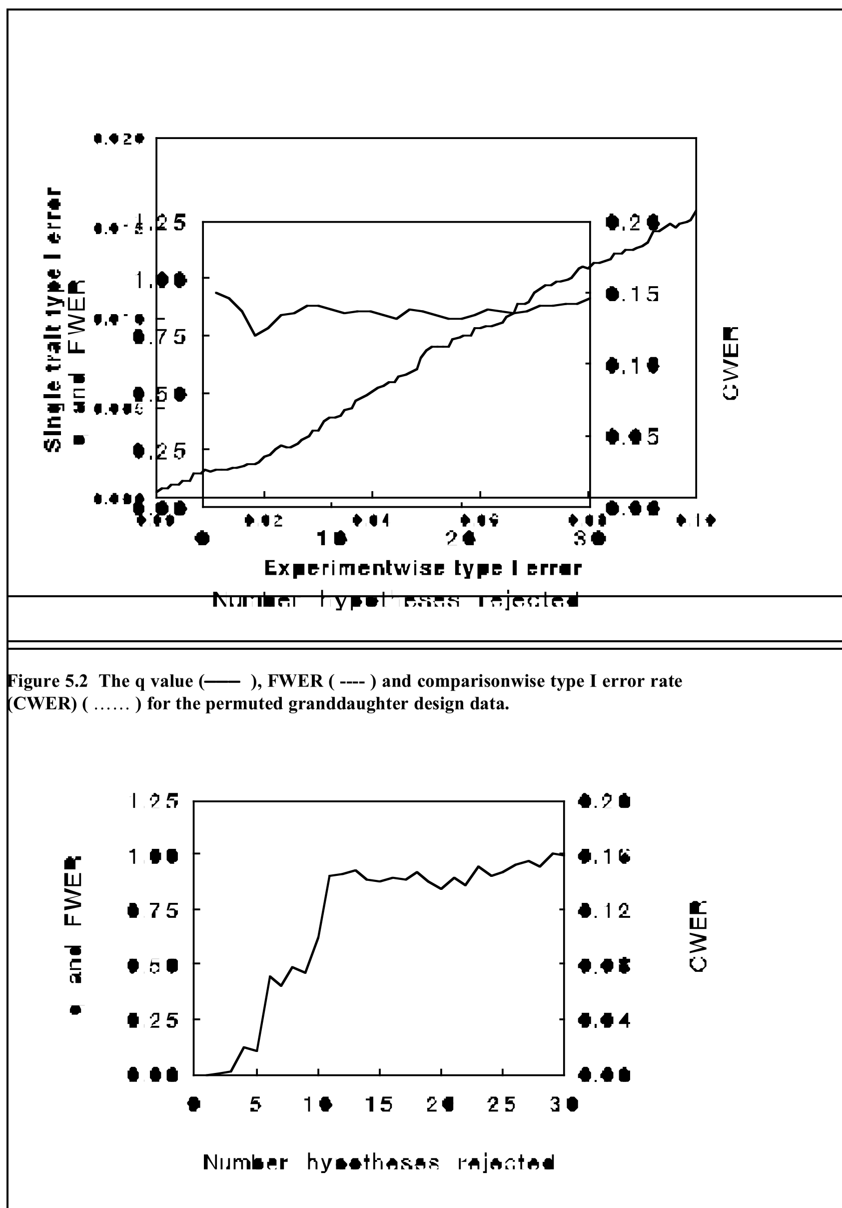

Figure 5.2 The q value (——), FWER (----) and comparisonwise type I error rate (CWER) (……) for the permuted granddaughter design data.

Figure 5.1 The q value (—), FWER (- - -) and comparisonwise type I error rate (CWER) (……) for analysis of the GD data.

Chapter 5. Multiple markers and multiple traits

5.1 Introduction

As we already noted in Chapter 1, it is now possible using DNA-level markers to obtain as many polymorphic markers as desired for any species of interest. Thus, it is now possible to conduct complete genome scans for QTL affecting any trait of interest. In the previous chapter we discussed power to detect segregating QTL, based on the assumption that each marker or marker bracket was tested separately. If the number of markers included in the analysis is large, the individual test type I error rate is no longer appropriate. For example, if 100 tests are performed, five should be "significant" at the 5% level purely by chance. The traditional approach to deal 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 strict so that the probability that *any* of the null hypotheses tested are erroneously rejected is below a specified low level, usually 0.05. For uncorrelated hypotheses, the FWER can be readily computed by the "Bonferroni adjustment". However, linked markers are correlated.

Although the vast majority of studies have considered multiple traits, nearly all studies have analyzed each trait separately. Only a few studies have considered the theoretical aspects of multiple trait QTL analysis (Korol et al., 1987, 1995, Weller et al., 1996). In this chapter, we will first consider the specific theoretical problems related to multiple trait analysis. We will then describe the methods that have been proposed to deal with these problems.

A third level of multiple comparisons in addition to multiple markers and traits, that will be considered is multiple pedigrees. For example in daughter and granddaughter designs should each family be analyzed separately, or should data be analyzed jointly over all families, even though the QTL are segregating in only some of the pedigrees.

5.2 Multiple markers and whole genome scans

The problem of multiple markers was first considered in detail by Lander and Botstein (1989). They presented analytical formula for two specific situations; a "sparse" map, and a "dense" map. In the former they assumed that the markers were sufficiently far apart that the individual tests could be considered independent. In this case the FWER can be computed by the "Bonferroni adjustment" (Simes, 1986) as follows:

$$\alpha = 1 - (1 - p)^n \qquad \{5.1\}$$

where α = FWER, p = comparisonwise type I error (CWER), and n = number of markers. For small α, p is approximately equal to α/n, the formula presented by Lander and Botstein (1989). In the latter case they assume that markers a sufficiently close so that all "sites"

along the chromosome are being tested for segregating QTL. In this case, the expected number of regions with a t value greater than the critical value for p, $\mu(T)$, can be computed as follows for either the BC or F-2 designs (Lander and Kruglyak, 1995):

$$\mu(T) = [C + 2GT^2]p \qquad \{5.2\}$$

Where C = number of chromosomes, G = genome length in Morgans, and T = the standard normal distribution value for p. For small values of $\mu(T)$, $\mu(T)$ tends to α. Lander and Botstein (1989) present a similar formula for likelihood ratio tests. For the full-sib design, it is necessary to multiply the term $2GT^2$ by a factor of 2. For a dense map scan of the bovine genome by the daughter design, a comparisonwise probability of approximately 4×10^{-5}, comparable to a t-value > 4, is required to obtain an FWER of 0.05. Requirement of such a stringent type I error results in a corresponding increase in the type II error. The alternative is replication of significant results on an independent sample. To deal with the problem of appropriate thresholds for declaration of significance, Lander and Kruglyak (1995) propose the following criteria:

1. Suggestive linkage - obtaining a test statistic with the CWER corresponding to $\mu(T) = 1$, or the expectation that a test statistic of this magnitude should occur no more that once by chance in a complete genome scan.

2. Significant linkage - obtaining a test statistic with a CWER required for $\mu(T) < 0.05$.

3. Highly significant linkage - obtaining $\mu(T) < 0.001$.

4. Confirmed linkage - significant linkage confirmed by obtaining $p<0.01$ on a second, independent study.

Lander and Kruglyak (1995) also propose that, unless there is a reason to focus *a priori* on a specific chromosomal region, type I errors should be based on complete genome scans, even if the number of markers actually analyzed was limited. They argue that even if the original marker spacing is quite wide, additional markers will be genotyped for those regions that display marginal significance. Thus, the whole genome is potentially under observation. Two other methods that provide alternative solutions to this problem will now be considered.

5.3 QTL detection by permutation tests

Churchill and Doerge (1994) proposed a method to empirically estimate FWER rejection thresholds that can be applied to a very wide range of experimental designs. Many different samples are generated from the actual data by "shuffling" the trait values with respect to the marker genotypes. Each individual genotyped is randomly assigned one of the trait values from the sample. Since 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 definition. The test statistics computed from these "permutation samples" are then used to construct the empirical distribution of the test statistic under the null hypothesis. The appropriate rejection threshold for any desired comparisonwise or experimentwise type I error can then be derived from the empirical distribution of the test statistic. This method has the advantage that no assumptions are made with respect to distributional properties of either the quantitative traits or the genetic markers. Rejection thresholds are computed based on the actual number and genomic distribution of markers genotyped. A disadvantage of this method is thresholds must be computed anew by permutation for each data set analyzed.

Churchill and Doerge (1994) computed CWER and FWER based on permutation tests for simulated data. This method can also be applied to the problem of multiple traits, considered in detail below. The fact that no assumptions are made with respect to the distribution of the test statistic under the null hypothesis is especially important for computation of the FWER. As demonstrated above, to obtain a reasonable FWER for a complete genome scan, a very small CWER is required. At these very low probabilities it is likely that minuscule divergence of the actual data distribution from the theoretical distribution may result in a significant divergence of the analytically computed probability from the actual probability for the specific data set analyzed. An example will be given below.

5.4 QTL detection based on the false discovery rate

Benjamini and Hochberg (1995) recently 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". Derivation of rejection thresholds based on controlling the FDR, and important properties of this method will be described. We will then present examples based on actual data.

Assume that m multiple comparisons are tested. For each null hypothesis; $H_1, H_2, \cdots, H_m$; a test statistic and the corresponding p-values; $P_1, P_2, \cdots, P_m$; are computed. Let $P_{(1)} \le P_{(2)} \le \cdots \le P_{(m)}$ 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 hypothesis rejected should be less than k. 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: $q^* \ge mP_{(i)}/i$. That is, out of k hypothesis 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 now be considered.

Comparison of FDR and FWER will be illustrated using results from a granddaughter design analysis of the US Holstein population. Sons of 18 US Holstein families were genotyped for 26 genetic markers and analyzed for seven traits. Daughter yield deviations (DYD) were analyzed by the following linear model:

$$Y_{ijk} = GS_i + A_{ij} + e_{ijkl} \qquad \{5.3\}$$

Where Y_{ijk} is the DYD (VanRaden and Wiggans, 1991) for k^{th} son of the ith grandsire with paternal allele j, GS_i is the effect of the ith grandsire, A_{ij} is the effect of the jth marker allele, progeny of the ith grandsire. For each marker-trait combination an F statistic was computed for the paternal marker allele effect nested within grandsire. Thus, 182 comparisons were tested. The comparisons with the ten smallest p-values are given in Table 5.1.

Table 5.1 Estimation of FDR for granddaughter design results.

i	Trait	Marker	F-value	p-value	Exp[1]	FWER	q
1	Fat %	A	6.79849	0.00000	0.0000	0.00000	0.00000
2	Fat	A	4.36708	0.00004	0.0078	0.00781	0.00392
3	Fat %	B	4.43349	0.00023	0.0415	0.04065	0.01383
4	Protein %	B	3.41650	0.00263	0.4786	0.38033	0.11964
5	SCS	C	4.15081	0.00272	0.4942	0.38996	0.09885
6	Milk	D	2.56941	0.01457	2.6517	0.92947	0.44195
7	Milk	B	2.65195	0.01545	2.8118	0.93990	0.40168
8	Fat	E	2.05714	0.02155	3.9213	0.98018	0.49016
9	Protein	D	2.38579	0.02274	4.1385	0.98405	0.45983
10	Herdlife	F	2.10262	0.03453	6.2847	0.99814	0.62847

[1] Expectation for the number of hypothesis rejected under the null hypothesis.

Assuming uncorrelated tests, the FWER was <0.05 for only three marker-trait combinations. Using Lander and Kruglyak's (1995) criteria of "suggestive linkage" (FWER<0.5), five null hypotheses would be rejected. However $q = mp_{(i)}/i$ is < 0.5 for nine marker-trait combinations. Thus, if these nine null hypotheses are rejected, it is expected that more than half will in fact be false. Since the objective at this point is to determine which markers should be repeated on independent samples, a relatively high FDR can be justified. With i=9, FWER=0.98. Thus these two criteria are not similar in this case. Note that unlike FWER, q is not monotonic. Where successive p-values are very close, q decreases with increasing p-values. Results for q, FWER, and p up to i=30 are plotted in Figure 5.1. For i>10, q and FWER are very close, with both close to unity. For i=10, p is still <0.05. Thus in this case, the criteria of controlling the FDR at 0.5 and comparisonwise type-I error of 0.05 give similar results.

These results were compared to the p-values computed from a typical permutation of the same genotype data against the trait data. The permutation results are plotted in Figure 5.2. Since the relationship between the markers and the traits after permutation is random by definition, no null hypotheses should be rejected, and FDR and FWER should be similar (Binyamini and Hochberg, 1995). For the lowest p-value, FWER was 0.38, and q was 0.94. Thus, this null hypotheses would be rejected by the criteria of suggestive linkage, but not with FDR controlled at any reasonable level.

For i values >5 both curves are flat with the FWER nearly equal to unity, and q at about 0.85. By theory, q should be unity, but this criteria is much more affected by random fluctuation than FWER. With i=10, p is still <0.05, which is close to the expected number of 9 (0.05*182 comparisons). Thus, by the criteria of comparisonwise α, 10 hypotheses would be rejected, which is almost the same as for the actual data, and illustrates how unreliable this criteria is. 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.
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 methods for controlling FWER, it is not necessary to assume 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. Although the true FDR <q, as i increases, the FDR approaches q. This will be true even if the hypotheses are correlated.
6. By controlling the FDR, the number of hypotheses rejected, *i. e.* 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 have 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.

Controlling the FDR is recommended primarily for a preliminary genomic scan. 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. A weakness of the FDR is that it tends to fluctuate widely for low i if the total number of hypotheses tested is very large.

5.5 Problems and solutions for multiple trait QTL analyses

The main problems with multiple trait QTL analysis were summarized by Weller et al. (1996), and will be mentioned here with some additions.

1. Most studies have determined statistical significance based on each marker-trait combination. As noted above, increasing the number of traits analyzed increases the probability that some markers will display statistical significance "by chance". This problem becomes more severe as the number of markers and traits increases.
2. If a significant effect is found associated with more than one trait, it is not clear whether several different QTL, or a single locus with correlated effects on several traits has been detected. This will be especially acute if some of the traits are highly correlated.
3. Several techniques have been suggested to increase statistical power per individual genotyped at the expense of individual phenotyped (Darvasi and Soller, 1992, 1994; Lebowitz et al., 1987). As noted in the previous chapter, some of these techniques are trait specific, for example selective genotyping and sample pooling. How will these techniques be affected, and what is the optimum strategy in a multiple trait analysis?

Two main methods have been proposed for multiple trait QTL analysis that alleviate some of the problems considered above. Korol et al. (1995) proposed a maximum likelihood multivariate normal analysis, and applied this method to simulated data sets. They showed that under certain conditions, power of QTL detection is increased as compared to univariate analyses methods. Since separate QTL effects are estimated for each trait, it is possible to determine whether the same locus is affecting both traits. However, with two traits and a single segregating QTL in a backcross population it is necessary to estimate at least eight parameters (a recombination frequency, two means for each trait, a variance for each trait, and a correlation coefficient). As the number of traits increases, the number of parameters that must be estimated increases exponentially.

Weller et al. (1996) proposed a canonical transformation of the original traits in order to derive an uncorrelated set of variables. The QTL analyses are then performed on the uncorrelated canonical variables. QTL effects on the actual traits can then be derived by reverse transformation. The advantages of this method are that any number of traits can be readily analyzed, with only a single trait analysis for each variable. Thus, analysis is relatively simple. Since the canonical variables are by definition uncorrelated, it is possible to compute the FWER as described above. It may also be possible to reduce the total number of traits analyzed, and thus increase the power of detection, by deleting canonical variables with low eigenvalues. Finally, since the canonical variables are uncorrelated, a QTL with correlated effects on two traits should affect only a single canonical variable.

This method was applied to daughter design data for milk, fat, and protein production of Israeli Holsteins (Weller et al., 1996). A significant QTL effect was found associated with milk and protein production, but not fat. Milk and protein production are highly correlated. By a canonical transformation, it was possible to reduce the number of variables from three to two. A significant effect was found associated only with one of the remaining variables, which was highly correlated with both milk and protein production. Thus, we conclude that a single QTL was affecting both traits.

The disadvantages of this method are, first, that infinitely many canonical transformations are possible, and significant effects found for one transformation may not be significant by another transformation. Second, a canonical transformation can be applied only if all traits are recorded on all individuals. Third, if the canonical transformation is applied to the phenotypic trait values, then the residual variances after correction for the QTL effects will no longer be uncorrelated. Finally, although it is possible to preform the reverse transformation, it is generally more useful to determine effects on the biological scale of interest, rather than some arbitrary scale of the canonical variables.

5.6 Determination of statistical significance for multitrait analyses

Both methods described above provide partial answers to the problem of determination of statistical significance in the multitrait situation. For a multivariate analysis it is possible to maximize the likelihood for the complete model and for a "restricted model" with equal means for all QTL genotypes for all traits. Significance of an effect can then be tested by a likelihood ratio test of the two hypotheses. Similarly it is possible to test the hypothesis that the QTL affects only one of the two traits. With the canonical transformation, each trait is analyzed separately, and a p-value is computed for each trait. The FWER can then be computed as described above.

Even if the individual traits are analyzed separately on the original, correlated scale, a FWER can still be computed empirically by a permutation test, as suggested by Churchill and Doerge (1994) for multiple linked loci. For multiple traits, the vector of trait values for each individual is permuted against the genotypes numerous times. For each permutation, a test statistic and its p-value under the null hypothesis are computed for each trait. The lowest p-value at each permutation is then selected, and these are ranked over all the permutations. The 5% lowest p-values over all traits is then an approximate 5% FWER. For correlated traits, this method should result in a higher p-value than computation of FWER assuming an equal number of uncorrelated traits.

We applied this method to a single marker and seven correlated traits for granddaughter design data considered above. The genotype data were permuted against the vector of daughter yield deviations (VanRaden and Wiggans, 1991) for the seven traits. F-values were computed for the seven traits at each permutation. The correlation matrix of the traits are given in Table 5.2, and the results of the permutation analysis are in Figure 5.3. The empirical comparisonwise type I error computed by ranking all 7000 F-values computed is compared

to the empirical FWER computed by ranking on the highest F-value of the seven traits at each permutation. The expected comparisonwise probabilities assuming six or seven independent traits are also plotted.

Table 5.2 Correlations among DYD for the seven traits analyzed.

	Milk	Fat	Protein	Fat %	Protein %	Herdlife	SCS[1]
Milk	1.	0.512	0.821	-0.456	-0.419	0.304	0.020
Fat		1.	0.633	0.537	0.122	0.214	-0.066
Protein			1.	-0.155	0.174	0.309	0.010
Fat %				1.	0.539	-0.075	-0.087
Protein %					1.	-0.028	-0.017
Herdlife						1.	-0.270
SCS							1.

-

[1] Somatic cell score

The correlations among milk, fat, and protein are all > 0.5, as is the correlation between fat and protein percentage. We therefore assumed that the empirical FWER for these seven traits would be considerably smaller than the theoretical FWER assuming seven uncorrelated traits. However, the empirical FWER was generally between the theoretical FWER computed for six or seven uncorrelated traits, and at some points even higher. The relatively low gain in reducing the number of traits can be explained by the fact that the empirical distributions for the individual traits are not exactly the same, and are not equal to the theoretical F-distribution. As noted above in Section 5.3, even a slight discrepancy may become important at very low p-values.

Figure 5.3 Nominal single-trait type I error as a function of the empirical experimentwise type I error (—), the experimentwise type I error assuming six independent traits (– –), and the experimentwise type I error assuming seven independent traits (----).

5. 7 Selective genotyping for multiple traits

As considered in the previous chapter, power to detect segregating QTL can be increased per individual genotyped by selectively genotyping those individuals with extreme values for the quantitative traits (Darvasi and Soller, 1992; Lebowitz et al., 1987; Lander and Botstein, 1989). If only the highest and lowest 5% of individuals are genotyped, it is possible to obtain equal power as compared to random genotyping with only one fourth as many genotypes. Although power is increased per individual genotyped, it is reduced per individual phenotyped. Since selective genotyping is trait specific, the question arises as to the effect of selective genotyping for one trait on correlated traits.

Darvasi and Soller (1992) demonstrated that the estimate of the QTL effect will be biased if only individuals genotyped are used to estimate the effect. They also derived a method to estimate the actual QTL effect as a function of observed effect and the proportion selected for genotyping. Results on simulated data are presented in Table 5.3 for single trait ML. All individuals with phenotypes are included in the analysis. For individuals with phenotypes, but without genotypes, the population genotype probabilities are assumed. For example, in a backcross, it is assumed that each of these individuals has a one half probability of each genotype. Estimates of the QTL parameters for the trait x, the trait under selection are unbiased.

However, if selective genotyping is applied to a single trait, but other correlated traits are also analyzed by single trait ML, then QTL effects associated with the correlated traits will be biased, even if all individuals with phenotypes are included in the analysis. In the example in Table 5.3, selective genotyping was performed relative to trait x, and the QTL was associated with this trait, but not the correlated trait, y. Although single trait ML was able to accurately estimate the effect on trait x and the QTL location, a "ghost" effect of nearly the same magnitude, and a power of nearly 0.5, was found associated with trait y. In the second row of Table 5.4, the segregating QTL was simulated for y, but not x, and selective genotyping was still relative to x. Although no effect was found associated with x, the effect associated with y was underestimated, and the power was only 0.22.

Table 5.3 ML single trait estimates of QTL parameters with selective genotyping[1]

Simulated effect	a_x	a_y	σ_x	σ_y	L_x	L_y	Power for x^2	Power for y
x	0.261 (0.005)	0.260 (0.010)	0.999 (0.001)	0.988 (0.001	48.11 (1.11)	54.16 (1.98)	0.89	0.48
y	-0.005	0.168	0.998	1.001	60.86 (2.89)	56.55	0.11	0.22

	(0.009)	(0.012)	(0.001)	(0.001)	(2.40)	

[1] Results are the mean and standard deviations (in parenthesis) of 200 simulated data sets for each set of parameters. For each data set 2000 individuals from a backcross population were simulated, with a QTL effect of a = 0.25 on either trait x or y at position 50 cM on the chromosome. The marked chromosome had a length of 120 cM, with markers spaced at 20 cM intervals. In both cases the 200 highest and lowest individuals for x were selected for genotyping. The correlation between x and y was 0.5, and the residual variance was $\sigma=1$ for both traits. Parameter estimates for a_x, a_y, σ_x, σ_y, and QTL location, L_x and L_y, were derived by single ML interval mapping including individuals with unknown genotypes.

[2] Empirical power to detect a segregating QTL by a likelihood ratio test with a type-1 error of 0.05.

Table 5.4 ML multiple trait estimates of QTL parameters with selective genotyping[1]

Simulated effect	a_x	a_y	σ_x	σ_y	Location	Power[2]
x	0.256 (0.006)	0.011 (0.012)	0.999 (0.001)	0.996 (0.001)	48.96 (1.33)	0.87
y	-0.004 (0.007)	0.264 (0.012)	0.999 (0.001)	0.994 (0.001)	55.81 (2.19)	0.45

[1] Data set were simulated as described for Table 5.3. Parameter estimates for a_x, a_y, σ_x, σ_y, and QTL location were derived by multitrait ML interval mapping including individuals with unknown genotypes.

[2] Empirical power to detect a segregating QTL for the trait with the true effect by a likelihood ratio test with a type-1 error of 0.05.

Results of the multivariate analyses for both situations are presented in Table 5.4. Unbiased estimates were obtained for the effects on both traits, whether an effect was simulated for trait x or y. Power of detection for an effect on x was similar for both analyses, but much greater for a true effect on y. Moreover, power of detection is increased as compared to random sampling whether the QTL is associated with the trait under selection or with the correlated trait (data not shown). Thus, for selective genotyping it is possible by multivariate ML to derive accurate estimates of QTL effects for both traits under selection and correlated traits. For correlated traits power is increased relative to either single trait ML with selective genotyping or power with random sampling.

5.8 Analysis of multiple pedigrees

With daughter and granddaughter designs the question arises whether the different families should be analyzed jointly or separately. For the full sib design, separate analysis of each family is generally not a viable option, because not enough data is available in single families to obtain reasonable power. A joint analysis over all families has the advantage that the number of tests is reduced, and thus a less restrictive CWER is required to obtain the desired FWER. However, if a QTL is segregating in only a small fraction of the families, then power may be reduced, as compared to a separate analysis of each family.

Analysis models must also be more complicated if several families are analyzed jointly. Georges et al. (1995) used an ML algorithm to estimate QTL parameters including chromosomal location for a granddaughter design, but analyzed each grandsire family separately. Knott et al. (1994) proposed a regression method in which information from all families is used to determine the QTL's map location, but a separate effect is estimated for each family. Thus, for the daughter design, each sire is considered heterozygous. Bovenhuis and Weller (1994) and Mackinnon and Weller (1995) analyzed all families jointly, but assumed that only two QTL alleles were segregating in the population. Thus, the expectation was that at least half of the families would be homozygous for the QTL.

For most populations of interest, the number of alleles segregating for a particular locus will be very low. If all alleles have equal fitness, then the effective number of alleles, k, at equilibrium can be estimated as a function of the effective population size, N_e, and the mutation rate, μ, as follows (Spiess, 1977):

$$k = 4 N_e\mu + 1 \qquad \{5.4\}$$

For example, with $N_e = 3x10^4$ and $\mu = 10^{-6}$, k = 1.12. The effective number of alleles is defined as the number of alleles required to obtain a given level of heterozygosity in the population if all alleles are of equal frequency. Thus, assuming that this mutation rate is more-or-less representative of mutations that are selectively neutral, but have a measurable effect on some trait of interest, the number of alleles segregating in populations of this size will rarely be more than two. Selection will further reduce the number of allele maintained in the population. Thus, mathematical models that assume two different QTL alleles in each pedigree cannot be justified biologically.

5.9 Summary

With multiple markers, and the possibility of complete genome scans, type I error rates for individual tests are virtually meaningless. Three methods were presented to deal with the problem of multiple comparisons; computation of error rates for complete genome scans, permutation tests, and controlling the false discovery rate. None of these methods completely solve the multiple comparison problem. Analysis of multiple traits presents additional problems that can be solved by either a multitrait analysis, which is computationally demanding, or canonical transformation. Various solutions have been presented to analyze multiple pedigrees, covering the range from a separate analysis of each family, to a joint analysis with the same allele segregating in all families, but again there is no uniformly "best" solution.

References

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

Bovenhuis, H. and Weller J. I. (1994) Mapping and analysis of dairy cattle quantitative trait loci by maximum likelihood methodology using milk protein genes as genetic markers. *Genetics* **137**: 267-280.

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

Darvasi, A. and Soller, M. (1992) Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus. *Theor. Appl. Genet.* **85**: 353-359.

Georges, M., Nielsen, D. Mackinnon, M., Mishra, A., Okimoto, R., Pasquino, A. T., Sargent, L. S., Sorensen, A., Steele, M. R., Zhao, X., Womack, J. E., and Hoeschele., I. (1995) Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. *Genetics* **139**: 907-920.

Knott, S. A., Elsen, J. M., and Haley, C. S. (1994) Multiple marker mapping of quantitative trait loci in half-sib populations. *Proc. 4th World Cong. Genet. Appl. Livest. Prod*. Guelph **21:**

Korol, A. B., Preygel, I. A., and Bocharnikova, N. I. (1987) Linkage between loci of quantitative characters and marker loci. 5. Combined analysis of several markers and quantitative characters. *Genetika* **23:** 1421-31.

Korol, A. B., Ronin, Y. I., and Kirzhner, V. M. (1995) Interval mapping of quantitative trait loci employing correlated trait complexes. *Genetics* **140**: 1137-1147.

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

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

Lebowitz, R. J., Soller, M., and Beckmann, J S. (1987) Trait-based analyses for the detection of linkage between marker loci and quantitative trait loci in crosses between inbred lines. *Theor. Appl. Genet.* **73**: 556-562.

Mackinnon, M. J. and Weller, J. I. (1995). Methodology and accuracy of estimation of quantitative trait loci parameters in a half-sib design using maximum likelihood. *Genetics* **141**: 755-770.

Simes, R. J. (1986) An improved Bonferroni procedure for multiple tests of significance. *Biometrika* **73**: 751-754.

Spiess, E. B. (1977) *Genes in Populations.* John Wiley and Sons, New York, NY.

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

Weller, J. I., Wiggans, G. R., VanRaden, P. M., and M. Ron, M. (1996) Application of a canonical transformation to detection of quantitative trait loci with the aid of genetic markers in a multi-trait experiment. *Theor. Appl. Genet.* **92**: 998-1002.