

A simple correction for multiple comparisons in interval mapping genome scans

JAMES M. CHEVERUD

Department of Anatomy & Neurobiology, Washington University School of Medicine, 660 S. Euclid Ave., St. Louis, MO, 63110, U.S.A.

Several approaches have been proposed to correct point-wise significance thresholds used in interval-mapping genome scans. A method for significance threshold correction based on the Bonferroni test is presented. This test involves calculating the effective number of independent comparisons performed in a genome scan from the variance of the eigenvalues of the observed marker correlation matrix. The more highly correlated the markers, the higher the variance of the eigenvalues and the lower the number of independent tests performed on a chromosome. This approach was evaluated by mapping 1000 normally distributed phenotypes along chromosomes of varying length and marker density in a population size of 500. Experiment-wise significance thresholds obtained from the simulation are compared to those calculated using the Bonferroni criterion and the newly developed measure of the effective number of independent tests in a genome scan. The Bonferroni calculation produced significance thresholds very similar to those obtained by simulation. The threshold levels for both Bonferroni and simulation analysis depended strongly on the marker density and size of chromosomes. There was a slight bias of about 1% in the thresholds obtained at the 5% and 10% point-wise significance levels. The method introduced here provides a relatively simple correction for multiple comparisons that can be easily calculated using standard statistics packages.

Keywords: Bonferroni correction, interval mapping, multiple comparisons, quantitative trait loci.

Introduction

Whole genome scans have been commonly used to detect Quantitative Trait Loci (QTLs), especially since Lander & Botstein (1989) introduced the concept of interval mapping. In interval mapping studies, a series of molecular marker genotypes are scored across the genome at some desired density, often a marker every 10–20 centiMorgans (cM), along each chromosome (Darvasi & Soller, 1994). Tests for linkage between chromosomal locations and a phenotype of interest are then performed at arbitrary map distances, usually in two centiMorgan intervals (Lander & Botstein, 1989), through each intermarker segment. One persistent and sometimes controversial element of QTL mapping approaches has been the determination of appropriate thresholds for identifying statistically significant linkage (Lander & Kruglyak, 1995, 1996; Curtis, 1996; Witte *et al.*, 1996). In statistical testing it is usual to reject the null hypothesis of no linkage if the probability of obtaining observed results under the null

hypothesis is less than a standard threshold, typically 5%. However, when many tests are performed addressing the same issue, such as linkage of a trait across a genome, we expect that fully 5% of the tests performed will produce observations ‘significantly different’ from the null hypothesis at the 5% level even when there is truly no linkage present. Such linkages are often referred to as ‘false positives’ because they falsely reject a true null hypothesis of no linkage. Although a multiple comparison problem occurs in any research design involving many comparisons, it is particularly important in genome-wide QTL mapping where the phenotype in question is tested repeatedly along the chromosomes.

Lander & Botstein (1989) carefully considered the question of appropriate significance thresholds for sparse and dense map designs in their original paper. They define a sparse map as one in which consecutive markers are well separated so that each marker is inherited independently and the number of independent comparisons is equal to the number of markers tested. In this situation, they recommend the commonly used Bonferroni correction for multiple comparisons in which the probability of a false positive decision in

any one of the many experimental tests is limited to 0.05. The value corresponding to a particular experiment-wide significance threshold (α) is usually obtained by dividing the nominal significance threshold for a single test (γ) by the number of independent tests (n), as in $\alpha = (\gamma/n)$. This is an approximation of a more exact equation given by Lynch & Walsh (1998) where

$$\alpha = 1 - (1 - \gamma)^{(1/n)} \quad (1)$$

Lander & Botstein (1989) also analytically derived an appropriate Bonferroni correction for a dense map, in which the number of intervals in a genomic scan approaches infinity. In this case they showed that the required significance level for individual tests approaches a nonzero limit independent of the number of markers and they provided the appropriate equation. Most genome studies lie between these two extreme examples, leaving many researchers still in doubt about an appropriate significance threshold for their genome scans. Lander & Botstein (1989) also proposed simulation as an appropriate means of identifying thresholds for intermediate designs.

Rebai *et al.* (1994) introduced the Davies approximation as an appropriate method for determining thresholds and provided equations useful for a variety of research designs. They found that the Davies equation gave a good approximation for thresholds as long as sample sizes were not too small. However, they pointed out that the ‘...formal calculations needed could be difficult to carry out ...’ (p. 238; Rebai *et al.*, 1994) and that if they were difficult in any given case, either numerical approximations or simulations could be used to obtain thresholds.

Churchill & Doerge (1994) proposed and validated a very general method for determining threshold levels with multiple comparisons. They recommended that one obtain empirical thresholds for rejecting the null model of no linkage by a permutation test. In a permutation test the phenotypic data are randomly shuffled relative to the genotypic data and the entire genome scan repeated on the shuffled data. Permuted phenotypes should bear no relation to their non-permuted genotypic counterparts and thereby simulate results expected under the null hypothesis. These simulations are repeated many times to obtain a distribution of linkage statistics that is expected under the null hypothesis. The experiment-wise significance threshold can then be obtained from the appropriate percentile of this distribution. If the observed linkage statistic exceeds the threshold, the null hypothesis of no linkage is rejected. Churchill & Doerge (1994) recommend at least 1000 shuffles to estimate a 0.05 threshold and 10 000 shuffles to estimate a 0.01 threshold.

The permutation test approach to setting significance thresholds is very robust and has several advantages. Primary among these is that it draws the threshold directly from the data being analysed. Peculiarities of the observed data, such as deviations of the phenotype from a normal distribution, biased allele frequencies, and patterns of missing data are maintained in the permuted data sets and are included in estimation of the thresholds obtained. However, this faithfulness to the observed data also leads to some disadvantages for this method. It is very computer intensive. Instead of performing the entire genome scan once with the observed data, it is necessary to repeat it 1000–10 000 times. A second disadvantage of this method is that the permutation testing needs to be performed separately for each phenotypic trait because each trait has its own distributional peculiarities in any given sample. This difficulty can be remedied, in part, by using multivariate QTL mapping approaches (Jiang & Zeng, 1995) but still can be very time-consuming.

Lander & Kruglyak (1995) addressed the issue of multiple comparisons in genome scans for a wide variety of research designs in human populations and in various model organisms, concentrating on correcting the threshold for a dense map, as in Lander & Botstein (1989). A dense map contains an infinite number of markers and infinitely small intermarker intervals. They point out the very real problem that many false positive linkages will be reported unless significance thresholds are adjusted for multiple comparisons. Because, in their words, they perceive a ‘glazed-eye indifference’ (Lander & Kruglyak, 1995; p. 241) in how biologists often view statistical issues, they provide significance threshold guidelines for a dense map research design and suggest that these be generally followed. The classification scheme includes criteria for suggestive linkage (false positive expected to occur one time at random in a genome scan), significant linkage (false positive expected to occur 0.05 times in a genome scan), highly significant linkage (false positive expected to occur 0.001 times in a genome scan) and confirmed linkage (significant linkage confirmed with new data). Only QTLs with confirmed linkage should be named. These thresholds are severe relative to most QTL mapping experiments in which a moderate density of one marker every 10–20 cM is more common than a truly dense map.

Lander & Kruglyak (1995) defend the assumption of a dense map based on research designs that may or may not be used in following up the initial results in the original mapping population. However, they may assume too much about individual research plans for the dense map assumption to be universally valid. In this paper, I present a general method of adjusting significance thresholds for multiple comparisons using the

Bonferroni procedure. This method allows the calculation of the number of independent tests performed in any interdependent set of tests either across independent chromosomes or across the whole genome, thus providing appropriate significance thresholds for genome scans.

Correction for multiple comparisons

The Bonferroni correction for multiple comparisons is given in eqn 1 above. It allows an experiment-wide threshold (α) to be calculated given the corresponding point-wise (single-test) threshold (γ) and the number of independent comparisons made (n). The point-wise threshold is determined by the researcher in advance and takes on an arbitrary value, although 0.05 is most commonly used. After settling on a point-wise significance level, one needs to determine how many independent tests are performed in order to calculate the experiment-wide threshold. The logic behind calculating the number of independent comparisons is described below.

Calculation of the number of independent comparisons contained within a set is based on measuring the correlations among independently scored markers along a chromosome, because it is the interdependence of these markers that makes for interdependent statistical tests of the null hypothesis of no linkage. Correlations are measures of marker interdependence and correcting the total number of tests performed by their level of interdependence provides the number of independent tests needed to apply a Bonferroni correction. It has been known for some time that the total amount of correlation among traits in a set can be measured by the variance of the eigenvalues derived from their correlation matrix (Cheverud *et al.*, 1983, 1989; Wagner, 1984a, 1990; Cheverud, 1989, 1996). Higher correlation among the traits leads to higher eigenvalue variance. For example, when there is no

correlation among traits, all of the eigenvalues of the correlation matrix are equal to one and the set of eigenvalues has no variance. On the other hand, if all traits are maximally correlated, the first eigenvalue is equal to the number of traits represented in the matrix while the rest of the eigenvalues are equal to zero. In this situation the variance is at its maximum and equal to the number of traits in the matrix (M). The variance of the eigenvalues will range between zero, when all traits are independent, and M , where M is the number of traits included in a matrix.

One can calculate the proportional reduction in the number of elements in a set by the ratio of the eigenvalue variance to its maximum ($V_{\lambda\text{obs}}/M$). A general equation for the effective number of independent traits (M_{eff}), rescaled to vary between 1.0 and M , is

$$M_{\text{eff}} = M(1 - (V_{\lambda\text{obs}}/M)^2). \quad (2)$$

To determine the number of independent tests in a genome scan, the number of traits (M) is the number of markers on a chromosome or in the genome as a whole and $V_{\lambda\text{obs}}$ is the observed variance of the eigenvalues of the correlation matrix for additive genotypic scores at each marker. One first obtains the marker correlation matrix and its eigenvalues, calculates the variance of these eigenvalues, and substitutes the observed eigenvalue variance into eqn 2. The effective number of independent comparisons (substituting M_{eff} for n) can then be used in eqn 1 to obtain the Bonferroni-corrected threshold

$$\alpha = 1 - \gamma^{(1/M_{\text{eff}})}. \quad (3)$$

An example based on the genome scans in the LG/J by SM/J mouse F_2 intercross reported by Cheverud *et al.* (2001) is presented in Table 1. We consider mouse chromosome 1 scored with eight markers in map positions indicated in Table 1. First, additive genotypic scores are calculated at each of the markers on a

Table 1 Microsatellite markers scored on chromosome 1 in the LG/J \times SM/J intercross (Cheverud *et al.*, 2001). Mapped marker positions are given in Haldane's centiMorgans (cM). Correlations of additive genotypic scores at the markers are given below the diagonal and the eight eigenvalues associated with this correlation matrix are given along the diagonal

Locus	Position (cM)	<i>D1Mit3</i>	<i>D1Mit20</i>	<i>D1Mit74</i>	<i>D1Mit7</i>	<i>D1Mit11</i>	<i>D1Mit14</i>	<i>D1Mit17</i>	<i>D1Mit155</i>
<i>D1Mit3</i>	0.0	3.63							
<i>D1Mit20</i>	9.4	0.82	1.79						
<i>D1Mit74</i>	29.5	0.64	0.65	1.13					
<i>D1Mit7</i>	41.4	0.48	0.51	0.75	0.47				
<i>D1Mit11</i>	52.1	0.32	0.35	0.57	0.75	0.37			
<i>D1Mit14</i>	73.3	0.12	0.12	0.27	0.45	0.60	0.25		
<i>D1Mit17</i>	115.2	0.08	0.10	0.15	0.15	0.19	0.33	0.18	
<i>D1Mit155</i>	131.7	0.02	0.02	0.06	0.03	0.09	0.35	0.61	0.17

chromosome. Additive genotypic scores take on a value of (+1) at one homozygote (0) at the heterozygote, and (-1) at the alternate homozygote (Haley & Knott, 1992). Then marker correlations are calculated as Pearson product moment correlations among the additive genotypic scores (see Table 1). The eigenvalues of this correlation matrix are obtained by principal components analysis, or more generally by spectral decomposition. The mean of these eigenvalues is 1.0 by definition and their variance is calculated in the ordinary fashion

$$V_{\lambda_{\text{obs}}} = \sum_{i=1}^M (\lambda_i - 1)^2 / (M - 1). \quad (4)$$

The eigenvalues of the marker correlation matrix for chromosome 1 are given along the diagonal of the matrix in Table 1. The variance of these eigenvalues is 1.266. Placing this value into eqn 3 yields an estimate of 6.89 independent markers (M_{eff}) on the chromosome. Placing this value into eqn 4 using a nominal point-wise significance threshold of 0.05 yields a chromosome-wide threshold value of 0.00725, or 2.14 when considered on a LOD scale $\{\text{LOD} = \log_{10}(1 / \text{Prob.})\}$. Thus, we expect probabilities less than 0.00725 to occur somewhere along this chromosome in 5% of the chromosome scans even when the null hypothesis of no QTL effect is true. This procedure can then be followed for each chromosome in succession. Genome-wide threshold values can be obtained by summing M_{eff} over all chromosomes when the chromosomes are in linkage equilibrium, as in an experimental F_2 intercross. Alternatively, a single correlation matrix can be constructed for all the markers across the whole genome and then the variance of the eigenvalues of this matrix obtained and substituted into eqn 2.

Importantly, different chromosomes in a single study will have different numbers and densities of markers and therefore should be tested against their own, chromosome-specific threshold. Also, the structure and amount of missing data varies among chromosomes and this structure is accounted for in calculating the effective number of independent tests. Lander & Kruglyak's (1995) dense map approximation does not differ by chromosome because, with infinite markers, the number of markers no longer affects the number of tests. Correction for a whole genome scan, corresponding to Lander & Kruglyak's (1995) criterion for significant linkage, can be obtained using the correlation matrix for all markers together or, in an F_2 intercross where chromosomes are in linkage equilibrium, by summing the effective number of markers across the chromosomes.

Simulation

Simulation data and analysis

The proposition that appropriate significance thresholds for linkage analysis can be obtained from eqn 3 above was tested by simulation. One thousand independent, random, normally distributed ($N(0,1)$) traits were generated for each of 500 individuals using the appropriate random value generator in SYSTAT 7.0 (Wilkinson, 1997). These traits were tested for deviations from a normal distribution and 5% of the cases differed from normality at the 0.05 level, as expected. Sets of 500 marker genotypes were produced for each chromosome tested. The genotype at the most proximal locus on each chromosome was chosen at random from the genotype distribution expected in an F_2 intercross population. Subsequent marker genotype values were obtained using the simulated recombination rate between markers and the genotype of the next most proximal marker until all markers on the chromosome were assigned genotypes. Chromosomes tested were 50, 75 and 100 cM long and had intermarker distances of 50 cM, 25 cM, 12.5 cM and 6.25 cM. Short intermarker distance implies higher density of markers.

Interval mapping was performed using the set correlation approach in SYSTAT 7.0 (Wilkinson, 1997). The random phenotype is the dependent variable whereas the additive (X_a) and dominance (X_d) genotypic scores are the independent variables (Haley & Knott, 1992). The additive genotype score (X_a) is the weighted probability that the genotype is AA or aa at the position of interest. The dominance genotype score is the weighted probability that the genotype is heterozygous (Aa) or homozygous (AA , aa) at the position of interest. The probabilities and genotype scores are obtained using the recombination rate between the position of interest and the flanking markers (Haley & Knott, 1992). Genotypic scores were estimated every 2 cM in the intervals between markers.

The 1000 traits serve as 1000 replicates of the null model in which the phenotypes are unlinked to the marker data. By performing interval-mapping analyses on this simulated data, it was possible to compare the analytical results obtained from eqn 3 with those obtained by the simulated data in an F_2 intercross composed of 500 individuals. For each trait at each length and marker density, 1000 chromosomes were interval mapped. For each replicate, the minimum probability of obtaining the observed result with no linkage was extracted from the data and collated into a distribution of probabilities expected under the null hypothesis. For example, the 5% threshold for significance accounting for multiple comparisons would be

equal to the 5th percentile, or 50th case in the ordered distribution of probabilities.

Simulation results

The significance thresholds obtained from the simulation and from the Bonferroni correction using the effective number of independent tests are presented in Table 2 for chromosome-wide 1%, 5%, and 10% levels. It is clear that the Bonferroni-based thresholds are quite close to the simulation-based values in each case. Both simulation and Bonferroni-corrected values decrease as

the density of the markers increases and as the length of the chromosome increases. The squared correlation between simulated and Bonferroni-calculated thresholds is very high, greater than 0.96.

However, there is a very slight bias towards less extreme probability thresholds from the Bonferroni correction relative to the simulation results. The bias in Bonferroni-based thresholds is very small and often not significantly different from zero given 1000 replicates. The percentile position of the Bonferroni-based threshold within the distribution of simulation results is a measure of this bias (see Table 2). Bias is small to

Chromosome length (cM)	Intermarker distance (cM)	Bonferroni-based threshold	Simulation-based threshold	Bias†
1% thresholds				
50	6.25	0.00171	0.00125	0.003
75	6.25	0.00109	0.00099	0.001
100	6.25	0.00080	0.00077	0.000
50	12.50	0.00269	0.00192	0.004
75	12.50	0.00189	0.00148	0.001
100	12.50	0.00143	0.00103	0.001
50	25.00	0.00376	0.00343	0.001
75	25.00	0.00284	0.00188	0.004
100	25.00	0.00227	0.00173	0.006
50	50.00	0.00533	0.00360	0.002
100	50.00	0.00361	0.00202	0.004
5% thresholds				
50	6.25	0.00871	0.00659	0.007
75	6.25	0.00557	0.00472	0.009
100	6.25	0.00407	0.00411	-0.010
50	12.50	0.01364	0.01109	0.014
75	12.50	0.00962	0.00693	0.010
100	12.50	0.00728	0.00554	0.015
50	25.00	0.01905	0.01763	0.007
75	25.00	0.01442	0.01016	0.012
100	25.00	0.01151	0.00831	0.024
50	50.00	0.02692	0.02135	0.012
100	50.00	0.01830	0.01340	0.019
10% thresholds				
50	6.25	0.01781	0.02025	-0.013
75	6.25	0.01140	0.01249	-0.004
100	6.25	0.00833	0.00957	-0.012
50	12.50	0.02782	0.02092	0.020
75	12.50	0.01966	0.01689	0.014
100	12.50	0.01491	0.01269	0.013
50	25.00	0.03874	0.03627	0.006
75	25.00	0.02940	0.02511	0.015
100	25.00	0.02349	0.01793	0.027
50	50.00	0.05452	0.04453	0.015
100	50.00	0.03723	0.02700	0.029

Table 2 Simulation results for experiment-wide 1%, 5% and 10% significance thresholds compared to thresholds based on Bonferroni corrections using the calculated number of independent comparisons

†Bias is the deviation of the percentile rank of the Bonferroni-based threshold in the Simulation-based distribution.

nonexistent at the 1% threshold. The average bias is 0.002. Bias is more apparent at the 5% and 10% thresholds, but is still only about 0.01.

Discussion

A simple method for obtaining appropriate thresholds for genome scans using the Bonferroni correction has been validated by simulation. Advantages of the method are its relative ease of use and its specificity to the research design actually carried out in any particular interval mapping study. Unlike the theoretical thresholds suggested by Lander & Kruglyak (1995), these thresholds are based directly on the amount and pattern of interdependence among the markers scored in a particular experiment. Patterns of missing genotypic data are accounted for in these calculations because the correlations are obtained directly from the marker data themselves. The proposed correction can also be directly applied in cases in which a series of covariates are used in the analysis, as in some applications of composite interval mapping (Zeng, 1994). If the covariates are independent of the markers composing the map of the chromosome under consideration, there will be no effect because this does not change the level and pattern of marker correlation along the mapped chromosome. In cases where covariates are correlated with mapped marker positions because of linkage disequilibrium, the covariates should be regressed out of the marker genotype scores before calculating the intermarker correlation matrix and its eigenvalues.

The proposed Bonferroni correction can be easily calculated for any experimental design because it depends directly on the observed marker correlations. Thus this approach should prove useful in various inbreeding and outbreeding designs because the effect of the design will be manifest at the level of the observed marker correlations. Modifications may be required when more than two alleles are present, but the principle remains the same. Marker correlation matrices and associated eigenvalues can be calculated with most statistical packages and the variance of these eigenvalues quickly calculated in a spreadsheet. Using this method, appropriate, experiment-specific significance thresholds can be easily calculated.

The relationship between the variance of the eigenvalues of the marker correlation matrix and the number of independent comparisons represented on a chromosome can also be used prospectively in designing genome-scanning experiments. The relationship between recombination rate and correlation can be approximated by

$$r = 1 - 2c, \quad (5)$$

where r is the intermarker correlation and c is the recombination rate. This correlation is sometimes referred to as the linkage parameter (Weir, 1996). Thus, published map positions can be used to calculate recombination rates and marker correlations so that the effects of multiple comparisons on chromosome-wide and genome-wide significance thresholds can be accounted for in determining the statistical power of experimental designs.

Churchill & Doerge's (1994) suggestion that appropriate thresholds be obtained by simulation depends even more fully on the characteristics of the data collected than the proposed Bonferroni correction because it can also take into account the potentially non-normal distribution of the mapped traits. The simulation performed here used normally distributed traits. However, Churchill & Doerge's (1994) approach also has the disadvantage of being very time consuming and computer intensive with a new simulation needed for every character studied. The Bonferroni correction described here is specific to an experiment, but does not have to be redone for every character. The loss in threshold accuracy potentially involved in performing the proposed Bonferroni correction is likely to be very small unless trait distributions within genotype classes deviate strongly from normality.

The multiple comparisons correction method proposed here is based on corrections for individual chromosomes. This was done because individual chromosomes vary in length and marker density in most interval mapping studies. Chromosome-specific corrections will differ from one another given differences in chromosome length and marker density. The appropriate chromosome-wide threshold for mouse chromosome 1 should be much higher than for mouse chromosome 19 because, even with the same marker density, the chromosomes are grossly different in length and therefore contain different numbers of multiple comparisons. Likewise, chromosomes of the same length but with varying marker densities also require different thresholds.

Genome-wide thresholds can also be obtained by this method. If the experimental population is from an F_2 intercross, different chromosomes are expected to be in linkage equilibrium. If this is the case, the effective number of independent tests across the whole genome can be obtained by summing the chromosome-specific values. More generally and even with linkage disequilibrium among chromosomes, all markers across the genome can be analysed jointly with a single variance of the eigenvalues obtained from the genome-wide intermarker correlation matrix.

No matter which of the methods discussed above is used for correcting point-wise significance levels for

multiple comparisons, some correction must be applied. The simulations reported here and in other papers (Lander & Kruglyak, 1995) clearly show that point-wise thresholds will be repeatedly exceeded in a genome scan even if there is no linkage between markers and phenotypes.

Acknowledgements

We gratefully acknowledge NIH grant DK52514 and NSF grant DEB-9726433 for support of the work presented here. We also appreciate the comments from D.C. Rao, Alan Templeton and the anonymous reviewers on an earlier version of this manuscript.

References

- CHEVERUD, J. 1989. A comparative analysis of morphological variation patterns in the papionins. *Evolution*, **43**, 1737–1747.
- CHEVERUD, J. 1996. Quantitative genetic analysis of cranial morphology in the cotton-top (*Saguinus oedipus*) and saddle-back (*S. fuscicollis*) tamarins. *J. Evol. Biol.*, **9**, 5–42.
- CHEVERUD, J., RUTLEDGE, J. AND ATCHLEY, W. 1983. Quantitative genetics of development: Genetic correlations among age-specific trait values and the evolution of ontogeny. *Evolution*, **37**, 895–905.
- CHEVERUD, J., WAGNER, G. AND DOW, M. 1989. Methods for the comparative analysis of variation patterns. *Syst. Zool.*, **38**, 201–213.
- CHEVERUD, J. M., VAUGHN, T. T., PLETSCHER, L. S., PERIPATO, A. C. ET AL. 2001. Genetic architecture of adiposity in the cross of LG/J and SM/J inbred mice. *Mamm. Genome*, **12**, 3–12.
- CHURCHILL, G. AND DOERGE, R. 1994. Empirical threshold values for quantitative trait mapping. *Genetics*, **138**, 963–971.
- CURTIS, D. 1996. Comment on Genetic dissection of complex traits. *Nature Genet.*, **12**, 356–357.
- DARVASI, A. AND SOLLER, M. 1994. Optimum spacing of genetic markers for determining linkage between marker loci and quantitative trait loci. *Theor. Appl. Genet.*, **89**, 351–357.
- HALEY, C. AND KNOTT, S. 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity*, **69**, 315–324.
- JIANG, C. AND ZENG, Z.-B. 1995. Multiple trait analysis of genetic mapping for quantitative trait loci. *Genetics*, **140**, 1111–1127.
- LANDER, E. AND BOTSTEIN, D. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, **121**, 185–199.
- LANDER, E. AND KRUGLYAK, L. 1995. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genet.*, **11**, 241–247.
- LANDER, E. AND KRUGLYAK, L. 1996. Reply to comments on Genetic dissection of complex traits. *Nature Genet.*, **12**, 357–358.
- LYNCH, M. AND WALSH, B. 1998. *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, MA.
- REBAI, A., GOFFNET, B. AND MANGIN, B. 1994. Approximate thresholds of interval mapping tests for QTL detection. *Genetics*, **138**, 235–240.
- WAGNER, G. 1984. On the eigenvalue distribution of genetic and phenotypic dispersion matrices: Evidence for a nonrandom organization of quantitative character variation. *J. Math. Biol.*, **21**, 77–95.
- WAGNER, G. 1990. A comparative study of morphological integration in *Apis mellifera*. *Z. Zool. Syst. Evolution-forsch.*, **28**, 48–61.
- WEIR, B. S. 1996. *Genetic Data Analysis II*. Sinauer Associates, Sunderland, MA.
- WILKINSON, L. 1997. *SYSTAT 7.0 for Windows*. SPSS Inc., Chicago.
- WITTE, J., ELSTON, R. AND SCHORK, N. 1996. Comment on Genetic dissection of complex traits. *Nature Genet.*, **12**, 355–356.
- ZENG, Z.-B. 1994. Precision mapping of quantitative trait loci. *Genetics*, **136**, 1457–1468.