Theor Appl Genet (1996) 93:1261–1266

J. -Z. Lin · K. Ritland

The effects of selective genotyping on estimates of proportion of recombination between linked quantitative trait loci

Received: 9 April 1996 / Accepted: 26 April 1996

Abstract Selective genotyping is the marker assay of only the more extreme phenotypes for a quantitative trait and is intended to increase the efficiency of quantitative trait loci (QTL) mapping. We show that selective genotyping can bias estimates of the recombination frequency between linked QTLs – upwardly when QTLs are in repulsion phase, and downwardly when QTLs are in coupling phase. We examined these biases under simple models involving two QTLs segregating in a backcross or F_2 population, using both analytical models and computer simulations. We found that bias is a function of the proportion selected, the magnitude of QTL effects, distance between QTLs and the dominance of QTLs. Selective genotyping thus may decrease the power of mapping multiple linked QTLs and bias the construction of a marker map. We suggest a large proportion than previously suggested (50%) or the entire population be genotyped if linked QTLs of large effects (explain > 10% phenotypic variance) are evident. New models need to be developed to explicitly incorporate selection into QTL map construction.

Key words Selctive genotyping · Quantitative trait loci · Genetic markers · Recombination frequency

Introduction

The techniques of mapping quantitative trait loci (QTLs) with genetic markers are now widely used in crop and animal breeding (Edwards et al. 1987; Paterson et al. 1988; Stuber et al. 1992; Dudly 1993; Tanksley 1993) and increasingly being used in evolutionary biology (Mitchell-Olds 1995; Bradshaw et al. 1995). The statistical power in detecting QTLs depends on the magnitude of their effects. To detect QTLs with large effects (explaining > 15% of the phenotypic variance), a relatively small number of progeny (ca. 100) may be sufficient to obtain reasonable power (Lander and Botstein 1989). Detection of QTLs with medium to small effects, however, requires hundreds to thousands of progeny (Soller et al. 1976; Soller and Genizi 1978; Lander and Botstein 1989; Darvasi et al. 1993). In many cases, the cost of determining the marker genotypes of such a large number of progeny may become a major limiting factor in QTL mapping (Darvasi and Soller 1992).

The selective genotyping procedure proposed by Lander and Botstein (1989) is an important technique to overcome this difficulty (Lander and Botstein 1989; Darvasi and Soller 1992, 1994). This technique, based on the genotyping of those individuals showing only the more extreme phenotypes, has, for example, been employed by Groover et al. (1994) to increase the power of detecting QTLs for wood specific gravity in loblolly pine. Using 48 extreme phenotype individuals, they identified five regions of the loblolly pine genome as harbouring QTLs influencing wood specific gravity. In our own work with the annual wildflower *Mimulus*, we generated a linkage map using 96 selectively genotyped individuals from a population of 247 backcross individuals (Lin and Ritland 1996) and simultaneously detected several QTLs for mating system characters (Lin and Ritland, in review).

The statistical power of selective genotyping for the purpose of detecting linkage between QTLs and markers was considered by Lander and Botstein (1989) and Darvasi and Soller (1992). They found that genotyping only the individuals from both the upper and lower 25% tails of the phenotypic distribution is practically as efficient, in terms of detecting QTLs, as genotyping

Communicated by G. Wenzel

J.-Z. Lin[1] (✉) · K. Ritland[2]
Department of Botany, University of Toronto, Toronto, Ontario, M5S 3B2, Canada

Present addresses:
[1] Division of Biological Sciences, University of Montana, MT 59812, USA
[2] Department of Forest Sciences, 2357 Main Mall, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

the entire phenotypic distribution. However, selective genotyping has the consequential result of also selectively assaying meiotic events. When individuals in a backcross or F_2 population which resemble either parent are selected for marker genotyping, one may indirectly select against recombinants, thereby reducing the effective recombination rate and the power to distinguish linked QTLs, or at least distorting their true linkages. The purpose of this paper is to identify and illustrate this bias of selective genotyping and to discuss its implications in mapping multiple QTLs.

Theory

Definitions and assumptions

In the following analysis, we will analyse two QTLs segregation in a backcross or F_2 population. The alternative alleles at these loci are denoted Q_1, q_1 at the first locus and Q_2, q_2 at the second. The capital lettered alleles both have effects in the positive direction, and the lower-case alleles have effects in the negative direction. The results we obtain will hold for traits determined by additional QTLs as long as those loci are not linked to the two QTLs of interest. The effects of QTLs are assumed to be additive among loci, and for the case of a segregating F_2, dominance is allowed. In the following, the subscript notation "1" refers henceforth to parameters for QTL Q_1, subscript "2" refers to parameters for QTL Q_2, subscript "U" refers to the upper tail of the phenotypic distribution, and subscript "L" to the lower tail.

Recombination between QTLs after selective genotyping in a backcross

Consider first the pair of QTLs in a backcross population. We assume that of the total proportion P of the selectively genotyped individuals, half are selected from the upper tail, and half from the lower tail, of the phenotypic distribution of a quantitative trait. To examine the effects of this selection, we first consider how selection in the upper tail modifies the QTL gene frequency. Fig. 1A illustrates the idealized phenotypic distributions of two alternative QTL genotypes. After selection a certain fraction of each genotype remains, as represented by the area shaded under each curve. This area can be used to compute the change of gene frequency, and this change can be equated to a trunction selection regime where the probability of QTL genotypes Qq and qq being selected for (or "viability") is 1 and $1-s$, respectively. Falconer (1989) computed the selection coefficient s by linearly approximating the difference in the area under the two respective curves, obtaining

$$s \approx i\frac{A}{\sigma_p} = i\sqrt{\frac{A^2}{\sigma_p^2}} = i\sqrt{v} \tag{1}$$

where i is the selection intensity which is related to the proportion selected P as given by Falconer (1989), σ_p is the total phenotypic standard deviation, A is the average effect of the QTL (Fisher 1941) and $v = A^2/\sigma_p^2$ is the proportion of phenotypic variance explained by the QTL. In a backcross population, $A = \mu_{Qq} - \mu_{qq}$, where μ_{Qq} and μ_{qq} are the means of QTL genotypes Qq and qq, respectively.

By symmetry of selective genotyping, in the lower tail of the distribution, the "viabilities" of Qq and qq are $1-s$ and 1, respectively. Now, if both QTLs are considered simultaneously, the backcross population is a mixture of four QTL genotypes $Q_1q_1Q_2q_2$, $Q_1q_1q_2q_2$, $q_1q_1Q_2q_2$ and $q_1q_1q_2q_2$, where subscripts denote loci. Their expected frequencies and "viabilities" are shown in Table 1, where the "viability" is calculated by averaging over two tails, assuming a multiplicative fitness function among loci. For example, the "viability" of $Q_1q_1Q_2q_2$ is $(1/2)[1+(1-s_1)(1-s_2)] = (1/2)(2-s_1-s_2+s_1s_2)$.

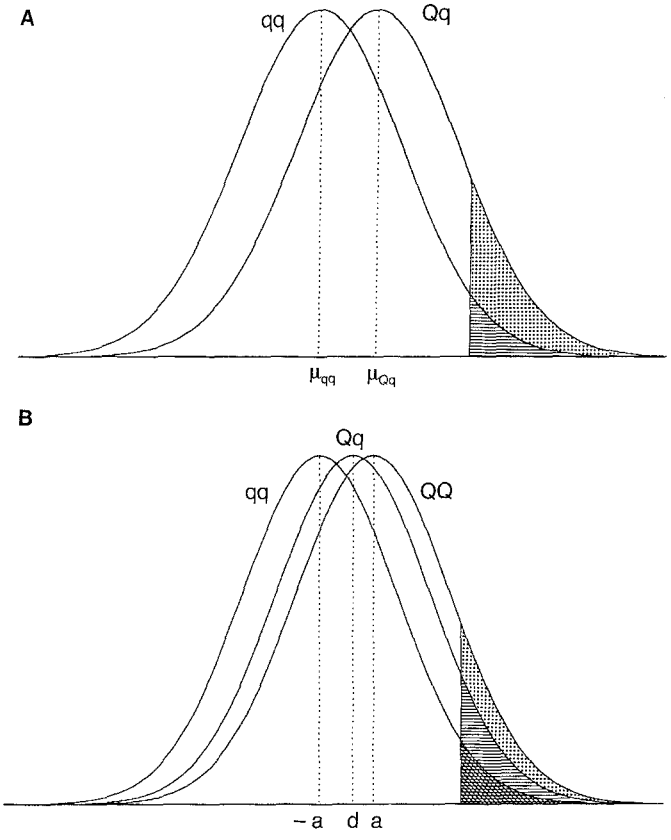

Fig. 1A, B Distribution of the genotypes of one QTL locus in a backcross (**A**) and F_2 population (**B**). The relative frequencies of genotypes after truncation selection (selective genotyping) in the upper tail are indicated by the *shaded* areas

The recombination frequency between QTLs after selective genotyping (denoted r' to distinguish from the true r) is obtained by the standard procedure of computing genotypic frequencies after selection. For QTLs in coupling phase, this is

$$r' = \frac{\frac{1}{2}r(2-s_1-s_2)}{\frac{1-r}{2}(2-s_1-s_2+s_1s_2)+\frac{r}{2}(2-s_1-s_2)} = \frac{r(2-s_1-s_2)}{2-s_1-s_2+(1-r)s_1s_2} \tag{2}$$

and for QTLs in repulsion phase it is

$$r' = \frac{\frac{r}{2}(2-s_1-s_2+s_1s_2)}{\frac{1-r}{2}(2-s_1-s_2)+\frac{r}{2}(2-s_1-s_2+s_1s_2)} = \frac{r(2-s_1-s_2+s_1s_2)}{2-s_1-s_2+rs_1s_2} \tag{3}$$

Recombination between QTLs after selective genotyping in an F_2 population

In an F_2 population, each QTL has three genotypes segregating, so dominance should be considered. Let the average effects of QTL

Table 1 Expected frequencies before selection and "viabilities" of the genotypes of two QTLs in coupling and repulsion in a backcross population

[a] Assuming the recurrent parent is $q_1 q_1 Q_2 Q_2$

Coupling		Repulsion[a]		"Viability"
Genotype	Frequency	Genotype	Frequency	
$Q_1 q_1 Q_2 q_2$	$(1/2)(1-r)$	$Q_1 q_1 Q_2 Q_2$	$(1/2)r$	$(1/2)(2-s_1-s_2+s_1 s_2)$
$Q_1 q_1 q_2 q_2$	$(1/2)r$	$Q_1 q_1 Q_2 q_2$	$(1/2)(1-r)$	$(1/2)(2-s_1-s_2)$
$q_1 q_1 Q_2 q_2$	$(1/2)r$	$q_1 q_1 Q_2 Q_2$	$(1/2)(1-r)$	$(1/2)(2-s_1-s_2)$
$q_1 q_1 q_2 q_2$	$(1/2)(1-r)$	$q_1 q_1 Q_2 q_2$	$(1/2)r$	$(1/2)(2-s_1-s_2+s_1 s_2)$

genotypes QQ, Qq and qq be a, d and $-a$, where d represents the dominance deviation. After selective genotyping, let the relative "viabilities" of QTL genotypes QQ, Qq and qq in the upper tail of a F_2 population be 1, $1-hs$ and $1-s$, respectively, where h represents the dominance coefficient ($h=1/2$ for no dominance, $h=-1$ or $+1$ for complete dominace, etc.) Figure 1B illustrates the idealized phenotypic distributions for each of the three QTL genotypes following truncation selection. To equate the shaded areas under each curve to selection intensity, we again make the linear approximation of Falconer (1989) as he did for homozygote selection (using W's to denote areas or fitness),

$$1-s=\frac{W_{qq}}{W_{QQ}}\approx\left(1-i\frac{2a}{\sigma_p}\right)$$

and by analogy, the approximation for heterozygote selection is

$$1-hs=\frac{W_{Qq}}{W_{QQ}}\approx\left(1-i\frac{a-d}{\sigma_p}\right)$$

Solving these jointly for h gives

$$h=(1/2)(1-d/a). \tag{4}$$

Now, if both QTLs are considered simultaneously, the F_2 population is a mixture of nine QTL genotypic subpopulations (Table 2). Let $\mathbf{F}$ and $\mathbf{W}$ be 9×1 vectors. Elements F_i contain the expected frequency of each QTL genotype before selection, and elements W_i contain the differential "viability" of each QTL genotype in the selected population. In computing the W_i, as before, fitness is multiplicative across loci and W_i is calculated by averaging over both tails. For example,

$$\begin{aligned}W_2=W(Q_1Q_1Q_2q_2)&=(1/2)[W(U,Q_1Q_1)\cdot W(U,Q_2q_2)\\&\quad+W(L,Q_1Q_1)\cdot W(L,Q_2q_2)]\\&=(1/2)(2-s_1-2h_2s_2+h_2s_1s_2).\end{aligned}$$

The recombination frequency after selective genotyping between QTLs in coupling phase is

$$r'=\frac{\frac{1}{2}\times\frac{1}{2}r(1-r)(W_2+W_4+W_6+W_8)+\frac{1}{4}r^2(W_3+W_7)+\frac{1}{2}r^2W_5}{\Sigma_{i=1}^{9}F_iW_i}$$

$$=\frac{r(s_1(s_2(h_1(2h_2r-r+1)+h_2(1-r))-2h_1-1)-s_2(2h_2+1)+4)}{s_1(s_2(2h_1(h_2(2r^2-2r+1)-r(r-1))+(1-r)(2h_2r-r+1))-2h_1-1)-s_2(2h_2+1)+4} \tag{5}$$

and for QTLs in repulsion, it is

$$r'=\frac{\frac{1}{2}\times\frac{1}{2}r(1-r)(W_2+W_4+W_6+W_8)+\frac{1}{4}r^2(W_1+W_9)+\frac{1}{2}r^2W_5}{\Sigma_{i=1}^{9}F_iW_i}$$

$$=\frac{r(s_1(s_2(h_1(2h_2r-r+1)+h_2(1-r)+r)-2h_1-1)-s_2(2h_2+1)+4)}{s_1(s_2(2h_1(h_2(2r^2-2r+1)-r(r-1))-r(2h_2(r-1)-r))-2h_1-1)-s_2(2h_2+1)+4} \tag{6}$$

where s is found using (Eq. 1) with $A=\mu_{QQ}-\mu_{qq}=2a$, and h is found using (Eq. 4).

Numerical results

To evaluate the expected levels of bias in recombination frequency between QTLs due to selective genotyping, we calculated the values of r' by solving Eqs. 2, 3, 5 and 6 by specifying P, r, v_1, v_2, h_1 and h_2. The relative bias in r is expressed as $(r'-r)/r\times 100\%$. Because the approximation in (Eq. 3) becomes less accurate as the effects of QTL becomes large (see Falconer 1989), we present our results by subtracting the bias for $P=1.0$ from each data point to correct this approximation bias.

Table 2 Expected frequencies before selection and "viabilities" of the genotypes of two QTLs in coupling and repulsion in an F_2 population

QTL genotype	Frequency		"Viability"
	Coupling	Repulsion	
$Q_1Q_1Q_2Q_2$	$(1/4)(1-r)^2$	$(1/4)r^2$	$(1/2)(2-s_1-s_2+s_1s_2)$
$Q_1Q_1Q_2q_2$	$(1/2)r(1-r)$	$(1/2)r(1-r)$	$(1/2)(2-s_1-2h_2s_2+h_2s_1s_2)$
$Q_1Q_1q_2q_2$	$(1/4)r^2$	$(1/4)(1-r)^2$	$(1/2)(2-s_1-s_2)$
$Q_1q_1Q_2Q_2$	$(1/2)r(1-r)$	$(1/2)r(1-r)$	$(1/2)(2-2h_1s_1-s_2+h_1s_1s_2)$
$Q_1q_1Q_2q_2$	$(1/2)[r^2+(1-r)^2]$	$(1/2)[r^2+(1-r)^2]$	$1-h_1s_1-h_2s_2+h_1h_2s_1s_2$
$Q_1q_1q_2q_2$	$(1/2)r(1-r)$	$(1/2)r(1-r)$	$(1/2)(2-2h_1s_1-s_2+h_1s_1s_2)$
$q_1q_1Q_2Q_2$	$(1/4)r^2$	$(1/4)(1-r)^2$	$(1/2)(2-s_1-s_2)$
$q_1q_1Q_2q_2$	$(1/2)r(1-r)$	$(1/2)r(1-r)$	$(1/2)(2-s_1-2h_2s_2+h_2s_1s_2)$
$q_1q_1q_2q_2$	$(1/4)(1-r)^2$	$(1/4)r^2$	$(1/2)(2-s_1-s_2+s_1s_2)$

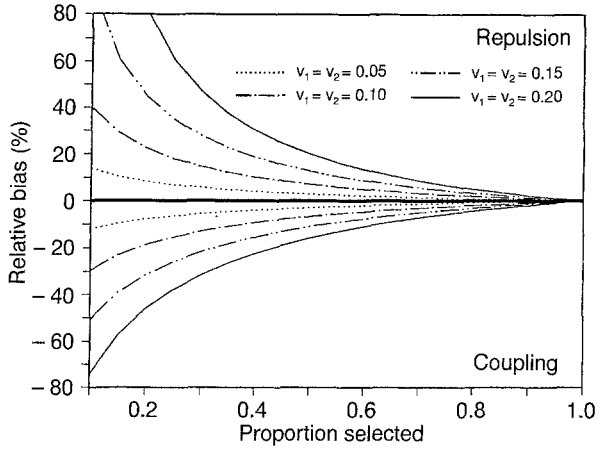

Fig. 2 Relative bias in r as a function of the proportion selected (P) with $r = 0.2$ for QTLs in coupling and repulsion in a backcross population

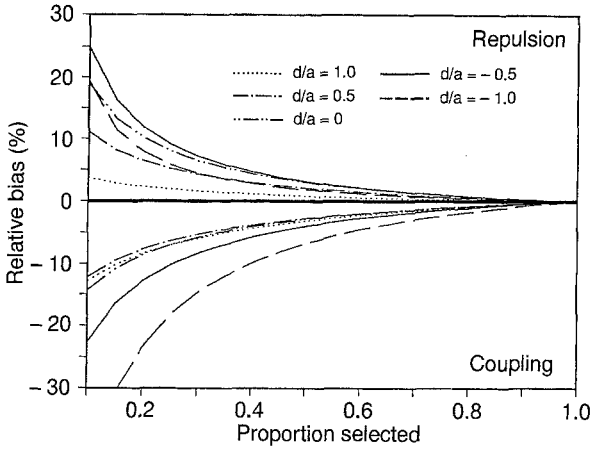

Fig. 4 Relative bias in r as a function of the proportion selected (P) with $r = 0.2$, $v_1 = v_2 = 0.15$ for QTLs with completely dominant ($d/a = 1$), partially dominant ($d/a = 0.5$), additive ($d/a = 0$), partially recessive ($d/a = -0.5$) or completely recessive ($d/a = -1$) gene effect in a F_2 population

Figure 2 shows the relative bias in r as functions of the proportion selected (P) with fixed value of $r = 0.2$ for QTLs of various effects in a backcross population. Selective genotyping causes downward bias in r between QTLs in coupling phase, and upward bias between QTLs in repulsion phase. The magnitude of the bias is slightly larger for QTLs in repulsion than QTLs in coupling and increases as the effects of the QTLs increase. As P increases the bias in r decreases considerably. For QTLs of relatively small effects (explains $\leqslant 10\%$ phenotypic variance), bias in r is less than 10% if 50% of the population is selected for genotyping. For QTLs of larger effects (explains $> 10\%$ phenotypic variance), however, a larger proportion must be selected for genotyping in order to keep the bias to less than 10%.

Figure 3 shows the relative bias in r as functions of r with $P = 0.5$ for QTLs of various effects in a backcross population. The bias decreases slightly as r increases. The magnitude of the bias is within 5% for QTLs of small effects ($v_1 = v_2 = 0.05$), but can be as high as 25% for QTLs of large effects ($v_1 = v_2 = 0.2$).

Fig. 3 Relative bias in r as a function of r with $P = 0.5$ for QTLs in coupling and repulsion in a backcross population

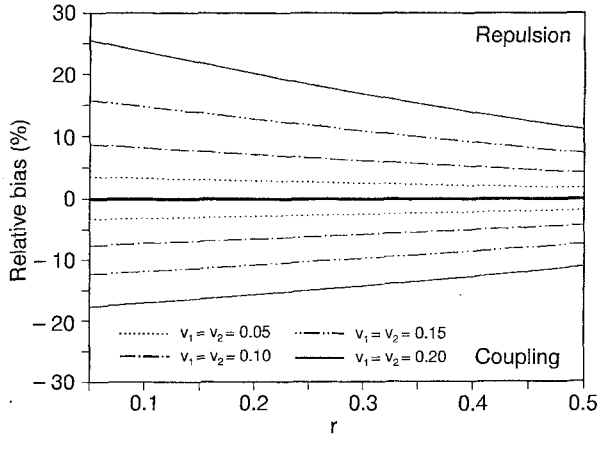

Figure 4 shows the relative bias in r as functions of P for QTLs with different dominance levels with $r = 0.2$, $v_1 = v_2 = 0.15$ in an F_2 population. For QTLs in repulsion, there are significant differences in magnitude of the bias in r between QTLs with different dominance levels. The largest bias in r is found for QTLs with partially recessive gene effect ($d/a = -0.5$), while the least is found for QTLs with completely dominant effect ($d/a = 1$). For QTLs in coupling, the largest bias in r is found for QTLs with recessive ($d/a = -1$) and partially recessive ($d/a = -0.5$) gene effect, while there is little difference between dominant ($d/a = 1$), partially dominant ($d/a = 0.5$) and additive QTLs ($d/a = 0$). The magnitude differences of the bias in r between different modes of gene effect decreases as the proportion selected increases.

Figure 4 also shows a significant magnitude difference of the bias between linkage phases of the QTLs for the same mode of gene effect. For additive genes ($d/a = 0$), there is a larger bias in r between QTLs in repulsion than those in coupling (see also Figs. 2 and 3). The opposite is found for dominant ($d/a = 1$) and partially recessive ($d/a = -0.5$) genes in that there is less bias in r between QTLs in repulsion than those in coupling. There is, however, very little difference between linkage phases for completely dominant and completely recessive genes.

A computer simulation was also conducted by randomly generating phenotypes in a backcross or F_2 population, then applying selective genotyping and estimating recombination rates. Phenotypes were generated by randomly choosing marker and QTL genotypes according to their probabilities under segregation and recombination, then adding a random environmental deviation. These simulations confirmed that the differences between the analytical and simulated results were close, generally within 5%.

Discussion

Our results show that selective genotyping can cause bias in the estimated recombination frequency between linked QTLs. The bias is caused by the differential survival of multilocus QTL genotypes. However, the sign of the bias depends upon the linkage phase of the QTLs. It is downward when QTLs are in coupling phase but upward when QTLs are in repulsion. In wide crosses, or crosses between distinct strains or taxa, we generally expect loci to be in coupling, so that more often the QTL linkage map is "shrunk" or reduced in total size.

The magnitude of this bias is a function of the proportion selected for genotyping, the magnitude of QTL effects, the recombination frequency between QTLs and the dominance level of the QTL effects. The bias is more serious for QTLs of large effects (explaining > 10% of the phenotypic variance, Figs. 2 and 3) and for QTLs with recessive and partially recessive gene effect (Fig. 4), but this bias decreases considerably as the proportion selected for genotyping increases (Figs. 2 and 4).

Lander and Botstein (1989) and Darvasi and Soller (1992) suggested that selecting 50% of a random population for genotyping is sufficient in most QTL mapping practice. Our results indicate that when linked QTLs are involved, the effective recombination between QTLs can still be seriously biased when 50% of the population is selected, particularly when the QTLs are of large effects (Figs. 2 and 3). We suggest that if evidence indicates linked QTLs of large effects for a trait (e.g. through preliminary examination of QTL locations), then a larger proportion or even the entire (random) population should be genotyped for the markers linked to the QTLs in order to minimize the bias in estimates of recombination frequency between QTLs.

One of the most challenging problems in mapping QTLs is to distinguish multiple linked QTLs (Lander and Botstein 1989; Zeng 1993, 1994; Jiang and Zeng 1995). Since selective genotyping causes downward bias of recombination estimates (and map distance) between linked QTLs in coupling, it may significantly decrease the power of mapping multiple linked QTLs, particularly if the QTLs are of large effects.

This bias due to selective genotyping may also extend to the construction of the linkage map, if one constructs the linkage map from the same selectively genotyped population (see Lin and Ritland 1996). While most QTL mapping projects use preexisting marker linkage maps, many workers cannot afford the time or expense of separately constructing a linkage map. This particularly applies to studies in evolutionary biology using species of no economic importance, such as monkeyflowers. QTL mapping in natural (wild) populations is in its infancy (Mitchell-Olds 1995). New inferences in evolutionary biology should be possible if we could efficiently examine QTLs involved in evolutionary divergence (Doebley and Stec 1991, 1993; Mitchell-Olds 1995; Bradshaw et al. 1995). In particular, it would be most efficient to jointly infer the linkage map and the QTL effects and locations in a single experiment with selective genotyping. New statistical models and computer programmes need to be developed that jointly incorporate classical QTL mapping algorithms with selective genotyping.

Acknowledgements We thank M. Soller, A. Darvasi, M. Daly, T. Mitchell-Olds, Y.-B. Fu, S. Steward, J. Dole and two anonymous reviewers for comments on an earlier version of this paper. This work is a part of the graduate project of JZL who was supported by scholarships from NSERC and University of Toronto and stipends from KR's NSERC operating grants.

References

Bradshaw HD Jr, Wilbert SM, Otto KG, Schemske DW (1995) Genetic mapping of floral traits associated with reproductive isolation in monkeyflowers (*Mimulus*). Nature 376:762–765

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

Darvasi A, Soller M (1994) Selective DNA pooling for determination of linkage between a molecular marker and a quantitative trait locus. Genetics 138:1365–1373

Darvasi A, Weinreb A, Minke V, Weller JI, Soller M (1993) Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. Genetics 134:943–951

Doebley J, Stec A (1991) Genetic analysis of the morphological differences between maize and teosinte. Genetics 129:285–295

Doebley J, Stec A (1993) Inheritance of the morphological differences between maize and teosinte: comparison of results for two F_2 populations. Genetics 134:559–570

Dudly JW (1993) Molecular markers in plant improvement: manipulation of genes affecting quantitative traits. Crop Sci 33:660–668

Edwards MD, Stuber CW, Wendel JF (1987) Molecular-marker-facilitated investigation of quantitative-trait loci in maize. I. Numbers, genomic distribution and types of gene action. Genetics 116:113–125

Falconer DS (1989) Introduction to quantitative genetics, 3rd edn. Longman, New York

Fisher RA (1941) Average excess and average effect of a gene substitution. Ann Eugen 11:53–63

Groover A, Devey M, Fiddler T, Lee J, Megraw R, Mitchell-Olds T, Sherman B, Williams C, Neale D (1994) Identification of quantitative trait loci influencing wood specific gravity in an outbred pedigree of loblolly pine. Genetics 138:1293–1300

Jiang C, Zeng Z-B (1995) Multiple trait analysis of genetic mapping for quantitative trait loci. Genetics 140:1111–1127

Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199

Mitchell-Olds T (1995) The molecular basis of quantitative genetic variation in natural populations. Trends Ecol Evol 10:324–328

Lin J-Z, Ritland K (1996) Construction of a genetic linkage map in the wild plant *Mimulus* using RAPD and isozyme markers. Genome 39:63–70

Paterson AH, Lander ES, Hewitt JD, Paterson S, Lincoln SE, et al. (1988) Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335:721–726

Soller M, Genizi A (1978) The efficiency of experimental designs for the detection of linkage between a marker locus and a locus affecting a quantitative trait in segregating populations. Biometrics 34:47–55

Soller M, Genizi A, Brody T (1976) On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. Theor Appl Genet 47:35–39

Stuber CW, Lincoln SE, Wolff DW, Helentjaris T, Lander ES (1992) Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. Genetics 132:823–839

Tanksley SD (1993) Mapping polygenes. Annu Rev Genet 27:205–233

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

Zeng Z-B (1994) Precision mapping of quantitative trait loci. Genetics 136:1457–1468