A New Approach to the Problem of Multiple Comparisons in the Genetic Dissection of Complex Traits

Joel Ira Weller,* Jiu Zhou Song,* David W. Heyen,[†] Harris A. Lewin[†] and Micha Ron*

**Institute of Animal Sciences, Agricultural Research Organization, the Volcani Center, Bet Dagan 50250, Israel and [†]Laboratory of Immunogenetics, Department of Animal Sciences, University of Illinois, Urbana, Illinois 61801*

Manuscript received June 1, 1998
Accepted for publication September 11, 1998

ABSTRACT

Saturated genetic marker maps are being used to map individual genes affecting quantitative traits. Controlling the "experimentwise" type-I error severely lowers power to detect segregating loci. For preliminary genome scans, we propose controlling the "false discovery rate," that is, the expected proportion of true null hypotheses within the class of rejected null hypotheses. Examples are given based on a granddaughter design analysis of dairy cattle and simulated backcross populations. By controlling the false discovery rate, power to detect true effects is not dependent on the number of tests performed. If no detectable genes are segregating, controlling the false discovery rate is equivalent to controlling the experimentwise error rate. If quantitative loci are segregating in the population, statistical power is increased as compared to control of the experimentwise type-I error. The difference between the two criteria increases with the increase in the number of false null hypotheses. The false discovery rate can be controlled at the same level whether the complete genome or only part of it has been analyzed. Additional levels of contrasts, such as multiple traits or pedigrees, can be handled without the necessity of a proportional decrease in the critical test probability.

MANY studies have demonstrated that individual loci affecting quantitative traits (QTL) can be detected via linkage to genetic markers (Sax 1923; Lander and Botstein 1989; Georges *et al.* 1995; Lander and Kruglyak 1995; Weller 1996). Traditionally, application of these techniques was limited by the paucity of segregating genetic markers in most species of economic interest. In the last decade this problem has been solved by the development of new classes of highly polymorphic DNA-level genetic markers, such as microsatellites. Moderate- to high-density genetic maps have been developed for a wide range of species (Lander and Kruglyak 1995; Weller 1996), and it is now possible to rapidly scan a complete genome for genes affecting any trait of interest.

Most studies have used either maximum likelihood or regression models to test for segregating QTL (Simpson 1989; Haley and Knott 1992; Martinez and Curnow 1992). Generally, a likelihood ratio or a t or F statistic is computed for each marker or pair of markers on the basis of the null hypothesis that no segregating loci affecting the trait of interest are linked to the genetic markers. Ordinarily, a null hypothesis is rejected when the probability of the test statistic under the null hypothesis is below a predetermined level, generally either 5 or 1%. However, if many markers or chromosomal segments are tested, several null hypotheses will meet these rejection criteria by chance (Lander and Botstein 1989; Lander and Kruglyak 1995). The traditional approach to dealing with multiple comparisons has been to control the "familywise (or experimentwise) error rate" (FWER), instead of controlling the "comparisonwise error rate" (CWER). The FWER is controlled by setting the rejection threshold sufficiently strictly so that the probability that *any* of the null hypotheses tested are erroneously rejected is below a specified low level, usually 0.05 (Lander and Botstein 1989; Lander and Kruglyak 1995). A demanding threshold for hypothesis rejection will result in the "acceptance" of many false null hypotheses, and many true effects will not be detected. As noted previously, determination of the appropriate rejection threshold is a choice "between Scylla and Charybdis" (Lander and Kruglyak 1995).

Among multiple comparison situations, determination of rejection thresholds for QTL detection is especially problematic. First, a putative QTL could lie anywhere along the genetic map. Thus, the possible number of "sites" (hypotheses) tested is virtually unlimited. On the other hand, all positions along a single chromosome are genetically linked and therefore not statistically independent. Second, several correlated traits are often analyzed using the same genotype information (Weller *et al.* 1997). Should multiple traits be analyzed jointly or separately? And how should the critical values for rejection be determined? Third, a number of pedigrees are often included in the analysis. Should the individual pedigrees be analyzed jointly or sepa-

Corresponding author: J. I. Weller, Institute of Animal Sciences, A.R.O., The Volcani Center, P.O. Box 6, Bet Dagan 50250, Israel. E-mail: weller@agri.huji.ac.il

rately? The number of tests is reduced, and sample size is increased if all pedigrees are analyzed jointly. Thus, a joint analysis should increase statistical power based on controlling the FWER (Weller *et al.* 1990). However, if QTL are homozygous in some pedigrees, then power may be reduced by joint analysis. Finally, because there is genetic variance for the traits under consideration, the null hypothesis of no QTL segregating throughout the genome is not realistic (Southey and Fernando 1998).

Two approaches have been suggested to control the FWER specifically for detection of QTL. Lander and Botstein (1989) developed analytical formula for two specific situations: a "dense" map and a "sparse" map. In the latter case, it is assumed that the genetic markers are spaced far enough apart so that they can be considered to be independently assorted, while in the former case they are sufficiently close so that all "sites" along the chromosome are being tested for segregating QTL. For a dense map scan of the human genome, a comparisonwise probability of 2×10^{-5} is required to obtain an FWER of 0.05!

Churchill and Doerge (1994) proposed empirically estimating FWER rejection thresholds by generating many different samples from the actual data by "shuffling" the trait values with respect to the marker genotypes. Because the trait value for each individual is now random with respect to marker genotypes, the null hypothesis of no linkage between the genetic markers and QTL is correct by design. The appropriate rejection threshold for any desired type I error is then derived from the empirical distribution of the test statistic. This method has the advantage of making no assumptions with respect to distributional properties of either the quantitative traits or the genetic markers. Rejection thresholds are computed on the basis of the actual number and genomic distribution of markers genotyped.

Contrary to Churchill and Doerge (1994), Lander and Kruglyak (1995) maintain that whole genome thresholds should be used even if only a limited number of markers have been genotyped. Because of the severity of whole genome FWER thresholds, Lander and Kruglyak (1995) propose several classifications for evaluating results. *Suggestive linkage,* the least restrictive criterion, is defined as "statistical evidence that would be expected to occur one time at random in a genome scan." This criterion is problematic for two reasons. First, if only a single null hypothesis is rejected by the criterion of suggestive linkage, there is no reason to assume that a QTL has been located. One event of suggestive linkage is expected by chance! Second, the null hypothesis of no segregating QTL throughout the genome is not realistic, as noted above. However, if *many* hypotheses are rejected by this criterion, then some, but not all, should be due to segregating QTL. Although the prevalence of actual segregating QTL is of paramount importance, the criteria proposed by Lander and Kruglyak (1995) ignore this question.

Clearly, with whole genome scans it is nearly impossible to obtain definitive results from a single experiment. Thus, several studies have suggested that only those effects that display statistical significance on two independent trials should be considered as "confirmed" (Ron *et al.* 1994; Lander and Kruglyak 1995). Therefore, the primary objective of any preliminary genome scan is to determine which chromosomal regions should be resampled.

Benjamini and Hochberg (1995) proposed controlling the "false discovery rate" (FDR) as an alternative to controlling the FWER for the general problem of multiple testing. They defined the FDR as: "The expected proportion of true null hypotheses within the class of rejected null hypotheses." We will demonstrate that controlling the FDR is more appropriate than controlling the FWER for multiple marker QTL detection. Derivation of rejection thresholds, based on controlling the FDR, and important properties of this method will be described. Examples from an actual QTL detection experiment that illustrate the advantages of controlling have been presented previously (Weller *et al.* 1997, 1998). In the current study we have extended the analyses based on actual data and also present results from application of the method to simulated data sets.

MATERIALS AND METHODS

Definition and properties of the false discovery rate: Assume that m, multiple comparisons, are tested. For each null hypothesis; $H_1, H_2, \ldots, H_m$, a test statistic and the corresponding P-values, $P_1, P_2, \ldots, P_m$, are computed. Let $P_{(1)} \leq P_{(2)} \leq \ldots \leq P_{(m)}$ be the ordered P-values and denote by $H_{(i)}$ the null hypothesis corresponding to $P_{(i)}$. If all null hypotheses are true, but k hypotheses, $H_{(1)}$ to $H_{(k)}$, are rejected, then the expectation of the number of hypotheses rejected should be approximately equal to the actual number of hypotheses rejected for any value of k. If, in fact, some of the null hypotheses are false, then the expectation of the number of hypotheses rejected should be $<k$. That is, the number of hypotheses rejected should be greater than the expectation under the null hypothesis. The expectation of the number of hypotheses rejected, assuming that all null hypotheses are true, is $mP_{(k)}$. Defining $q = mP_{(i)}/i$, Benjamini and Hochberg (1995) prove that the FDR can be controlled at some level, $q*$, by determining the largest i for which $q* \leq mP_{(i)}/i$. That is, out of k hypotheses rejected, it is expected that the proportion of erroneously rejected hypotheses is no greater than $q*$. Illustrative examples and important properties of the FDR will be considered below.

Analysis of actual data: In the "granddaughter design" sons of sires heterozygous for genetic markers are genotyped, and records of the granddaughters are compared to detect a segregating QTL (Weller *et. al.* 1990; Ron *et al.* 1994; Georges *et al.* 1995). A total of 1555 sons of 18 U.S. grandsires were genotyped for 128 microsatellites. There was at least one marker on all 29 autosomes. Each bull genotyped had at least 10 daughters with milk records. The dependent variable was the daughter-yield deviation of each sire, which is a weighted mean of the daughter records corrected for fixed effects (VanRaden and Wiggans 1991). A linear model, including the

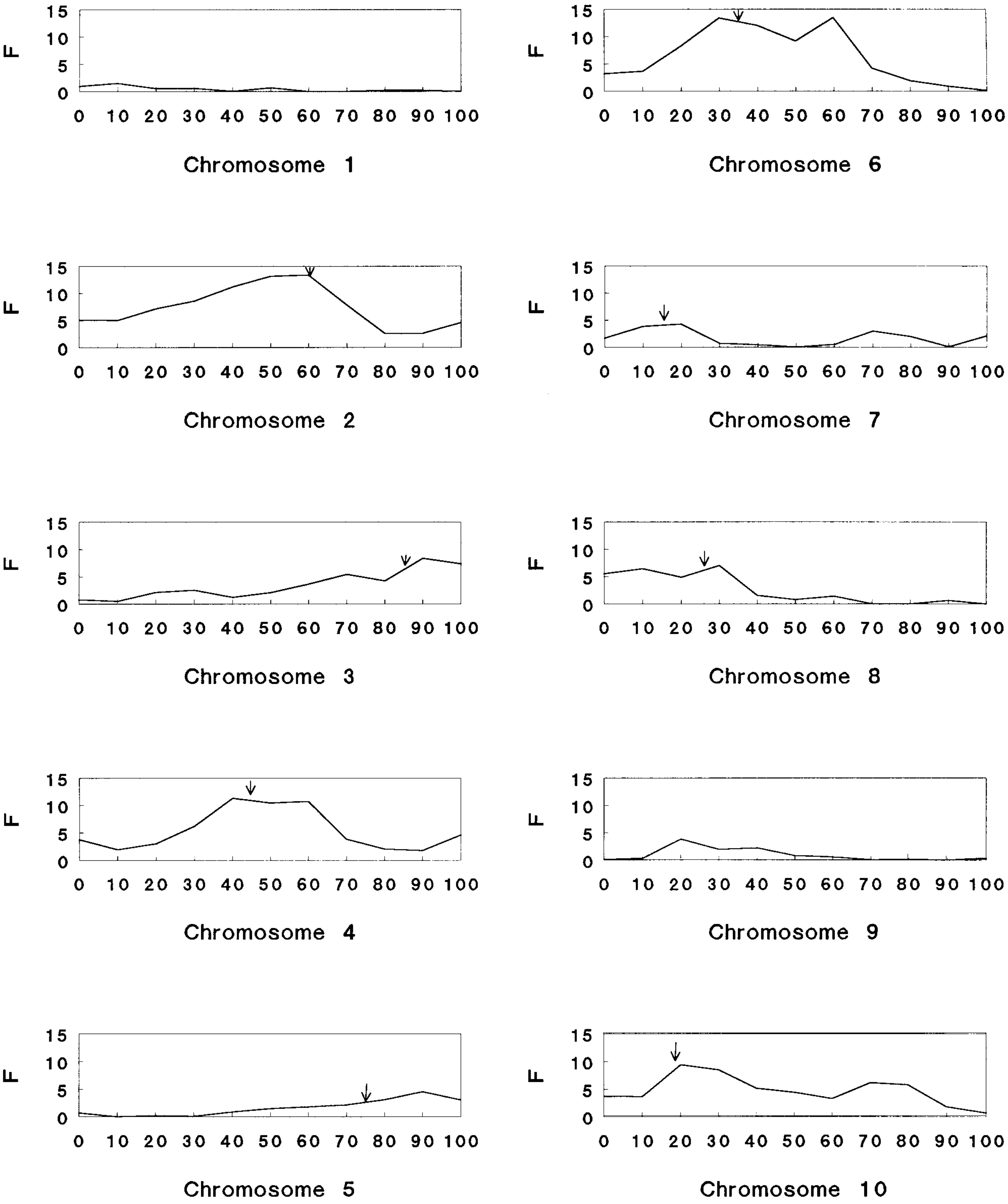

Figure 1.—*F*-values for a simulated genome scan of backcross population. The simulated genome consisted of 10 chromosomes, each of 100 cM. Locations of the eight simulated QTL are denoted by arrows.

effect of the sire and paternal allele within sire, was used to analyze each marker for seven quantitative traits: milk, fat, and protein production, fat and protein percentage, milk somatic cell score, and herd life. The total number of analyses was 7 traits × 128 markers = 896. A significant *F*-value for the paternal allele effect nested within a grandsire is indicative of a segregating QTL affecting the trait in question linked to the marker (Weller *et. al.* 1990).

Even if a particular *F*-value was significant over all families, not all grandsires need be heterozygous for the linked QTL. Individual within family contrasts were also tested by *t*-test on a subset of 26 markers for those families with grandsires

TABLE 1

Locations and effects of the four QTL in simulated backcross populations

Chromosome	Location (cM)	Effect
3	60	0.32
4	87	0.26
5	51	0.32
6	17	0.29

A backcross population of 500 individuals with 10 chromosomes, each of length 100 cM, was simulated. Eleven equally spaced markers were genotyped on each chromosome. Four loci affecting the quantitative trait were simulated at random throughout the genome. An effect of between 0.2 and 0.4 residual standard deviations was simulated for each QTL. No QTL were simulated on the remaining 6 chromosomes. FDR results are presented in Figures 8, 9, and 10.

heterozygous for each marker. A total of 1309 paternal allele contrasts were included in this analysis. Thus, in addition to multiple QTL sites along the genome there are the additional levels of multiple traits and multiple families.

Both the *F*- and *t*-test analyses were also performed on permuted data (Churchill and Doerge 1994). Within each family, the seven trait values were permuted against the vector of genotypes. Therefore, the correlations among traits, the grandsire effects, and any linkages among markers were maintained, while relationships among markers and the trait values were randomized. The FDR was computed as described above. The CWER was assumed equal to $P_{(i)}$, the probability of the test statistic under the null hypotheses for comparison *i*. The FWER appropriate for each value of $P_{(i)}$, was computed as the probability of rejecting at least one hypothesis, assuming a Poisson distribution for the number of rejected hypotheses with an expectation of $mP_{(i)}$, where *m* is the total number of tests.

Analysis of simulated data: Simulated data were generated for a cross between two inbred lines assumed to differ in both genetic markers and QTL alleles. All markers and QTL were assumed codominant. Each parental line was assumed homozygous for all loci. The F_1, which is heterozygous for segregating QTL and markers, was backcrossed to one of the parental lines to produce the BC_1 generation. The simulated genetic map consisted of 10 chromosomes each with a length of 100 cM. The BC individuals were genotyped for 11 evenly spaced markers on each chromosome, with one marker at each end of each chromosome and 10 cM between markers. Zero recombination interference was assumed, and recombination probabilities between markers and between markers and QTL were computed using the Haldane (1919) mapping function. Each chromosome was considered to consist of 100 "sites" with a probability of slightly more than 1% recombination between sites, as determined by the Haldane mapping function.

Zero or one segregating QTL was simulated per chromosome. Locations of segregating QTL were determined by sampling from a uniform distribution corresponding to the chromosome length. QTL substitution effects were determined by sampling from a uniform distribution between 0 and 0.4. If a value <0.2 residual standard deviations was obtained, then the substitution effect was set at 0. Thus, the expectation was for five segregating QTL per genome. The trait value for each BC_1 individual was computed as the sum of its QTL effects and a random residual computed by sampling from a normal distribution with a variance of unity. Presence of segregating QTL was tested by a one-way ANOVA for each of the 110 markers in the genome. Because markers were linked, these tests were not independent. Results are presented for two typical simulated populations of 500 individuals each, with eight and two segregating QTL, respectively. The distribution of *F*-values by chromosome for the population with eight QTL is given in Figure 1.

Five simulated populations with four segregating QTL given in Table 1 were also analyzed. In addition, five populations simulated with two QTL set to zero and five populations simulated with all four QTL set to zero were also analyzed. The 110 tests of each set of five populations were ordered by $P_{(i)}$. For each *i* the means of the five simulations were computed for $P_{(i)}$, the FWER, *q*, and the true FDR (TFDR), computed as the number of tests among all tests with probability $<P_{(i)}$ not genetically linked to a simulated QTL divided by *i*.

TABLE 2

Computation of the false discovery rate for granddaughter design analyses of variances across families

i	Traits	Chromosome	Marker	*F*	$P_{(i)}$	$E_{(i)}$	FWER	*q*
1	Fat %	14	15	7.157	10^{-8}	10^{-5}	10^{-5}	10^{-5}
2	Fat %	3	1	5.295	0.00003	0.025	0.024	0.012
3	Fat yield	14	15	4.146	0.00009	0.077	0.074	0.026
4	Protein %	2	4	5.279	0.00042	0.378	0.315	0.094
5	Protein %	3	8	4.246	0.00091	0.818	0.559	0.163
6	SCS	22	1	3.819	0.00101	0.907	0.596	0.151
7	SCS	22	2	4.590	0.00124	1.112	0.671	0.159
8	Fat %	3	8	3.880	0.00194	1.734	0.823	0.217
9	Milk yield	7	3	3.466	0.00231	2.068	0.874	0.230
10	SCS	23	1	4.218	0.00242	2.166	0.885	0.217

Seven milk production traits were analyzed for 128 markers in the U.S. Holstein population. *i*, number of null hypothesis ranked by descending *P*; SCS, somatic cell score; Chromosome, bovine chromosome number; marker, anonymous designation of DNA-microsatellite; *F*, *F*-value for the effect of paternal marker allele nested within grandsire; $P_{(i)}$, probability of corresponding *F*-value; $E_{(i)}$, expected number of rejected true null hypotheses with rejection probability $P_{(i)}$, computed as $896P_{(i)}$; FWER, experimentwise type I error rate, computed as the probability of rejecting at least one hypothesis assuming a Poisson distribution for the number of rejected hypotheses with an expectation of $E_{(i)}$; *q*, expected false discovery rate when rejecting *i* hypotheses $= 896P_{(i)}/i$.

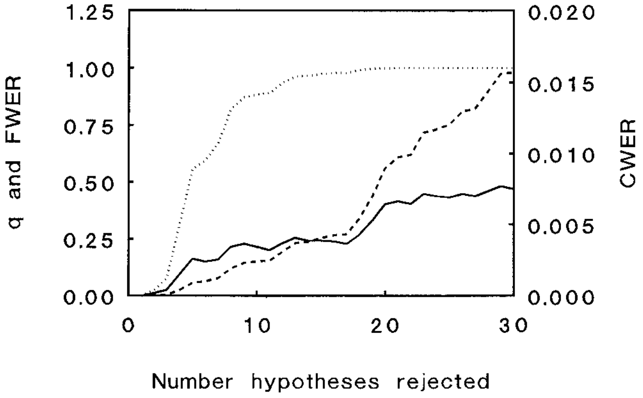

Figure 2.—The false discovery rate (q) (—), experimentwise type I error rate (FWER) (· · ·), and comparisonwise type I error rate (CWER) (- - -) for the F-values computed from the across-family analysis of the actual granddaughter design data for each of 896 marker-trait combinations.

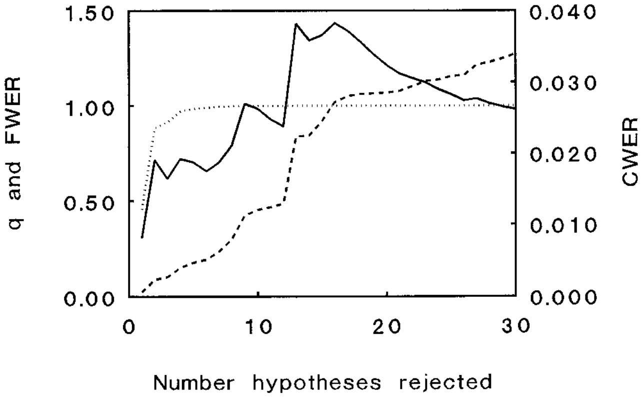

Figure 3.—The false discovery rate (q) (—), experimentwise type I error rate (FWER) (· · ·), and comparisonwise type I error rate (CWER) (- - -) for the F-values computed from the across-family analysis of the granddaughter design data from a typical random permutation of trait values within families.

RESULTS

The F-values and their corresponding probabilities from the granddaughter design analysis with the 10 smallest probabilities out of the 896 marker-trait association tests computed are given in Table 2. The expected number of rejected hypotheses assuming no segregating QTL, the FWER, and q values are also presented. Assuming uncorrelated tests, only two F-values have a FWER <0.05. Using the criterion of "suggestive linkage" (FWER $<$ 0.5 for a complete genome scan) only four null hypotheses would be rejected.

If all 10 hypotheses are rejected, q, and thus FDR, are still <0.25, even though FWER $=$ 0.88. Thus, seven or eight marker-trait combinations should represent "true" effects and can be expected to repeat on a second population sample. Unlike FWER, $q = mP_{(i)}/i$ is not monotonic. For example, as i increases from 5 to 6 and from 9 to 10, *q decreases*. A decrease in q occurs when the increase in successive probabilities is low.

Results for q, FWER, and CWER, computed as the individual F probabilities up to $i = 30$, are plotted in Figure 2. For $i > 20$, $q > 0.4$. For $i = 30$, CWER is still <0.02. Thus, in this case, the criterion of controlling the FDR at 0.5 and CWER at 0.02 give similar results.

Results computed from a typical permutation of the genotype data against the trait data are plotted in Figure 3. Because the relationships among the markers and the traits have been randomized, no null hypotheses should be rejected. For the lowest F probability, FWER was 0.45, and q was 0.31. Thus, 1 hypothesis would be rejected with FWER controlled at 0.5, but not with FDR controlled at any reasonable level. For i values >5, the FWER is nearly equal to unity. By theory, the expectation of q is unity for all values of i, but this criterion is much more affected by random fluctuation than FWER. q is nearly equal to unity for $i = 9$, but then rises to nearly 1.5 before settling down to close to unity by $i = 30$. With $i = 9$, CWER is still 0.01, which is almost exactly the expectation by chance (0.01 × 896 comparisons). Thus, by the criterion of CWER $<$ 0.01, 9 hypotheses would be rejected, as compared to 17 for the actual data (Figure 2), this illustrates how unreliable the CWER criterion is.

Results from separate analysis of all within-family contrasts are plotted in Figure 4 for $i \leq 30$. The number of comparisons was 1309. Only two hypotheses would be rejected by the criterion of FWER $<$ 0.5. There is a peak for q with $i = 5$. If only five hypotheses are rejected, $q = 0.69$. However, for $i = 11$, $q = 0.47$. As in the previous example, FDR can be *decreased* by rejecting more hypotheses. At $i = 30$, $P = 0.025$, which is very close to the expected level of 0.022 (30/1309).

Results for within-family comparisons from a typical permutation of the data are plotted in Figure 5. As expected, the FWER is close to 0.5 for the lowest t-test probability and increases to nearly unity for $i > 2$. q is

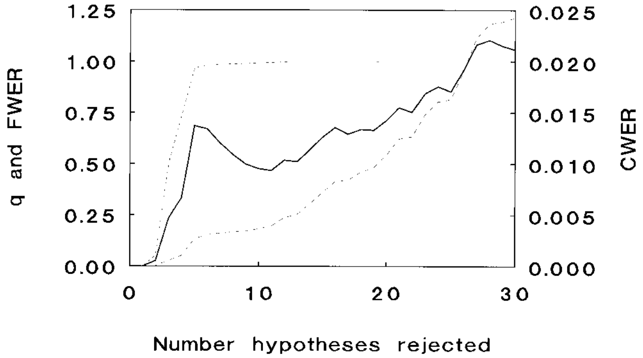

Figure 4.—The false discovery rate (q) (—), experimentwise type I error rate (FWER) (· · ·), and comparisonwise type I error rate (CWER) (- - -) for the t-values computed from the within-family analysis of the actual granddaughter design data for each of 1309 family-marker-trait combinations.

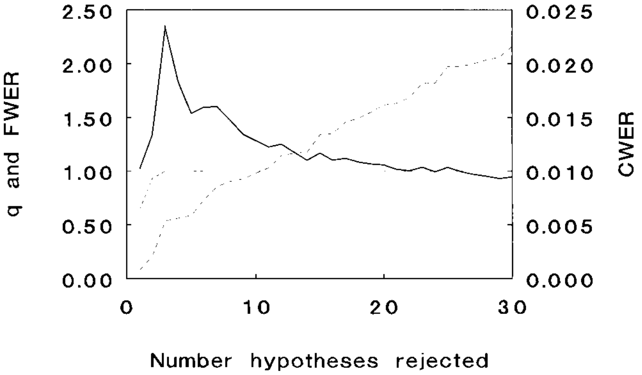

Figure 5.—The false discovery rate (q) (—), experimentwise type I error rate (FWER) (· · ·), and comparisonwise type I error rate (CWER) (- - -) for the t-values computed from the within-family analysis of the granddaughter design data from a typical random permutation of trait values within families.

nearly equal to unity for $i = 1$, increases rapidly, but then settles close to unity for $i > 15$. As in the previous case, this is probably due to random fluctuation. The graph of the CWER with permuted data is very similar to the plot in Figure 4 with the actual data. This example again demonstrates that controlling the CWER is virtually meaningless if many hypotheses are tested.

The FDR, FWER, and CWER computed from the F probabilities from the backcross population simulated with eight QTL are plotted in Figure 6 for i up to 50. Although only eight QTL were simulated, many more markers gave relatively high test values because of linkage. The expectation for the effect associated with a given marker as estimated by a linear model is $1 - 2r$ of the simulated effect, where r is the recombination frequency between the marker and the simulated QTL (Weller *et al.* 1990). At $i = 50$, FDR = 0.18, FWER = 0.999, and CWER = 0.08. However, with $i = 16$, FWER = 0.44. For all these 16 tests, the expectation of the estimated effect was >0.13 standard deviations and these can therefore be considered "true" effects. With $i = 50$, the expectation was <0.1 standard deviations in only five cases, or 10%. As noted above, all QTL simulated had effects >0.2. Thus, q controlled the FDR as predicted, while many true effects would have been missed by the "weak" criterion of FWER < 0.5.

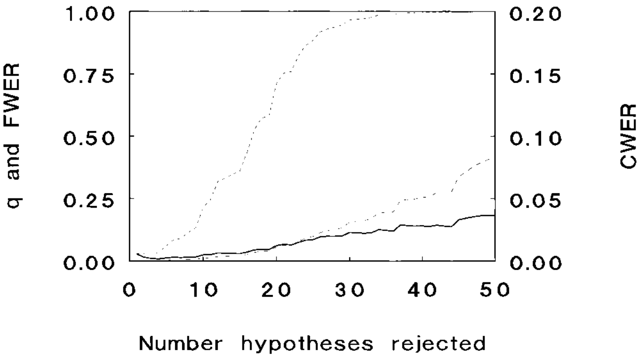

Figure 6.—The false discovery rate (q) (—), experimentwise type I error rate (FWER) (· · ·), and comparisonwise type I error rate (CWER) (- - -) for a simulated whole genome scan of a backcross population. Eight segregating QTL were simulated.

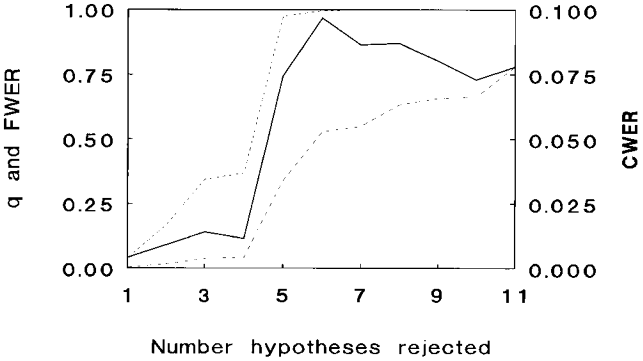

Figure 7.—The false discovery rate (q) (—), experimentwise type I error rate (FWER) (· · ·), and comparisonwise type I error rate (CWER) (- - -) for a simulated whole genome scan of a backcross population. Two segregating QTL were simulated.

Results for the whole genome scan with only two simulated effects is plotted in Figure 7. The four markers with the lowest probabilities were all linked to the QTL with a simulated effect of 0.31 standard deviations located on chromosome *6*. Expectations of the effects associated with these markers were all >0.18 standard deviations. The six next lowest probabilities were all associated with markers located on chromosomes without simulated QTL. The second QTL simulated on chromosome *9* with an effect of 0.24 standard deviations was not detected. For $i = 4$, $q = 0.11$, FWER = 0.37, and CWER = 0.004. At $i = 5$, $q = 0.74$, FWER = 0.98, and CWER = 0.034. Again the FDR is controlled by the q statistic, while spurious effects are detected by controlling CWER. With few true effects, FWER and q are similar, and controlling FWER or FDR yields similar results.

Results are presented in Figures 8, 9, and 10 for i up to 30 for the means of CWER, FWER, q, and TFDR from five runs simulated with four, two, and zero segregating QTL, respectively. Because these curves are the means of five runs, the q values and FDR are nearly monotonic, except for q with no QTL simulated. As predicted, TFDR is $<q$ except for a few low values of i in Figure 10. Thus q was able to control the actual FDR, in the sense that TFDR $\leq q$. However, the expectations for some of the "true" effects were relatively small, because of incomplete linkage, and would therefore have relatively low power of detection on a repeat sample. As expected with repeat random samples, the probability rankings for the individual markers were different among runs. With four and two QTL simulated, three and one null hypotheses, respectively, would have been rejected by

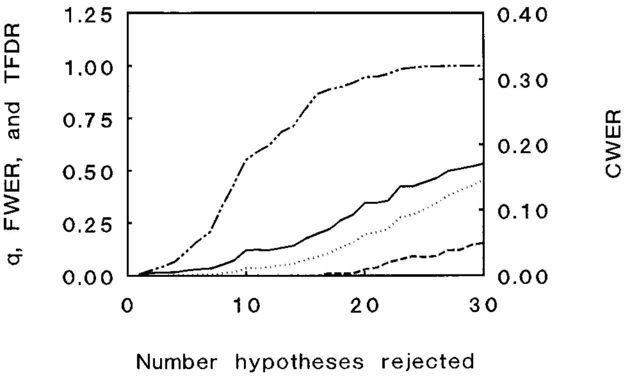

Figure 8.—The mean estimated false discovery rate (*q*) (—), experimentwise type I error rate (FWER) (— · · —),the comparisonwise type I error rate (CWER) (· · ·), and the true FDR (TFDR) (- - -) for whole genome scans of five backcross populations with four segregating QTL simulated as shown in Table 1. TFDR, the number of tests among the set of rejected hypotheses for which no linked QTL was simulated, divided by *i*.

the criteria of FWER < 0.05. Thus, not all of the four or two QTL simulated would have been detected by the criterion of FWER < 0.05. With four and two QTL, the *q* values reached 0.5 for $i = 27$ and $i = 25$, respectively. For these *i* values the TFDR values were 0.12 and 0.34, respectively. With no QTL simulated (Figure 10) TFDR $= 1$ for all *i* values, and *q* was >1 except for a few low values of *i*. For $i = 6$, CWER $= 0.05$, which was very close to the expected value of $6/110 = 0.055$. For $i > 7$ the FWER and TFDR curves overlap.

DISCUSSION

Although theoretically an infinite number of chromosomal sites can be tested between linked markers, power with interval mapping is not greater than power obtained by testing each marker individually (Simpson 1992; Darvasi *et al.* 1993). Thus, the number of hypotheses tested can be considered a direct function of the number of markers genotyped.

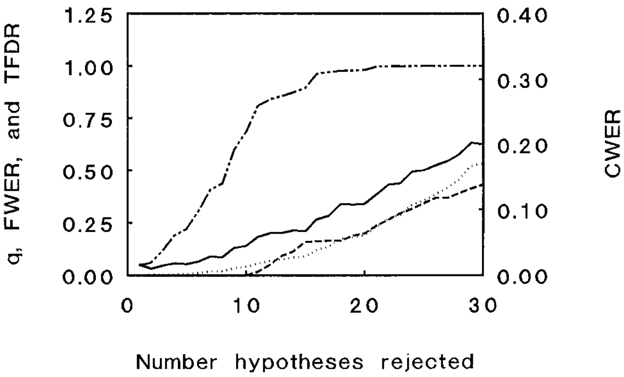

Figure 9.—The mean estimated false discovery rate (*q*) (—), experimentwise type I error rate (FWER) (— · · —), the comparisonwise type I error rate (CWER) (· · ·), and the true FDR (TFDR) (- - -) for whole genome scans of five backcross populations with two segregating QTL simulated.

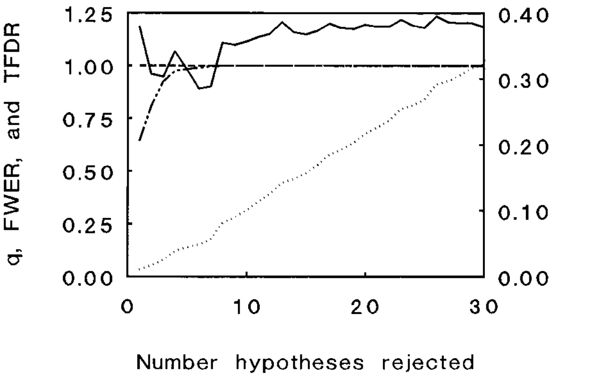

Figure 10.—The mean estimated false discovery rate (*q*) (—), experimentwise type I error rate (FWER) (— · · —), the comparisonwise type I error rate (CWER) (· · ·), and the true FDR (TFDR) (- - -) for whole genome scans of five backcross populations with no segregating QTL simulated.

The examples presented demonstrate the following important properties of the FDR.

1. If all null hypotheses are true, controlling FDR is equivalent to controlling FWER (Benjamini and Hochberg 1995).
2. If some of the null hypotheses are false, then the FDR is smaller than the FWER. The difference between the two criteria increases with increase in the number of false null hypotheses. Thus, any procedure that controls the FDR at a given level will also control the FWER at this level.
3. Unlike FWER, FDR can be controlled without the assumption that relationships among the test statistics are known. As demonstrated, the FDR can be readily controlled both for multiple linked markers and linked traits.
4. Even though $P_{(i)}$ increases monotonically with *i*, *q* does not. Thus, it may be necessary sometimes to *increase i* to control the FDR at the desired level.
5. If there are QTL segregating, then the true FDR $<$ *q*. The discrepancy increases with increase in the genomic density of segregating QTL.
6. Through control of the FDR, the number of hypotheses rejected, that is, the QTL detected, is a function of the actual number of segregating QTL in the population, but not if either the FWER or CWER are controlled.
7. The dilemma of the appropriate rejection criterion for a partial genome scan is solved. The FDR can be controlled at the same level whether the complete genome or only part of the genome has been analyzed.
8. Additional levels of contrasts, such as multiple traits or multiple populations, can be handled without the necessity of a proportional increase in the critical test value.

Thus, controlling the FDR has several advantages over controlling either the FWER or CWER for preliminary detection of segregating QTL. Controlling the FDR is recommended primarily for a preliminary genomic scan. Methods to optimize these scans with respect to marker spacing have been described previously (Darvasi and Soller 1994). A second, independent experiment will be required to determine which hypotheses tentatively rejected by the first analyses represent actual segregating QTL. A further advantage of the FDR is that an accurate prediction has been made of the proportion of hypotheses rejected in the first analyses that represent true effects. Weaknesses of the FDR are that it tends to fluctuate widely for low *i* if the total number of hypotheses tested is very large and that with many true effects *q* overestimates TFDR. Further study is required to determine whether the information derived from the preliminary analysis can be used to augment probabilities obtained in the confirmation study.

Other methods for determining rejection rates with multiple testing, most based on Bonferroni-type procedures, are summarized by Benjamini and Hochberg (1995). They also considered the advantages and disadvantages of these methods. Benjamini and Hochberg (1995) proposed to apply the FDR after data collection. Recently, Southey and Fernando (1998) developed a method based on the same principles to estimate a rejection threshold prior to data collection. Thus, they require a well-defined alternative hypothesis with a known prior probability and must set the required power. Theoretically, their method can be used to decide how large an experiment is required to obtain a given power. However, prior information on QTL effects is generally very vague.

This study was supported by the U.S.-Israel Binational Agricultural Research and Development fund (BARD) and the Israel Milk Marketing Board. We thank G. R. Wiggans, P. M. VanRaden, and Yang Da for providing data used in this study.

LITERATURE CITED

Benjamini, Y., and Y. Hochberg, 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Statist. Soc. B. **57:** 289–300.

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

Darvasi, A., and M. Soller, 1994 Optimum spacing of genetic markers for determining linkage between marker loci and quantitative trait loci. Theor. Appl. Genet. **89:** 351–357.

Darvasi, A., A. Vinreb, V. Minke, J. I. Weller and M. Soller, 1993 Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. Genetics **134:** 943–951.

Georges, M., D. Nielsen, M. Mackinnon, A. Mishra, R. Okimoto *et al.*, 1995 Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. Genetics **139:** 907–920.

Haldane, J. B. S., 1919 The combination of linkage values, and the calculation of distances between the 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.

Lander, E. S., and D. Botstein, 1989 Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics **121:** 185–190.

Lander, E., and L. Kruglyak, 1995 Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nature Genet. **11:** 241–247.

Martinez, O., and R. N. Curnow, 1992 Estimating the locations and the size of the effects of quantitative trait loci using flanking markers. Theor. Appl. Genet. **85:** 480–488.

Ron, M., M. Band, A. Yanai and J. I. Weller, 1994 Mapping quantitative trait loci with DNA microsatellites in a commercial dairy cattle population. Anim. Genet. **25:** 259–264.

Sax, K., 1923 The association of size differences with seed-coat pattern and pigmentation in *Phaseeolus vulgaris*. Genetics **8:** 552–560.

Simpson, S. P., 1989 Detection of linkage between quantitative trait loci and restriction fragment length polymorphisms using inbred lines. Theor. Appl. Genet. **77:** 815–819.

Simpson, S. P., 1992 Correction: detection of linkage between quantitative trait loci and restriction fragment length polymorphisms using inbred lines. Theor. Appl. Genet. **85:** 110.

Southey, B. R., and R. L. Fernando, 1998 Controlling the proportion of false positives among significant results in QTL detection. Proc. 6th World Cong. Genet. Appl. Livest. Prod. **26:** 221–224.

VanRaden, P. M., and G. R. Wiggans, 1991 Derivation, calculation, and use of national animal model information. J. Dairy Sci. **74:** 2737–2746.

Weller, J. I., 1996 Introduction to QTL detection and marker-assisted selection, pp. 259–275 in *Beltsville Symp. Agric. Res. XX: Biotechnology's Role in the Genetic Improvement of Farm Animals*. American Society of Animal Science, Savoy, IL.

Weller, J. I., Y. Kashi and M. Soller, 1990 Power of "daughter" and "granddaughter" designs for genetic mapping of quantitative traits in dairy cattle using genetic markers. J. Dairy Sci. **73:** 2525–2537.

Weller, J. I., J. Z. Song, Y. I. Ronin and A. B. Korol, 1997 Experimental designs and solutions to multiple trait comparisons. Anim. Biotechnol. **8:** 107–122.

Weller, J. I., J. Z. Song, D. W. Heyen, H. A. Lewin and M. Ron, 1998 A new approach to the problem of multiple comparisons for detection of quantitative trait loci. Proc. 6th World Cong. Genet. Appl. Livest. Prod. **26:** 229–232.

Communicating editor: C. Haley