Consequences of Recombination Rate Variation on Quantitative Trait Locus Mapping Studies: Simulations Based on the *Drosophila melanogaster* Genome

Mohamed A. F. Noor, Aimee L. Cunningham and John C. Larkin

Department of Biological Sciences, Louisiana State University, Baton Rouge, Louisiana 70803

Manuscript received January 15, 2001
Accepted for publication July 5, 2001

ABSTRACT

We examine the effect of variation in gene density per centimorgan on quantitative trait locus (QTL) mapping studies using data from the *Drosophila melanogaster* genome project and documented regional rates of recombination. There is tremendous variation in gene density per centimorgan across this genome, and we observe that this variation can cause systematic biases in QTL mapping studies. Specifically, in our simulated mapping experiments of 50 equal-effect QTL distributed randomly across the physical genome, very strong QTL are consistently detected near the centromeres of the two major autosomes, and few or no QTL are often detected on the *X* chromosome. This pattern persisted with varying heritability, marker density, QTL effect sizes, and transgressive segregation. Our results are consistent with empirical data collected from QTL mapping studies of this species and its close relatives, and they explain the "small X-effect" that has been documented in genetic studies of sexual isolation in the *D. melanogaster* group. Because of the biases resulting from recombination rate variation, results of QTL mapping studies should be taken as hypotheses to be tested by additional genetic methods, particularly in species for which detailed genetic and physical genome maps are not available.

QUANTITATIVE trait locus (QTL) mapping has recently become a standard tool for unraveling the genetic basis of phenotypic variation in natural populations, tests of adaptation, and directing marker-assisted selection of agronomically important traits. Coupling sophisticated statistical analyses with molecular genetic data, many QTL mapping studies have identified precise genomic regions contributing to differences between strains or species. The results of these studies have also provided minimum estimates of the number of genes contributing to observed phenotypic differences. These estimates are necessarily minimum because loci contributing subtle effects have a lower probability of being detected (Otto and Jones 2000).

Low recombination rates may cause multiple independent genetic factors contributing to a trait to resemble a single QTL of large effect. Regional differences in relative rates of recombination have been documented in numerous taxa (Larkin and Woolford 1983; Begun and Aquadro 1992; Tanksley *et al.* 1992; Nachman and Churchill 1996; Copenhaver *et al.* 1998; Nachman *et al.* 1998; Stephan and Langley 1998; Hamblin and Aquadro 1999). Centromeric and telomeric regions in particular are often associated with very low recombination rates. Consequently, character differences caused by numerous genes spread homogeneously or randomly across the genome may appear to be caused by genes of large effect in centromeric or telomeric regions. However, this effect is dependent on the number of genes as well as on the recombination rate. A reduced number of genes in centromeric and telomeric regions could offset the bias resulting from a low recombination rate.

Here, we use simulation results to demonstrate that variance in recombination rate across a genome can cause systematic biases in the interpretation of mapping studies. The recent completion of the *Drosophila melanogaster* genome project (Adams *et al.* 2000) and the published rates of recombination along cytological bands of this species (Kindahl 1994) provided us the opportunity to test the impact of recombination rate variation on QTL mapping studies in this species. The mapping methods themselves are not biased, but variation in gene density per centimorgan creates the appearance of strong QTL in regions of low recombination and weak or no QTL in regions of high recombination, even when QTL are of equal effect and distributed randomly across the genome. We conclude that, in organisms with small numbers of chromosomes such as *D. melanogaster*, phenotypic differences between strains that are highly polygenic can easily produce patterns consistent with few QTL of major effect due to clustering of multiple QTL in regions of high gene density per centimorgan. We also show that this variation could cause the *D. melanogaster X* chromosome to harbor fewer and weaker observable QTL than the autosomes, consistent with several genetic studies of this species group.

Corresponding author: Mohamed A. F. Noor, Department of Biological Sciences, Life Sciences Bldg., Louisiana State University, Baton Rouge, LA 70803. E-mail: mnoor@lsu.edu

MATERIALS AND METHODS

We obtained the number of predicted mRNA coding sequences on the basis of the genome annotation (also referred to as "genes") in each of the 100 *D. melanogaster* cytological bands and the recombination rates within each of these bands from published sources (Kindahl 1994; Adams *et al.* 2000). In total, ~14,000 mRNA coding sequences were contained in the database: 2415 on the *X* chromosome, 5301 on chromosome *2*, and 6293 on chromosome *3*. Using these numbers, we manually estimated the number of coding sequences per centimorgan across the three major *D. melanogaster* chromosomes, whose lengths are ~73 cM (the *X* chromosome), 110 cM (chromosome *2*), and 110 cM (chromosome *3*). This conversion was performed by counting the number of coding sequences and recombination fractions simultaneously until a complete centimorgan was reached. If a cytological band was >1 cM, we inferred that the proportion of a centimorgan contained within the band was equal to the proportion of the total number of genes within the band present in that centimorgan. Because the dot chromosome encodes <100 transcripts, it was excluded from this study.

Following this procedure, we designed two computer programs (available upon request) to randomly assign 50 QTL affecting a hypothetical trait difference. One program assigned QTL at random with respect to the 14,000 coding sequences [*r*andom-by-*c*oding (RC)], and the other assigned QTL at random with respect to recombinational position [*r*andom-by-*r*ecombination (RR)]. RR QTL assignments are typical of the procedure used by most simulated QTL mapping studies (*e.g.*, Otto and Jones 2000).

The output of these programs was used as input for the QTL Cartographer suite of programs (Basten *et al.* 1999) to simulate the results of backcross mapping studies if 1000 progeny were scored. We used QTL Cartographer to evaluate (1) single marker linear regressions on the data and (2) a composite interval mapping (Zeng 1994) procedure on the simulated results with a window size of 10 cM and using up to five background markers. Significance was estimated using 1000 permutations of the genotypes relative to phenotypes. In our baseline program, we simulated 50 QTL of equal effect in the same direction, no dominance, a heritability of 0.3, and with markers spaced every 5 cM. Equal effect QTL were used merely to illustrate the extent of biases, and this genetic structure was not assumed to be realistic. Subsequently, we evaluated an exponential distribution of QTL effects, effects in opposite directions (transgressive segregation), a heritability of 0.7, and a marker spacing of every 1 cM. Parameters were varied individually. Both RC- and RR-simulated QTL distributions were used for all of these variations. Statistical tests comparing the results from RC and RR simulations were performed using StatView software (SAS Institute, Cary, NC). One hundred simulations for each set of conditions were run unless otherwise specified.

Throughout this article we use the word QTL in two contexts. We refer to "true QTL" as those that were assigned by our RC and RR programs. Unless otherwise specified, there were exactly 50 true QTL in all of our simulations. We refer to "predicted QTL" as those loci that QTL Cartographer subsequently detected to have a significant association with the phenotypic variance. With an infinite number of markers and an infinite sample size, QTL Cartographer should have identified 50 predicted QTL in most of our simulations.

RESULTS

Gene density per centimorgan: The distribution of coding sequences per centimorgan in the *D. melanogaster* genome is highly skewed, as can be seen for the three main chromosomes in Figure 1. Although there were typically fewer coding sequences close to the centromeres, the suppression in recombination extended beyond the centromeres far enough to create high gene densities per centimorgan in the centromeric region. Indeed, along chromosome *3*, there was a 20-fold difference in gene density per centimorgan between some points along the chromosome. For example, 1990 genes were within 5 cM of the centromere of chromosome *3*, resulting in >14% of the total number of genes occupying <3.5% of the recombinational genome. Further, the mean coding sequence density per centimorgan was also higher along the two autosomes (chromosome *2*, 48.0 genes/centimorgan; chromosome *3*, 58.5 genes/centimorgan) than along the *X* chromosome (32.2 genes/centimorgan, Mann-Whitney *U*-test, $P < 0.0001$ in both comparisons). This difference was still statistically sig-

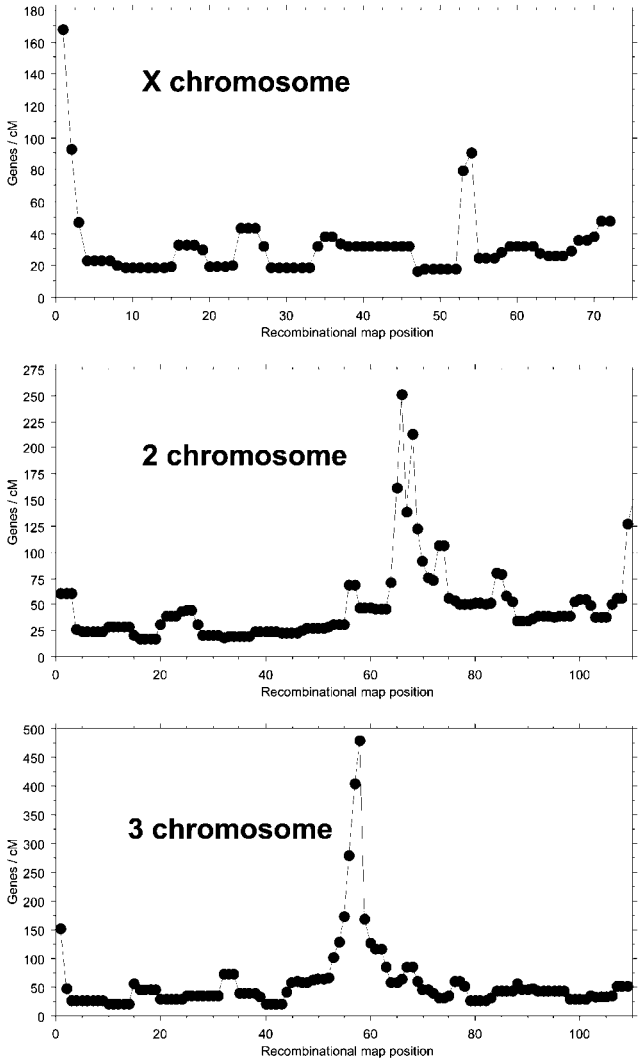

Figure 1.—Variation in number of genes at various recombinational positions in the genome of *D. melanogaster*. The centromeres are at positions 33 (*X* chromosome), 66 (chromosome *2*), and 73 (chromosome *3*).

nificant if the 5 cM surrounding the centromeres of the two autosomes was excluded.

Basic simulation results: To examine the consequences of a skewed distribution of coding sequences relative to the recombinational map on the detection of true QTL, we took a simulation approach. In all experiments, 50 true QTL loci were distributed randomly across the genome, using two different models to control placement of the QTL. In one set of simulations, true QTL were distributed randomly on the basis of the recombinational map (RR simulations). This set of simulations represents the assumptions inherent in most current simulated QTL mapping studies (*e.g.*, Otto and Jones 2000). In the second set of simulations, true QTL were distributed randomly on the basis of the distribution of coding regions on the physical map (RC simulations). This set of distributions reflects the more realistic assumption that each coding region has an equal probability of being a gene that affects the trait of interest in the absence of any specific knowledge of gene function. QTL Cartographer (Basten *et al.* 1999) was then used to identify predicted QTL that could be detected by these standard methods. If the bias in the distribution of coding regions per centimorgan were sufficiently high to significantly distort QTL detection, then we expected to identify fewer predicted QTL in the analyses of RC simulations than in the analyses of RR simulations, and we expected that strong predicted QTL would be detected more often in regions of higher gene density per centimorgan than regions of low gene density per centimorgan in the RC simulations.

The results of these simulated mapping experiments were unambiguous. Almost identical results were obtained for both single marker linear regression analyses and composite interval mapping (CIM), so we focus here on the latter. First, in RC simulations, the strongest predicted QTL detected across the genome were within 5 cM of the centromere of chromosomes *2* or *3* (usually *3*) in 46 of 100 simulated data sets (Table 1 and Figure 2). This is almost three times more frequent than the 14 of 100 RR simulations in which the strongest predicted QTL was associated with these two centromeric regions (chi-square, $P < 0.0001$). The reason for this difference is that, in RC simulations, many true QTL (sometimes up to 7 out of 50) were placed within 1 cM of each autosomal centromere because of the much higher gene density per centimorgan in these regions. This clustering was interpreted by the mapping algorithms as a single strong predicted QTL in these regions. In both simulations, no noticeable clustering of predicted QTL was observed on the *X* chromosome, consistent with its more homogeneous distribution of genes per centimorgan. The LOD score of the strongest predicted QTL was between 10.0 and 55.7 in RC simulations and between 10.8 and 37.3 in RR simulations (Mann-Whitney *U*-test, $P < 0.0001$).

CIM estimated 5–28 predicted QTL (mean 14.7) contributing to the RC-simulated phenotype and 11–26 QTL (mean 17.9) in RR simulations (Mann-Whitney *U*-test, $P < 0.0001$). While this difference may not appear dramatic, 20% of RC simulations estimated 10 or fewer predicted QTL—fewer than observed in any of the RR simulations. This difference came from the relative absence of predicted QTL on the *X* chromosome and chromosome *2* in RC simulations *vs.* RR simulations (see Table 1; Mann-Whitney *U*-test, $P < 0.0001$). Similarly, CIM did not identify any *X* chromosomal QTL in 10% of RC simulations, but all of the analyses of RR simulations identified at least one *X* chromosomal QTL (Fisher's exact test, $P = 0.0015$).

Varying QTL effect sizes, heritability, and marker density: To explore the robustness of this pattern, we allowed the true QTL to have effect sizes drawn from an exponential distribution rather than all having equal effects, as in the simulations described above. This effect size distribution may be expected for alleles fixed during the process of adaptation (Orr 1998a, 1999). The results of these simulations did not differ noticeably from the results of the simulations of true QTL bearing equal effect size for any of the parameters tested for either RR or RC simulations (Table 1). Having effects drawn from an exponential distribution did not make predicted QTL differ in strength, placement, or abundance from the basic simulations above.

We also simulated a phenotype having a heritability of 0.7 rather than 0.3. This higher heritability allowed for the detection of a significantly greater number of predicted QTL on all chromosomes together or individually in both RC and RR simulations as compared with the simulated phenotype with a heritability of 0.3 (Mann-Whitney *U*-test, $P < 0.0001$). RC and RR simulations still differed significantly in overall number of QTL detected (Mann-Whitney *U*-test, $P = 0.0003$), although this difference was proportionately much smaller (Table 1). Nonetheless, the autosomes were still consistently more likely to have predicted QTL associated with greater effects than the *X* chromosome in RC simulations. It was very rare that the strongest predicted QTL would be associated with the *X* chromosome in RC simulations (see Table 1). In 13 of the RR and 53 of the RC simulations, the strongest QTL detected were again situated within 5 cM of the centromere of chromosomes *2* or *3* (chi-square, $P < 0.0001$).

We also varied the marker density from every 5 cM in the basic simulation to every 1 cM. This change had an almost identical effect to increasing the heritability of the phenotype. The results of these simulations are also presented in Table 1.

In all of these simulation variants, predicted QTL on the *X* chromosome are typically still less abundant and weaker in RC simulations than in RR simulations. Correspondingly, the strongest predicted QTL in RC simulations are typically within 5 cM of the centromere of

TABLE 1

Comparison of composite interval mapping results in RC and RR simulations

	RC	RR
Heritability = 0.3, markers every 5 cM, equal distribution of QTL effects		
Total QTL	14.7 ± 0.5 (5–28)	17.9 ± 0.4 (11–26)
X chromosome QTL	2.7 ± 0.2 (0–10)	4.1 ± 0.2 (1–9)
Chromosome *2* QTL	5.1 ± 0.2 (1–13)	6.8 ± 0.3 (3–17)
Chromosome *3* QTL	6.9 ± 0.3 (2–16)	7.0 ± 0.3 (2–14)
No *X* QTL	10/100	0/100
QTL on both *2* and *3* bigger than *X*	86/100	40/100
Largest QTL on *X*	3/100	23/100
Largest QTL near centromere	46/100	14/100
Heritability = 0.3, markers every 5 cM, exponential distribution of QTL effects		
Total QTL	15.1 ± 0.5 (8–34)	17.1 ± 0.4 (10–27)
X chromosome QTL	2.8 ± 0.2 (0–10)	4.3 ± 0.2 (0–10)
Chromosome *2* QTL	6.2 ± 0.3 (1–15)	6.5 ± 0.3 (2–14)
Chromosome *3* QTL	6.1 ± 0.3 (2–15)	6.4 ± 0.3 (2–15)
No *X* QTL	11/100	1/100
QTL on both *2* and *3* bigger than *X*	85/100	47/100
Largest QTL on *X*	3/100	27/100
Largest QTL near centromere	46/100	8/100
Heritability = 0.7, markers every 5 cM, equal distribution of QTL effects		
Total QTL	30.0 ± 0.5 (18–41)	32.1 ± 0.4 (24–45)
X chromosome QTL	6.6 ± 0.2 (2–13)	7.9 ± 0.2 (3–13)
Chromosome *2* QTL	12.0 ± 0.3 (5–19)	11.9 ± 0.3 (6–19)
Chromosome *3* QTL	11.4 ± 0.3 (3–20)	12.3 ± 0.3 (6–18)
QTL on both *2* and *3* bigger than *X*	89/100	42/100
Largest QTL on *X*	2/100	33/100
Largest QTL near centromere	53/100	13/100
Heritability = 0.3, markers every 1 cM, equal distribution of QTL effects		
Total QTL	28.5 ± 0.8 (12–48)	32.9 ± 0.9 (13–57)
X Chromosome QTL	5.6 ± 0.4 (0–22)	8.0 ± 0.4 (3–20)
Chromosome *2* QTL	10.9 ± 0.5 (3–27)	12.3 ± 0.6 (1–29)
Chromosome *3* QTL	11.5 ± 0.5 (3–29)	12.6 ± 0.6 (4–29)
No *X* QTL	6/100	0/100
QTL on both *2* and *3* bigger than *X*	77/100	36/100
Largest QTL on *X*	1/100	30/100
Largest QTL near centromere	52/100	10/100

Means ± standard errors or counts are given with the ranges presented parenthetically. In addition to numbers of predicted QTL, we evaluated how often the predicted QTL of largest effect size on both major autosomes was associated with more of the phenotypic variance than the QTL of largest size on the *X* chromosome. We also note how often the predicted QTL of largest effect was within 5 cM of the centromere of one of the two major autosomes.

one of the two major autosomes, while those of RR simulations are more randomly distributed.

***X* chromosomal *vs.* autosomal QTL:** In the simulations above, we observed fewer predicted QTL on the *X* chromosome in RC simulations relative to RR simulations. This difference is consistent with the proportion of coding sequences on the *X* chromosome: 17.2% of QTL should be on the *X* chromosome according to gene density, while RC simulations estimated 18.4% (2.7/14.7, from Table 1) of predicted QTL to be on the *X* chromosome. Similarly, RR simulations predicted that 22.9% (4.1/17.9) of predicted QTL would be on the *X* chromosome, while 24.9% of the recombinational length of the genome is along the *X* chromosome. These observations may suggest a more severe bias in the inferred effects of QTL due to clustering in centromeric regions rather than on which chromosome a QTL is likely to be detected.

However, we expanded our simulations to evaluate the likelihood of detecting strong true QTL on the *X* chromosome *vs.* autosomes in the RC and RR models. We again randomly placed 50 true "background" QTL with small effects (37 having additive effects of 1 and 13 having additive effects of 2) along with three "larger-effect" QTL with effect sizes twice as large as the largest of the background QTL (all having additive effects of

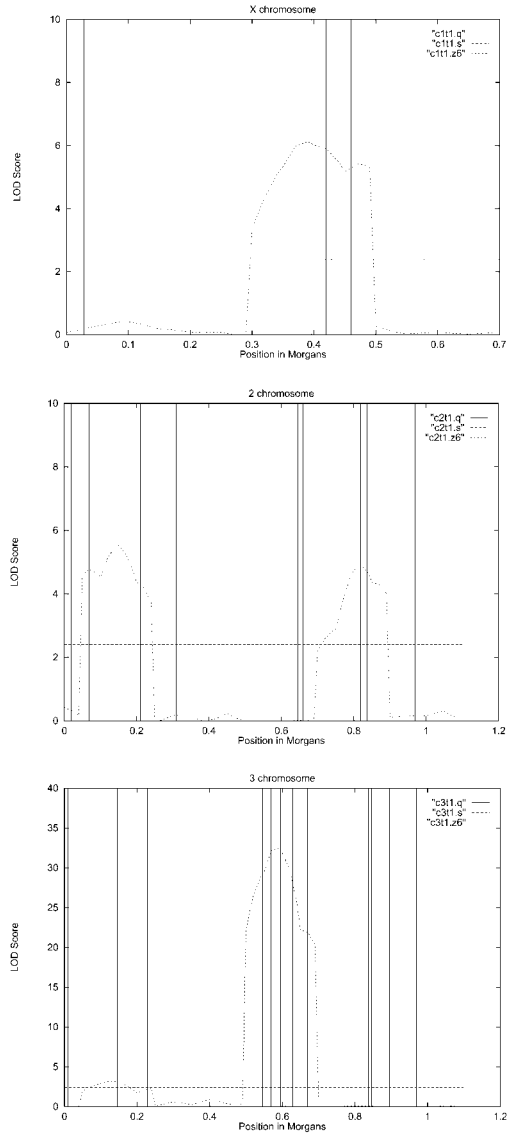

FIGURE 2.—Composite interval mapping results from a randomly selected RC-simulated mapping experiment (denoted .z6 in inset). Vertical lines (denoted by .q) indicate the locations of one or more true QTL. The threshold for statistical significance of this experiment (denoted .s) was LOD 2.4.

4). A total of 172 RC simulations and 188 RR simulations were executed, the results were analyzed using composite interval mapping (ZENG 1994), and the number of predicted QTL that were within 3 cM of the true larger-effect QTL were counted. Among RC simulations, 30 of 80 (37%) true large-effect *X* chromosome QTL were detected, and 213 of 436 (49%) true large-effect autosomal QTL were detected. Among RR simulations, 67 of 122 (55%) true large-effect *X* chromosome QTL were detected, and 237 of 442 (54%) true large-effect autosomal QTL were detected. Overall, true QTL were thus more likely to be detected in RR than in RC simulations (chi-square, $P = 0.025$). More specifically, RC and RR simulations differed significantly (chi-square, $P = 0.015$) in the likelihood of detecting large-effect *X* chromosome QTL but not large-effect autosomal QTL. This observation suggests that variation in recombination rate within *D. melanogaster* may cause true *X* chromosomal quantitative trait loci to be less likely to be detected than true autosomal loci.

Transgressive segregation: Several mapping studies have identified predicted QTL with effects opposite in direction to the difference observed between the strains being tested. To evaluate the effects of transgressive segregation on the difference between RC and RR simulations, we simulated three additional types of situations: true QTL having a 50% probability of having effects in either direction, true QTL having a 66% probability of being in one direction, and true QTL having an 83% probability of being in one direction. These were then compared with our basic simulation in which 100% of true QTL had effects in one direction.

The results of these simulations are presented in Table 2. Generally, as more true QTL had effects in opposite directions, fewer predicted QTL were detected. Interestingly, the difference between the *X* chromosome and autosomes in RC *vs.* RR simulations decayed with increasing transgressive segregation. However, even in the simulations where only one-half of the true QTL had effects in one direction, the largest predicted QTL were typically near the autosomal centromeres in RC simulations, significantly more often than in RR simulations (Table 2; Fisher's exact test, $P = 0.0002$).

All genes are QTL: As a final approach, we assigned each of the ~14,000 mRNA coding sequences from the genome of *D. melanogaster* to contribute equally to the simulated phenotype of interest. The result is presented in Figure 3. Composite interval mapping predicted two distinct predicted QTL peaks on the third chromosome, each associated with ~10% of the phenotypic variance; two smaller peaks on the second chromosome; and one region on the *X* chromosome significantly associated with the phenotypic variance. Single marker linear regressions detected strong associations between markers and phenotype across both of the autosomes, but again detected only two regions of the *X* chromosome as being significantly associated with the phenotypic variance.

DISCUSSION

QTL mapping studies necessarily provide minimum numbers of genes that contribute to differences between strains, and all of our simulations corroborate that fact. More significantly, two general patterns emerged from our simulation studies of the effect of recombination rate variation on QTL mapping studies in *D. melanogaster*: biases in inferred effect sizes of QTL and biases on which chromosomes QTL are likely to be detected. First, regions of low recombination (in this study, primarily centromeric regions) are likely to harbor the strongest apparent QTL. Given that all true QTL effects were equal in our simulations, the reason for this ten-

TABLE 2

Comparison of CIM results in RC and RR simulations when there is transgressive segregation

	RC	RR
100% effects in one direction		
Total QTL	14.7 ± 0.5 (5–28)	17.9 ± 0.4 (11–26)
No *X* QTL	10/100	0/100
QTL on both *2* and *3* bigger than *X*	86/100	40/100
Largest QTL near centromere	46/100	14/100
83% effects in one direction		
Total QTL	13.4 ± 0.4 (6–33)	15.7 ± 0.4 (5–27)
No *X* QTL	21/100	3/100
QTL on both *2* and *3* bigger than *X*	76/100	44/100
Largest QTL near centromere	50/100	8/100
66% effects in one direction		
Total QTL	11.2 ± 0.4 (4–23)	12.6 ± 0.4 (6–21)
No *X* QTL	21/100	11/100
QTL on both *2* and *3* bigger than *X*	67/100	41/100
Largest QTL near centromere	38/100	6/100
50% effects in one direction		
Total QTL	11.9 ± 0.4 (4–24)	11.7 ± 0.4 (4–23)
No *X* QTL	13/100	13/100
QTL on both *2* and *3* bigger than *X*	53/100	45/100
Largest QTL near centromere	22/100	4/100

Means ± standard errors or counts are given with the ranges presented parenthetically. In addition to numbers of predicted QTL, we evaluated how often the predicted QTL of largest effect size on both major autosomes was associated with more of the phenotypic variance than the QTL of largest size on the *X* chromosome. We also note how often the predicted QTL of largest effect was within 5 cM of the centromere of one of the two major autosomes.

dency is that multiple independent true QTL were often clustered in these regions, whereas in regions of high recombination, single true QTL were more isolated from others. This tendency does not result from the properties of QTL mapping algorithms but is instead an artifact of variation in gene density per centimorgan itself. Second, we found a tendency for regions of the *X* chromosome to harbor weaker apparent QTL (or none at all) than autosomal regions. This tendency results in part at least from the lower overall number of genes per centimorgan across the *X* chromosome, and it was amplified by the particularly high density of genes per centimorgan in the centromeric regions of the two autosomes.

Our observation that regions of low recombination should often have strong predicted QTL is intuitive but should not be trivialized. The assumption of homogeneity in gene density is explicit in virtually all simulation studies of QTL mapping, explained simply as "true QTL were randomly assigned to genomic locations," and referring to assignments identical to our RR simulations. The assumption is also implicit in empirical QTL mapping studies, as the observation of a single QTL associated with much of the genetic variance is often interpreted as a single or small number of genes associated with a disproportionate effect, rather than as the location of a region of low recombination. Our simulations under a perhaps unrealistic model, that of 50 small-but-equal-effect true QTL, often predicted single QTL of large effect in centromeric regions. Without explicit knowledge of gene density per centimorgan, no claims can be made as to whether characters are "polygenic" or "oligogenic" on the basis of QTL mapping results. If a genome has been sequenced and regional recombination rates have been estimated, then perhaps one can design a mapping protocol that would "correct" for variation in gene density per centimorgan and yield more accurate estimates of QTL numbers and effect sizes.

Many investigators interpret the "infinitesimal" model of Fisher (1958) to predict "QTLs of equal estimated magnitude distributed uniformly across the genome" (Bradshaw *et al.* 1998, p. 380). However, the presence of recombination rate variation would not support such a prediction, even if a character were truly infinitely polygenic. Instead, some regions of the genome would be associated with fairly large effects, and others would be associated with weak or no effects, which is generally what is observed. This conclusion was upheld in our simulated mapping experiment where every coding sequence of *D. melanogaster* was assigned as a true QTL, and yet parts of the genome were not significantly associated with the phenotypic variation.

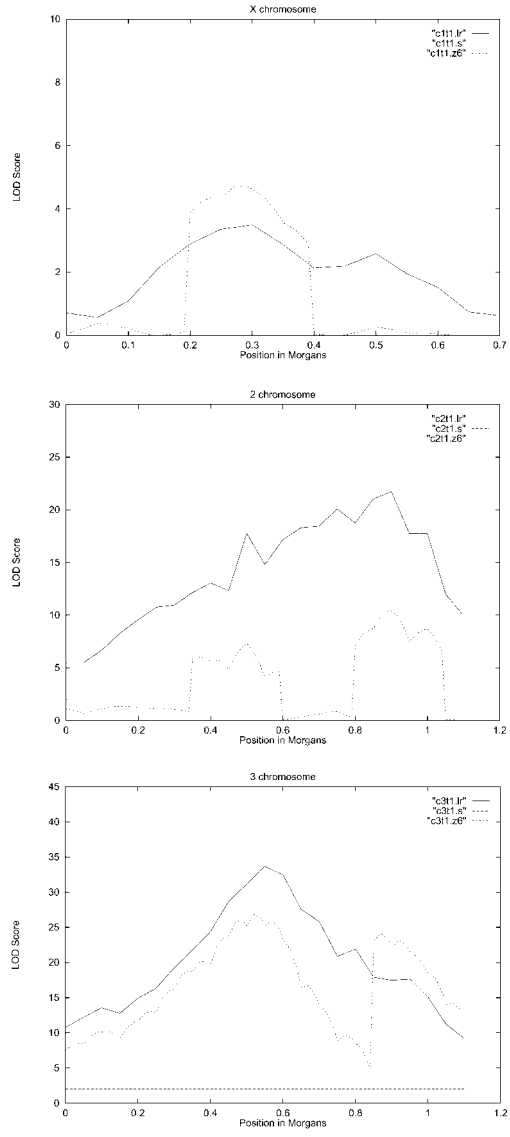

FIGURE 3.—Composite interval (denoted .z6) and single marker linear regression (denoted .lr) results of a simulated mapping experiment in which every coding sequence in the *D. melanogaster* genome contributes equally to the phenotype being studied. The threshold for statistical significance of this experiment (denoted .s) was just over LOD 2.0. Note the differences in the scale of the *y*-axis.

Recent empirical results comparing mapping studies that used recombinational linkage to molecular markers *vs.* physical mapping methodologies (*e.g.*, deficiency mapping) are consistent with the predictions of our study. For example, PASYUKOVA *et al.* (2000) used deficiency mapping to repeat an earlier genetic study of *D. melanogaster* longevity by NUZHDIN *et al.* (1997) that used recombinant inbred lines. Within several of the regions identified by longevity by NUZHDIN *et al.* (1997), PASYUKOVA *et al.* (2000) identified four or five QTL. This observation and our simulation results illustrate that intervals to which QTL are mapped can represent the combined effect of multiple contributing loci, particularly in regions of low recombination. This effect would also be true if there are inverted regions that differentiate strains being studied (*e.g.*, NOOR *et al.* 2001; WILLIAMS *et al.* 2001).

Our second observation, that the *X* chromosome of *D. melanogaster* will typically have smaller effects associated with it than the autosomes because of variation in gene density per centimorgan, has not previously been suggested. However, this finding corresponds with the "small *X*-effect" suggested by several investigators in smaller-scale genetic studies. The *X* chromosome was not significantly associated with female sexual isolation in studies of the species of this group, although strong autosomal associations were frequently identified (COYNE 1989, 1992). This observation spurred some authors to invoke differential selection pressures affecting the two sexes as a mechanism to explain this small *X*-effect (HOLLOCHER *et al.* 1997). Similar results were also obtained in studies of the genetics of male secondary sexual characters, where *X* chromosome effects should be easier to identify than autosomal ones due to hemizygosity and dosage compensation. In a study of the genetics of seven male secondary sexual character differences between *D. simulans* and *D. mauritiana* (TRUE *et al.* 1997), *X* chromosomal regions tended to be more weakly associated with phenotypes than autosomal regions: *X* chromosome QTL consistently also had the weakest or next-weakest associations with phenotypic variance for each character studied. In a comparable study of the genetics of sexual trait differences between *D. simulans* and *D. sechellia* (MACDONALD and GOLDSTEIN 1999), QTL on the *X* chromosome were smaller on average than QTL on either of the autosomes for each of the four additive traits tested. The two QTL having the strongest association with each trait were always autosomal. Several *X* chromosomal QTL were identified, but such loci would be easier to identify than comparable autosomal loci in males because of hemizygosity and dosage compensation.

The biases presented here may represent something of a worst-case scenario: *D. melanogaster* has a very small number of chromosomes. Increasing the number of chromosomes necessarily increases the overall amount of recombination in the genome through independent assortment. Hence, if the number of chromosomes increases without a corresponding increase in the number of genes [as, for example, noted in the similar number of transcripts between *D. melanogaster* and *Homo sapiens* (LANDER *et al.* 2001)], the biases will likely decrease.

Finally, the biases that we observed were most strikingly different between our RC and RR simulations in cases where most QTL had effects in the same direction. This tendency would be particularly common in genetic studies of adaptive traits (ORR 1998b). Indeed, recent work by JONES (1998) found very strong QTL for *D. sechellia*'s resistance to toxins in morinda fruit to be strongly associated with the pericentromeric regions of the third chromosome and much weaker QTL on the

X chromosome. Jones (1998) interpreted this result as consistent with an oligogenic basis to this trait, but our simulations suggest it could also be polygenic.

QTL mapping can lead to identification of individual genes contributing to a trait (*e.g.*, Doebley *et al.* 1995; Long *et al.* 1998; Frary *et al.* 2000). Nevertheless, several important implications emerge from these results. First, one cannot extrapolate from the observation that QTL are restricted to certain regions to the suggestion that few (or any) "major" genes contribute to the observed trait difference in the absence of detailed genetic and physical maps. Regions adjacent to centromeres may be especially problematic. Second, as genome sequences become available for agronomically important species, correcting for the number of genes per centimorgan across the genome may allow for more accurate QTL localization. Strong QTL detected in regions of high recombination are more likely to be single genes of large effect than are QTL in regions of low recombination, thus making them also more likely to be isolated and cloned. Finally, we hope that this study will underscore the usefulness of genome sequencing efforts and a high-quality genetic map in organisms being examined for quantitative traits. Both are often lacking in nontraditional experimental species, causing difficulty in the interpretation of QTL mapping results. Our results underscore that QTL maps should be treated as hypotheses to be tested by additional genetic methods.

This research was supported by National Institutes of Health grant GM-58060 to M.A.F.N. (subcontracted through J. Hey at Rutgers University), National Science Foundation (NSF) grant DEB9980797 to M.A.F.N., NSF grant IBN9728047 to J.C.L., and a Howard Hughes Medical Institute summer undergraduate research fellowship to A.L.C.

LITERATURE CITED

Adams, M. D., S. E. Celnicker, R. A. Holt, C. A. Evans, J. D. Gocayne *et al.*, 2000 The genome sequence of *Drosophila melanogaster*. Science **287:** 2185–2195.

Basten, C. J., B. S. Weir and Z-B. Zeng, 1999 *QTL Cartographer: A Reference Manual and Tutorial for QTL Mapping*. North Carolina State University, Raleigh, NC (http://statgen.ncsu.edu).

Begun, D. J., and C. F. Aquadro, 1992 Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. Nature **356:** 519–520.

Bradshaw, H. D., K. G. Otto, B. E. Frewen, J. K. McKay and D. W. Schemske, 1998 Quantitative trait loci affecting differences in floral morphology between two species of monkeyflower (Mimulus). Genetics **149:** 367–382.

Copenhaver, G. P., W. E. Browne and D. Preuss, 1998 Assaying genome-wide recombination and centromere functions with *Arabidopsis* tetrads. Proc. Natl. Acad. Sci. USA **95:** 247–252.

Coyne, J. A., 1989 Genetics of sexual isolation between two sibling species, Drosophila simulans and Drosophila mauritiana. Proc. Natl. Acad. Sci. USA **86:** 5464–5468.

Coyne, J. A., 1992 Genetics of sexual isolation in females of the Drosophila simulans species complex. Genet. Res. Camb. **60:** 25–31.

Doebley, J., A. Stec and C. Gustus, 1995 *teosinte branched1* and the origin of maize: evidence for epistasis and the evolution of dominance. Genetics **141:** 333–346.

Fisher, R. A., 1958 *The Genetical Theory of Natural Selection*. Dover, New York.

Frary, A., T. C. Nesbitt, S. Grandillo, E. Knaap, B. Cong *et al.*, 2000 fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. Science **289:** 85–88.

Hamblin, M. T., and C. F. Aquadro, 1999 DNA sequence variation and the recombinational landscape in *Drosophila pseudoobscura*: a study of the second chromosome. Genetics **153:** 859–869.

Hollocher, H., C. T. Ting, M. L. Wu and C.-I. Wu, 1997 Incipient speciation by sexual isolation in *Drosophila melanogaster*: extensive genetic divergence without reinforcement. Genetics **147:** 1191–1201.

Jones, C., 1998 The genetic basis of *Drosophila sechellia*'s resistance to a host plant toxin. Genetics **149:** 1899–1908.

Kindahl, E. C., 1994 Recombination and DNA polymorphism on the third chromosome of *Drosophila melanogaster*, Ph.D. Thesis, Cornell University, Ithaca, NY.

Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody *et al.*, 2001 Initial sequencing and analysis of the human genome. Nature **409:** 860–921.

Larkin, J. C., and J. L. Woolford, 1983 Molecular cloning and analysis of the *CRY1* gene: a yeast ribosomal gene. Nucleic Acids Res. **11:** 403–420.

Long, A. D., R. F. Lyman, C. H. Langley and T. F. C. Mackay, 1998 Two sites in the *Delta* gene region contribute to naturally occurring variation in bristle number in *Drosophila melanogaster*. Genetics **149:** 999–1017.

Macdonald, S. J., and D. B. Goldstein, 1999 A quantitative genetic analysis of male sexual traits distinguishing the sibling species *Drosophila simulans* and *D. sechellia*. Genetics **153:** 1683–1699.

Nachman, M. W., and G. A. Churchill, 1996 Heterogeneity in rates of recombination across the mouse genome. Genetics **142:** 537–548.

Nachman, M. W., V. L. Bauer, S. L. Crowell and C. F. Aquadro, 1998 DNA variability and recombination rates at X-linked loci in humans. Genetics **150:** 1133–1141.

Noor, M. A. F., K. L. Grams, L. A. Bertucci, Y. Almendarez, J. Reiland *et al.*, 2001 The genetics of reproductive isolation and the potential for gene exchange between *Drosophila pseudoobscura* and *D. persimilis* via backcross hybrid males. Evolution **55:** 512–521.

Nuzhdin, S. V., E. G. Pasyukova, C. L. Dilda, Z-B. Zeng and T. F. C. Mackay, 1997 Sex-specific quantitative trait loci affecting longevity. Proc. Natl. Acad. Sci. USA **94:** 9734–9739.

Orr, H. A., 1998a The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. Evolution **52:** 935–949.

Orr, H. A., 1998b Testing natural selection *vs.* genetic drift in phenotypic evolution using quantitative trait locus data. Genetics **149:** 2099–2104.

Orr, H. A., 1999 The evolutionary genetics of adaptation: a simulation study. Genet. Res. Camb. **74:** 207–214.

Otto, S. P., and C. D. Jones, 2000 Detecting the undetected: estimating the total number of loci underlying a trait in QTL analyses. Genetics **156:** 2093–2107.

Pasyukova, E. G., C. Vieira and T. F. C. Mackay, 2000 Deficiency mapping of quantitative trait loci affecting longevity in *Drosophila melanogaster*. Genetics **156:** 1129–1146.

Stephan, W., and C. H. Langley, 1998 DNA polymorphism in Lycopersicon and crossing-over per physical length. Genetics **150:** 1585–1603.

Tanksley, S. D., M. W. Ganal, J. P. Prince, M. C. deVicente, M. W. Bonierbale *et al.*, 1992 High density molecular linkage maps of the tomato and potato genomes. Genetics **132:** 1141–1160.

True, J. R., J. Liu, L. F. Stam, Z-B. Zeng and C. C. Laurie, 1997 Quantitative genetic analysis of divergence in male secondary sexual traits between *Drosophila simulans* and *Drosophila mauritiana*. Evolution **51:** 816–832.

Williams, M. A., A. G. Blouin and M. A. F. Noor, 2001 Courtship songs of *Drosophila pseudoobscura* and *D. persimilis*. II. Genetics of species differences. Heredity **86:** 68–77.

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

Communicating editor: Z-B. Zeng