Virginie Orgogozo, Karl W. Broman and David L. Stern

SUPPLEMENTAL MATERIAL

***D. sechellia* backcross:** We phenotyped 226 backcross progeny flies. The hypothesis that ovariole number is normally distributed could not be rejected ($p>0.25$). We genotyped the 47 flies with the highest and the 47 flies with the lowest mean ovariole number using 14 molecular markers (Supplemental Table 2). The largest distance between markers was 39cM, thus allowing the detection of QTLs with an effect of 1 ovariole or more (calculation not shown, adapted from DARVASI and SOLLER 1992; LYNCH and WALSH 1998; SOLLER *et al.* 1976; SOLLER and GENIZI 1978)).

First *D. simulans* backcross: We first phenotyped 313 backcross progeny flies. An initial interval mapping analysis using the five morphological markers *f, nt, pm, st* and *e* indicated that a main QTL resides between *st* and *e*. To increase resolution within this region, 70 flies that were recombinants between *st* and *e* were added to the analysis. Then, 26 *st* - *e* recombinant flies with the lowest mean ovariole number, 26 *st* - *e* recombinant flies with the highest mean ovariole number, 21 non *st* - *e* recombinant flies with the lowest mean ovariole number and 21 non *st* - *e* recombinant flies with the highest mean ovariole number were genotyped for 10 molecular markers (Supplemental Table 3). The hypothesis that ovariole number is normally distributed could not be rejected ($p>0.05$). The largest distance between markers was 42.5cM, thus allowing the

detection of QTLs with an effect of 1 ovariole or more (calculation not shown, adapted from DARVASI and SOLLER 1992; LYNCH and WALSH 1998; SOLLER *et al.* 1976; SOLLER and GENIZI 1978)).

When we found that the region on the right of *e* was likely to contain a QTL, we also genotyped the *e-slo* recombinants from the first *D. simulans* backcross for the markers *nos* and *cpo*. These data were included in the total analysis of the first and second backcrosses.

Second *D. simulans* backcross: 1,038 *st - e* recombinant flies were phenotyped. All were genotyped for the 5 morphological markers. Among these, the 496 flies with the highest and lowest mean ovariole numbers were genotyped for 3 molecular markers (*acc004516*, *Ga47A* and *rpk*) and a random subset of these (95 flies) were also genotyped for the marker *grh*. Then, individuals that displayed recombination events between *rpk* and *e*, or between *acc004516* and *Ga47A*, were genotyped for additional intervening markers (Supplemental Table 1, Supplemental Table 3).

Composite interval mapping: Composite interval mapping (ZENG *et al.* 1994) was performed using the R/qtl program (BROMAN *et al.* 2003), with the multiple imputation method of Sen and Churchill (SEN and CHURCHILL 2001), to identify genomic regions containing QTLs affecting ovariole number, while simultaneously controlling for the effects of QTLs located on other chromosomes using background markers. We did not use background markers located on the same chromosome as the test position because the results are greatly influenced by the choice of window sizes. Instead, we performed two- and three-dimensional scans on each chromosome while accounting for QTLs identified on the other chromosomes.

Mean ovariole number differed slightly between the first and second *D. simulans* backcross experiments. To correct for this difference, an additive covariate representing the backcross was included in the model for the analysis of the pooled data. We thus assumed that the effects of the QTLs were the same in both backcrosses, but allowed that the average ovariole number could be shifted between the two backcrosses.

Hypotheses were tested by calculating a LOD score, $LOD=\log_{10}(L_1/L_0)$, where L_1/L_0 is the ratio of the likelihood under the alternative hypothesis (there is a QTL at the locus) to the null hypothesis (there is no QTL at the locus). The LOD score measures the strength of the evidence for the presence of a QTL at the tested location, compared to there being no segregating QTL at this position. The LOD score was initially evaluated at every centiMorgan (cM); for the final analysis of QTL locations on chromosome 3, calculations were performed at every 0.25 cM. The experiment-wise significance level was determined by permutation (CHURCHILL and DOERGE 1994).

Similar composite interval mapping results were obtained with the QTL Cartographer version 1.17e software (BASTEN *et al*. 1997) (not shown). Maximum-likelihood interval mapping assumes that the phenotypic data are normally distributed. However, the ovariole data are slightly leptokurtic for the second *D. simulans* backcross. We therefore explored the use of a Box-Cox transformation, defined as

$y_{transformed} = \frac{y^{\lambda} - 1}{\lambda}$ where y is the positive value to transform and λ is the transformation parameter, to transform the data so that they follow a normal distribution, with y equal to the standardized ovariole number plus 10 (so that this number is positive) and $\lambda = 2$ (using the {car} package in the R software (R DEVELOPMENT CORE TEAM 2004)). Composite interval mapping, as implemented by the QTL Cartographer version 1.17e software

(BASTEN *et al*. 1997), gave identical results with transformed or untransformed data. All results are therefore presented for untransformed data.

LITERATURE CITED

BASTEN, C. J., B. S. WEIR and Z. Z. B., 1997 *QTL cartographer: A reference manual and tutorial for QTL mapping*, Dep. of Statistics, NC State Univ., Raleigh.

BROMAN, K. W., H. WU, S. SEN and G. A. CHURCHILL, 2003 R/qtl: QTL mapping in experimental crosses. Bioinformatics **19:** 889-890.

CHURCHILL, G. A., and R. W. DOERGE, 1994 Empirical Threshold Values for Quantitative Trait Mapping. Genetics **138:** 963-971.

DARVASI, A., and M. SOLLER, 1992 Selective Genotyping for Determination of Linkage between a Marker Locus and a Quantitative Trait Locus. Theoretical and Applied Genetics **85:** 353-359.

LYNCH, M., and B. WALSH, 1998 *Genetics and analysis of quantitative traits*. Sinauer, Sunderland, Mass.

R DEVELOPMENT CORE TEAM, 2004 R: A language and environment for statistical computing, pp. R Foundation for Statistical Computing, Vienna, Austria.

SEN, S., and G. A. CHURCHILL, 2001 A statistical framework for quantitative trait mapping. Genetics **159:** 371-387.

SOLLER, M., T. BRODY and A. GENIZI, 1976 Power of Experimental Designs for Detection of Linkage between Marker Loci and Quantitative Loci in Crosses between Inbred Lines. Theoretical and Applied Genetics **47:** 35-39.

SOLLER, M., and A. GENIZI, 1978 Efficiency of Experimental-Designs for Detection of Linkage between a Marker Locus and a Locus Affecting a Quantitative Trait in Segregating Populations. Biometrics **34:** 47-55.

ZENG, Z. B., A. DARVASI, A. WEINREB, V. MINKE, J. I. WELLER *et al.*, 1994 Precision mapping of quantitative trait loci. Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. Genetics **136:** 1457-1468.