Effect of Within-Strain Sample Size on QTL Detection and Mapping Using Recombinant Inbred Mouse Strains

J. K. Belknap


Increasing the number of mice used to calculate recombinant inbred (RI) strain means increases the accuracy of determining the phenotype associated with each genotype (strain), which in turn enhances quantitative trait locus (QTL) detection and mapping. The purpose of this paper is to examine quantitatively the effect of within-strain sample size \( n \) on additive QTL mapping efficiency and to make comparisons with \( F_2 \) and backcross (BC) populations, where each genotype is represented by only a single mouse. When 25 RI strains are used, the estimated equivalent number of \( F_2 \) mice yielding the same power to detect QTLs varies inversely as a function of the heritability of the trait in the RI population \( h_{RI} \). For example, testing 25 strains with \( n = 10 \) per strain is approximately equivalent to 160 \( F_2 \) mice when \( h_{RI} = 0.2 \), but only 55 when \( h_{RI} = 0.6 \). While increasing \( n \) is always beneficial, the gain in power as \( n \) increases is greatest when \( h_{RI} \) is low and is much diminished at high \( h_{RI} \) values. Thus, when \( h_{RI} \) is high, there is little advantage of large \( n \), even when \( n \) approaches infinity. A cost analysis suggested that RI populations are more cost-effective than conventional selectively genotyped \( F_2 \) populations at \( h_{RI} \) values likely to be seen in behavioral studies. However, with DNA pooling, this advantage is greatly reduced and may be reversed depending on the values of \( h_{RI} \) and \( n \).

KEY WORDS: Recombinant inbred strain; quantitative trait locus; chromosome mapping; quantitative genetics; mouse.

INTRODUCTION

A common question in the design of experiments is to how to allocate subjects to cells in an experimental design to obtain the highest relative efficiency (RE), where RE reflects the power to detect treatment effects (smallest error term) when different designs of the same sample size are compared (Sokal and Rohlf, 1995). RE can also be expressed as the ratio of sample sizes for experimental designs yielding the same power. For example, if we are limited to 100 mice, is it better to test 100 recombinant inbred (RI) strains with one mouse per strain or 25 strains with four mice per strain? In general, the first is the better choice (Sokal and Rohlf, 1995; Knapp and Bridges, 1990), but when every available strain is routinely tested, as is usually the case in the mouse, there is no option concerning the number of RI strains. The only choice is the selection of the within strain sample size, or \( n \). In this paper, the RE at varying values of \( n \) is examined when the number of strains is fixed at the maximum number available, and comparisons are made with segregating \( F_2 \) or BC populations.

For QTL mapping studies in the mouse, the major disadvantage of RI strains compared to segregating populations is the limited number of gen-
otypes available. Since each genotype is represented by a single RI strain, the largest existing RI sets, the B × D, A × B/B × A, and LS × SS sets, are limited to no more than 26 or 27 distinct genotypes (Taylor, 1995; DeFries et al., 1989). Since one or two strains per set often reproduce too poorly to include in most RI studies, 25 strains is a reasonable upper limit for the existing mouse RI sets. In contrast, segregating (F₂ or BC) populations can involve any number of genotypes, since each genotype is represented by a single RI strain, the largest existing RI sets, the B × D, A × B/B × A, and LS × SS sets, therefore, \( N_{RI} = n \times N_{str} \). When \( n \) is unequal, average \( n \) can be substituted.

Quantitative Genetic Considerations. When data from individual mice (not strain means) are used in quantitative genetic analyses, we can partition the variance in the usual way for an RI population as follows. This partitioning is similar to that in a segregating population when there is no dominance variation [Eq. (A)]: \( V_P = V_A + V_E \), and the heritability \( h^2_A \) is given by \( V_A/(V_A + V_E) \), where \( V_P \) is the phenotypic (trait) variance, \( V_A \) is the additive genetic component of variance, and \( V_E \) is the environmental component of variance. The value of \( h^2_A \) can be estimated in several ways. The first is to use \( R^2 \) from a one-way ANOVA by RI strain, or \( SS_{mean}/SS_{total} \). The second way is to use components of variance between and within strains calculated from the same one-way ANOVA (Hegmann and Possidente, 1981; Belknap et al., 1996). The third method, and perhaps the simplest, is to use the variance of strain means to estimate \( V_A \) and divide by \( V_P \). In this case, adjustments are often needed to correct for the fact that the variance of strain means contains a portion of \( V_E \) (Hegmann and Possidente, 1981).

\[ V_{QTL} \] is the additive genetic variance due to a QTL and is calculated in an RI population as \( (M_{A_1} - M_{A_2})^2/4 \), where \( M_{A_1} - M_{A_2} \) is the difference in phenotypic means between the two homozygote classes at a QTL or closely linked marker. In Falconer’s terminology, \( M_{A_1} - M_{A_2} \) is equal to twice the average effect of a single gene substitution (Falconer and Mackay, 1996). One-half of this value, or \( (M_{A_1} - M_{A_2})^2/8 \), gives an estimate of \( V_{QTL} \) to be expected in an F₂ population for the same QTL, and half of the F₂ estimate gives the expected BC estimate (Kearsey and Pooni, 1996). To determine \( h^2_{QTL} \), these \( V_{QTL} \) estimates are divided by \( V_P \), the phenotypic variance in each population.

While \( V_A \) in an RI population can be expected to be double that in a comparable F₂ and quadruple that in a BC (Kearsey and Pooni, 1996), what can we expect for the heritability? The heritabilities in each population for the same trait will also differ approximately in proportion to \( V_A \), that is, \( h^2_A \approx 2h^2_{F2} \approx 4h^2_{BC} \), if \( V_P \) remains about the same in all three populations. However, \( V_P \) may not be equal, especially when \( h^2_A \) is large. The twofold greater value of \( V_A \) in RI vs. F₂ populations (i.e., \( V_{A(RI)} = 2V_{A(F2)} \)) can be expected to cause \( V_P \) to be larger in RI populations by an amount equal to \( 1/2V_{A(RI)} \).
Thus, \( V_{RF2} = V_{RRI} - 1/2V_{AR2} \). From this, a more accurate estimate of \( h_{Rt}^2 \) from RI data can be obtained by taking into account the expected inequality in \( V_s \), as follows [Eq. (B)]:

\[
  h_{Rt}^2 = 1/2V_{AR2}/(V_{RRI} - 1/2V_{AR2}) = 1/2h_{Rt}^2(1 - 1/2h_{Rt}^2).
\]

[For a BC, \( h_{tBC}^2 = 1/4h_{Rt}^2(1 - 3/4h_{Rt}^2) \)]. Therefore the ratio of \( h_{Rt}^2/h_{Rt}^2 \) will be \( 2(1 - 1/2h_{Rt}^2) \) rather than 2. The same is also true at the QTL level; the heritability of a QTL (\( h_{QTL}^2 \)), or \( V_{QTL}/V_e \), in an RI population, is expected to be somewhat less than double that in a comparable \( F_2 \) and quadruple that in a comparable BC. More accurately, when the inequality of \( V_s \) noted above is accounted for [Eq. (C)],

\[
  h_{QTL,RI}^2 = 1/2h_{QTL,AR2}(1 - 1/2h_{Rt}^2) = 1/2h_{QTL,AR2}(1 - 1/2h_{Rt}^2).
\]

The above equations assume that each individual mouse is a data point for ANOVA, and that \( V_e \) is approximately the same in RI, \( F_2 \), and BC populations. The equality of \( V_e \) assumption may be reasonable for some behavioral traits and not for others (e.g., Hyde, 1973).

**Effect of Within-Strain Sample Size, n.** What happens when \( n > 1 \) and strain means are used in the analysis rather than individual mice? In this case, the genotypic value is the mean of measurements on \( n \) mice per strain, providing a more accurate assessment of the phenotype associated with each genotype compared to \( n = 1 \). For RI strains, the variance partitioning based on strain means (\( \bar{x} \)) is as follows [Eq. (D)]:

\[
  V_{px} = V_A + V_e/n \quad \text{(Soller and Beckmann, 1990)},
\]

and the heritability (\( h_{Rt}^{-n} \)) is given by \( V_A/(V_A + V_e/n) \), where \( V_{px} \) is the phenotypic variance of strain means and \( h_{Rt}^{-n} \) is the heritability of strain means. \( h_{Rt}^{-n} \) reflects the degree to which the variance of strain means, \( V_{px} \), is due to genetic sources of variation. As \( n \) increases, the contribution of the environmental component to \( V_{px} \) decreases by a factor of \( 1/n \) (Soller and Beckmann, 1990). This has the effect of increasing the heritability based on strain means in an RI population as a function of \( n \). As \( n \) becomes very large, \( h_{Rt}^{-n} \) approaches 1.0 because the contribution of the environmental component of variance is approaching zero, causing \( V_{px} \) to approach \( V_A \) in value. When this happens, the variance of phenotypic strain means, \( V_{px} \), provides a good estimate of \( V_A \). However, in many reports in the literature, it is often overlooked that this estimate is biased upward when \( h_{Rt}^{-n} \) is considerably less than unity. Corrections for this bias can be made by multiplying \( V_{px} \) by the estimate of \( h_{Rt}^{-n} \) taken from Fig. 1, which eliminates this source of bias. Much the same is true for standard (non-RI) inbred strains when the analysis is carried out within and between strains in the same manner.

The same relationship between \( h_{Rt}^2 \) and \( h_{Rt}^{-n} \), outlined above also applies to the heritability of a QTL based on individual mice compared to that calculated from strain means. The heritability of a QTL is the proportion of the phenotypic variance due to a QTL and is given by \( h_{QTL}^2 = V_{QTL}/V_e = V_{QTL}/(V_A + V_e) \) when data from individual mice are the basis for the analysis. When strain means are used, the heritability of a QTL becomes [Eq. (E)]:

\[
  h_{QTL,RI}^2 = V_{QTL}/V_{px} = V_{QTL}/(V_A + V_e/n),
\]

which increases as \( n \) increases in a directly parallel manner to \( h_{Rt}^{-n} \). In other words, as \( n \) increases, the proportionate increase in \( h_{Rt}^{-n} \) and \( h_{QTL,RI}^2 \) will be the same; thus the present analysis applies to both.

Using Eqs. (A) and (D) above, the relationship between \( h_{Rt}^{-n} \) and \( h_{Rt}^2 \) is shown in Fig. 1 as a function of \( n \). Six values of \( n \) are plotted: 1, 4, 6, 10, 15, and \( \infty \). The ratio of \( h_{Rt}^{-n}/h_{Rt}^2 \) represents the gain in heritability in an RI set when strain means are
used compared to when individual mice are used in the analysis, as in a segregating population. This ratio is plotted in Fig. 2 (inset) as a function of \( h_{RI}^2 \).

The same \( h_{RI}^2 / h_{QTL}^2 \) ratios in Fig. 2 (inset) also apply to the heritability of a QTL, or \( h_{QTL}^2 \). This is because the ratio \( h_{RI}^2 / h_{QTL}^2 \) is equal to the ratio \( h_{QTL}^2 / h_{QTL}^2 \), i.e., the effect of increasing \( n \) is proportionately the same for both the heritability of all QTLs in the aggregate and the heritability of an individual QTL.

**Equivalent F₂ Sample Size.** From the \( h_{RI}^2 / h_{QTL}^2 \) ratio, it is possible to estimate the approximate equivalent \( F₂ \) sample size, \( N_{F₂} \), when 25 RI strains are tested with varying \( n \). This estimated \( N_{F₂} \) will have the same power to detect additive effects of a QTL as the RI population. [Methods for calculating power for \( F₂ \) populations are given by Darvasi and Soller (1992, 1994), and for RI populations by Belknap et al. (1996).] Since \( N_{F₂} \) required is proportional to the heritability of a QTL (Lander and Botstein, 1989; Soller and Beckmann, 1990; Belknap et al., 1996), this estimate was calculated by multiplying the appropriate \( h_{QTL}^2 / h_{QTL}^2 \) ratio (which equals the \( h_{QTL}^2 / h_{QTL}^2 \) ratio) from Fig. 2 (inset) times 25 (\( N_{strain} \)) times 2(1 - 1/2\( h_{QTL}^2 \)), the ratio of \( h_{QTL}^2 / h_{QTL}^2 \) when \( n = 1 \) for both. Thus [Eq. (F)], \( N_{F₂} = 2(1 - 1/2h_{QTL}^2)h_{QTL}^2/2h_{QTL}^2N_{strain} \). These \( N_{F₂} \) estimates are shown in Fig. 2 as a function of \( h_{QTL}^2 \). The plotted values presume that all phenotyped mice are also genotyped for both RI and \( F₂ \) populations. If selective genotyping is used in an \( F₂ \) population, where only the extreme tails of the trait distribution are genotyped to reduce cost (Lander and Botstein, 1989; Darvasi and Soller, 1992), the equivalent \( N_{F₂} \) will need to be larger than shown in Fig. 2 to offset the loss in power due to the restricted sample (Lander and Botstein, 1989). For example, with a selection fraction (both tails) of 10, 20, 30, 40, or 50% of the total population, the plotted \( N_{F₂} \) values shown in Fig. 2 should be increased by a factor of 2.3-, 1.5-, 1.3-, 1.15-, or 1.08-fold, respectively (Darvasi and Soller, 1992).

**Relative Efficiency Ratio (RE).** An important question is the relative efficiency (RE) of \( F₂ \) vs. RI populations, which can be expressed as the ratio of \( N_{F₂}/N_{RI} \) when the power to detect additive QTLs is the same in both populations. Thus, \( F₂ \) populations are more efficient (require fewer mice) than RI populations when RE is <1.0 and less efficient when RE >1.0. This ratio can be calculated by determining the ratio of the estimated equivalent \( F₂ \) \( N \) for an RI population (Fig. 2) divided by the actual \( N \) used in the RI population. These results are shown in Fig. 3 in the case of no selective genotyping. When selective genotyping is used, the plotted RE values should be multiplied by a factor that varies with the selection fraction, as noted in the previous section.

For example, in the absence of selective genotyping, the estimated equivalent \( F₂ \) \( N \) when \( h_{RI}^2 = 0.4 \) and \( n = 6 \) for 25 RI strains is 80 (from Fig. 2), while the actual RI \( N \) is 25 × 6 = 150. The RE for the RI experiment is thus 80/150 = 0.53, or 53%. Thus, an RI population under these conditions is only about half as efficient (requires twice as many mice) as an \( F₂ \) of the same power. If selective genotyping is used in the \( F₂ \) where (for ex-
ample) 1/8 of the population is genotyped (1/16 at each tail), the plotted \(N_{F2}\) values should be increased by 2.0-fold to offset loss of power (Darvasi and Soller, 1992), yielding \(N_{F2} = 160\). In this case, \(RE = 160/150 = 1.07\), thus RI has a slight advantage (requires 7% fewer mice). As can be seen in Fig. 3, \(RE\) is determined by \(h_{RI}\), \(n\), and the \(F_2\) selection fraction. \(RE\) (or \(N_{F2}/N_{RI}\) when power is equal) is increased when either \(n\), \(h_{RI}\), or the selection fraction is reduced in magnitude, and vice versa. The same considerations for \(h_{RI}\) also apply to the heritability of a QTL (not shown).

The newest genotyping cost-saving measure for segregating populations is DNA pooling. The extreme ends of the trait distribution are genotyped, as in conventional selective genotyping, except that DNA pooled from all individuals in an extreme tail is genotyped rather than individual mouse samples. Darvasi and Soller (1994) discuss the cost benefit and genetic implications of this approach in the general case. Since the PCR reaction is only semiquantitative, a loss in accuracy occurs in estimating allele frequencies in each tail, resulting in an increase in both Type I (false positive) and Type II (false negative) errors (reduced power) that can vary widely from marker to marker. A further power loss occurs because interval mapping, as implemented by standard programs such as MapMaker QTL, cannot be used with pooled data. (If a high density of markers is used, this power loss is minimal, but the additional genotyping required diminishes the cost savings.) To offset these shortcomings, (1) DNA pooling can be used as a low-cost preliminary screen of the genome, followed by individual genotyping for only those markers (say, 15%) showing evidence of QTL linkage (e.g., Taylor and Phillips, 1996), or (2) a larger \(F_2\) population can be grown and phenotyped to offset the loss in power. (In a cost analysis example given below, it is assumed that a 20% larger population is needed.) With either option, the overall cost remains roughly the same in our example (calculations not shown).

Relative Cost per Mouse Ratio (RC). In addition to the relative efficiency, which reflects the number of \(F_2\) vs. RI mice needed to obtain the same power, the relative cost per mouse, expressed as a ratio (\(F_2/RI\)), is also important in answering questions about the overall cost efficiency of \(F_2\) vs. RI populations. We refer to this ratio as RC, in parallel with RE. Table I shows the costs per mouse in \(F_2\) and RI populations based on data from our laboratory, both with and without \(F_2\) selective genotyping, and with or without DNA pooling. The optimal selection fraction to minimize overall cost has been estimated by Darvasi and Soller (1992) in the general case. For the data shown in Table I, where the ratio of genotyping cost to all other costs in the absence of selective genotyping was just over 13 ($163/$12.30), 1/8 is about the optimal fraction, or the extreme 1/16 at each end of the trait distribution. The “bottom line” from Table I is as follows: the total cost for an RI mouse was $12.30 (breeding, raising, and phenotyping); for an \(F_2\) mouse without selective genotyping, $175.30; for an \(F_2\) mouse with optimal selective genotyping, $31.80; and for an \(F_2\) mouse with DNA pooling coupled with selective genotyping, $13.40. The relative cost ratio per mouse (\(F_2/RI\), or RC) is therefore 14.25 without selective genotyping, 2.59 with conventional selective genotyping, and 1.09 with DNA pooling and selective genotyping combined.
Relative Total Cost Ratio (RTC). The total cost (TC) of an experiment is given by the number of mice multiplied by the cost per mouse. The relative total cost ratio (RTC) of \( F_2/RI \) is thus \( \text{RTC} = \frac{\text{RE} \times RC}{RC} \) when the power to detect QTLs is the same in both populations. For example, when \( h^2_{\text{RI}} = 0.4 \) (\( h^2_{\text{F2}} = 0.25 \)) and \( n = 10 \), typical values for many behavioral traits, \( F_2 \) populations without selective genotyping will be \( \text{RTC} = \frac{0.35 \times 14.25}{1.09} \) (from Table I), or five times as costly compared to RI populations of the same power.

When selective genotyping is used, RC declines to 2.59 in our example (Table I). In this case, \( \text{RTC} = \frac{\text{RE}}{\text{RC}} \times 2 \) (2.59). The factor of 2 for RE is to offset the loss in power when selective genotyping of this magnitude (1/8) is practiced. Multiplying by 2 gives the RE value expected if no selective genotyping was used. (This is necessary because the RE values shown in Fig. 3 presume no selective genotyping.) For example, when \( h^2_{\text{RI}} = 0.4 \) and \( n = 10 \), RT = RE (0.35 from Fig. 3) \( \times 2.59 \) (from Table I) = 1.8. Therefore, \( F_2 \) populations will be 1.8 times as costly compared to RI populations of the same power. For \( n = 6 \), RTC rises to almost threefold. Figure 4 shows RTC values as a function of \( h^2_{\text{RI}} \) for \( n = 6 \) and 10, based on RE values taken from Fig. 3 and RC from Table I. Note that RTC will increase (\( F_2 \))
will cost relatively more than equivalent RI populations) when either $h_{RI}^2$ or $n$ decreases, and vice versa.

With selective genotyping and DNA pooling, $RC$ declines to 1.09 in our example (Table I). As before, $RTC = RE \times RC$, or $RE$ (from Fig. 3 × 2.4) × $RC$ (1.09). The factor of 2.4 reflects the greater number of mice needed to offset the loss in power due to selective genotyping (2.0-fold) and DNA pooling (1.2-fold) used together.

The horizontal dashed line in Fig. 4 shows $RTC = 1.0$, when both populations have the same cost for the same power. Where this line intersects the curved lines for $n = 6$ or 10 gives the equicost values of $h_{RI}^2$, where both populations are equal in power and cost. For conventional selective genotyping, this occurs when $h_{RI}^2 = 0.90$ or 0.65 for $n = 6$ or 10, respectively. $F_2$ populations will be more cost-effective (cost less) than RI populations above the equicost $h_{RI}^2$ values and less cost-effective below them. For selective genotyping coupled with DNA pooling, the corresponding “break-even” $h_{RI}^2$ values are 0.55 ($h_{F2}^2 = 0.38$) and 0.37 ($h_{RI}^2 = 0.23$) for $n = 6$ and 10, respectively, which fall within the range often seen for behavioral traits. Thus, the question of which population is more cost-effective will depend on observed values of $h_{RI}^2$ for a given $n$.

**DISCUSSION**

Typical values of $n$ reported in the RI behavioral literature are 6–10, and $h_{RI}^2$ is often in the range of 0.3 to 0.6 (reviewed by Belknap et al., 1997a), which is approximately equivalent to $h_{F2}^2$ of 0.18 to 0.43 from Eq. (B). The approximate equivalent $F_2$ $N$ conferring the same power for this frequently encountered range is about 50 to 115 mice when 25 RI strains are used. For QTL mapping purposes, this estimated $F_2$ $N$ is capable of detecting only the very largest QTLs at Lander and Schork (1994) significance levels (Belknap et al., 1997a). To detect smaller effect QTLs reliably, newer strategies have been developed involving RI and additional mapping populations (e.g., Johnson et al., 1992; Plomin and McClearn, 1993; Belknap et al., 1996, 1997a, b; Crabbe et al., 1994). The most common strategy is a two-step approach, where an RI population is used as a preliminary screen of the genome for provisional QTLs at relatively relaxed $\alpha$ levels of (usually) .01 or .05 (Step 1), followed by confirmation testing of each RI-implicated chromosomal region in a large $F_2$ population derived from the same progenitors (Step 2). The advantages and disadvantages of this two-step approach have been recently reviewed (Belknap et al., 1996, 1997a).

Increasing $n$ increases the power to detect QTLs, but the relative efficiency diminishes rapidly as $n$ exceeds 4, especially at higher values of $h_{RI}^2$ (Figs. 2 and 3). For example, the increase in the estimated equivalent $F_2$ $N$ as $n$ is increased from 4 to 10 (a 2.5-fold increase) is 1.7-fold (70%) at $h_{RI}^2$. 

![Fig. 4. The relative total cost (RTC = RE × RC) of an F_2 vs. RI experiment yielding the same power to detect additive QTLs based on RE from Fig. 3 and RC from the cost analysis shown in Table I. The results are shown for both conventional selective genotyping (main body) and selective genotyping and DNA pooling combined (inset). Plotted values assume within strain sample sizes (n) of either 6 or 10. The dashed horizontal lines show the equicost conditions (i.e., RTC = 1.0), where both populations have equal power and equal cost. The results shown are typical in our analysis of over 30 behavioral traits but can vary widely from laboratory to laboratory or trait to trait. For example, if genotyping costs are double those shown in Table I, then the plotted relative total cost values (Y axis) should be multiplied by 1.65 (conventional selective genotyping) or 1.24 (DNA pooling). In contrast, if genotyping costs are half those shown in Table I, the correction factors are 0.71 and 0.94, respectively. Note that the total costs associated with DNA pooling are relatively little affected by fluctuations in genotyping costs compared to conventional selective genotyping.](image-url)
= 0.1 but is only about 4% when \( h^2_{RI} = 0.8 \). When \( h^2_{RI} = 0.3 \) to 0.6, as is often the case in the behavioral literature, the increase in equivalent \( F_2 \) \( N \) is only about 1.1- to 1.3-fold (10–30%) in response to the 2.5-fold (150%) increase in \( n \) (and \( N \)). Moreover, inspection of Figs. 2 and 3 shows that the added burden of testing 15 or more mice per strain, compared to only 6 or 10, for traits with high \( h^2_{RI} \) (say, >0.5) is probably not economically justified. While increasing \( n \) under these conditions is not efficient in terms of animal numbers, perhaps a more important question is, Is it cost-effective? This question is discussed below.

Since QTL mapping is inherently a large-scale enterprise, the costs per trait are high and often beyond the resources of many laboratories. This, in turn, inhibits progress. For this reason, the study of relative costs of one experimental design vs. another for QTL detection is an especially important consideration. As an example of a cost analysis of \( F_2 \) vs. RI populations, data from our laboratory are presented in Table I. They roughly follow the cost efficiency analysis explicated by Sokal and Rohlf (1995).

We compared the cost of an individual RI mouse vs. an \( F_2 \) mouse for QTL detection using the cost structure shown in Table I. While all phenotyped RI mice are also genotyped, this is generally not the case in segregating populations. A common practice to minimize \( F_2 \) genotyping costs is to employ conventional selective genotyping, where only individual mice at the extreme ends of the trait distribution are genotyped (Lander and Botstein, 1989; Darvasi and Soller, 1992). While effective in dramatically reducing costs, selective genotyping has several disadvantages that must also be considered. Mapping accuracy and power are somewhat reduced (Darvasi and Soller, 1992; Darvasi, 1997) and the newer and more powerful multiple regression-based QTL analyses, e.g., Jansen (1993), Zeng (1994), Manly and Cudmore (1996) and Basten et al. (1996), cannot be used. Also, the assessment of interactions among QTLs is weakened, as is the analysis of linked QTLs (Lin and Ritland, 1966). Finally, QTL results emerging from selective vs nonselective genotyping can, at times, be surprisingly different when the selection fraction is small, as observed, for example, by Gershenfeld et al. (1997) using a selection fraction of 0.12 (0.06 in each tail). This raises questions about whether a highly restricted sample is (1) increasing the sampling error to serious levels or (2) magnifying the spurious effects of experimental artifacts that cause extreme scores for reasons unrelated (or poorly related) to genotype (phenocopies). Thus, there are several reasons to avoid selective genotyping, especially when the selection fraction is small.

Generally, whether RI populations are more cost-effective (less costly) than a comparable \( F_2 \) will depend on the relative efficiency ratio, or RE (which depends on \( h^2_{RI} \), \( n \), and the selection fraction) and the relative cost ratio per mouse, or RC (which depends on the costs of genotyping relative to the other costs, and the selection fraction), all of which can vary widely from trait to trait and laboratory to laboratory. Darvasi and Soller (1992) discuss the cost implications of selective genotyping in the general case. For the cost data shown in Table I, the equicost value of \( h^2_{RI} \), when RI and \( F_2 \) populations of equal power are also equal in cost, was 0.65 for \( n = 10 \) and 0.90 for \( n = 6 \), when optimal (for cost) conventional selective genotyping was practiced (Fig. 4). Since most behavioral traits will have \( h^2_{RI} \) values less than the equicost value, RI populations generally will be more cost-effective than \( F_2 \) populations with similar RC values to our example, even when selective genotyping is optimized to reduce cost. The cost advantage of RI over segregating populations is severalfold at low \( h^2_{RI} \) and disappears as \( h^2_{RI} \) reaches the equicost value. When \( h^2_{RI} = 0.4 \) (\( h^2_{F_2} = 0.25 \)) and \( n = 10 \), for example, typical values in our experience, the cost advantage is just under twofold from Fig. 4 (the \( F_2 \) population is almost twice as costly) under the cost conditions shown in Table I, a major difference. This difference is even larger with smaller \( n \); for example, it is almost threefold when \( n = 6 \).

When selective genotyping and DNA pooling of each tail is used, equicost \( h^2_{RI} \) is 0.55 (\( h^2_{F_2} = 0.38 \)) for \( n = 6 \) and 0.37 (\( h^2_{F_2} = 0.23 \)) for \( n = 10 \), values which fall in the range typically seen for behavioral traits. Thus, when DNA pooling is used under the cost conditions of our example (Table I), \( F_2 \) compared to RI populations will be more cost-effective (cost less) for traits with heritabilities above this equicost value and less cost-effective below them.

The conclusion drawn above concerning relative total cost strictly hold only for traits with a relative cost ratio per mouse (RC) similar to that used in our example (Table I). However, our example is typical of our experience with over 30 traits subjected to QTL analyses. Actual costs from
other laboratories can easily be substituted for the values shown in Table I to obtain more accurate cost evaluations for a particular experiment. (A MathCad worksheet is available from the author for this purpose.)

Throughout this paper, dominance variation in the \( F_2 \) has been ignored, since this source does not exist in the RIs. However, dominance provides another source of QTL information that can increase power to detect QTLs showing dominance, thus increasing \( F_2 \) power and cost-effectiveness. On the other hand, the opposite can occur for QTLs showing no dominance, because the assessment of both additive and dominance effects in the QTL analysis (e.g., MapMaker QTL) requires a twofold more stringent \( p \) value as the threshold for statistical significance than do additive effects alone (Lander and Schork, 1994), which effectively reduces the power and \( F_2 \) cost-effectiveness.

Overall, the use of RI populations to gain QTL information in behavioral studies has much to recommend it for cost as well as other reasons noted elsewhere (Bailey, 1981; Belknap et al., 1996, 1997a; Plomin and McClearn, 1993). This is especially true if the within-strain sample size, \( n \), is reasonably adjusted for the expected heritability (i.e., using smaller \( n \) when \( h^2_0 \) is high, and vice versa), which can greatly reduce RI costs relative to \( F_2 \) conferring equal power. However, of all the costs considered, those of genotyping are likely to be most affected by advances in technology, which will likely make \( F_2 \) and other segregating populations more attractive economically than they are at present.

ACKNOWLEDGMENTS

This work was supported by a VA Merit Review Program from the Department of Veterans Affairs and by NIH Grants AA06243, AA10760, DA05228, and DA10913.

REFERENCES


Edited by David Fulker