essentially a parametric bootstrap.

We have shown that, under the null hypothesis, the test statistics are functions of certain zero-mean Gaussian processes over the genome positions and the realizations from the Gaussian processes can be generated by Monte Carlo simulations. In practice, the resampling procedure is as follows:

1. Sample G_i, $i = 1, 2, \ldots, n$, from $N(0, 1)$.
2. Calculate $U^*(d) = \Sigma_{i=1}^{n} \hat{U}_i(d) G_i$, $W^*(d) = U^{*T}(d) \hat{V}^{-1}(d) U^*(d)$ and $S^* = \max_d W^*(d)$.
3. Repeat steps 1 and 2 a large number of times, say R times.
4. For a given genome-wide type I error rate α, calculate the $100(1 - \alpha)$th percentile of the R values of the S^*. If the observed value of the LRT exceeds this threshold, then reject the null hypothesis.

The above calculations are based on the score function and the observed information matrix from the original data. These quantities are evaluated once and used repeatedly in step 2. Since it does not involve refitting the model in each iteration, the proposed method is computationally much more efficient than the permutation method. This is important with complex breeding designs and sophisticated QTL models (*e.g.*, CIM and MIM), where the likelihood calculations are time consuming.

Extensions to MIM/CIM: In this section, we show how to apply the resampling method to MIM and CIM. Suppose

that we are searching for multiple QTL in a backcross population. Given the genotypes of K QTL, the normal regression model takes the form

$$y = \mu + \sum_{1 \le k \le K} x_k \gamma_k + \sum_{1 \le j \ne k \le K} x_j x_k \gamma_{jk} + e, \qquad (6)$$

where x_k is the QTL genotype indicator variable, which takes the value -1 or 1 when the kth QTL is heterozygote or homozygote, respectively, μ is the grand mean, γ_k is the main effect of the kth QTL, γ_{jk} is the interaction between the jth and kth QTL, and e is a zero-mean normal error with variance σ^2.

As in the case of the single-QTL analysis, the QTL genotypes are generally unobservable but the conditional probabilities of the QTL can be calculated given flanking markers. This results in the following mixture-model likelihood for K putative QTL loci $d_1, \ldots, d_K$,

$$l(\theta; d_1, \ldots, d_K) = \sum_{i=1}^{n} l_i(\theta; d_1, \ldots, d_K), \qquad (7)$$

where $l_i(\theta; d_1, \ldots, d_K) =$

$$\log\Bigg[\sum_{a_1, \ldots, a_K \in \{-1, 1\}} \pi_i(a_1, \ldots, a_K; d_1, \ldots, d_K) \phi$$

$$\times \left\{ \frac{y_i - \mu - \Sigma_{1 \le k \le K} a_k \gamma_k - \Sigma_{1 \le j \ne k \le K} a_k a_j \gamma_{jk}}{\sigma} \right\} \Bigg]$$

and $\pi_i(a_1, \ldots, a_K; d_1, \ldots, d_K) = \Pr(x_1 = a_1, \ldots, x_K = a_K|$subject i's marker genotypes), which is the conditional probability of the joint genotypes of K QTL given the marker genotypes of the ith subject. Let $\theta = (\beta, \eta)$, where $\beta = (\gamma_1, \ldots, \gamma_K, \gamma_{12}, \ldots, \gamma_{K-1,K})$ and $\eta = (\mu, \sigma^2)$. We test the null hypothesis $H_0: \beta = 0$ against the alternative hypothesis $H_1: \beta \ne 0$. Note that for MIM, the profile likelihood is calculated in K-dimensional space $(d_1, \ldots, d_K)$. Once the likelihood is obtained, the resampling procedure above can be applied to the resulting score test statistic with $d = (d_1, \ldots, d_K)$.

For CIM, the model is essentially the same as MIM except that, for a given putative QTL position d, $(x_1, \ldots x_{K-1})$ corresponding to the selected marker genotypes are known and only x_K corresponding to the putative QTL genotype is unobservable. Also, in CIM the interaction terms are generally ignored. Thus for CIM, our mixture model will have likelihood (1) but $l_i(\theta; d) = \log[\Sigma_{a_K \in \{-1, 1\}} \pi_i(a_K; d) \phi((y_i - \mu - \Sigma_{1 \le k \le K-1} x_k \gamma_k + a_K \gamma_K)/\sigma)]$ and $\pi_i(a_K; d) = \Pr(x_K = a_K|$subject i's marker genotypes), the conditional probability of the genotypes of the putative QTL given the marker genotypes of the ith subject. In this situation, $\beta = \gamma_K$ and $\eta = (\mu, \gamma_1, \ldots, \gamma_{K-1}, \sigma^2)$.

SIMULATION STUDIES

Simulations were conducted to study the behavior of the proposed method in an F_2 population. One chromosome with a total length of 100 cM was simulated. The

Empirical type I error and power of the proposed and theoretical methods in an F_2 population and an advanced intercross F_3 population

Setup					Proposed[f]		Dense-map[g]		Sparse-map[h]	
Dist[a]	Pop[b]	a[c]	d[d]	Δ[e]	α = 0.05	α = 0.01	α = 0.05	α = 0.01	α = 0.05	α = 0.01
Normal	F_2	0	0	2	0.058 (0.040)[i]	0.012 (0.008)	0.026	0.004	0.051	0.010
				10	0.056 (0.045)	0.012 (0.010)	0.016	0.003	0.063	0.015
				20	0.051 (0.041)	0.010 (0.006)	0.009	0.001	0.069	0.015
	F_3	0	0	2	0.060 (0.044)	0.013 (0.009)	0.023	0.004	0.053	0.011
				10	0.054 (0.045)	0.012 (0.009)	0.011	0.002	0.064	0.016
				20	0.052 (0.043)	0.011 (0.007)	0.007	0.002	0.079	0.019
	F_2	0.5	0	2	0.978 (0.970)	0.922 (0.909)	0.955	0.870	0.974	0.916
				10	0.977 (0.973)	0.928 (0.918)	0.939	0.845	0.979	0.935
				20	0.982 (0.978)	0.938 (0.926)	0.937	0.845	0.986	0.952
	F_2	0	0.7	2	0.968 (0.954)	0.902 (0.874)	0.940	0.841	0.965	0.895
				10	0.977 (0.969)	0.923 (0.902)	0.935	0.833	0.980	0.931
				20	0.980 (0.969)	0.927 (0.901)	0.926	0.828	0.985	0.944
	F_2	0.4	0.4	2	0.968 (0.954)	0.895 (0.877)	0.936	0.839	0.963	0.888
				10	0.972 (0.966)	0.910 (0.895)	0.924	0.812	0.975	0.920
				20	0.977 (0.969)	0.924 (0.906)	0.922	0.816	0.982	0.942
χ^2	F_2	0	0	2	0.064 (0.038)	0.017 (0.008)	0.023	0.005	0.047	0.010
				10	0.054 (0.040)	0.013 (0.009)	0.013	0.003	0.052	0.013
				20	0.054 (0.039)	0.015 (0.009)	0.011	0.002	0.064	0.016

[a] Error distribution.
[b] Mapping population (F_2 or F_3).
[c] Additive effect.
[d] Dominant effect.
[e] Marker distance (in centimorgans).
[f] Type I error/power based on the proposed thresholds.
[g] Type I error/power based on the dense-map thresholds.
[h] Type I error/power based on the sparse-map thresholds.
[i] The values in parentheses are type I error/power based on PIEPHO's (2001) quick method.

markers were evenly spaced with a marker distance of 2, 10, or 20 cM. The null and alternative models were simulated to investigate the type I error and power. Under the null hypothesis, the trait was randomly sampled from the standard normal distribution. Under the alternative, a QTL was simulated at 40 cM with different additive and dominant effects. We set the sample size to 200. We simulated 10,000 data sets for each combination of the marker distance and QTL effects. For each simulated data set, we set $R = 10,000$ and $\alpha = 0.05$ or $\alpha = 0.01$. The step width of the QTL scan is set to 1 cM for all simulations. To compare the resampling method with the theoretical method, we also calculated the thresholds on the basis of the dense-map and sparse-map approximations of DUPUIS and SIEGMUND (1999) as well as the corresponding type I error and power. The results are summarized in Tables 1 and 2. To demonstrate the generality of the proposed method, simulations were also performed on an advanced intercross F_3 (see Tables 1 and 2).

The thresholds based on the proposed method match the empirical thresholds reasonably well, and the thresholds are similar when the data are generated from the null and alternative hypotheses. The corresponding

tests have proper type I error and power. The theoretical thresholds based on the dense-map assumption are too conservative while those based on the sparse-map approximation tend to be too liberal, especially for sparse maps. The results based on the method of PIEPHO (2001) are also included in Table 2. As mentioned before, Piepho's method is generally conservative when the marker density is high. In contrast, the proposed method is somewhat on the liberal side in small samples with dense maps. We may combine the proposed method with Piepho's method when the marker density is relatively high.

To further assess the proposed method, we simulated a backcross population in searching for multiple QTL. Again, one chromosome with a total length of 100 cM was simulated. Markers are evenly distributed with a distance of 2, 10, or 20 cM. The sample size is 300. A single QTL is located at 20 cM.

When mapping multiple QTL, the analysis is done either sequentially so that we search for the next most significant QTL after accounting for the effects of the identified QTL or jointly so that we search all QTL simultaneously. In the former, we search for a new gene conditional on previously identified genes.

Setup		Thresholds				Empirical type I error	
		Empirical		Proposed			
Model[a]	Δ[b]	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
1	2	12.13	15.66	12.06	15.74	0.055	0.012
	10	10.66	14.24	10.76	14.29	0.048	0.010
	20	10.10	13.73	10.17	13.71	0.049	0.010
2	2	10.84	14.48	11.17	14.65	0.058	0.011
	10	9.94	13.50	10.27	14.04	0.057	0.013
	20	9.39	12.91	9.37	12.95	0.05	0.010
3	2	11.20	14.68	10.86	14.50	0.058	0.011
	10	10.27	14.04	9.96	13.52	0.057	0.013
	20	9.37	12.95	9.41	12.93	0.049	0.010
4	2	17.80	21.42	17.31	21.44	0.055	0.011
	10	15.96	19.81	15.64	19.68	0.053	0.011
	20	15.10	18.47	14.51	18.51	0.059	0.011

[a] Models 1–3 are sequential tests, where search for the second QTL is conditional on the identified first QTL. Under both models 1 and 2, we search the second QTL on the same chromosome of the first QTL. However, in model 1, the locus of the first QTL is fixed at its true location while in model 2, the first QTL locus is treated as unknown and the marker closest to the locus with the maximum LOD is chosen as the estimated position of the first QTL. Model 3 is similar to model 1, except that we search the second QTL on a different chromosome of the first QTL. Model 4 fits the MIM model with two QTL fitted simultaneously.
[b] Marker distance (in centimorgans).

For the sequential analysis, either we assumed that the position of the first QTL is known (at 20 cM), and given this QTL, we searched for the second QTL, or we assumed that the position of the first QTL is unknown and the marker closest to the locus with the maximum LOD is selected as the locus for the first QTL. Regardless of the method used to choose the position of the first QTL, we tested the null hypothesis H_0: $\gamma_2 = \gamma_{12} = 0$ under model (6) against the alternative hypothesis H_1: $\gamma_2 \neq 0$ or/and $\gamma_{12} \neq 0$ across the whole chromosome. We treated x_1 as fixed and calculated the profile likelihood at all possible loci for the second QTL.

If the putative QTL locus is very close to the primary QTL, the collinearity between x_1 and x_2 will be very strong, which may result in relatively high LOD scores in a region very close to the primary QTL. To investigate this, we simulated another chromosome that is also 100 cM long and searched the second QTL only on the second chromosome.

The above two cases are examples of the CIM analysis in which only one marker, instead of several, is used as the covariate in the analysis. To show the strength of the proposed method in multiple-QTL mapping (MIM), where the computational demand for permutation tests is very high, we also simulated two 100-cM chromosomes and fit a two-QTL model to investigate how the type I errors are controlled under the global null hypothesis of no QTL present. For simplicity, we restricted our profile likelihood calculation to one QTL on each chromosome.

The results of the sequential analysis under $\gamma_1 = 1$ and $\gamma_2 = \gamma_{12} = 0$ are summarized in Table 3. The proposed thresholds are again close to the empirical levels and have proper control of the type I error regardless of whether the first QTL locus is fixed at its true position or selected with the results of the single-QTL interval mapping. Additional simulations (not shown) demonstrate that the resampling thresholds for data generated under alternatives with two QTL are similar to those of Table 3, so that the resampling method yields adequate power.

As shown in Table 3, when we search for the second gene on a different chromosome from the chromosome where the first gene resides, the thresholds are slightly lower than when we search for the second gene on the same chromosome as that of the first gene. This suggests that to retain the power to detect genes not linked to the primary gene, we may partition the whole genome into two groups, one linked with the primary QTL and one unlinked with the primary QTL. The LOD scores within each group can then be compared to the corresponding threshold. We can also exclude a small region, say 10 cM to the left and to the right of the primary QTL to break down the high collinearity between x_1 and x_2, as in the case with CIM.

For MIM, we fit model (6), where neither x_1 nor x_2 is observed and the profile likelihood is calculated for all possible locus combinations of the first and second QTL. The number of testing positions for MIM is on the order of L^K, where L is the total number of loci in

TABLE 4

Empirical type I error of the proposed method in an F$_2$ population with missing genotype data

Setup:	$n^b = 100$		$n = 200$	
Δ^a	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
2	0.066	0.015	0.058	0.011
10	0.065	0.015	0.056	0.010
20	0.055	0.012	0.054	0.012

The average missing genotype rate at each marker is 10%.
^{a} Marker distance (in centimorgans).
^{b} Sample size.

the single-QTL analysis and K is the total number of QTL fitted in MIM, which is a dramatic increase relative to CIM. We performed 1000 simulations for MIM. As shown in Table 3, the proposed method works reasonably well for MIM mapping and the type I errors are well controlled.

To investigate the robustness of the proposed method, we also simulated situations with smaller sample sizes, missing marker genotypes, and χ_1^2-distributed traits. The results for χ_1^2 traits are presented in the bottom three rows of Tables 1 and 2. The type I error is only slightly inflated for $\alpha = 0.05$. As shown in Table 4, the performance of the proposed method is also fairly insensitive to missing genotype data and small sample sizes. With sample size 100 and 10% missing marker genotypes, the type I error is still close to the nominal level.

APPLICATION TO A DROSOPHILA BACKCROSS

We use a Drosophila data set (ZENG *et al.* 2000) to compare the permutation procedure with the proposed method. Two closely related allopatric species, *Drosophila simulans* and *D. mauritiana*, differ dramatically in the size and shape of the posterior lobe of the male genital arch. To investigate the genetic architecture of the morphometric difference between the two species, female *D. simulans* were crossed to males of *D. mauritiana* to generate an F$_1$ population. The F$_1$ females were backcrossed to parental line *D. simulans* and 299 backcross males were produced. A morphometric descriptor, referred to as PC1 by ZENG *et al.* (2000), is the average over both sides of the first principal components of the Fourier coefficients of the posterior lobe and is used to quantify both the size and shape variation. There are 42 markers unevenly distributed on the X chromosome and on chromosomes 2 and 3. Interval mapping was performed across all three chromosomes. The step size of the QTL scan was 1 cM. Threshold calculations were based on 10,000 permutations and resamples. Our recorded running time showed that the proposed method is several hundred times faster than the permutation

procedure (the recorded CPU times of the resampling and permutation procedures running on an IBM Blade-Center HS20 machine are 13 and 6000 sec, respectively).

The derived 95% thresholds are 10.08 and 9.96 from the proposed and permutation methods, respectively. The corresponding 99% thresholds are 13.49 and 13.46. The two procedures result in very similar thresholds, but the proposed method takes far less computing time. The LOD score profile of the original data and the estimated 95% threshold are plotted in Figure 1. The genetic signals on all three chromosomes are very strong. As suggested in ZENG *et al.* (2000), as many as 19 different QTL controlling this morphometric descriptor may exist. For these complicated real data, where some of our assumptions, such as normality, are likely to fail, the thresholds from permutation and our proposed procedures agree very well. Since the permutation procedure is known to be robust to those violations, this real example further demonstrates the usefulness of the proposed method. To further compare the permutation and proposed method, we provided a QQ-plot (Figure 2) of the permutation- and the resample-based null distribution estimates of the maximum profile likelihood-ratio test statistic. The two estimated distributions match rather well up to the 99.5th percentile. The discrepancy in the tails of the distributions may be due, in part, to the limited number of resamplings and permutations. To improve the accuracy of the estimates of the null distribution in the tail, a larger number of resamplings and permutations are necessary. For comparison, we also calculated the 95 and 99% thresholds by PIEPHO's (2001) method, which are 11.23 and 14.44, respectively. Those thresholds are slightly larger than both the permutation-based and our resampling-based thresholds.

DISCUSSION

In this article, we propose a new empirical method to calculate the threshold for QTL mapping. The method is far more efficient than the popular permutation procedure since the proposed method needs to maximize the likelihood of the observed data only once with no need to maximize the likelihood in each resampling iteration any more. For standard interval mapping with simple crosses, the resampling method is several hundred times faster than the permutation procedure. Furthermore, the proposed method is applicable to more complicated designs and models that cannot be handled by the permutation procedure. For example, for MIM where the model selection is involved, the bootstrap resampling method of ZENG *et al.* (1999) is applicable to the linear regression model but may not be applicable to nonlinear models, such as logistic regression and Poisson regression. The proposed method also avoids the derivation of parameters in the Ornstein-Uhlenbeck diffusion approximations, which can be a difficult task when the model is complicated.
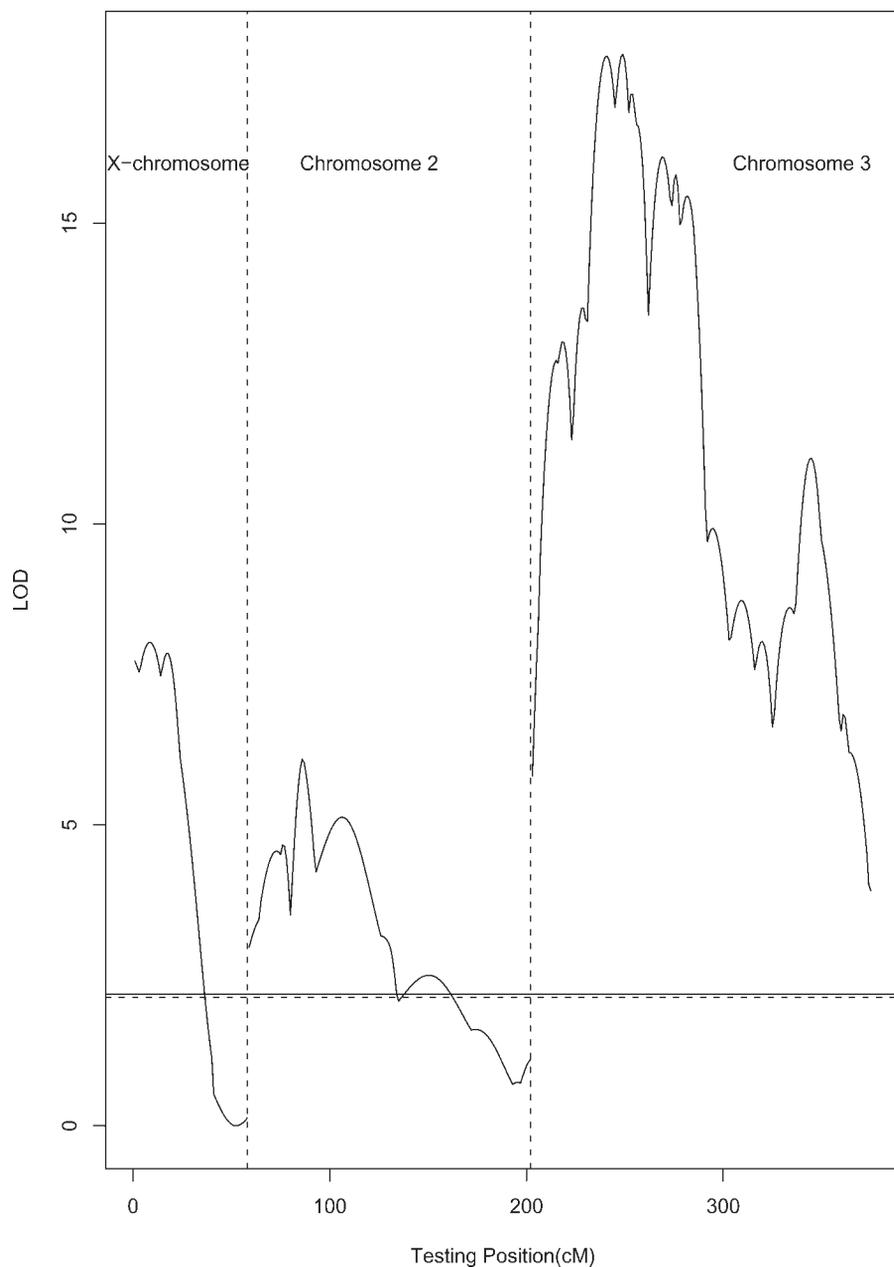
FIGURE 1.—The LOD profile for chromosomes X, 2, and 3 from interval mapping. The solid horizontal line is the 95% resampling threshold, which is almost identical to the 95% permutation threshold (the dashed horizontal line).

The computational advantage of the proposed method over the permutation procedure depends on how complex the original model is. The more complicated the model is, the more there is to be gained from the proposed method. In the Drosophila analysis, where a simple interval-mapping model was fitted on three chromosomes, there was a decrease in computing time in the order of 10^2. If more complicated models, such as multiple-QTL mapping or CIM, are used, where maximization via the EM algorithm is more time consuming, the proposed method may be thousands of times faster than the permutation procedure. With the recent efforts to map the gene expression levels of thousands of genes via microarrays (LAN *et al.* 2003), an efficient way to compute thresholds in a large number of screens is critical, even with the current trend in computing power.

The simulations indicated that for simple interval mapping with F_2 or backcross, either the restricted or unrestricted estimator of η can be used and the two estimators tend to give very similar thresholds. However, for the two-gene model, we found that the unrestricted estimator of η works slightly better than the restricted one. For this reason, we suggest the use of the unrestricted estimator of the nuisance parameters in evaluating the thresholds, and the simulation results presented in this article are based on an unrestricted estimator of η.

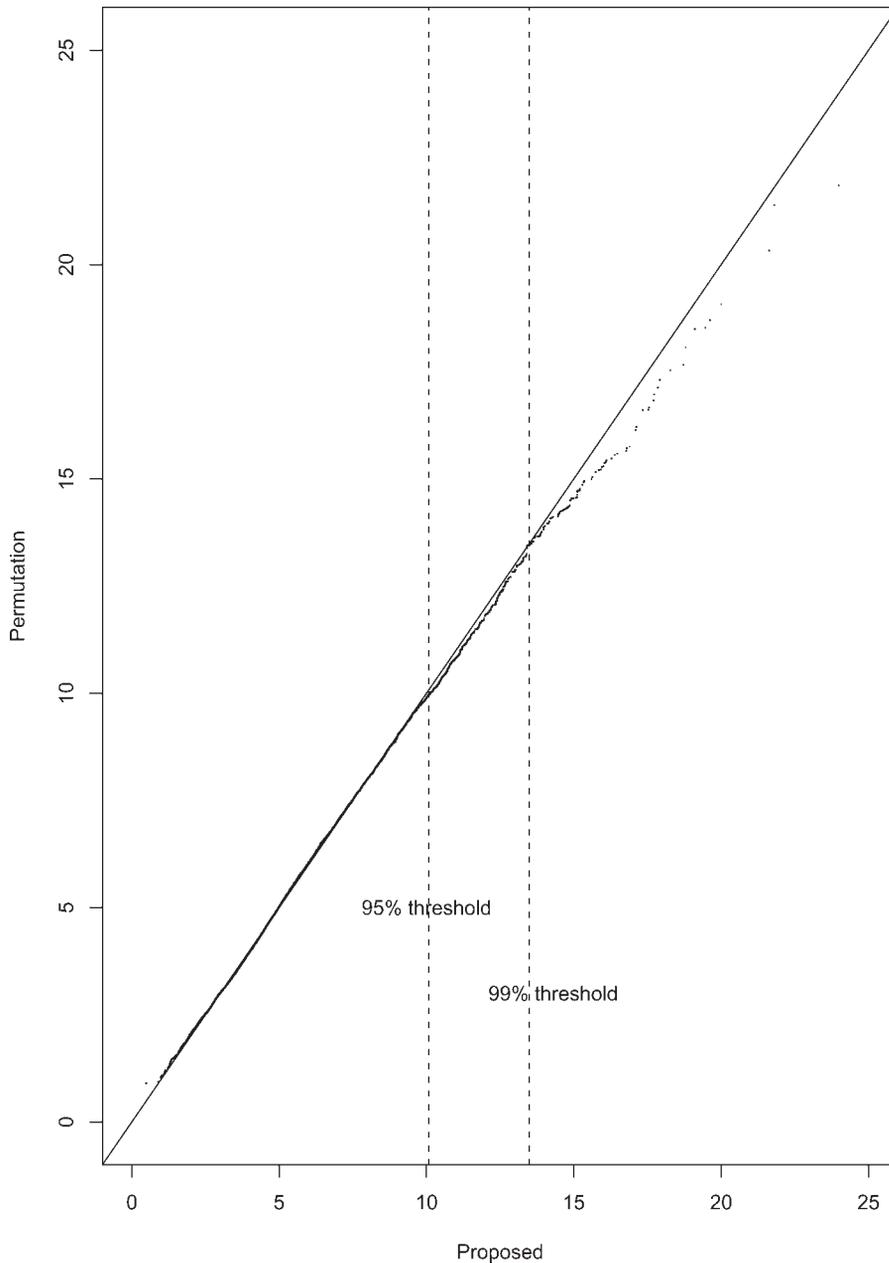The simulations also showed that the proposed method is robust to nonnormality as well as missing

FIGURE 2.—The QQ-plot of the estimated null distributions from the proposed method and the empirical permutation method in the Drosophila data analysis. The two vertical dashed lines are the 95 and 99% thresholds from the proposed method. The solid diagonal line represents the situation when the two estimated distributions were identical. The actual estimated distributions are represented by the points along this diagonal line.

data. Though the normal model is used to fit the nonnormal chi-square data, the empirical type I error from the proposed method is reasonably controlled at the targeted level. However, it is unclear how this method will work for data with segregation distortion, which is a complex phenomenon. Due to different mechanisms of segregation distortion, it is difficult to predict the performance of the method. If the existence of segregation distortion is suspected, a simple solution is to remove those markers that are in segregation distortion from the analysis. Including markers in segregation distortion in the analysis will result in a distorted map estimate and may give biased mapping results, regardless of the method used to compute the thresholds. Last, in contrast to PIEPHO's (2001) method, which is

generally conservative when the marker density is high, the proposed method is somewhat on the liberal side in small samples with dense maps. We may consider combining the proposed method with Piepho's method when the marker density is relatively high.

LITERATURE CITED

BASTEN, C. J., B. S. WEIR and Z-B. ZENG, 1997   *QTL Cartographer: A Reference Manual and Tutorial for QTL Mapping*. Department of Statistics, North Carolina State University, Raleigh, NC.

BROMAN, K. W., 2001   Review of statistical methods for QTL mapping in experimental crosses. Lab Anim. **30:** 44–52.

CHURCHILL, G. A., and R. W. DOERGE, 1994   Empirical threshold values for quantitative trait mapping. Genetics **138:** 963–971.

COX, D., and C. HINKLEY, 1974   *Theoretical Statistics*. Chapman & Hall, London.

DAVIES, R. B., 1977 Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika **64:** 247–254.

DAVIES, R. B., 1987 Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika **74:** 33–43.

DEMPSTER, A. P., N. M. LAIRD and D. B. RUBIN, 1977 Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. **39:** 1–38.

DOERGE, R. W., 2001 Mapping and analysis of quantitative trait loci in experimental populations. Nat. Rev. Genet. **3:** 43–52.

DUPUIS, J., and D. SIEGMUND, 1999 Statistical methods for mapping quantitative trait loci from a dense set of markers. Genetics **151:** 373–386.

JANSEN, R. C., and P. STAM, 1994 High resolution of quantitative traits into multiple quantitative trait in line crosses using flanking markers. Genetics **136:** 1447–1455.

KAO, C. H., and Z-B. ZENG, 1997 General formulas for obtaining the maximum likelihood estimates and the asymptotic variance-covariance matrix in QTL mapping when using the EM algorithm. Biometrics **53:** 653–665.

KAO, C. H., Z-B. ZENG and R. D. TEASDALE, 1999 Multiple interval mapping for quantitative trait loci. Genetics **152:** 1203–1216.

LAN, H., J. P. STOEHR, S. T. NADLER, K. L. SCHUELER, B. S. YANDELL *et al.*, 2003 Dimension reduction for mapping mRNA abundance as quantitative traits. Genetics **164:** 1607–1614.

LANDER, E. S., and D. BOTSTEIN, 1989 Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics **121:** 185–199.

LIN, D. Y., L. J. WEI and Z. YING, 1993 Checking the Cox model with cumulative sums of martingale-based residuals. Biometrika **80:** 557–572.

LINCOLN, S. E., M. J. DALY and E. S. LANDER, 1993 *A Tutorial and Reference Manual for MAPMAKER/QTL.* Whitehead Institute, Cambridge, MA.

LYNCH, M., and B. WALSH, 1998 *Genetics and Analysis of Quantitative Traits.* Sinauer Asociates, Sunderland, MA.

MANLY, K. F., and J. M. OLSON, 1999 Overview of QTL mapping software and introduction to Map Manager QT. Mamm. Genome **10:** 327–334.

PIEPHO, H. P., 2001 A quick method for computing approximate thresholds for quantitative trait loci detection. Genetics **157:** 425–432.

REBAI, A., B. GOFFINET and B. MANGIN, 1994 Approximate thresholds of interval mapping tests for QTL detection. Genetics **138:** 235–240.

REBAI, A., B. GOFFINET and B. MANGIN, 1995 Comparing power of different methods for QTL detection. Biometrics **51:** 87–99.

ZENG, Z-B., 1993 Theoretical basis of separation of multiple linked gene effects on mapping quantitative trait loci. Proc. Natl. Acad. Sci. USA **90:** 10972–10976.

ZENG, Z-B., 1994 Precision mapping of quantitative traits loci. Genetics **136:** 1457–1468.

ZENG, Z-B., C. H. KAO and C. J. BASTEN, 1999 Estimating the genetic architecture of quantitative traits. Genet. Res. **74:** 279–289.

ZENG, Z-B., L. LIU, L. F. STAM, C. H. KAO, J. M. MERCER *et al.*, 2000 Genetic architecture of a morphological shape difference between two Drosophila species. Genetics **154:** 299–310.

ZOU, F., B. S. YANDELL and J. P. FINE, 2001 Statistical issues in the analysis of quantitative traits in combined crosses. Genetics **158:** 1339–1346.

Communicating editor: C. S. HALEY