Selection with Recurrent Backcrossing to Develop Congenic Lines for Quantitative Trait Loci Analysis

William G. Hill

Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh, EH9 3JT, Scotland, United Kingdom

Manuscript received July 2, 1997
Accepted for publication November 10, 1997

ABSTRACT

Sewall Wright suggested that genes of large effect on a quantitative trait could be isolated by recurrent backcrossing with selection on the trait. Loci [quantitative trait loci (QTL)] at which the recurrent and nonrecurrent lines have genes of different large effect on the trait would remain segregating, while other loci would become fixed for the gene carried by the recurrent parent. If the recurrent line is inbred and the backcrossing and selection is conducted in a series of replicate lines, in each of which only one backcross parent is selected for each generation, the lines will become congenic to the recurrent parent except for the QTL of large effect and closely linked regions of the genome, and these regions can be identified using a dense set of markers that differ between the parental lines. Such lines would be particularly valuable for subsequent fine-scale mapping and gene cloning; but by chance, even QTL of large effect will be lost from some lines. The probability that QTL of specified effect remain segregating is computed as a function of its effect on the trait, the intensity of selection, and the number of generations of backcrossing. Analytical formulas are given for one or two loci, and simulation is used for more. It is shown that the method could have substantial discriminating ability and thus potential practical value.

WITH modern molecular methods, it is becoming possible to map quantitative trait loci (QTL), those regions of the genome that affect traits with polygenic expression. The precision with which the QTL can be located depends on the populations available, as well as the size and design of the experiment and the statistical methods adopted. To understand the genetic control of a trait, the mapping of QTL is only the start; it is subsequently necessary to identify, if possible, whether a QTL actually comprises a single genetic locus and then the actual gene(s) involved. This has not yet been accomplished for a continuous trait, although Alpert and Tanksley (1996) obtained a clone containing a QTL in tomato. Precise mapping and cloning require well-defined stocks. Multigeneration backcross lines, more specifically congenic lines, for which all the genome except the region of interest comes from an inbred line, are likely to be of particular value for QTL identification (Démant and Hart 1986; Tanksley and Nelson 1996). Sets of congenic lines can be used to obtain narrow intervals for a QTL, providing its effect is sufficiently large that genotypes can be assigned accurately (Darvasi 1997).

While it is desirable to isolate QTL with both small and large effects, in practice the former cannot be mapped with sufficient precision to be useful for subsequent location and cloning; in any case, more information about biological processes are likely to come from a study of those with large effect. A method that can be used to isolate genes (*i.e.*, QTL) of large effect was proposed long ago by Wright (1952). He suggested that recurrent backcrossing be practiced in which the nonrecurrent line is, say, of high performance and the recurrent line is of low performance, with the parents of the backcross individuals selected each generation for high performance of the trait of interest. Backcrossing leads to a halving of frequency each generation of genes that do not affect the trait and are not linked to those that do. Genes with large effect on the trait are, however, more likely to be present in the selected individuals, so their expected frequency is >0.5 and the backcross line should eventually remain segregating only for genes of large effect. In simple terms, a gene would need to confer a twofold higher fitness to survive indefinitely in a large backcross population. In practice, lines are finite in size, so even genes with very large effect may be lost, albeit slowly. While it may be feasible to maintain one large backcross line with several parents each generation, this provides only one replicate and requires that selection be practiced between and within families, so between-family environmental covariances, important in species such as mice, increase the errors in selection.

The recurrent backcrossing enables all but very closely linked markers and QTL to recombine, thereby facilitating detailed mapping using the dense maps now available, for example, in the mouse (Dietrich *et al.* 1996). The method is likely to be of most use in species such

Corresponding author: William G. Hill, Institute of Cell, Animal and Population Biology, University of Edinburgh, West Mains Road, Edinburgh, EH9 3JT, Scotland, UK. E-mail: w.g.hill@ed.ac.uk

as Drosophila, mice, and Arabidopsis, which have a short generation interval and for which inbred lines are available. It has been proposed by Snell (1958) as a means of identifying histocompatibility genes, and it has been used by Beebe *et al.* (1997) to identify loci associated with resistance to *Leishmania major* in mice. The principles and methods, particularly for detection of loci associated with a disease with all-or-none expression, have been investigated by N. J. Schork, A. M. Beebe, B. Thiel, P. St. Jean, and R. L. Coffman, (unpublished data), in which the scheme generally involves taking sublines from individuals who show resistance so that a pedigree of resistant individuals is built up. The method is then analogous to that of identity-by-descent (ibd) mapping (Houwen *et al.* 1994; Guo 1995; Charlier *et al.* 1996), in which regions of chromosomes in related individuals with extreme phenotypes for a quantitative trait carrying putative QTL are identified by a genome-wide scan, and for which relevant theory has been developed (Thomas *et al.* 1994; N. J. Schork, A. M. Beebe, B. Thiel, P. St. Jean, and R. L. Coffman, unpublished data). The use of backcrossing with selection to identify QTL by their linkage to markers has similarities to the selection scheme practiced by G. Bulfield from an inbred cross in which changes in marker frequency are monitored in replicate lines to infer QTL position (Keightley and Bulfield 1993; Keightley *et al.* 1996; Ollivier *et al.* 1997), but differs in that the backcross selected lines are congenics and immediately useful for precision mapping and gene cloning.

A formal backcrossing scheme, which has the particular benefit that it leads to the production of several independent lines that are congenic for small but different parts of the genome, is to maintain a series of separate single-family lines during the backcrossing, with each family maintained by only one selected parent each generation. The presence of QTL of large effect is then detected by undertaking a genome-wide scan of molecular markers only after several generations of backcrossing and only on the one selected individual of each line, and by identifying regions that have remained segregating in many of the families. Subsequently, these regions that remain segregating can be investigated more finely by further developing lines by continued backcrossing, but now maintaining segregating the marker or flanking markers that are identified as close to QTL in the initial screen and by progeny testing within families to identify the marker-associated effect.

In this paper, the properties of the method of backcrossing with selection to develop such independent congenic lines are examined, for example, in terms of the probability that QTL remain segregating as a function of the size of its effect on the trait, the number of generations of backcrossing, and its linkage to other QTL. Ways to analyze and interpret the data are considered.

ANALYSIS

It is assumed that backcrossing is practiced recurrently to a completely inbred line. The nonrecurrent and recurrent lines are assumed to be homozygous for different alleles at QTL of interest. A total of M families (lines) are maintained independently. In each generation, in each backcross family, n individuals are recorded for the quantitative trait, and the highest scoring individual is selected and backcrossed again to the recurrent line. Backcrossing and selection are continued for t generations, F_1 being generation 0. Variation that is not explained by the QTL, because of within-family environmental and residual genetic deviations in the trait (likely to be significant only in early backcross generations) is assumed to be normally distributed, N(0,1). Effects, a, of QTL are expressed in units of this within-family (mainly environmental) standard deviation.

Single locus: This simple example serves as a reference for others. It is assumed that a QTL with allele A from the nonrecurrent and A' from the recurrent parent confers an increase in the heterozygote of a SD on the trait, and that it is unlinked to any other QTL affecting the trait. It is necessary to compute the probability $P(n,a)$ that the offspring of generation $t + 1$ selected from a heterozygous backcross parent of generation t is itself heterozygous for the locus. This probability can be readily computed using binomial probabilities for the number k of heterozygous offspring among the n recorded and order statistics for the probability that the highest scoring offspring is heterozygous. For example, if one of the k heterozygotes has phenotypic value x for the trait with probability $\phi(x)dx$ and is the highest in the family, it implies that $k - 1$ heterozygotes and $n - k$ homozygotes have performance less than x, the probabilities for each individual being $\Phi(x)$ and $\Phi(x + a)$, respectively, where $\phi(x)$ and $\Phi(x)$ denote the density and cumulative distribution functions, respectively, of the standardized normal distribution. Then, using the method of Hill (1969), the probability an AA' heterozygote is selected is as follows:

$$P(n,a) = \sum_{k=1}^{n} \frac{n!}{k!(n-k)!2^n} k \int_{-\infty}^{\infty} [\Phi(x)]^{k-1}[\Phi(x+a)]^{n-k}\,\phi(x)dx,$$

which reduces to

$$P(n,a) = (n/2^n)\int_{-\infty}^{\infty} [\Phi(x) + \Phi(x+a)]^{n-1}\,\phi(x)dx. \quad (1)$$

In the simplified Equation 1, for each of the n individuals in the family, there is a probability of one-half that the individual is an AA' heterozygote and, if it is AA', a probability of $[\Phi(x)/2 + \Phi(x + a)/2]^{n-1}$ that all other family members have a lower phenotypic value. For $a = 0$, the integral in Equation 1 reduces to $2^{n-1}/n$ and $P(n,a) = 1/2$, as expected for a neutral gene, since the highest scoring individual is equally likely to be AA' or $A'A'$; and for $a \to \infty$, the integral reduces to $(2^n - 1)/$

TABLE 1

Probability $P(n,a)$ that a heterozygote with a QTL of effect a SD units is selected from the offspring of a backcross mating, with selection of the best one from n

a	n: 2	3	4	6	8	12	20
0	0.5	0.5	0.5	0.5	0.5	0.5	0.5
0.125	0.5176	0.5264	0.5321	0.5395	0.5444	0.5508	0.5581
0.25	0.5351	0.5526	0.5640	0.5786	0.5882	0.6007	0.6150
0.5	0.5691	0.6036	0.6257	0.6538	0.6718	0.6948	0.7203
1	0.6301	0.6952	0.7352	0.7831	0.8115	0.8446	0.8769
1.5	0.6778	0.7667	0.8186	0.8751	0.9045	0.9340	0.9574
2	0.7107	0.8160	0.8743	0.9311	0.9560	0.9763	0.9882
$\rightarrow\infty$	0.75	0.875	0.9375	0.9844	0.9961	0.9998	1.0000

n and $P(n,a) = 1 - (1/2)^n$, which is the probability that at least one of the offspring is a heterozygote and available to be selected. Also, the probability that a homozygote $A'A'$ is selected is given by $P(n - a) = 1 - P(n,a)$. Equation 1 can readily be evaluated using Simpson's rule numerically. Results are given in Table 1, and these and later results obtained using order statistics were checked by Monte Carlo simulation.

These values can be approximated for small values of a, by

$$P(n,a) \sim 0.5 + ia/4, \tag{2}$$

where i is the standardized selection intensity for the normal distribution with finite numbers recorded. For $n = 2$, 3, 4, 6, 8, 12, and 20, $i = 0.56$, 0.85, 1.03, 1.27, 1.42, 1.63, and 1.87, respectively (Falconer and Mackay 1996). Comparing these values with those in Table 1, it is seen that the approximation is adequate for $a \leq 0.5$.

Repeated backcrossing: In the first few generations of backcrossing, the variance is likely to be inflated by segregation at other loci in a way that the accuracy of selection and probabilities of retention will be lower than those shown in Table 1, where the effect of the gene is expressed in terms of the within-family environmental SD. This problem is visited again later, but first, let us assume for simplicity that the (unit) within-family variance remains constant. The probability $P(n,a)_t$ that the QTL remains segregating for t generations in a line maintained with one parent and the same selection intensity each generation is therefore

$$P(n,a)_t = [P(n,a)]^t. \tag{3}$$

If the numbers available for selection, n_t, vary among generations, Equation 3 has to be replaced by the product of the appropriate values or, equivalently, the harmonic mean of $P(n_t,a)$ used. Unless n_t varies greatly, simply using an approximate mean of n in Equation 3 suffices. For example, if n_t takes values of 2, 3, 4, 5, and 6 in successive generations and $a = 1$, the probability the QTL is retained for five generations is 0.19, whereas inserting $n = 4$ in each generation in Equation 3 gives 0.21.

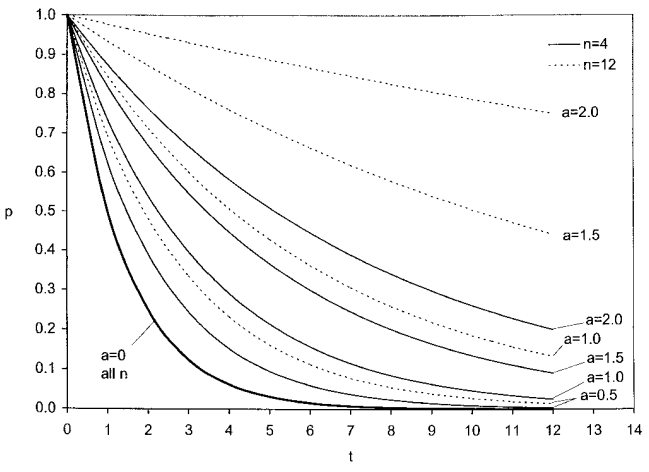

Figure 1.—Probability that a QTL of effect a remains segregating for t generations of backcrossing when the best individual of n is selected each generation: $P(n,a)_t$ plotted against t for $a = 0$, 0.5, 1, 1.5, and 2, and $n = 4$ and 12.

Some examples using Equation 3 are given in Figure 1. The main point is that the differences between the probability of continued segregation of a QTL of large effect and a QTL of small effect becomes wider as more generations of backcrossing are undertaken and more intense selection is practiced. Of course, if backcrosssing is continued too long, even those of very large effect are lost. In principle, there is some intermediate optimum time for discriminating among QTL of specified effects, if such can be defined.

Since a number of independent replicate lines can be kept, it is useful to reconsider these results in terms of the distribution of the number of lines in which a QTL of specified effect would be segregating. If M lines are maintained, then the expected number in which there is segregation at generation t is $MP(n,a)_t$. The actual number segregating, m, has a binomial distribution, but the Poisson distribution gives an adequate approximation and results are more readily generalized. Hence, we assume that m has a Poisson distribution with parameter $MP(n,a)_t$. Some examples are given in Table

TABLE 2

Probability distribution of the number *m* of a total of *M* lines in which a QTL of effect *a* SD units remains segregating for *t* generations with selection of the best one from *n* each generation

M		10			20			10		
n, *t*		4, 4			4, 4			12, 8		
a	*m*	≥1	≥2	≥4	≥1	≥2	≥4	≥1	≥2	≥4
0		0.465	0.130	0.004	0.713	0.355	0.038	0.038	0.001	0.000
0.125		0.558	0.197	0.010	0.799	0.476	0.079	0.081	0.003	0.000
0.25		0.636	0.269	0.020	0.868	0.600	0.147	0.156	0.013	0.000
0.5		0.784	0.453	0.070	0.953	0.810	0.368	0.419	0.104	0.002
1		0.946	0.789	0.335	0.997	0.980	0.834	0.925	0.730	0.262
1.5		0.989	0.938	0.656	1.000	0.999	0.978	0.997	0.979	0.829
2		0.997	0.980	0.834	1.000	1.000	0.997	1.000	0.998	0.964
→∞		1.000	0.996	0.949	1.000	1.000	1.000	1.000	0.999	0.989

2, for experiments in which 10 or 20 lines are maintained. For example, if 10 lines are maintained with selection of one from $n = 4$ for $t = 4$ generations, a QTL with effect of 0.25 SD has a probability of <2% of remaining segregating in four or more lines, whereas a QTL with effect of 1.5 SD or more, has a <1% chance of being lost from all 10 lines and a 66% chance of remaining segregating in four or more lines.

Correction for background genetic variation: With segregation at other loci than that being analyzed, there will be additional variation within families, particularly in early generations. With additive genes and genetic variation V_A caused by background genetic variation in the F_2, and assuming for simplicity that the background variation is caused by very many unlinked loci, each of very small effect, there will be $V_A/2$ in the first backcross and $V_A/2^t$ in the tth backcross. There will also be variance $a^2V_E/4$ caused by the QTL under consideration in the first backcross, which with additive gene action implies $a^2V_E/2$ in the F_2, where V_E is the within-family environmental variance. Consider a simple case, where $V_A = V_E$ and $a = 1$, so the total genetic variance in the F_2 would be $3V_E/2$, and the within-family environmental plus background genetic variance in the backcross would be $[1 + (1/2)^t]V_E$. Hence, the effect of the QTL in environmental SD units would be 0.816, 0.894, 0.942, ... in backcross generations $t = 1, 2, 3,...$ For example, $P(4, 0.816) = 0.698$, $P(4, 0.894) = 0.714$, and $P(4, 0.942) = 0.724$, whereas $P(4,1) = 0.735$. Hence, the probability that the QTL remains segregating to generations 1, 2, 3, and 4 is 0.697, 0.498, 0.361, and 0.263, respectively, whereas the equivalent values assuming that $P(4,1)$ is appropriate each generation are 0.735, 0.541, 0.397, and 0.292, respectively. It seems that the quantitative differences are rather small, and that the simple calculations are adequate unless the segregation variance, V_A, is much larger than V_E (*i.e.*, heritability in the F_2 considerably in excess of one-half) and if, because of the selection, it declines much more slowly than one-half per generation, *i.e.*, there are other QTL of large effect or linked in coupling phase.

Two loci: The usefulness of the method depends not only on keeping QTL of large effect segregating, but also on losing those of small or negative effect that may be maintained by linkage to a QTL of large effect. Let us consider a model where there are two additive (*i.e.*, nonepistatic) QTL, A_1 and A_2, on the same chromosome, with effects a_1 and a_2 within family standard deviations, respectively. The recombination fraction is r between the loci. In any generation where the backcross parent is a double heterozygote, $A_1A_2/A_1'A_2'$, there are four possible offspring genotypes selected in the next generation: the double heterozygote with both A_1 and A_2 present, *i.e.*, $A_1A_2/A_1'A_2'$ with probability $P^{12}(n,a_1,a_2)$, or the single heterozygote with only A_1, *i.e.*, $A_1A_2'/A_1'A_2'$ with probability $P^{12'}(n,a_1,a_2)$, or only A_2, or both lost. An extension of Equation 1 can be used. For example, the probability that the double heterozygote is selected is given by

$$P^{12}(n,a_1,a_2) = [n(1-r)/2^n]\int_{-\infty}^{\infty}[(1-r)\Phi(x+a_1+a_2) + r\Phi(x+a_2) + r\Phi(x+a_1) + (1-r)\Phi(x)]^{n-1}\phi(x)dx. \quad (4)$$

Similar equations apply for sampling the other genotypes, the term in $(1 - r)$ before the integral being replaced by r for recombinant types. If only one QTL

TABLE 3

Probabilities of selection of each alternative genotype for a two-locus model with additive effects a_1 between alleles A_1 and A_1' at locus 1 and a_2 between A_2 and A_2' at locus 2, and recombination fraction r between the loci—selection of the best one from n

n	4				12			
r	A_1A_2	A_1A_2'	$A_1'A_2$	$A_1'A_2'$	A_1A_2	A_1A_2'	$A_1'A_2$	$A_1'A_2'$
				a_1: 0.5	a_2: 0.5			
0.5	0.380	0.242	0.242	0.136	0.468	0.222	0.222	0.089
0.2	0.596	0.095	0.095	0.214	0.704	0.083	0.083	0.131
0.1	0.666	0.047	0.047	0.240	0.776	0.040	0.040	0.144
0.05	0.701	0.023	0.023	0.252	0.811	0.020	0.020	0.150
0	0.735	0.000	0.000	0.265	0.845	0.000	0.000	0.155
				a_1: 0.75	a_2: 0.25			
0.5	0.377	0.305	0.183	0.135	0.456	0.319	0.139	0.086
0.2	0.594	0.120	0.072	0.214	0.697	0.122	0.052	0.129
0.1	0.665	0.060	0.036	0.239	0.772	0.060	0.026	0.143
0.05	0.700	0.030	0.018	0.252	0.808	0.030	0.013	0.149
0	0.735	0.000	0.000	0.265	0.845	0.000	0.000	0.155
				a_1: 1.0	a_2: 0.0			
0.5	0.368	0.368	0.132	0.132	0.422	0.422	0.078	0.078
0.2	0.588	0.147	0.053	0.212	0.676	0.169	0.031	0.124
0.1	0.662	0.074	0.026	0.238	0.760	0.084	0.016	0.140
0.05	0.698	0.037	0.013	0.252	0.802	0.042	0.008	0.148
0	0.735	0.000	0.000	0.265	0.845	0.000	0.000	0.155
				a_1: 1.5	a_2: −0.5			
0.5	0.336	0.476	0.064	0.124	0.317	0.611	0.017	0.054
0.2	0.567	0.201	0.026	0.206	0.602	0.282	0.009	0.108
0.1	0.650	0.102	0.013	0.235	0.717	0.148	0.005	0.130
0.05	0.692	0.051	0.007	0.250	0.779	0.076	0.002	0.142
0	0.735	0.000	0.000	0.265	0.845	0.000	0.000	0.155

remains segregating, in subsequent generations its behavior in that line is described as would be for single loci (1).

Examples are given in Table 3 of the probabilities for a series of examples in which the sum of the effects of the two loci are the same ($a_1 + a_2 = 1$), but their relative sizes differ ($a_1 = 0.5$, 0.75 and 1.5). Results for $a_1 = 1$ and $a_2 = 0$ can be obtained from Table 1 by noting that $P^{12}(n,1,0) = (1 - r)P(n,1)$, where $P(n,1)$ is given by Equation 1. If linkage is loose, it is seen that the extreme cases of equal effects ($a_1 = a_2$) and $a_2 < 0$ give quite different outcomes, but as linkage becomes tight, the survival probability of the double heterozygote depends little on the relative size of effects of the two loci.

These probabilities can be approximated, providing values of ia_1 and ia_2 are not too large, for example:

$$P^{12}(n,a_1,a_2) \sim (1 - r)[1 + i(a_1 + a_2)/2]/2,$$
$$P^{12'}(n,a_1,a_2) \sim r[1 + i(a_1 - a_2)/2]/2, \quad (5)$$

and similarly for $P^{1'2}(n,a_1,a_2)$. For example, with $r = 0.2$, these approximations give values of $P^{12}(4, 0.75, 0.25) = 0.606$, $P^{12'}(4, 0.75, 0.25) = 0.126$, and $P^{1'2}(4, 0.75, 0.25) = 0.074$, compared with the exact values of 0.594, 0.120, and 0.072, respectively, in Table 3.

The examples in Table 3 are for nonepistatic loci, *i.e.*, with additive effects in heterozygotes over loci. Equation 4 changes in a straightforward way if this is not the case, and probabilities that one or both of the loci continue to segregate change correspondingly. For example, consider the case where the double heterozygote is 1 SD superior to each single heterozygote and the double homozygote. For complete linkage ($r = 0$), results are therefore the same as in each example in Table 3, whereas for $r = 0.2$, $P^{12} = 0.6376$, $P^{12'} = P^{1'2} = 0.0604$, and $P^{1'2'} = 0.2416$, and for unlinked loci ($r = 0.5$), $P^{12} = 0.4502$, and $P^{12'} = P^{1'2} = P^{1'2'} = 0.1832$, *i.e.*, single heterozygotes are less likely to be selected than in the additive case.

To consider the passage over several generations of each of the genotypic classes, it is necessary to include the probabilities of the single locus segregants. We construct the 3×3 transition matrix **B**, for which the rows and columns identify the following states: (1) A_1 and A_2, (2) A_1 but not A_2, and (3) A_2 but not A_1; the elements

TABLE 4

Probabilities of segregation after *t* generations of alternative genotypes for a two-locus model with additive effects a_1 between alleles A_1 and A_1' at locus 1 and a_2 between A_2 and A_2' at locus 2, with recombination fraction *r* between the loci—selection of the best one from *n* each generation

n	4			4			12		
t	4			8			8		
r	A_1A_2	$A_1'A_2$	A_1A_2'	A_1A_2	$A_1'A_2$	A_1A_2'	A_1A_2	$A_1'A_2$	A_1A_2'
				a_1: 0.5	a_2: 0.5				
0.2	0.126	0.087	0.087	0.016	0.024	0.024	0.060	0.054	0.054
0.05	0.241	0.027	0.027	0.058	0.011	0.011	0.186	0.023	0.023
				a_1: 0.75	a_2: 0.25				
0.2	0.125	0.126	0.056	0.016	0.043	0.013	0.055	0.117	0.021
0.05	0.240	0.040	0.018	0.058	0.018	0.006	0.182	0.047	0.010
				a_1: 1	a_2: 0				
0.2	0.120	0.172	0.034	0.014	0.071	0.006	0.043	0.215	0.007
0.05	0.238	0.054	0.012	0.057	0.029	0.003	0.172	0.087	0.004
				a_1: 1.5	a_2: −0.5				
0.2	0.104	0.276	0.011	0.011	0.152	0.001	0.017	0.048	0.001
0.05	0.223	0.089	0.004	0.053	0.061	0.001	0.136	0.217	0.001

$b_{i,j}$ specify the transition probability from state i at generation t to state j at generation $t + 1$. (Alternatively, a 4×4 matrix can be used, with the fourth row and column denoting the case where neither A_1 nor A_2 are segregating; because this is an absorbing state, the computation of segration probabilities are not affected.) Hence, using the single locus formulas from the preceding section,

$$\mathbf{B} = \begin{pmatrix} P^{12}(n,a_1,a_2) & P^{12'}(n,a_1,a_2) & P^{1'2}(n,a_1,a_2) \\ 0 & P(n,a_1) & 0 \\ 0 & 0 & P(n,a_2) \end{pmatrix}. \quad (6)$$

Assuming the population starts in the state where both A_1 and A_2 are segregating, from Equation 6, it can be shown that at generation t,

$$P^{12}(n,a_1,a_2)_t = b_{11}^t, \quad (7a)$$

$$P^{1}(n,a_1,a_2)_t = b_{12}(b_{11}^t - b_{22}^t)/(b_{11} - b_{22}), \text{ and} \quad (7b)$$

$$P^{2}(n,a_1,a_2)_t = b_{13}(b_{11}^t - b_{33}^t)/(b_{11} - b_{33}). \quad (7c)$$

For the example given in Table 3, results for segration probabilities for four and eight generations are given in Table 4. Although the probability that both loci remain segregating is not greatly affected by the relative magnitude of the gene effects (and no probabilities of retention are high in this example because the total effect is only $a_1 + a_2 = 1$ and $n = 4$), the QTL of smaller or negative effect has a low probability of remaining segregating alone for many generations, so the method does have some discriminating power.

Marker segregation: The previous analyses have been restricted to the fate of the QTL, but their segregation has to be detected by means of molecular markers. The calculations for individual markers are straightforward and are simply a special case of the two-locus analysis given above. Let us assume that A_1 is the QTL and A_2 is the marker, *i.e.*, $a_1 = a$ and $a_2 = 0$, with the recombination fraction between the loci equal to r. Then the elements of **B** are given by

$$b_{11} = (1 - r)P(n,a),\ b_{12} = rP(n,a),\ b_{13} = \mathrm{r}[1 - P(n,a)],$$
$$b_{22} = P(n,a),\ b_{33} = \tfrac{1}{2}$$

where $P(n,a)$ is given by Equation 1. An example for a marker locus is given as part of Table 3 ($a_1 = 1$, $a_2 = 0$). The probability, from Equations 7a and 7c, respectively, that the marker remains segregating in coupling with the QTL is

$$P^{12}(n,a,0)_t = [(1 - r)P(n,a)]^t,$$

the probability it remains segregating without the QTL is

$$P^{1'2}(n,a,0)_t = [r(1 - P(n,a)]\{(\tfrac{1}{2})^t - [(1 - r)P(n,a)]^t\}/$$
$$\{\tfrac{1}{2} - [(1 - r)P(n,a)\},$$

and their sum, $P^{12}(n,a,0)_t + P^{1'2}(n,a,0)_t$, is the overall probability the marker is retained. Because the probabilities that QTL are retained for many generations are already small unless the QTL has an effect as large as 2 SD or so (Figure 1), it is clear that only very tightly linked markers are likely to be of value in QTL detection. An illustration is given in Table 4 ($a_1 = 1$, $a_2 = 0$). Hence, the marker analysis after backcrossing needs to be done with very closely spaced markers in a genome-wide scan.

If there are two QTL, the fate of alleles at a marker locus, say A_3, depends on whether it is between or outside this pair. If A_3 is outside the interval A_1–A_2, then its probability of segregation is given by expanding the

TABLE 5

Probabilities of segregation of multiple QTL and markers, computed using Monte Carlo simulation with 500 replicates*

L	50			100			200			400		
a	0.1	0.2	0.4	0.1	0.2	0.4	0.1	0.2	0.4	0.1	0.2	0.4
Position												
0, *20* end markers	0.18	0.32	0.46	0.11	0.20	0.26	0.09	0.10	0.15	0.07	0.08	0.09
1, *19* end QTL	0.20	0.34	0.50	0.12	0.23	0.32	0.11	0.14	0.18	0.08	0.11	0.15
2, *18* markers	0.21	0.36	0.54	0.13	0.25	0.32	0.11	0.14	0.22	0.08	0.09	0.14
3,..., *17* mid QTL	0.21	0.42	0.62	0.13	0.30	0.43	0.12	0.18	0.30	0.08	0.12	0.16
4,..., *16* mid markers	0.21	0.42	0.62	0.13	0.30	0.41	0.12	0.18	0.28	0.08	0.11	0.18

* SE (probability estimate) <2%, dependent on actual probability and map position.

A chromosome of map length *L*cM comprises 21 equally spaced loci, numbers *0* and *20* being at the ends. There are QTL of effect *a* in coupling at positions *1*,*3*,..., *19*, and markers with no effect at positions *0*, *2*, ..., *20*. Results given are averages over the loci indicated. $n = 4$, $t = 4$.

formulas given in Equation 3. If A_3 lies outside the interval, but nearer to A_2, for example, the probability that all three loci remain segregating is given by $(1 - r_{23})P^{12}(n,a_1,a_2)$. If A_3 lies between A_1 and A_2, then, for example, the probability that all three loci remain segregating is $[(1 - r_{13})(1 - r_{23})/(1 - r_{12})]P^{12}(n,a_1,a_2)$. The matrix has to be formally extended to consider seven possible classes (all three loci, three pairs, and three singles), but the calculations are straightforward.

Multiple linked QTL: The preceding analysis is concerned with the outcome of backcrossing when there are only one or two QTL on the chromosome affecting the trait under selection. An alternative model is that the difference in performance between the recurrent and nonrecurrent parent lines are caused by many QTL of (mainly) small effect on each chromosome. Some examples have been considered using the Monte Carlo simulation, with a model of a chromosome of map length *L*cM and typically 21 loci simulated at equal spacing, with the most distant at the ends of the chromosome. Ten of these loci, numbers *1* (*i.e.*, at position 0.05*L*), *3*,..., *19*, were assumed to be QTL of equal effect, and the remaining 11 loci, numbers *0*, *2*,..., *20*, were assumed to be markers with no effect on the trait. The probabilities that individual QTL remain segregating do not differ greatly whether or not they are near the middle or the end of the chromosome, unless the chromosome is of length 200 cM or more. Similarly, the probability that individual markers remain segregating is little different from that of the QTL between which they lie. Hence, only summary figures are given for the examples in Table 5. If the chromosome is short ($L <$ 100), the probability that loci on it remain segregating is quite substantial, even if the individual loci have effects of 0.4 SD or less. If it is long ($L > 200$), the probability of continued segregation is a little higher than for neutral genes (6% in the example of $t = 4$ generations), even if the individual effects are as large as 0.4 SD and the total difference between the ancestral chromosomes is 4 SD. In general, there will be greater discrimination if selection is more intense and the numbers of generations are longer than the examples in Table 5 ($n = 4$, $t = 4$). A model with very many QTL of very small effect in coupling would give similar results to those in Table 5, as seen by the similar segregation probabilities for QTL and markers.

Unless the chromosome is very long ($L > 200$) and individual QTL are all of large effect (say $a > 1$), an alternative extreme model in which there is complete repulsion of QTL, *i.e.*, alternating positive and negative effects on the trait along the chromosome with no net effect of the QTL together on the chromosome, would behave in a very similar way to the case in which all loci are neutral. In general, there will be greater discrimination if selection is more intense and the numbers of generations are longer than the examples in Table 5 ($n = 4$, $t = 4$).

Interspersed *inter se* matings: QTL of small effect, particularly if they are partially recessive to that in the recurrent parent, have a low probability of retention. It is possible to increase the probability of QTL segregation by increasing the strength of the selection in increasing frequency relative to that of backcrossing in reducing frequency. One possible method, which fits within the independent family (line) structure discussed here, is to intersperse a generation of *inter se* mating between each generation of backcrossing, *i.e.*, to allow two generations of selection per generation of backcrossing. For simplicity, it is assumed that the same family size is used in each case: in the backcrossing generations, the best male (or female) from *n* of that sex is selected; in the *inter se* generation, the best male from *n* males recorded and the best female from *n* females recorded are selected and mated. The important difference from the previous analysis is that now matings can be made between two heterozygotes so that homozygotes for the QTL of interest may then be selected for the next backcross generation.

In the previous analysis of backcrosses alone, the quantity a was used for simplicity to define the homozygote–heterozygote difference; a fuller definition is now required, and the notation of Falconer and Mackay (1996) is adopted, adding an asterisk to distinguish values from those above: the genotypic values (in within family SD units) are AA a^*, AA' d^*, and $A'A'$ $-a^*$. Hence, in the previous analysis, $a = a^* + d^*$. If the parents of the *inter se* generation are a homozygote and a heterozygote, the calculations given in Equation 1 still apply. If the mating is between two heterozygotes, the probabilities that the individual selected for the next backcross generation is AA, AA', or $A'A'$ are readily computed by extending the formula, as exemplified by Equation 4. Let these probabilities be, for example, $P^{AA}(n,a^*,d^*)$. The full two-generation process can be described by a transition matrix C, in which the rows and columns denote the genotype of the backcross parent, AA for the first and AA' for the second (a third row and column could be added for the absorbing state when the backcross parent is $A'A'$):

$$\mathbf{C} = \begin{pmatrix} P^{AA}(n,a^*,d^*) & P^{AA'}(n,a^*,d^*) \\ [P(n,a^* + d^*)]^2\, P^{AA}(n,a^*,d^*) & [P(n,a^* + d^*)]^2\, \{P^{AA'}(n,a^*,d^*) + 2[1 - P(n,a^* + d^*)]\} \end{pmatrix}.$$

In the first row of **C**, because the backcross parent is AA, the mating for the *inter se* generation is always $AA' \times AA'$, so the probabilities c_{11} and c_{12} refer to the selection of AA and AA', respectively, from among their offspring. In the second row, where the backcross parent is the AA' heterozygote, the probability is $[P(n,a^* + d^*)]^2$ that both the selected male and female for the *inter se* mating are heterozygote, which leads immediately to c_{21} and the first term in c_{22}. The second term in c_{22} is the product of the probabilities, $2P(n,a^* + d^*)[1 - P(n,a^* + d^*)]$, that the *inter se* mating is between a homozygote and heterozygote, and $P(n,a^* + d^*)$, that a heterozygous offspring is selected from the mating.

In Table 6, examples are given assuming that $n = 4$ individuals are recorded per family in both the backcross and *inter se* mating generations, for different degrees of dominance. The value of t refers to the number of backcross generations completed, and the probabilities given are that the QTL is still segregating, *i.e.*, the individual selected from the *inter se* generation is either AA or AA'. For reference, to define the rate of loss after a few generations, values of $1 - \lambda$, where λ is the larger eigenvalue of **C**, are also given and can be compared directly with the probabilities $1 - P(n,a)$, which equal $1 -$ eigenvalue for a 1×1 matrix), from Table 1 when only backcrossing with selection is practiced.

The greatest benefits from the *inter se* mating arise, of course, when the QTL of high value is recessive and therefore neutral during the selection among backcrossed individuals. In the additive case, the interspersed generation does not half the rate of loss (*i.e.*, $1 - \lambda$) unless the effect of the QTL is quite large. Because the time (*i.e.*, total number of generations) required to reduce the probability of segregation of neutral background genes is doubled, it is moot whether there is benefit in inserting the *inter se* generation. Maintaining more replicate lines or increasing the number of individuals from each family recorded for the quantitative trait, each generation, family size may make more efficient use of resources. Where reproductive rate limits selection intensity, for example in mice, it might be increased by keeping two litters from each female or by selecting only among males that have been given multiple matings.

DISCUSSION

To convey some "feel" for the results, examples of simulated data sets are given in Table 7, in each case for a single chromosome of length 1 Morgan and following eight generations of backcrossing with selection of the best individual from 12 recorded. The different models simulated represent different distributions of the QTL effects, ranging from $a_i = 0.2$ to 2, but with the same total difference in effect, $\Sigma_i a_i = 2$, between these chromosomes (as heterozygotes) from the recurrent and nonrecurrent backcross lines. The variances in the F2 would differ among the models (unless there was no recombination), being largest with only one QTL differentiating the lines. As expected from the previous analyses, there is segregation at one or more of the markers in the majority of the replicate backcross lines. It is also seen that the pattern differs somewhat according to the number of QTL accounting for the line difference on this chromosome, but that many marker configurations can appear for two or more quite different distributions of QTL effects, and that identifying whether one or more QTL are responsible for the marker effects seen is unlikely to be feasible from the marker distribution alone. As in other methods of QTL mapping, it is difficult to distinguish between one QTL and a pair of closely linked QTL (Haley and Knott 1992; Jansen 1993; Zeng 1993).

Putative evidence for a QTL in the region of a marker comes from finding that the marker is present in more replicate lines than expected by chance. If there is no QTL in the region, the probability that the marker remains segregating is $1/2_t$, and if M lines are maintained, the probability that it is found in m of them is given by the Poisson distribution with parameter $M/2^t$. For example, with four generations of backcrossing, the probability that it is found in three or more lines is $<5\%$ (0.026, extending results of Table 2). Hence, in such an experiment, applying a site-by-site type I error, further attention should be given to regions that are found segregating in three or more lines. In an experiment run for eight generations with 10 lines, any region

TABLE 6

Probabilities (Prob: *t*) that a QTL is retained with *t* of each of alternating generations of backcrossing and *inter se* mating, with one individual of each sex selected from *n* = 4 recorded in each case.

a^*	d^*	Prob: $t = 4$	Prob: $t = 8$	Rate of loss
0	0	0.0521	0.0033	0.5000 (0.500)
		Dominant ($a = 2a^*$)		
0.125	0.125	0.1007	0.0119	0.4139 (0.436)
0.25	0.25	0.1754	0.0351	0.3311 (0.374)
0.5	0.5	0.3905	0.1658	0.1929 (0.265)
1	1	0.7869	0.6342	0.0525 (0.126)
		Additive ($a = a^*$)		
0.125	0	0.0803	0.0075	0.4464 (0.468)
0.25	0	0.1182	0.0160	0.3938 (0.436)
0.5	0	0.2242	0.0552	0.2955 (0.374)
1	0	0.5052	0.2672	0.1472 (0.265)
		Part recessive ($a = a^*/2$)		
0.125	−0.0625	0.0718	0.0060	0.4620 (0.484)
0.25	−0.125	0.0969	0.0107	0.4234 (0.468)
0.5	−0.25	0.1643	0.0296	0.3484 (0.436)
1	−0.5	0.3430	0.1237	0.2252 (0.374)
		Recessive ($a = 0$)		
0.125	−0.125	0.0643	0.0048	0.4768 (0.500)
0.25	−0.25	0.0800	0.0073	0.4507 (0.500)
0.5	−0.5	0.1227	0.0166	0.3938 (0.500)
1	−1	0.2287	0.0558	0.2971 (0.500)

The rate of loss of the QTL is given by $1 - \lambda$, where λ is the eigenvalue of the transition matrix C, and ( ) gives the rate of loss for backcrossing alone with the same QTL effects, *i.e.*, AA a^*, AA' d^*, $A'A'$ $-a^*$, with effect in the backcrossing generation $a = a^* + d^*$.

remaining segregating should be considered further. This does not take into account the multiple testing problem, which is straightforward if all markers are on different chromosomes, or if the markers are essentially unlinked because they are widely separated on fewer chromosomes; the Bonferroni correction can then be applied to give an experiment-wide error of specified value, albeit at the risk of considerable reduction in power. For example, taking 20 unlinked markers, one per mouse chromosome, the overall type I error if only markers found segregating in four or more of 10 congenic backcross lines are examined further is ~8% (0.004×20, from Table 2); more precisely, the probability that a locus remains segregating is $(1/2)^8$, and the probability that any of the 20 remain segregating is $1 - [1 - (1/2)^8]^{20} = 7.5\%$. A more sophisticated analysis is required to find critical values when the distribution of chromosome lengths and actual position of markers are taken into account, but it is quite straightforward to obtain genome-wide critical values such as those that are used in QTL mapping from one-generation crosses (Lander and Botstein 1989), for example, by a permutation test. It can be shown that for a chromosome with k markers, the recombination fractions between the adjacent markers being $r_1, r_2, \ldots, r_{k-1}$, then the probability that at least one remains segregating is $\sim(1/2)^t[k - \Sigma_i(1 - r_i)^t]$, a result obtained by using transition probability matrices such as **B** (Equation 6) or the methods of Visscher and Thompson (1995) and by ignoring double recombinants. In any case, if the parental lines differ in mean performance and if there is evidence of segregation variance in the initial backcross or an F_2 of the lines, it is moot whether the genome-wide null hypothesis of no effects is relevant. It might be more appropriate to assume an infinitesimal model (Visscher and Haley 1996), with the variance distributed equally among and within the chromosomes: it seems likely, however, since the net difference between the two lines is therefore likely to be small around any marker, that the probabilities of continued segregation of each marker will be little higher than in the neutral case unless the variance of aggregate effects among marker intervals is large.

It is possible, at least in principle, to estimate the effects of a QTL located near a marker from the number of replicate lines in which it is segregating. Considering first individual QTL, the expected proportion of lines in which it is segregating is given by $P(n,a)_t$, which can be equated to the actual proportion m/M (this is the maximum likelihood estimator). If there is no recombination between the marker and QTL, an estimate, $\hat{a}$, of the effect can then be obtained by trial and error,

TABLE 7

Simulated examples of patterns of markers for different models of QTL distribution with the same total effect on a chromosome of length 100 cM, with 11 equally spaced markers at 0, 10, ..., 100 cM.

1 QTL, $a = 2$ at 35 cM	2 QTL, $a = 1$ at 25, 45 cM	5 QTL, $a = 0.4$ at 15, 25, ..., 55 cM	10 QTL, $a = 0.2$ at 5, 15, ..., 95 cM
000**00**000000 × 4[a]	00**0000**00000 × 7	**000000**0000 × 5	**00000000000** × 13
001**11**000000	01**1111**11111	0**000010**0000	**10001111111**
001**10**000000	00**1111**00000	0**000000**1110	**01111000000**
000**01**000000	00**0010**00000	0**0011110**111	**11111110000**
111**11**000000	11**100**11111	0**011111**1100	**00111101100**
001**11**000000	00**1110**00000	0**0011111**100	**00000001110**
000**01**110000	00**0001**10000	1**1110000**000	**11111110000**
010**00**000000	01**1110**00000	0**1101110**000	**00100110000**
000**01**000000	00**1110**00000	0**0110000**000	
000**11**000000	00**0001**00000	1**1000000**000	
000**11**100000	01**1111**00000	1**0000000**000	
000**10**000000	00**1110**00000	1**111110**000	
000**11**000000	01**001**100000	1**1000000**000	
001**11**000000	00**0111**10001	1**1111000**000	
011**10**000000		0**000001**1000	
001**11**000000		0**0010000**000	
000**10**000000			

[a] Number of occurrences of haplotype of recurrent parent. Results for other replicates are given in the ordered sampled. Boldface shows flanking markers for the QTL.

$n = 12$, $t = 8$ and $M = 20$ lines. Symbols show the marker haplotype of selected individual; 11111111111 is the marker haplotype of the nonrecurrent and 00000000000 that of the recurrent inbred parent.

evaluating Equations 1 and 3. As an approximation, Equations 2 and 3 can be used together to give $[(1 + i\hat{a}/2)/2]^t = m/M$ or $\hat{a} = (2/i)[2(m/M)^{1/t} - 1]$. Because the estimate is obtained from a realization of the Poisson distribution with a small parameter value, the standard error of the estimate is likely to be of the same size as the estimate itself, which can be considered as no more than a guide. The estimate also has to be corrected for recombination between the marker and the QTL. In principle, but beyond the scope of this paper, interval mapping (Lander and Botstein 1989) could be used to combine the data on markers, but sampling errors will remain large. Information on the effect of the QTL can also be obtained directly from segregation analysis within the backcross families at the end of the backcrossing phase, those retaining a QTL of large effect having both higher mean and higher within family variance than those in which it is lost. Further precision can be obtained by typing progeny for the marker(s) near the putative QTL.

A QTL can be mapped solely using the marker information on single individuals in each family at the end of the backcrossing phase. Consider the example in the first column of Table 7, and assume, as was actually the model simulated, that there was only one QTL on the chromosome, and that the lines in which the QTL was segregating could be identified from their mean and variance. In the 20 replicate lines, markers *1*, *2*, ..., *7* were segregating in, respectively, 1, 3, 7, 12, 11, 2, and 1 lines, and none of markers *8–10* remained segregating, so the most likely QTL position is between markers *4* and *5*. Among the 18 lines segregating at markers *4* or *5*, the QTL (let us assume identified from phenotypes for the trait) was segregating at 17; of these 18 lines, nine were recombinants between marker *4* (map position 30 cM) and the QTL, and five were recombinants between the QTL and marker *5* (map position 40 cM). Hence, the maximum likelihood estimate of QTL position can be shown to be at 36.7 cM, close to its actual position at 35 cM. The precision was achieved by typing solely 20 genotypes, but obviously, this is an extreme example of a QTL of very large effect retained segregating over many (eight) generations. More generally, the precision will be a function of the number of lines in which the QTL remains segregating, those in which it and the markers are lost providing none, and the number of generations for which backcrossing is continued, the effective recombination rate being $1 - (1 - r)^t$. There is, however, a trade-off because increasing the number of generations leads to a reduction in probability of segregation and an increase in the effective recombination rate.

It would be possible to obtain more information about QTL positions and effects if the marker screening were conducted during the backcrossing program, and this would also enable decisions to be made as to when to cease backcrossing to optimize the trade-off between recombination and loss of QTL. The records of individual animals for the quantitative trait and for each marker can be combined to provide further information using

maximum likelihood methods that are computationally feasible using Markov chain Monte Carlo methods, such as Gibbs sampling, and have been used in an analysis of recurrent backcross lines in which a set of different marker regions were maintained segregating (Rance *et al.* 1997). The simple scheme analyzed here, in which marker information is collected only at the end, is more appropriate for laboratory species with short generation intervals than for commercial animal or crop species. In such cases, it is likely to be preferable to collect marker data each generation, and perhaps to also choose individuals for backcrossing on their marker genotype as well as on phenotype for the trait of interest. This leads to more complicated design and interpretation problems than are discussed here.

If only QTL on autosomes are to be identified, it should not matter greatly whether a male or female is selected for the next generation of backcrossing. If, however, QTL on the sex chromosome are to be located, then females should obviously be selected in the backcross line; if map lengths in females are greater than in males, there is an added benefit in doing so. There are other practical issues that have not been considered. For example, there is a risk that the selected individual in a line is infertile or that none of the required sex are available for selection. In such cases, it may be appropriate to initiate many more lines than are expected to be maintained, or to sacrifice the simplicities of having completely independent backcross lines by drawing sublines from surviving lines to maintain numbers. The analysis can be clearly developed further.

With recurrent backcrossing and selection to only one line, QTL that are recessive in the nonrecurrent parent will be missed. Although alternating backcrossing and *inter se* mating alleviates this problem, net selective pressures on recessives remain small. An alternative, if both lines are inbred, is to practice backcrossing and selection in two reciprocal sets of lines, differing in which is used as the recurrent parent. The probability that the QTL is maintained segregating in each type of line and over the whole set can readily be computed from the methods given here.

It is important to note that the method discussed in this and other studies (*e.g.*, N. J. Schork, A. M. Beebe, B. Thiel, P. St. Jean, and R. L. Coffman, unpublished data) on the use of recurrent backcrossing paper is essentially a prescreening procedure for QTL detection, not a finishing point. When the recurrent backcrossing with selection on phenotype for the quantitative trait is completed, and regions of the genome which the marker analysis indicates that QTL for the trait are likely to be present, more detailed analysis is needed; the congenic lines, however, provide a useful starting point. For example, further backcrossing can be practiced while maintaining segregating-only specific short marker intervals. Segregation within the families and QTL and marker recording can be undertaken to confirm QTL effects and more closely map their position. Progeny testing of recombinants can be used to make this procedure more accurate (P. D. Keightley, personal communication).

For precise mapping of QTL, opportunity is needed for substantial recombination between linked QTL and between them and markers. The advantage of multiple generation schemes is that, in effect, recombination fractions are increased roughly in proportion to the number of generations, and similarly, the average length of chromosome retained around a marker is reduced in inverse proportion to the number of generations. The method discussed here has advantages and disadvantages over others for QTL location. In the most conventional, QTL are mapped by recording phenotypes and marker genotypes in an F_2 or backcross of a line cross. An additional experiment is subsequently required to isolate these further, which can involve introgression or retention of marked segments by backcrossing; this contrasts with the backcrossing with selection proposed here in that it is the markers that are identified in the backcrossing rather than the phenotypes. Even so, a number of lines have to be retained for each QTL because the precise relation between marker and QTL position is not known. The use of markers, however; has the benefit that QTL of smaller effect can be retained, and as marker and trait data are collected throughout, information on the location accumulates during the backcrossing phase, but at the expense of a lot of recording, compared to backcrossing with selection on the trait. An alternative approach is to proceed to QTL mapping of more advanced generations of the cross, for example, F_3, F_4,..., before undertaking the QTL analysis; however, this may still require backcrossing to establish congenic lines for gene cloning. (Recurrent backcrossing rather than *inter se* mating is not feasible without selection because most QTL would be lost.) A further alternative for multigeneration analysis is the use of selected lines and identification of QTL, preferably replicated by changes in marker frequencies between high and low lines (Keightley and Bulfield 1993); however, this still requires subsequent backcrossing if congenic lines are needed for precise mapping and cloning.

The objective of this paper is not to show that the use of recurrent backcrossing with selection with several independent single family backcross lines is optimal in any broad way. It is indeed clear that if more effort were expended on recording markers during the backcrossing phase and associating them with performance of the trait, and perhaps using this information to generate sublines on a dynamic basis, further precision could be obtained; the analysis of such a scheme, however, is beyond the scope of this paper. The aim is merely to suggest a method with relatively low input of effort that may lead to fairly clear identification of QTL of large effect and simultaneous production of congenic lines

that may be useful for further, more detailed analysis. The idea of backcrossing with selection is old indeed, and its successful use to identify QTL for disease resistance has been reported (Beebe *et al.* 1997). The only novelty is in the use of independent sublines and the quantification of the probabilities that QTL are retained. N. J. Schork, A. M. Beebe, B. Thiel, P. St. Jean and R. L. Coffman (unpublished data) discuss methods for analysis of nonindependent backcross lines with selection and use of markers.

I am grateful to Philippe Baret, Peter Keightley, Sara Knott, Peter Visscher, Zhao-Bang Zeng, and two anonymous referees for helpful comments, and to the Biotechnology and Biological Sciences Research Council for financial support.

LITERATURE CITED

Alpert, K. D., and S. D. Tanksley, 1996 High-resolution mapping and isolation of a yeast artificial chromosome containing fw2.2: a major fruit weight quantitative trait locus in tomato. Proc. Natl. Acad. Sci. USA **93:** 15503–15507.

Beebe, A. M., S. Mauze, N. J. Schork and R. L. Coffman, 1997 Serial backcross mapping of multiple loci associated with resistance to Leishmania major in mice. Immunity **6:** 551–557.

Charlier, C., F. Farnir, P. Berzi, P. Vanmanshoven, B. Brouwers *et al.*, 1996 Identity-by-descent mapping of recessive traits in livestock: application to map the bovine *syndactyly* locus to chromosome 15. Genome Res. **6:** 580–589.

Darvasi, A., 1997 Interval-specific congenic strains (ISCS): an experimental design for mapping a QTL into a 1-centimorgan interval. Mamm. Genome **8:** 163–167.

Démant, P., and A. A. M. Hart, 1986 Recombinant congenic strains— a new tool for analyzing traits determined by more than one gene. Immunogenetics **24:** 416–422.

Dietrich, W. F., J. Miller, R. Steen, M. A. Merchant, D. Damron-Boles *et al.*, 1996 A comprehensive genetic map of the mouse genome. Nature **380:** 149–152.

Falconer, D. S., and T. F. C. Mackay, 1996 *Introduction to Quantitative Genetics.* Ed. 4, Longman, Harlow, UK.

Guo, S. W., 1995 Proportion of genome shared identical by descent by relatives: concept, computation and applications. Am. J. Hum. Genet. **56:** 1468–1476.

Haley, C. S., and S. A. Knott, 1992 A simple regression method for interval mapping in line crosses. Heredity **69:** 315–324.

Hill, W. G., 1969 On the theory of artificial selection in finite populations. Genet. Res. Camb. **13:** 143–163.

Houwen, R. H. J., S. Baharloo, K. Blankenship, P. Raeymaekers, J. Juyn *et al.*, 1994 Genome screening by searching for shared segments: mapping a gene for benign recurrent intrahepatic cholestasis. Nature Genet. **8:** 380–386.

Jansen, R. C., 1993 Interval mapping of multiple quantitative loci. Genetics **135:** 205–211.

Keightley, P. D., and G. Bulfield, 1993 Detection of quantitative trait loci from frequency changes at marker loci under selection. Genet. Res. **62:** 195–203.

Keightley, P. D., T. Hardge, L. May and G. Bulfield, 1996 A genetic map of quantitative trait loci for body weight in the mouse. Genetics **142:** 227–235.

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

Ollivier, L., L. A. Messer, M. F. Rothschild and C. Legault, 1997 The use of selection experiments for detecting quantitative trait loci. Genet. Res. **69:** 227–232.

Rance, K. A., S. C. Heath and P. D. Keightley, 1997 Mapping quantitative trait loci for body weight on the X chromosome in mice. II. Analysis of congenic backcrosses. Genet. Res. **70:** 125–133.

Snell, G. D., 1958 Histocompatibility genes of the mouse. II. Production and analysis of isogenic resistant lines. J. Natl. Cancer Inst. **21:** 843–877.

Tanksley, S. D., and J. C. Nelson, 1996 Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor. Appl. Genet. **92:** 191–203.

Thomas, A., M. H. Skolnick and C. M. Lewis, 1994 Genomic mismatch scanning in pedigrees. IMA J. Math. Appl. Med. Biol. **11:** 1–16.

Visscher, P. M., and S. C. Haley, 1996 Detection of putative quantitative trait loci in line crosses under infinitesimal genetic models. Theor. Appl. Genet. **93:** 691–702.

Visscher, P. M., and R. Thompson, 1995 Haplotype frequencies of linked loci in backcross populations derived from inbred lines. Heredity **75:** 644–649.

Wright, S., 1952 The genetics of quantitative variability, pp. 5–14 in *Quantitative Inheritance*, edited by E. C. R. Reeve and C. H. Waddington. Her Majesty's Stationery Office, London, UK.

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

Communicating editor: Z-B. Zeng