

# *hairy*: A Quantitative Trait Locus for *Drosophila* Sensory Bristle Number

Charles Robin,\* Richard F. Lyman,<sup>†</sup> Anthony D. Long,<sup>‡</sup> Charles H. Langley<sup>§</sup>  
and Trudy F. C. Mackay<sup>†,1</sup>

\*Centre for Environmental Stress and Adaptation Research, Department of Genetics, University of Melbourne, Victoria, 3010, Australia, <sup>†</sup>Department of Genetics and W. M. Keck Center for Behavioral Biology, North Carolina State University, Raleigh, North Carolina 27695-7614, <sup>‡</sup>Department of Ecology and Evolutionary Biology, University of California, Irvine, California 92697-2525 and <sup>§</sup>Center for Population Biology and Section of Evolution and Ecology, University of California, Davis, California 95616-8554

Manuscript received February 15, 2002

Accepted for publication May 13, 2002

## ABSTRACT

Advances in medicine, agriculture, and an understanding of evolution depend on resolving the genetic architecture of quantitative traits, which is challenging since variation for complex traits is caused by multiple interacting quantitative trait loci (QTL) with small and conditional effects. Here, we show that the key developmental gene, *hairy* (*h*), is a QTL for *Drosophila* sternopleural bristle number, a model quantitative trait. Near-isoallelic lines (NIL) for the *h* gene region exhibited significant variation in bristle number and failed to complement a *hairy* mutation. Sequencing 10 *h* alleles from a single population revealed 330 polymorphic sites in ~10 kb. Genotypes for 25 of these and 14 additional sites in the flanking regions were determined for the 57 NIL and associated with variation in bristle number in four genetic backgrounds. A highly significant association was found for a complicated insertion/deletion polymorphism upstream of the transcription start site. This polymorphism, present in 17.5% of the *h* alleles, was associated with an increase of 0.5 bristle and accounted for 31% of the genetic variance in bristle number in the NIL.

**M**OST traits of medical, agricultural, and evolutionary significance vary continuously in natural populations, due to the segregation of multiple quantitative trait loci (QTL), with individually small effects that are sensitive to the genetic, sexual, and external environments. Understanding the genetics of such quantitative traits begins with a genome scan for QTL and ends with the molecular definition of functional QTL alleles. In recent years, the availability of dense polymorphic marker linkage maps and robust statistical methods for estimating map positions and effects of QTL by linkage to these markers has generated an explosion of QTL maps for morphological, disease susceptibility, behavioral, and fitness-related traits in multiple species (MACKAY 2001). However, only a few QTL with large effects have been mapped to genetic loci using traditional positional cloning methodology (CORMIER *et al.* 1997; DOEBLEY *et al.* 1997; FRARY *et al.* 2000; FRIDMAN *et al.* 2000; EL-ASSAL *et al.* 2001). The difficulty arises because replicated QTL genotypes are necessary to estimate effects of QTL alleles; if QTL effects are small, the phenotype of a single

individual is not a reliable indicator of the QTL genotype.

In model organisms for which whole-genome sequence is available, it is possible to nominate positional candidate genes contained in the QTL interval and systematically test whether molecular polymorphisms in these candidate genes are associated with the quantitative trait phenotype (HORIGAWA *et al.* 2000) in natural populations. However, QTL intervals can contain many positional candidate genes (MACKAY 2001), some of which clearly affect the quantitative trait phenotype, but most of which are either of unknown function or might have unknown pleiotropic effects on the trait. In genetic model organisms in which controlled crosses can be made and mutations at positional candidate genes generated, quantitative complementation tests of QTL alleles to mutant and wild-type alleles of candidate genes in the QTL region can narrow the number of positional candidate genes targeted for such association studies (MACKAY 2001).

This strategy has the highest likelihood of success for quantitative traits for which many candidate genes have been defined by mutations that affect the phenotypic expression of the trait. Numbers of abdominal and sternopleural bristles in *Drosophila* are classic model quantitative traits (MACKAY 1996), with large amounts of naturally segregating genetic variation. Further, these bristles are external mechanosensory organs of the peripheral nervous system (PNS), and the many genes known to affect PNS development (CAMPOS-ORTEGA 1993; JAN and JAN 1993; KANIA *et al.* 1995; SALZBERG *et*

Sequence data from this article have been deposited with the EMBL/GenBank data libraries under the following accession nos.: *D. melanogaster* lines R105, AY055833; R107, AY055834; R6, AY055835; R2, AY055836; R19, AY055837; R53, AY055838; R48, AY055839; R95, AY055840; R74, AY055841; R24, AY055842; and *D. simulans* *h* region (Dsimhairy), AY055843.

<sup>1</sup>Corresponding author: Department of Genetics, Campus Box 7614, North Carolina State University, Raleigh, NC 27695-7614. E-mail: trudy\_mackay@ncsu.edu

*al.* 1997) are candidate bristle number QTL. The map positions of bristle number QTL often include these candidate genes (LONG *et al.* 1995; GURGANUS *et al.* 1998, 1999; NUZHIDIN *et al.* 1999), and complementation tests of QTL alleles to candidate gene mutations have implicated several of these loci as the genes corresponding to bristle number QTL (LONG *et al.* 1996; LYMAN and MACKAY 1998; GURGANUS *et al.* 1999; LYMAN *et al.* 1999). Subsequent association studies showed that molecular variation at three of these genes, *scabrous* (LYMAN *et al.* 1999), *Delta* (LONG *et al.* 1998), and the *achaete-scute* complex (LONG *et al.* 2000), was indeed associated with naturally occurring phenotypic variation in bristle number.

The *Drosophila* gene *hairy* (*h*) encodes a basic helix-loop-helix transcriptional repressor and exhibits a high degree of pleiotropy (RUSHLOW *et al.* 1989). As one of the first zygotically expressed genes, it functions as a pair-rule gene demarcating the first segmental boundaries in the developing embryo (INGHAM *et al.* 1985a). As the name suggests, *h* also plays a role in bristle development, with the classic mutants having supernumerary bristles distributed across the epidermis (INGHAM *et al.* 1985b). The expression pattern of *Hairy* in the wing, leg, and eye imaginal discs and the genetic interactions observed between it and *achaete*, another helix-loop-helix protein, suggest that *h* functions to regulate the spatial pattern of PNS development (CARROLL and WHYTE 1989). Previous studies have shown that QTL affecting variation in bristle number map to the *h* gene region (3-26.5 cM, 66D; SHRIMPTON and ROBERTSON 1988; LONG *et al.* 1995; GURGANUS *et al.* 1999) and fail to complement *h* mutations (LONG *et al.* 1996; GURGANUS *et al.* 1999).

Here, we report the results of further tests of the hypothesis that *h* is a bristle number QTL. We constructed a panel of near-isoallelic lines (NIL) for the *h* gene region and tested the extent to which they contributed to naturally occurring variation in bristle number and for failure to complement an *h* mutation. We also conducted tests for association of molecular polymorphisms in the *h* gene region with variation in bristle number. In outbred populations, the power to detect associations between polymorphic molecular markers and quantitative trait phenotypes depends on the magnitude of the effect of the causal molecular variant [the quantitative trait nucleotide (QTN)], the sample size, and the strength of linkage disequilibrium between the QTN and the markers used in the association test (LONG and LANGLEY 1999; NIELSEN and WEIR 1999). If only a subset of polymorphisms are genotyped, the optimal spacing of markers is the physical distance corresponding to the historical recombination parameter  $4Nc$  ( $= R$ ) (HUDSON 1987; LONG and LANGLEY 1999). Unfortunately,  $R$  varies by at least an order of magnitude among different gene regions and even within genes (MIYASHITA and LANGLEY 1988; AGUADÉ *et al.* 1989; MIYA-

SHITA 1990; LONG *et al.* 1998, 2000) and cannot be predicted in advance. Therefore, we sequenced a subset of *h* alleles to guide the choice of molecular markers used in the association study.

## MATERIALS AND METHODS

**Drosophila stocks:** Fifty-seven isogenic third chromosomes, derived from independent isofemale lines collected from Raleigh, North Carolina, were substituted into the *Samarkand* (*Sam*) homozygous genetic background (LYMAN and MACKAY 1998).

An allele of *hairy* ( $h^1$ ) was introgressed into *Sam* by 20 generations of backcrossing. After the initial cross of the *Sam* and  $h^1$  stocks (G0), *Sam/h^1* males and females were mated *inter se* (G1).  $h^1$  homozygous males from this cross were backcrossed to *Sam* females (G2). G1 and G2 crosses were repeated 19 more times, after which the backcross stock *Sam; h^1* (BC20) was established.

Two independent *h* region NIL were constructed for each of the chromosome substitution lines by 10 generations of backcrossing females heterozygous for the wild (+) and  $h^1$  (BC20) chromosomes to *Sam; h^1* (BC20) males. Homozygous NIL were derived by crossing *Sam; TM6B, h^1 D^3/h^1 females to  $+/h^1$  (BC20) males at G11; mating *Sam; TM6B, h^1 D^3/+* females and males *inter se* at G12; and eliminating the balancer chromosome at G13.*

All stocks were maintained on cornmeal-agar-molasses medium at 25°.

**Bristle number phenotypes:** Abdominal and sternopleural bristle numbers were scored as previously described (LYMAN and MACKAY 1998) in four genetic backgrounds: (i) homozygous chromosome 3 substitution lines; (ii) homozygous *h* region NIL; (iii) *h* region NIL as heterozygotes against *Sam; h^1* (BC20); (iv) *h* region NIL as heterozygotes against *Sam*. Both bristle traits were recorded on 10 males and 10 females from each of two replicate vials per line (or per replicate NIL).

**hairy sequence:** DNA sequence data were obtained for 10 kb including the *hairy* gene for 10 of the homozygous *D. melanogaster* alleles and one *D. simulans* allele. PCR primers were designed to amplify partially overlapping 2- to 3.5-kb segments of the *hairy* gene region. Several 50- $\mu$ l reactions from each primer pair were pooled for each line to minimize the contribution of polymerase errors to sequence variation and purified using Qiaquick columns (QIAGEN, Valencia, CA). PCR products were sequenced directly from both strands with internal primers and ABI big dye terminator chemistry. Sequence quality was assessed using PhredPhrap software (EWING *et al.* 1998). Sequences were aligned using Macvector and VectorNTI programs. Chromatograms were checked for singleton polymorphic sites and manually edited where necessary.

**Polymorphism genotyping:** Restriction map polymorphism in a 29-kb region including the *h* locus was evaluated using three 6-base-cutter restriction enzymes (*EcoRI*, *HindIII*, *BamHI*), exactly as described previously (LONG *et al.* 1998). A total of 25 polymorphic sites at intermediate frequency in the sample of 10 sequenced alleles was genotyped for the remaining 47 alleles by three methods.

The genotypes of 23 polymorphic sites were determined by pyrosequencing, a "sequence by synthesis" technique that monitors the release of pyrophosphate via a light-producing enzyme cascade (RONAGHI *et al.* 1998). Pyrosequencing was performed on a PSQ96 Pyrosequencer using reagents and protocols supplied by Pyrosequencing AB. The distribution of polymorphisms at *h* was such that multiple sites could often be scored in a single pyrosequencing reaction extending up

**TABLE 1**  
**Analysis of variance of bristle number for chromosome 3 substitution lines**

| Source <sup>a</sup> | d.f. | Sternopleural bristles |           |               | Abdominal bristles |           |            |
|---------------------|------|------------------------|-----------|---------------|--------------------|-----------|------------|
|                     |      | MS                     | F         | $\sigma^{2b}$ | MS                 | F         | $\sigma^2$ |
| S                   | 1    | 27.2                   | 4.38*     | Fixed         | 2834               | 130****   | Fixed      |
| L                   | 56   | 229                    | 29.6****  | 5.53          | 141                | 6.30****  | 2.97       |
| S × L               | 56   | 6.21                   | 1.55*     | 0.11          | 21.8               | 5.48****  | 0.89       |
| V(L)                | 57   | 5.53                   | 1.38 (NS) | 0.08          | 4.62               | 1.16 (NS) | 0.03       |
| S × V(L)            | 57   | 4.00                   | 1.32 (NS) | 0.10          | 3.97               | 1.05 (NS) | 0.02       |
| Error               | 2279 | 3.03                   |           | 3.03          | 3.78               |           | 3.78       |

\*\*\*\* $P < 0.0001$ ; \* $0.01 < P < 0.05$ ; NS,  $P > 0.05$ .  
<sup>a</sup> See text for explanation.  
<sup>b</sup> Variance component.

to 27 nucleotides from the sequencing primer, designed to anneal 1–5 bases from the first polymorphic site.

Genotypes of marker AG646-7GC were determined by allele-specific PCR. Two primers were designed such that the 3' base corresponded to the alternate polymorphic states. One primer was synthesized with 12 random nucleotides at the 5' end. These primers were used in a PCR reaction including a reverse primer that bound to the complementary strand 201 bp away at 58°. The allelic state of the product was determined by electrophoresis on a 3% Metaphor gel.

The genotype of C8591T was scored by digesting PCR products with *Sco*FI and assaying for digestion on a 1.8% agarose gel.

**Data analysis:** Quantitative genetic variation in bristle number for each of the four genotypes, as well as quantitative complementation of *h* region NIL to mutant and wild-type *h* alleles, was assessed by mixed-model analyses of variance (ANOVA). The model for partitioning the variance in bristle number for the whole chromosome 3 substitution lines was

$Y = \mu + L + S + L \times S + V(L) + S \times V(L) + \text{error}$ , where *L* and *S* are, respectively, the cross-classified random and fixed effects of line and sex, *V* is replicate vial, and parentheses indicate nested effects. The ANOVA models for homozygous *h* NIL and crosses of the NIL to wild-type and mutant *h* alleles were  $Y = \mu + L + S + L \times S + R(L) + S \times R(L) + V(L \times BC) + S \times V(L \times BC) + \text{error}$ , where *R* denotes the two replicate NIL per original chromosome substitution line. The model for the complementation test included cross (*C*) as an additional fixed factorial effect:  $Y = \mu + C + S + C \times S + L + C \times L + S \times L + S \times C \times L + R(L) + C \times R(L) + S \times R(L) + C \times S \times R(L) + V(C \times L \times R) + S \times V(C \times L \times R) + \text{Error}$ .

Estimates of  $\theta$  (WATTERSON 1975) and  $\pi$  (NEI and TAJIMA 1981) and tests for departure from neutrality (HUDSON *et al.* 1987; TAJIMA 1989; McDONALD and KREITMAN 1991) were applied to the DNA sequence data using DnaSP software (ROZAS and ROZAS 1999). Significance of pairwise linkage disequilibria was determined using Fisher's exact test. Hudson's estimator (HUDSON 1987) of the historical recombination parameter, *R*, was computed using DnaSP.

Associations between molecular polymorphisms and bristle phenotypes were assessed by three-way factorial ANOVA of line means, according to the model  $Y = \mu + M + G + S + M \times G + M \times S + G \times S + M \times G \times S + \text{Error}$ , where *M*, *G*, and *S* denote the fixed effects of molecular marker allele, background genotype, and sex, respectively. Reduced models were also run to test marker associations within each genotype.

All ANOVAs and *F*-ratio tests of significance were computed using SAS software.

**RESULTS AND DISCUSSION**

**Quantitative genetic analyses:** We extracted 57 third chromosomes from the Raleigh population and substituted them into the highly inbred *Sam* background. There was highly significant variation among lines for both bristle traits (Table 1, Figure 1). Assuming random mating, the genetic variance ( $V_G$ ) for bristle number is estimated by  $\sigma_L^2/2 + \sigma_{SL}^2$ , where  $\sigma_L^2$  and  $\sigma_{SL}^2$  are, respectively, the among-line and sex × line variance components (LYMAN and MACKAY 1998). The heritability ( $h^2$ ) attributable to segregation of third chromosome bristle number QTL is  $V_G/[V_G + V_E]$ , where  $V_E$ , the environmental variance, is estimated by the within-line variance com-

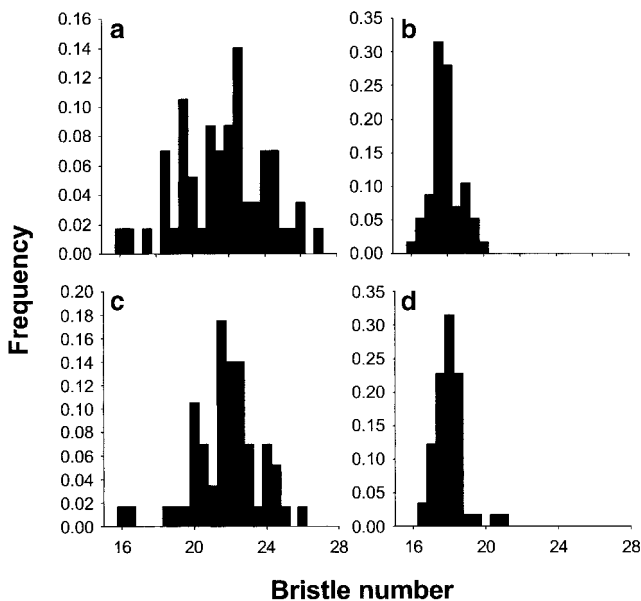


FIGURE 1.—Distributions of bristle number line means. (a) Sternopleural bristle number, homozygous chromosome 3 substitution lines. (b) Sternopleural bristle number, *h* NIL. (c) Abdominal bristle number, homozygous chromosome 3 substitution lines. (d) Abdominal bristle number, *h* NIL.

TABLE 2  
Analysis of variance of bristle number for *h* NIL

| Source <sup>a</sup>               | d.f. | Sternopleural bristles |           |                         | Abdominal bristles |           |                         |
|-----------------------------------|------|------------------------|-----------|-------------------------|--------------------|-----------|-------------------------|
|                                   |      | MS                     | <i>F</i>  | $\sigma^2$ <sup>b</sup> | MS                 | <i>F</i>  | $\sigma^2$ <sup>b</sup> |
| <i>S</i>                          | 1    | 112                    | 29.7****  | Fixed                   | 3717               | 473****   | Fixed                   |
| <i>L</i>                          | 56   | 49.4                   | 4.06****  | 0.48                    | 50.3               | 2.07**    | 0.33                    |
| <i>S</i> × <i>L</i>               | 56   | 3.83                   | 1.15 (NS) | 0.01                    | 7.95               | 1.00 (NS) | 0.00                    |
| <i>R(L)</i>                       | 55   | 11.7                   | 2.18***   | 0.16                    | 24.2               | 2.65****  | 0.39                    |
| <i>S</i> × <i>R(L)</i>            | 55   | 3.33                   | 1.17*     | 0.07                    | 7.94               | 2.47****  | 0.24                    |
| <i>V(R</i> × <i>L)</i>            | 110  | 4.02                   | 2.02****  | 0.10                    | 4.42               | 1.37*     | 0.06                    |
| <i>S</i> × <i>V(R</i> × <i>L)</i> | 110  | 1.99                   | 0.98 (NS) | 0.00                    | 3.21               | 0.94 (NS) | 0.00                    |
| Error                             | 3985 | 2.02                   |           | 2.02                    | 3.41               |           | 3.41                    |

\*\*\*\* $P < 0.0001$ ; \*\*\* $0.0001 < P < 0.001$ ; \*\* $0.001 < P < 0.01$ ; \* $0.01 < P < 0.05$ ; NS,  $P > 0.05$ .

<sup>a</sup> See text for explanation.

<sup>b</sup> Variance component.

ponent ( $\sigma_{\bar{L}}^2$ ). Estimates of these parameters were similar for the two bristle traits: For sternopleural bristle number,  $V_G = 2.88$  and  $h^2 = 0.47$ ; for abdominal bristle number  $V_G = 2.37$  and  $h^2 = 0.39$ .

The extent to which variation in the *h* gene region contributed to segregating variation in bristle number was assessed by constructing two independent NIL for each of the wild-derived *h* alleles. The NILs are expected to differ only in the 20 cM encompassing *h* (7% of the genome, 18% of the third chromosome; CROW and KIMURA 1970). There was highly significant variation in sternopleural and abdominal bristle number among the NIL (Table 2, Figure 1). The proportion of the total third chromosome genetic variance attributable to the *h* gene region (estimated by the ratio of the among-line variance component for the *h* NIL to the among-line variance of the whole chromosome) was 8.7% for sternopleural bristle number and 7.0% for abdominal bristle number. The *h* gene region did not contribute to sex-specific variation in abdominal bristle number. These data confirm earlier results that mapped bristle number QTL to this region (LONG *et al.* 1995; GURGANUS *et al.* 1998, 1999; NUZHIDIN *et al.* 1999).

To evaluate the contribution of *h* to the variation among lines, we crossed each NIL to *h*<sup>1</sup>, which had been introgressed by 20 generations of backcrossing into *Sam*, and to *Sam*, containing a wild-type *h* allele. The NIL failed to complement *h* for sternopleural, but not abdominal, bristle number (Table 3). This is also consistent with previous complementation tests showing failure of high and low selected chromosomes to complement *h* alleles for sternopleural, but not abdominal, bristle number (LONG *et al.* 1996; GURGANUS *et al.* 1999). Failure to complement indicates a genetic interaction between the sternopleural bristle number QTL and *h*, but cannot discriminate whether the interaction is allelic or epistatic. Association studies can provide corroborating evidence that *h* is a bristle number QTL.

**Molecular population genetics of *h*:** We sequenced

10.08 kb, including the *h* transcription unit from 10 *Drosophila melanogaster* alleles from the Raleigh population and one *D. simulans* allele. *h* is highly polymorphic within *D. melanogaster*. We observed 279 single-nucleotide polymorphisms (SNPs) and 51 length variants (indels) in the sample of 10 alleles; 158 SNPs and 23 indels were shared by two or more lines. Estimates of nucleotide diversity based on the number of segregating sites ( $\theta$ , WATTERSON 1975) and the average number of nucleotide differences between pairs of sites ( $\pi$ , NEI and TAJIMA 1981) were, respectively, 0.01022 and  $0.01016 \pm 0.00087$  (SD).

Three of the alleles were nearly identical, with only two nucleotides different between them in the 10 kb sequenced (Figure 2B). The three alleles that shared the “10-kb haplotype” had 13 polymorphic sites that distinguished them from the other sequences, and comparison with the *D. simulans* allele suggests all 13 are derived in the 10-kb haplotype lineage. One of these polymorphic sites was the only amino acid polymorphism observed in the sample—an alanine/serine polymorphism at position 5410 of the aligned sequence (Figure 2B). Given the number of segregating sites in the sample and the degree of recombination observed (see below), simulations indicate that this partitioning of the variation among alleles is unlikely under a neutral model ( $P < 0.03$ ; HUDSON *et al.* 1994) and suggest that these alleles may have recently risen to high frequency in this population. Of course, rejection of the null hypothesis does not imply selection as the cause of the increase in frequency of the 10-kb haplotype; recent admixture could also account for this pattern.

Application of several other tests of selection failed to detect any other significant departures from neutrality. The sequence of the *D. simulans* allele showed that the distribution of 257 diverged sites (only one of which was a replacement change) closely followed that of polymorphism (Figure 2C). In the coding region, *D. simulans* was divergent from *D. melanogaster* for 1 replacement and

**TABLE 3**  
Quantitative complementation test

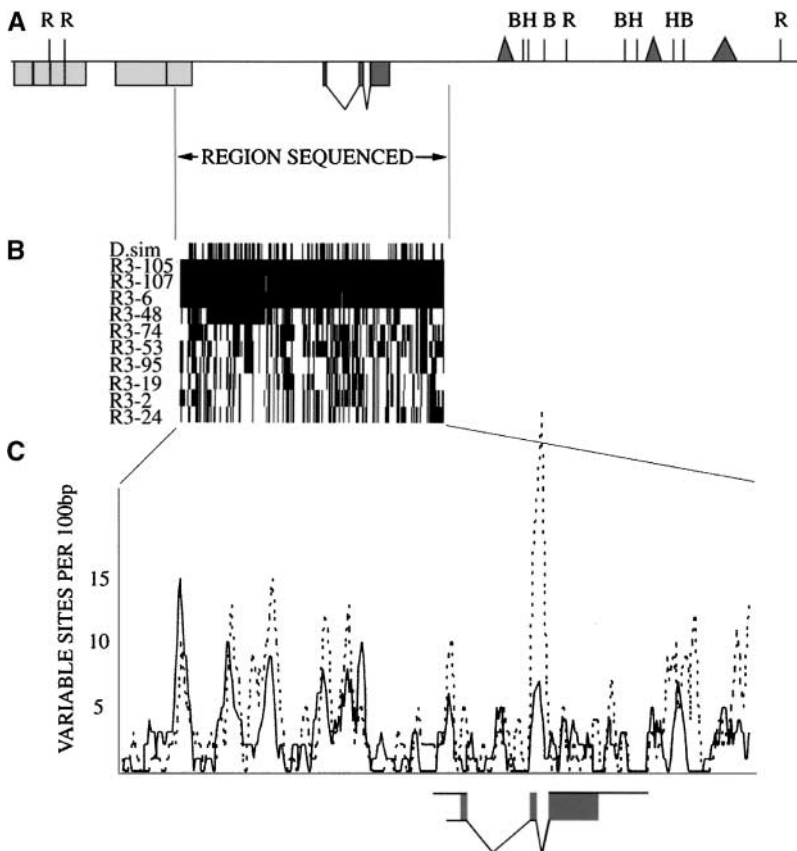
| Source <sup>a</sup> | d.f  | Sternopleural bristles |           | Abdominal bristles |           |
|---------------------|------|------------------------|-----------|--------------------|-----------|
|                     |      | MS                     | F         | MS                 | F         |
| C                   | 1    | 797                    | 112****   | 919                | 75.6****  |
| S                   | 1    | 342                    | 137****   | 9902               | 1235****  |
| C × S               | 1    | 6.69                   | 2.55 (NS) | 25.8               | 5.67*     |
| L                   | 54   | 31.8                   | 2.64***   | 40.4               | 1.73*     |
| C × L               | 54   | 7.08                   | 1.63*     | 12.2               | 1.28 (NS) |
| S × L               | 54   | 2.50                   | 1.08 (NS) | 8.02               | 0.98 (NS) |
| S × C × L           | 54   | 2.62                   | 1.53*     | 4.55               | 1.15 (NS) |
| R(L)                | 55   | 8.24                   | 2.63**    | 20.3               | 1.62*     |
| C × R(L)            | 55   | 3.44                   | 1.25 (NS) | 8.88               | 1.24 (NS) |
| S × R(L)            | 55   | 1.40                   | 0.82 (NS) | 7.60               | 1.93***   |
| C × S × R(L)        | 55   | 1.59                   | 0.93 (NS) | 4.49               | 1.14 (NS) |
| V(C × L × R)        | 220  | 2.74                   | 1.60***   | 7.15               | 1.82****  |
| S × V(C × L × R)    | 220  | 1.71                   | 0.98 (NS) | 3.94               | 1.00 (NS) |
| Error               | 7920 | 1.75                   |           | 3.94               |           |

\*\*\*\* $P < 0.0001$ ; \*\*\* $0.0001 < P < 0.001$ ; \*\* $0.001 < P < 0.01$ ; \* $0.01 < P < 0.05$ ; NS,  $P > 0.05$ .

<sup>a</sup> See text for explanation.

13 synonymous substitutions, and *D. melanogaster* was polymorphic at 1 replacement and 16 synonymous sites, consistent with neutrality [Yates-corrected  $G = 0.37$ ,  $P = 0.54$  (McDONALD and KREITMAN 1991)]. The test of HUDSON *et al.* (1987), which compares polymorphism and divergence between two loci, was applied to the 5'

and 3' halves of the *h* sequence and failed to reject the neutral model ( $\chi^2 = 0.59$ ,  $P = 0.44$ ). Values of  $\theta$  and  $\pi$  are expected to be equal under neutrality. TAJIMA's (1989) *D* statistic, which tests the significance of the difference in the estimates of these quantities, was  $-0.027$ , not significantly different from zero ( $P > 0.1$ ).



**FIGURE 2.**—Molecular variation in the *hairy* gene region. (A) Restriction map variation in a 29-kb region encompassing the *h* transcription unit (black boxes below the line) and the *cis* elements controlling stripe expression (gray boxes below the line; RIDDIHOUGH and ISH-HOROWICZ 1991). R, *EcoRI*; H, *HindIII* (H); B, *BamHI*. Deletion polymorphisms are shown as triangles. (B) DNA sequences for 10.08 kb including *h* for 10 *D. melanogaster* alleles from the Raleigh population (R3-X, where X designates the allele number) and one *D. simulans* (D. sim) allele. Sites that occur more than once in the *D. melanogaster* alleles are shown in the form of a condensed alignment where sites that are in the same state as R3-105 are black. (C) The distribution of nucleotide polymorphism within *D. melanogaster* (solid line) and fixed differences between *D. melanogaster* and *D. simulans* (dotted line) across the region sequenced. Polymorphism and divergence were computed at the midpoint of adjacent 100-bp sliding windows.

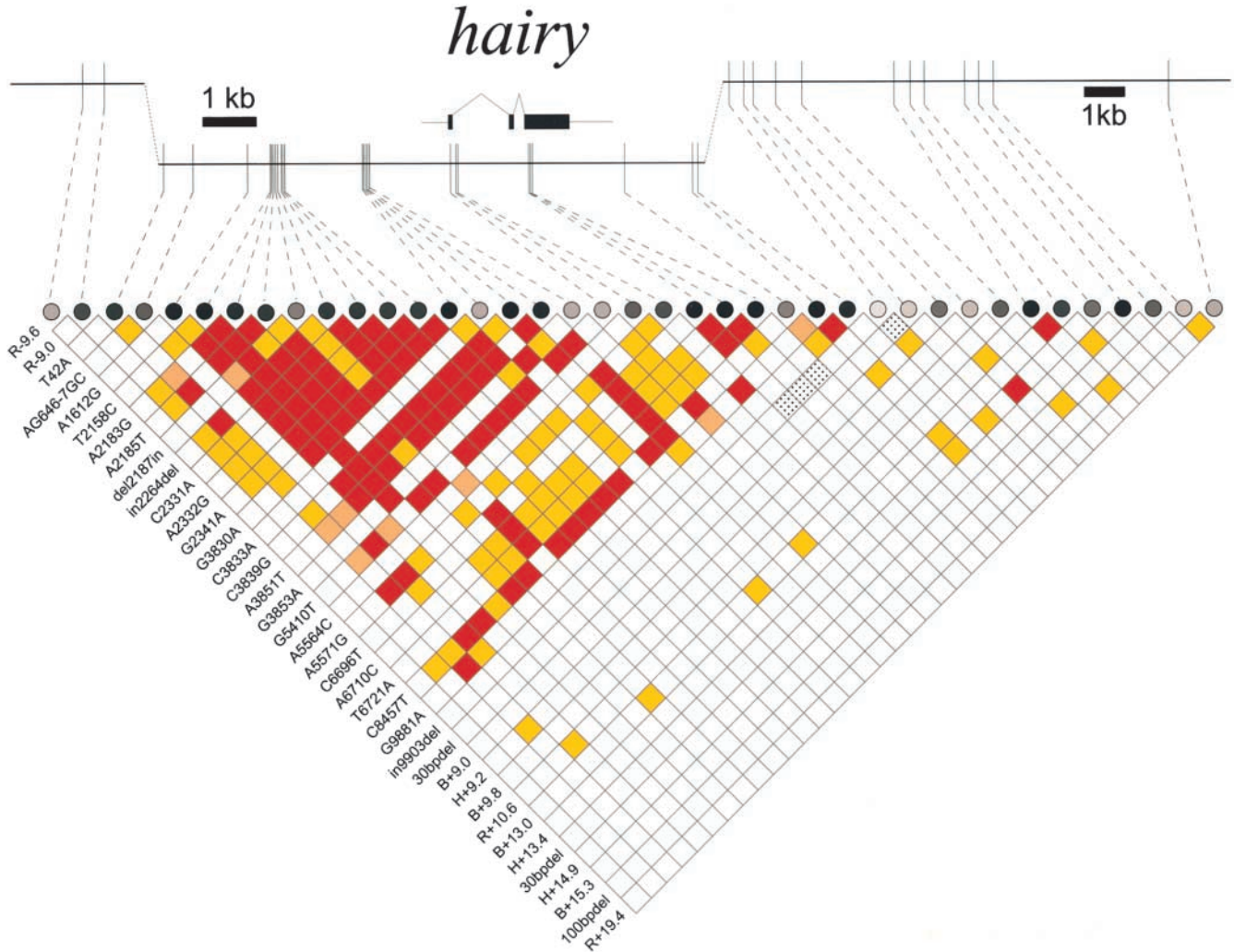


FIGURE 3.—Linkage disequilibrium in and around the *hairy* gene (intron/exon structure shown by lines/solid bars) is illustrated as a matrix of pairwise comparisons where the locations of polymorphic markers are represented graphically across the top and numerically along the bottom left edge. Sites scored by restriction fragment length polymorphisms are shown on the top line and are numbered with reference to transcript origin and cutter type (R = *EcoRI*, H = *HindIII*, B = *BamHI*). Sites detected by sequencing are shown below the gene structure on the bottom line and are labeled such that the most common state is followed by the position in the alignment and then the rarer state. Position 1 of the alignment corresponds to position 8,591,371 of the 3L sequence of Berkeley Drosophila Genome Project genome sequence release 2.0 (<http://www.fruitfly.org/>). Note the change in physical scale between the top and bottom lines, indicated by the solid bars. Linkage disequilibrium was assessed using Fisher's exact test. *P* values (uncorrected for multiple tests) are denoted as red ( $P < 0.005$ ), orange ( $0.005 < P < 0.01$ ), yellow ( $0.01 < P < 0.05$ ), white ( $P > 0.05$ ). Stippled diamonds indicate missing data. The shading of the circles at the top of each column increases with the expected heterozygosity of the polymorphic site.

The estimate of *R* (HUDSON 1987) from the nucleotide sequence data was 65.5, or 0.0066 between adjacent base pairs in the 10.08 kb of aligned sequence; *i.e.*, linkage disequilibrium (LD) is expected between sites separated by 154 bp. We genotyped 25 SNPs and indels spanning the 10.08 kb of aligned sequence, including the amino acid polymorphism and two other polymorphic sites that differentiated the common haplotype from the other sequences in the 47 remaining lines. In addition, a restriction map survey of all 57 lines using three 6-cutter restriction enzymes uncovered two restriction site polymorphisms in 4.2 kb 5' of the sequenced region and 12 polymorphisms 14 kb 3' of the sequenced

region (Figure 2A). There was considerable LD between the polymorphic sites in the sequenced region, consistent with an average spacing of 403 bp, but not between the more widely spaced restriction map polymorphisms (Figure 3).

**Association mapping:** ANOVA was used to assess marker associations with bristle number across all four genetic backgrounds. Six sites were associated with variation in abdominal bristle number at nominal ( $0.01 < P < 0.05$ ) significance levels (not shown). Fourteen sites were associated with variation in sternopleural bristle number: 10 at  $0.01 < P < 0.05$ , 3 at  $0.01 < P < 0.001$ , and 1 at  $P < 0.0001$  (Figure 4a). However, markers

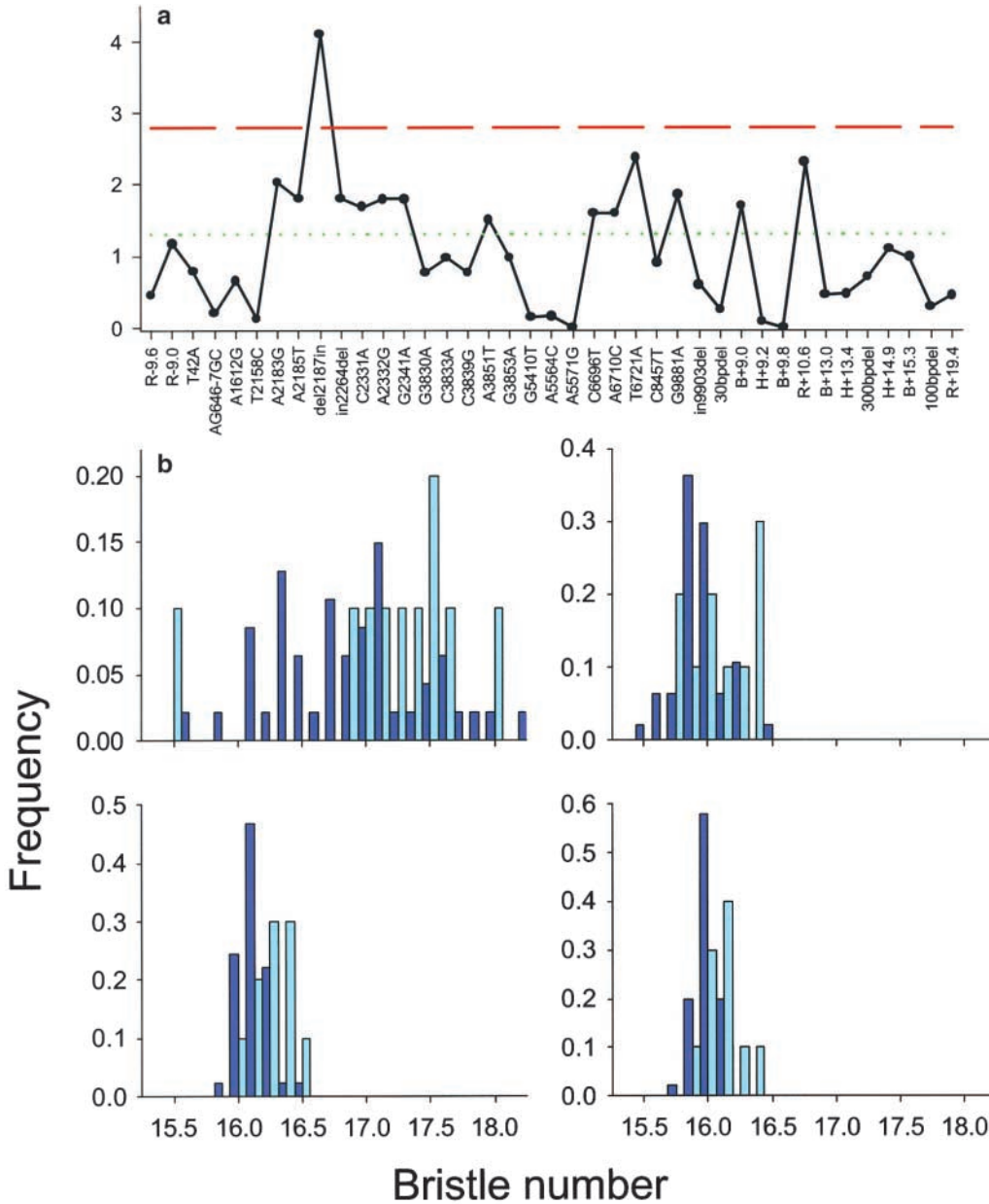


FIGURE 4.—Association of sternopleural bristle number with molecular variation in the *hairy* gene region. (a) *P* values [transformed to  $\log(1/P)$ , y-axis] from ANOVA tests of association of bristle number for each of the polymorphic molecular markers in the *h* gene region (*x*-axis). Marker designations are as in Figure 3. The bottom dotted line is the nominal  $P < 0.05$  significance threshold, and the top dashed line indicates the experiment-wise  $P < 0.05$  threshold given by the Bonferroni correction for multiple tests. (b) Histograms showing the distributions of bristle number between del2187 (dark blue) and in2187 (light blue) haplotypes. Top left, chromosome 3 substitution lines; top right, homozygous *h* region NIL; bottom left, NIL/*h*<sup>1</sup>; bottom right, NIL/*Sam*.

A2185T, in2264del, A2332G, and G2341A had the same haplotype for all 57 lines, as did SNPs G3830A and C3839G, G5410T and A5564C, C6696T and A6710C, and B + 13.0 and H + 13.4. Therefore, there were only 32 different (but still partly correlated) tests for association. With a conservative Bonferroni correction for multiple tests, only sites with a *P* value of 0.00156 or less can be considered to be significant. One site associated with variation in sternopleural bristle number, del2187in, met this criterion ( $F_{1,432} = 15.84$ ,  $P = 0.000081$ ; Table 4). No further sites became significant when this site was fixed and the ANOVAs were repeated on residuals. Thus, consistent with the quantitative complementation test results, molecular variation at *h* is associated with sternopleural, but not abdominal, bristle number.

2187in is a complicated insertion/SNP present in 17.5% of the alleles, where the common sequence ATAAAAAAAA has been replaced by TATACATAGTATAGTATATATAGT. Comparison with *D. simulans* shows del2187 is the ancestral state. The presence of 2187in is associated with an increase of 0.64 sternopleural bristles across all genetic backgrounds, with no significant interactions with genetic background or sex (Table 4). Differences of 0.54, 0.51, and 0.42 bristle between the in2187 and del2187 alleles were significant in the homozygous NIL, NIL/*h*<sup>1</sup>, and NIL/*Sam* genotypes, respectively; but a difference of 1.14 bristle was not significant in the chromosome 3 substitution lines (Table 5, Figure 4b). The fraction of the among-line genetic variance ( $V_L$ ) associated with del2187in is given by the ratio of the variance component attributable to this marker to the

**TABLE 4**  
**Analysis of variance of sternopleural bristle number**  
**line means for del2187in**

| Source <sup>a</sup> | d.f. | MS   | F         |
|---------------------|------|------|-----------|
| G                   | 3    | 243  | 137****   |
| M                   | 1    | 27.9 | 15.8****  |
| S                   | 1    | 6.73 | 3.82 (NS) |
| G × M               | 3    | 1.77 | 1.00 (NS) |
| G × S               | 3    | 0.17 | 0.10 (NS) |
| M × S               | 1    | 0.02 | 0.01 (NS) |
| G × M × S           | 3    | 0.04 | 0.02 (NS) |
| Error               | 432  | 1.76 |           |

\*\*\*\*  $P < 0.0001$ ; NS,  $P > 0.05$ .

<sup>a</sup> See text for explanation.

total  $V_L$  for each genotype and ranged from 12% in the chromosome 3 substitution lines to 73% in the NIL/*Sam* heterozygotes (Table 5).

We were able to detect a significant effect of ~0.5 bristle in a sample of only 57 alleles because 40 (whole third chromosomes) to 80 (NIL and crosses) individuals were scored per line, enabling precise estimation of genotypic values of each line. Further, the NILs varied for only 7% of the total genome. It is illustrative to calculate the potential contribution of del2187in to naturally occurring variation in sternopleural bristle number in a random breeding population. Estimates of phenotypic ( $V_P$ ) and additive genetic ( $V_A$ ) variances of sternopleural bristle number from wild-caught flies reared under standard laboratory conditions are 4 and 1.4, respectively (MACKAY 1980; GURGANUS *et al.* 1999). Assuming strict additivity,  $V_A$  attributable to a locus with a genotypic effect ( $a$ , one-half the difference between the means of the homozygous genotypes) of 0.25 and allele frequencies  $p = 0.825$  (del2187) and  $q = 0.175$  (in2187) is  $V_A = 2pqa^2 = 0.018$  (FALCONER and MACKAY 1996). Thus, this locus would account for only 1.3% of

the additive genetic variance and 0.45% of the phenotypic variance.

However, an additive model may not be appropriate, as in2187 appears to be dominant to del2187. Mean sternopleural bristle numbers of del2187 homozygotes were 17.5 in the NIL homozygotes and 17.7 in the NIL/*Sam* heterozygotes, whereas the mean bristle numbers of the in2187 homozygotes (NIL) and of in2187/del2187 heterozygotes (NIL/*Sam*) were both 18.1. Further, the effect of this marker was similar in all genetic backgrounds (Tables 4 and 5), which is not expected under strict additivity. Assuming complete dominance of in2187,  $V_A = 8p^3qa^2 = 0.049$  (FALCONER and MACKAY 1996), or 3.5% of the additive genetic and 1.2% of the phenotypic variance.

If one were to repeat this study using flies sampled directly from nature, what would be the minimum sample size necessary to detect a significant effect of del2187in? Statistical theory gives the minimum number ( $n$ ) of individuals necessary to detect a difference,  $\delta$ , between two marker class means as  $n \geq 2(z_\alpha + z_\beta)^2 / (\delta/\sigma_P)^2$  (SOKAL and ROHLF 1981), where  $\alpha$  and  $\beta$  are the type I and type II error rates,  $z$  is the ordinate of the normal distribution corresponding to the subscript, and  $\sigma_P$  is the within-marker-class phenotypic standard deviation. Evaluating this expression with  $\alpha = 0.05$  and  $\beta = 0.1$  gives  $n \geq 1344$  for  $\delta = 0.25$  and  $n \geq 336$  for  $\delta = 0.5$ . The former case is appropriate for strict additivity, where the difference of interest is between homozygous del2187 and heterozygous del2187/in2187 genotypes at Hardy-Weinberg expected frequencies of 0.68 and 0.29, respectively. To obtain 1344 heterozygotes, one would need to sample >4600 individuals. The latter case is appropriate for dominance of in2187. With a combined frequency of in2187 homozygotes and del2187/in2187 heterozygotes of 0.32, again assuming Hardy-Weinberg equilibrium, a total of 1050 would ensure adequate representation of the less frequent genotypes.

**TABLE 5**  
**Effect of del2187in in four genetic backgrounds**

| Genotype <sup>a</sup> | $a^b$ | $F_{1,110}^c$ | $\sigma_M^2^d$ | $\sigma_E^2^d$ | % $V_P^e$ | % $V_G^f$ |
|-----------------------|-------|---------------|----------------|----------------|-----------|-----------|
| C3                    | 1.14  | 3.68 (NS)     | 0.64           | 5.79           | 10        | 12        |
| NIL                   | 0.54  | 7.48**        | 0.15           | 0.65           | 19        | 31        |
| NIL/ $h^1$            | 0.51  | 14.46***      | 0.13           | 0.29           | 31        | 66        |
| NIL/ <i>Sam</i>       | 0.42  | 14.32***      | 0.09           | 0.20           | 31        | 73        |

\*\*\*  $0.0001 < P < 0.001$ ; \*\*  $0.001 < P < 0.01$ ; NS,  $P > 0.05$ .

<sup>a</sup> C3, chromosome 3 substitution lines; NIL, homozygous  $h$  near-isoallelic lines; NIL/ $h^1$ , NIL heterozygotes against the  $h^1$  allele; NIL/*Sam*, NIL heterozygotes against the *Sam* wild-type  $h$  allele.

<sup>b</sup> Difference in mean sternopleural bristle number between in2187 and del2187 marker genotypes.

<sup>c</sup>  $F$ -ratio test of significance of difference in mean bristle number between in2187 and del2187.

<sup>d</sup> Between ( $\sigma_M^2$ ) and within ( $\sigma_E^2$ ) marker genotype variance components.

<sup>e</sup> Percentage of phenotypic variance attributable to in2187del.

<sup>f</sup> Percentage of genetic variance attributable to in2187del.

These numbers are underestimates. The phenotypic variance of wild flies will be greater than that of lab-reared flies as a consequence of greater environmental variance in nature (COYNE and BEECHAM 1987). Further, allelic effects of the markers are likely to vary, depending on the environmental conditions (GURGANUS *et al.* 1998; GEIGER-THORNSBERRY and MACKAY 2002), such that the average allelic effect across environments may be reduced. Such considerations highlight the challenge of testing marker-phenotype associations in nature, even for “simple” complex traits, such as *Drosophila* sensory bristle number, and are relevant to association study designs for mapping genes affecting complex human diseases (LANDER 1996; COLLINS *et al.* 1997).

The rapid decay of LD in *Drosophila* regions of high recombination and polymorphism is a highly favorable situation for association mapping of QTL to the level of genes. There is no significant LD between del2187in and markers 5' of AG646-7GC and 3' of C8457T, which localizes the sternopleural bristle number QTL to this 7.8-kb region, in which *h* is the only gene. However, del2187in is not necessarily the QTN causing the variation in bristle number, since it is possible that a site that was not genotyped in this sample, in strong LD with del2187in, is the causal polymorphism. Del2187in is in strong LD with markers spanning 1.7 kb (from T2158C to A3851T) 5' to the beginning of the *h* coding sequence. Of the 94 variable sites in this interval detected by sequencing 10 alleles, only 11 were genotyped in the sample of 57 alleles. Among the 10 sequenced alleles, 3 of the 94 sites, SNPs at positions C2298A, C2414T, and C3118T of the aligned sequence, respectively, had the identical haplotype to del2187in. Outside this 1.7-kb region, only one other polymorphism, another complicated SNP/indel beginning at position 8962, 1.4 kb 3' of the end of the transcription unit, had the same haplotype as del2187in among the sequenced alleles.

These data illustrate how critical it is to utilize the correct density of markers, relative to historical recombination, in association study designs. If the genotype of del2187in had not been determined in this sample, none of the associations would have reached the stringent level of statistical significance required to account for multiple tests. It follows that additional QTN affecting bristle number might have been revealed had the marker density been greater. Resolving which polymorphic site(s) causes variation in phenotypes will ultimately require genotyping all variable sites on large samples of alleles, to eliminate the possibility of hidden causal QTN and to detect informative recombinants. In *Drosophila* regions of high recombination and polymorphism, this requirement currently restricts the utility of linkage disequilibrium mapping in outbred populations to mapping QTN within candidate genes. While *h* was a clear candidate gene affecting bristle number, many QTL map to regions containing no obvious candidate

genes. With the ultimate availability of stocks containing targeted disruptions of all known and predicted genes in *Drosophila* (SPRADLING *et al.* 1999) and other model organisms, quantitative complementation of QTL alleles to mutations of all genes in the region to which QTL map provides a reliable, rapid, and cost-effective method for nominating candidate genes for further study (LYMAN and MACKAY 1998; LONG *et al.* 1998; LYMAN *et al.* 1999).

This is one of a growing number of examples indicating that variation in noncoding regions is likely to be responsible for quantitative genetic variation (MACKAY 2001), which in turn can motivate functional studies to define regulatory motifs (in this case, regulatory sites for expression of *h* in the PNS). Intermediate frequency polymorphisms associated with quantitative traits are not likely to be maintained by mutation-selection balance (LONG *et al.* 1998, 2000), further motivating large-scale future studies with new designs to detect hallmarks of positive or balancing selection at this locus.

This work was supported by National Institutes of Health grant GM 45146 to T.F.C.M. This is a publication of the W. M. Keck Center for Behavioral Biology.

#### LITERATURE CITED

- AGUADÉ, M., N. MIYASHITA and C. H. LANGLEY, 1989 Reduced variation in the *yellow-achaete-scute* region in natural populations of *Drosophila melanogaster*. *Genetics* **122**: 607–615.
- CAMPOS-ORTEGA, J. A., 1993 Early neurogenesis in *Drosophila melanogaster*, pp. 1091–1129 in *The Development of Drosophila melanogaster*, Vol. 2, edited by M. BATE and A. MARTINEZ ARIAS. Cold Spring Harbor Laboratory Press, Plainview, NY.
- CARROLL, S. B., and J. S. WHYTE, 1989 The role of the *hairy* gene during *Drosophila* morphogenesis—stripes in imaginal discs. *Genes Dev.* **3**: 905–916.
- COLLINS, F. S., M. S. GUYER and A. CHAKRAVARTI, 1997 Variations on a theme: cataloging human DNA sequence variation. *Science* **278**: 1580–1581.
- CORMIER, R. T., K. H. HONG, R. B. HALBERG, T. L. HAWKINS, P. RICHARDSON *et al.*, 1997 Secretory phospholipase Pla2g2a confers resistance to intestinal tumorigenesis. *Nat. Genet.* **17**: 88–91.
- COYNE, J. A., and E. BEECHAM, 1987 Heritability of two morphological characters within and among natural populations of *Drosophila melanogaster*. *Genetics* **117**: 727–787.
- CROW, J. F., and M. KIMURA, 1970 *An Introduction to Population Genetics Theory*. Harper & Row, New York.
- DOEBLEY, J., A. STEC and L. HIBBARD, 1997 The evolution of apical dominance in maize. *Nature* **386**: 485–488.
- EL-ASSAL, S. E. D., C. ALONSO-BLANCO, A. J. M. PEETERS, V. RAZ and M. KOORNNEEF, 2001 A QTL for flowering time in *Arabidopsis* reveals a novel allele of *CRY2*. *Nat. Genet.* **29**: 435–440.
- EWING, B., L. HILLIER, M. C. WENDL and P. GREEN, 1998 Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. *Genome Res.* **8**: 175–185.
- FALCONER, D. S., and T. F. C. MACKAY, 1996 *Introduction to Quantitative Genetics*, Ed. 4. Addison Wesley Longman, Harlow, Essex, UK.
- FRARY, A., T. C. NESBITT, A. FRARY, S. GRANDILLO, E. VAN DER KNAAP *et al.*, 2000 *fw2.2*: a quantitative trait locus key to the evolution of tomato fruit size. *Science* **289**: 85–88.
- FRIDMAN, E., T. PLEBAN and D. ZAMIR, 2000 A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene. *Proc. Natl. Acad. Sci. USA* **97**: 4718–4723.
- GEIGER-THORNSBERRY, G. L., and T. F. C. MACKAY, 2002 Association of single nucleotide polymorphisms at the *Delta* locus with geno-

- type-environment interaction for sensory bristle number in *Drosophila melanogaster*. Genet. Res. (in press).
- GURGANUS, M. C., J. D. FRY, S. V. NUZHIDIN, E. G. PASYUKOVA, R. F. LYMAN *et al.*, 1998 Genotype-environment interaction for quantitative trait loci affecting sensory bristle number in *Drosophila melanogaster*. Genetics **149**: 1883–1898.
- GURGANUS, M. C., S. V. NUZHIDIN, J. W. LEIPS and T. F. C. MACKAY, 1999 High-resolution mapping of quantitative trait loci for sternopleural bristle number in *Drosophila melanogaster*. Genetics **152**: 1585–1604.
- HORIGAWA, Y., N. ODA, N. J. COX, X. LI, M. ORHO-MELANDER *et al.*, 2000 Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. Nat. Genet. **26**: 163–175.
- HUDSON, R. R., 1987 Estimating the recombination parameter of a finite population model without selection. Genet. Res. **50**: 245–250.
- HUDSON, R. R., M. KREITMAN and M. AGUADÉ, 1987 A test of neutral molecular evolution based on nucleotide data. Genetics **116**: 153–159.
- HUDSON, R. R., K. BAILEY, D. SKARECKY, J. KWIATOWSKI and F. J. AYALA, 1994 Evidence for positive selection in the *Superoxide dismutase (SOD)* region of *Drosophila melanogaster*. Genetics **136**: 1329–1340.
- INGHAM, P. W., K. R. HOWARD and D. ISH-HOROWICZ, 1985a Transcription of the *Drosophila* segmentation gene *hairy*. Nature **318**: 439–444.
- INGHAM, P. W., S. M. PINCHIN, K. R. HOWARD and D. ISH-HOROWICZ, 1985b Genetic analysis of the *hairy* locus in *Drosophila melanogaster*. Genetics **111**: 463–486.
- JAN, Y. N., and L. Y. JAN, 1993 The peripheral nervous system, pp. 1207–1244 in *The Development of Drosophila melanogaster*, Vol. 2, edited by M. BATE and A. MARTINEZ ARIAS. Cold Spring Harbor Laboratory Press, Plainview, NY.
- KANIA, A., A. SALZBERG, M. BHAT, D. D'EVELYN, Y. HE *et al.*, 1995 *P*-element mutations affecting embryonic peripheral nervous system development in *Drosophila melanogaster*. Genetics **139**: 1663–1678.
- LANDER, E. S., 1996 The new genomics: global views of biology. Science **274**: 536–539.
- LONG, A. D., and C. H. LANGLEY, 1999 Power of association studies to detect the contribution of candidate genetic loci to complexly inherited phenotypes. Genome Res. **9**: 720–731.
- LONG, A. D., S. L. MULLANEY, L. A. REID, J. D. FRY, C. H. LANGLEY *et al.*, 1995 High resolution mapping of genetic factors affecting abdominal bristle number in *Drosophila melanogaster*. Genetics **139**: 1273–1291.
- LONG, A. D., S. L. MULLANEY, T. F. C. MACKAY and C. H. LANGLEY, 1996 Genetic interactions between naturally occurring alleles at quantitative trait loci and mutant alleles at candidate loci affecting bristle number in *Drosophila melanogaster*. Genetics **114**: 1497–1510.
- LONG, A. D., R. F. LYMAN 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.
- LONG, A. D., R. F. LYMAN, A. H. MORGAN, C. H. LANGLEY and T. F. C. MACKAY, 2000 Both naturally occurring insertions of transposable elements and intermediate frequency polymorphisms at the *achaete-scute* complex are associated with variation in bristle number in *Drosophila melanogaster*. Genetics **154**: 1255–1269.
- LYMAN, R. F., and T. F. C. MACKAY, 1998 Candidate quantitative trait loci and naturally occurring phenotypic variation for bristle number in *Drosophila melanogaster*: the *Delta-Hairless* gene region. Genetics **149**: 983–998.
- LYMAN, R. F., C. LAI and T. F. C. MACKAY, 1999 Linkage disequilibrium mapping of molecular polymorphisms at the *scabrous* locus associated with naturally occurring variation in bristle number in *Drosophila melanogaster*. Genet. Res. **74**: 303–311.
- MACKAY, T. F. C., 1980 Genetic variance, fitness, and homeostasis in varying environments: an experimental check of the theory. Evolution **34**: 1219–1222.
- MACKAY, T. F. C., 1996 The nature of quantitative genetic variation revisited: lessons from *Drosophila* bristles. BioEssays **18**: 113–121.
- MACKAY, T. F. C., 2001 The genetic architecture of quantitative traits. Annu. Rev. Genet. **35**: 303–339.
- MCDONALD, J. H., and M. KREITMAN, 1991 Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature **351**: 652–654.
- MIYASHITA, N., 1990 Molecular and phenotypic variation of the *Zw* locus region in *Drosophila melanogaster*. Genetics **125**: 407–419.
- MIYASHITA, N., and C. H. LANGLEY, 1988 Molecular and phenotypic variation of the *white* locus region in *Drosophila melanogaster*. Genetics **120**: 199–212.
- NEI, M., and F. TAJIMA, 1981 DNA polymorphism detectable by restriction endonucleases. Genetics **97**: 145–163.
- NIELSEN, D. M., and B. S. WEIR, 1999 A classical setting for associations between markers and loci affecting quantitative traits. Genet. Res. **74**: 271–277.
- NUZHIDIN, S. V., C. L. DILDA and T. F. C. MACKAY, 1999 The genetic architecture of selection response: inferences from fine-scale mapping of bristle number quantitative trait loci in *Drosophila melanogaster*. Genetics **153**: 1317–1331.
- RIDDIHUGH, G., and D. ISH-HOROWICZ, 1991 Individual stripe regulatory elements in the *Drosophila hairy* promoter correspond to maternal, gap and pair-rule genes. Genes Dev. **5**: 840–854.
- RONAGHI, M., M. UHLÉN and P. NYRÉN, 1998 Sequencing method based on real-time pyrophosphate. Science **281**: 363–365.
- ROZAS, J., and R. ROZAS, 1999 DnaSP version 3: an integrated program for molecular population genetics and molecular evolutionary analysis. Bioinformatics **15**: 174–175.
- RUSHLOW, C. A., A. HOGAN, S. M. PINCHIN, K. M. HOWE, M. LARDELLI *et al.*, 1989 The *Drosophila hairy* protein acts in both segmentation and bristle patterning and shows homology to N-myc. EMBO J. **8**: 3095–3103.
- SALZBERG, A., S. N. PROKOPENKO, Y. HE, P. TSAI, M. PAL *et al.*, 1997 *P*-element insertion alleles of essential genes on the third chromosome of *Drosophila melanogaster*: mutations affecting embryonic PNS development. Genetics **147**: 1723–1741.
- SHRIMPTON, A. E., and A. ROBERTSON, 1988 The isolation of polygenic factors controlling bristle score in *Drosophila melanogaster*. II. Distribution of third chromosome bristle effects within chromosome sections. Genetics **118**: 445–459.
- SOKAL, R. R., and F. J. ROHLF, 1981 *Biometry*. W. H. Freeman and Company, New York.
- SPRADLING, A. C., D. STERN, A. BEATON, E. J. RHEM, T. LAVERTY *et al.*, 1999 The Berkeley *Drosophila* genome gene disruption project: single *P*-element insertions mutating 25% of vital *Drosophila* genes. Genetics **153**: 135–177.
- TAJIMA, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123**: 585–595.
- WATTERSON, G., 1975 On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. **7**: 256–276.