

Short communication

A note on the effect of within-strain sample sizes on QTL mapping in recombinant inbred strain studies

W. E. Crusio

Brudnick Neuropsychiatric Research Institute, University of Massachusetts Medical School, Department of Psychiatry, Worcester, MA, USA

Corresponding Author: Wim E. Crusio, Brudnick Neuropsychiatric Research Institute, University of Massachusetts Medical School, Department of Psychiatry, 303 Belmont Street, Worcester, MA 01604, USA. E-mail: wim.crusio@umassmed.edu

This note explores the effect of within-strain sample sizes on the correlations between a phenotype and a molecular-genetic marker in a battery of inbred strains. It is shown that the maximum correlation possible between a molecular marker and a behavioral or neuronal phenotype equals the additive-genetic correlation. How close the strain correlation will approach the additive-genetic correlation depends only on heritability and within-strain sample sizes. The equations derived can be used to optimize designs of studies attempting to localize Quantitative Trait Loci utilizing Recombinant Inbred Strains, provided information about the heritability of the character under study is available.

Keywords: Environmental variance, genetic correlation, quantitative genetics

Received 7 February 2004, revised 18 March 2004, accepted for publication 26 March 2004

Despite much effort, mapping the genes underlying quantitative neurobehavioral variation remains an elusive undertaking. The difficulties follow from the facts that most phenotypes studied by behavioral neurogeneticists have only moderate heritabilities and are influenced by multiple genes, conventionally called polygenes (Mather 1949; Mather & Jinks 1982). The genetic loci of these polygenes are called Quantitative Trait Loci (QTL). Recently, the development of several new or improved genetic tools justifies some guarded optimism regarding the feasibility of not only mapping QTL, but of actually identifying the underlying

polygenes. For instance, elegant crossing methods have been developed to employ congenic (knockout) strains to confirm QTL and refine their locations (Bolivar *et al.* 2001) and the genomes of several important inbred strains have now been sequenced.

Recombinant Inbred Strains (RIS) have been used for many years to dissect behavioral and neural phenotypes genetically (Belknap & Crabbe 1992; Crabbe *et al.* 1983; Desforges *et al.* 1989; Gora-Maslak *et al.* 1991; Peirce *et al.* 2003; Phillips *et al.* 2002). One of the most frequently used sets of RIS has been the BXD set, derived from the inbred strains C57BL/6J and DBA/2J. The main reason for this popularity have been the facts that the progenitor strains are by far the most often used inbred strains and that the BXD set is one of the larger RIS sets available. Obviously, if an RIS set encompasses more strains, then its mapping sensitivity (in terms of percentage of the genetic variation explained) and precision (in terms of size of the confidence interval) will increase. A very promising recent advance therefore has been the development of additional BXD RI strains by Peirce, Lu, Williams and collaborators (Peirce *et al.* 2004). These new strains increase the number of available BXD RIS to about 80, making it feasible to map QTL explaining as little as 10% of the genetic variation to within a 1–2 Mb interval (Peirce *et al.* 2004).

A question that has received comparatively little attention concerns the choice of the optimal within-strain sample size for a given RIS study. Belknap and colleagues have investigated Type I and Type II error rates and the effect of within-strain sample size (Belknap *et al.* 1996; Belknap 1998). However, when considering sample sizes, these studies were mainly concerned with comparing the efficiency and cost-effectiveness of an RIS study with a design using an F2 between two inbred strains. Here I address the problem of the effect of within-strain sample size using a different approach. As a first step to a more comprehensive statistical power analysis, I explore the effect of within-strain sample sizes on the correlation between strain means and a molecular-genetic marker. As mapping studies using RIS are based on such correlations, the following considerations should thus be helpful in optimizing RIS designs.

This present treatment uses some formulae presented earlier (Crusio 2000) and uses the notation of Falconer (Falconer 1960; Falconer & Mackay 1996). For readers

more familiar with the notation followed by Mather and Jinks (1982), some comparisons between the two notation systems can be found in Table 1. In what follows, two implicit assumptions have been made. First, we assume the absence of epistatic interactions between genes. Second, it is assumed that the molecular-genetic marker used is linked directly to the QTL. If the latter assumption is violated, strain correlations will be lower than expected. The consequences of violations of the first assumption will be explored later.

In the case of a set of inbred strains, of which RIS are just a special case, dominance deviations are absent and the heritability of a trait therefore equals

$$h^2 = V_A / (V_A + V_E)$$

Estimates of V_A and V_E can be obtained from an analysis of variance of a set of inbred strains. Within-strain variance equals V_E and the variance of the strain means equals $2V_A + V_E/n$, where n is the number of subjects used per strain (Crusio 2000; Hegmann & Possidente 1981). If we now express V_E in terms of h^2 and V_A , we obtain

$$V_E = (1 - h^2) / 2h^2 * 2V_A$$

For ease of representation, we introduce a symbol α , such that

$$\alpha = (1 - h^2) / 2h^2$$

and

$$V_E = 2\alpha V_A$$

The correlation between two variables x and y is given by

$$r = \text{cov}(x, y) / \sqrt{V_x V_y}$$

Analogously, the genetic correlation between two phenotypes x and y equals

$$r_A = V_{Axy} / \sqrt{V_{Ax} V_{Ay}}$$

where V_{Axy} is the additive-genetic covariance between characters x and y . In the specific case where character y is a molecular marker, its additive-genetic variance V_{Ay} equals 1 and this equation reduces to

$$r_A = V_{Axy} / \sqrt{V_{Ax}}$$

Similarly, the correlation between the means of a set of inbred strains (or RIS) equals

$$r_{\text{strain}} = (2V_{Axy} + V_{Exy}/n) / \sqrt{(2V_{Ax} + V_{Ex}/n)(2V_{Ay} + V_{Ey}/n)}$$

In the specific case where character y is a molecular marker with therefore perfect heritability, V_{Exy} and V_{Ey} will be zero and the previous formula reduces to

$$r_{\text{strain}} = 2V_{Axy} / \sqrt{2(2V_{Ax} + V_{Ex}/n)}$$

By substituting $2\alpha V_{Ax}$ for V_{Ex} , we finally obtain

$$r_{\text{strain}} = r_A / \sqrt{(n + \alpha)/n}$$

The implications of this last formula are as follows. First, the maximum correlation possible between a molecular marker and a behavioral or neuronal phenotype equals the additive-genetic correlation, in case n approaches infinity and/or heritability approaches unity (cf. also Fig. 1 in Belknap 1998). Note that the additive-genetic correlation in this particular case equals the square root of the percentage of the genetic variance (the effect size) explained by the molecular marker (i.e. the QTL). Second, how close the strain correlation will approach the additive-genetic correlation depends only on the heritability (as the size of α depends only on h^2) and n , the number of subjects used per strain (Fig. 1), but not on the effect size of the QTL. Let us take, for example, three characters with heritabilities around 0.10, 0.25 or 0.50,

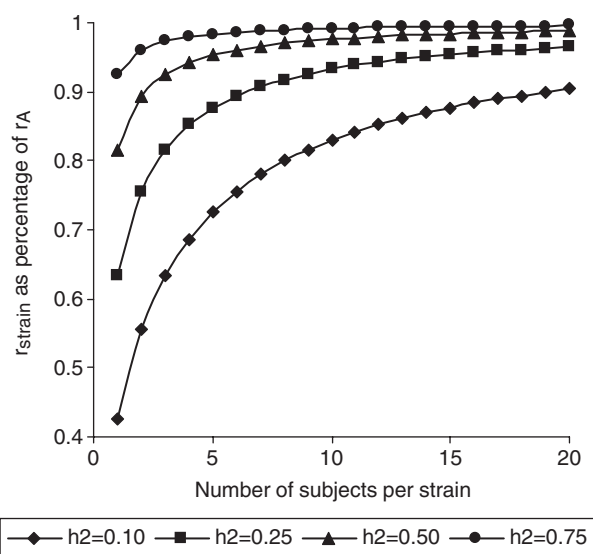


Figure 1: Correlation between inbred strain means and a molecular marker (QTL) (r_{strain}) relative to the additive-genetic correlation (r_A) as a function of within-strain sample sizes and heritability.

Table 1: Comparison of the notation systems used by Mather and Jinks (Mather 1949; Mather & Jinks 1982) and Falconer (Falconer 1960; Falconer & Mackay 1996)

Description of statistic	Mather and Jinks	Falconer
Additive-genetic correlation	r_D	r_A
Additive-genetic effect	d	a
Dominance deviation	h	d
Additive-genetic variance	D	$2V_A$
Environmental variance	E	V_E

respectively. The first two values are in the range of what is often obtained for behavioral characteristics in animal studies, whereas the latter value is in the range of values generally obtained for morphological characteristics. If it would now be deemed desirable that r_{strain} would equal about 90% of the maximum attainable (r_A), we would need within-strain sample sizes of around 20, 7 or 2 subjects per strain, respectively. In order for r_{strain} to approach 95% of r_A , these sample sizes would become 41, 14 and 5. Note that the sample sizes obtained agree very well with the values presented by Belknap (1998). Third, especially in cases where heritabilities are low, the correlation between strains will be much lower than the additive-genetic correlation if small sample sizes are used per strain (Fig. 2).

Whether a particular QTL can be mapped in a particular study will depend on four factors: the number of subjects used per strain, the heritability of the character under study, the number of RIS and the percentage of the genetic variation explained by the QTL. The precision of mapping will depend on the number of RIS and the number of chromosomal break points fixed in them. However, in all cases, statistical power will be maximal when the correlation between strain means and the molecular-genetic marker is maximal. If heritabilities are known, then the above formulae

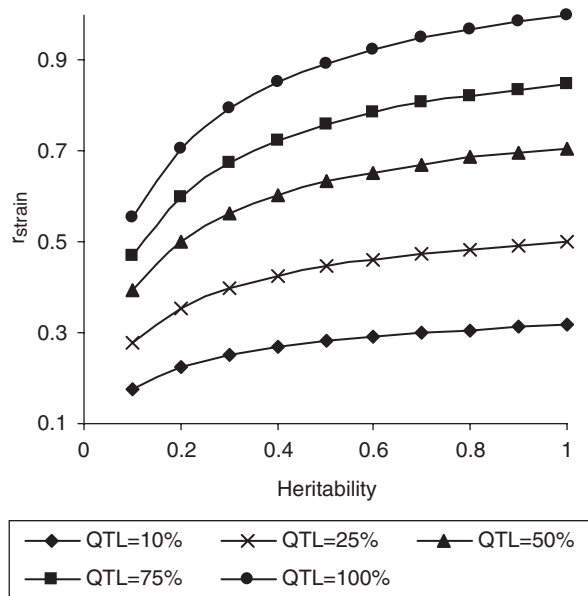


Figure 2: Correlation between inbred strain means and a molecular marker (QTL) (r_{strain}) as a function of heritability for different QTL effect sizes. Within-strain sample sizes: $n = 2$.

will allow the optimization of experimental designs, regardless of the number of available RIS or the effect size of a QTL.

References

- Belknap, J.K. (1998) Effect of within-strain sample size on QTL detection and mapping using recombinant inbred mouse strains. *Behav Genet* **28**, 29–38.
- Belknap, J.K. & Crabbe, J.C. (1992) Chromosome mapping of gene loci affecting morphine and amphetamine responses in BXD recombinant inbred mice. *Ann N Y Acad Sci* **654**, 311–323.
- Belknap, J.K., Mitchell, S.R., O'Toole, L.A., Helms, M.L. & Crabbe, J.C. (1996) Type I and type II error rates for quantitative trait loci (QTL) mapping studies using recombinant inbred mouse strains. *Behav Genet* **26**, 149–160.
- Bolivar, V.J., Cook, M.N. & Flaherty, L. (2001) Mapping of quantitative trait loci with knockout/congenic strains. *Genome Res* **11**, 1549–1552.
- Crabbe, J.C., Kosobud, A., Young, E.R. & Janowsky, J.S. (1983) Polygenic and single-gene determination of responses to ethanol in BXD/Ty recombinant inbred mouse strains. *Neuro-behav Toxicol Teratol* **5**, 181–187.
- Crusio, W.E. (2000) An introduction to quantitative genetics. In Jones, B.C. & Mormède, P. (eds), *Neurobehavioral Genetics: Methods and Applications*. CRC Press, Boca Raton, FL, pp. 13–30.
- Desforges, C., Venault, P., Dodd, R.H., Chapouthier, G. & Roubertoux, P.L. (1989) Beta-carboline-induced seizures in mice: genetic analysis. *Pharmacology, Biochemistry, Behavior* **34**, 733–737.
- Falconer, D.S. (1960) *Introduction to Quantitative Genetics*. Oliver and Boyd, Edinburgh, UK.
- Falconer, D.S. & Mackay, T.F.C. (1996) *Introduction to Quantitative Genetics*, 4th edn. Longman, Harlow, UK.
- Gora-Maslak, G., McClearn, G.E., Crabbe, J.C., Phillips, T.J., Belknap, J.K. & Plomin, R. (1991) Use of recombinant inbred strains to identify quantitative trait loci in psychopharmacology. *Psychopharmacology* **104**, 413–424.
- Hegmann, J.P. & Possidente, B. (1981) Estimating genetic correlations from inbred strains. *Behav Genet* **11**, 103–114.
- Mather, K. (1949) *Biometrical Genetics. The Study of Continuous Variation*. Methuen, London, UK.
- Mather, K. & Jinks, J.L. (1982) *Biometrical Genetics. The Study of Continuous Variation*, 3rd edn. Chapman & Hall, London, UK.
- Peirce, J.L., Chesler, E.J., Williams, R.W. & Lu, L. (2003) Genetic architecture of the mouse hippocampus: identification of gene loci with selective regional effects. *Genes, Brain Behav* **2**, 238–252.
- Peirce, J.L., Lu, L., Gu, J., Silver, L.M. & Williams, R.W. (2004) A new set of 46 BXD recombinant inbred lines from advanced intercross populations. *BMC Genet* **5**, 7.
- Phillips, T.J., Belknap, J.K., Hitzemann, R.J., Buck, K.J., Cunningham, C.L. & Crabbe, J.C. (2002) Harnessing the mouse to unravel the genetics of human disease. *Genes, Brain Behav* **1**, 14–26.

Edited by Daniel Goldowitz