# Associations Between DNA Sequence Variation and Variation in Expression of the *Adh* Gene in Natural Populations of *Drosophila melanogaster*

**Cathy C. Laurie,[1] Jamie T. Bridgham and Madhusudan Choudhary[2]**

*Department of Zoology, Duke University, Durham, North Carolina 27706*

Manuscript received May 14, 1991
Accepted for publication June 29, 1991

ABSTRACT

A large part of the genetic variation in alcohol dehydrogenase (ADH) activity level in natural populations of *Drosophila melanogaster* is associated with segregation of an amino acid replacement polymorphism at nucleotide 1490, which generates a difference in electrophoretic mobility. Part of the allozymic difference in activity level is due to a catalytic efficiency difference, which is also caused by the amino acid replacement, and part is due to a difference in the concentration of ADH protein. A previous site-directed *in vitro* mutagenesis experiment clearly demonstrated that the amino acid replacement has no effect on the concentration of ADH protein, nor does a strongly associated silent polymorphism at nucleotide 1443. Here we analyze associations between polymorphisms within the *Adh* gene and variation in ADH protein level for a number of chromosomes derived from natural populations. A sequence length polymorphism within the first intron of the distal (adult) transcript, ∇1, is in strong linkage disequilibrium with the amino acid replacement. Among a sample of 46 isochromosomal lines analyzed, all but one of the 14 Fast lines have ∇1 and all but one of the 32 Slow lines lack ∇1. The exceptional Fast line has an unusually low level of ADH protein (typical of Slow lines) and the exceptional Slow line has an unusually high level (typical of Fast lines). These results suggest that the ∇1 polymorphism may be responsible for the average difference in ADH protein between the allozymic classes. A previous experiment localized the effect on ADH protein to a 2.3-kb restriction fragment. DNA sequences of this fragment from several alleles of each allozymic type indicate that no other polymorphisms within this region are as closely associated with the ADH protein level difference as the ∇1 polymorphism.

MANY studies of a variety of enzyme-coding loci in Drosophila show that natural populations harbor extensive genetic variability affecting the quantitative level, tissue distribution and developmental pattern of specific enzymatic activities (see review by LAURIE-AHLBERG 1985). Many of these variants are *cis*-dominant, map very close to or within the structural gene and affect enzyme concentration (*e.g.*, CHOVNICK *et al.* 1980; SHAFFER and BEWLEY 1983; MARONI and LAURIE-AHLBERG 1983; BEWLEY 1981; DICKINSON 1978). These variants may represent "regulatory" polymorphisms, *i.e.*, substitutions that affect enzyme expression without altering the amino acid sequence of the protein. Recent developments in technology, such as rapid DNA sequencing techniques, site-directed *in vitro* mutagenesis and *P* element-mediated germline transformation, provide the tools necessary for determining the precise molecular basis of these naturally occurring variants. Nevertheless, the task is difficult because of the high levels of sequence polymorphism in natural populations and because of the fact that many of these polymorphisms are in linkage disequilibrium (*e.g.*, KREITMAN 1983). This means that any particular allele that has an effect on enzyme expression is likely to differ from any other allele in the population at many sites. Determining which polymorphisms have a phenotypic effect is time-consuming, but feasible (LAURIE-AHLBERG and STAM 1987; CHOUDHARY and LAURIE 1991). Here we report on a continuing effort to investigate the molecular basis of genetic variation affecting the expression of alcohol dehydrogenase in natural populations of *Drosophila melanogaster*.

The alcohol dehydrogenase enzyme (EC 1.1.1.1; ADH) of *D. melanogaster* is encoded by a single gene (*Adh*), which produces two alternative transcripts from two distinct, tandem promoters (Figure 1; BENYAJATI *et al.* 1983). The two transcripts show a developmentally specific pattern of expression; the distal transcript is the most abundant form in adult tissues and the proximal transcript is the most abundant form in larval tissues up until late third instar (BENYAJATI *et al.* 1983; SAVAKIS, ASHBURNER and WILLIS 1986). Deletion mutagenesis and *P*-element transformation have been used to determine the sequences necessary for a normal pattern and level of *Adh* expression. These studies show that transcription is regulated by

[1] To whom reprint requests should be addressed.
[2] Present address: Department Ecology and Evolutionary Biology, Rice University, P.O. Box 1892, Houston, Texas 77251.

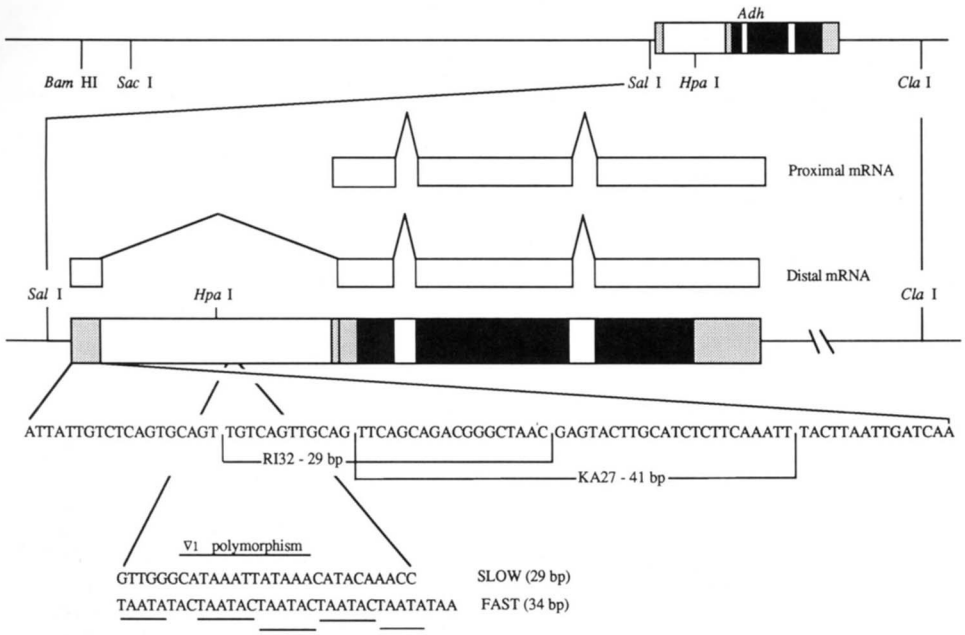


FIGURE 1.—The transcribed portion of the *Adh* gene is represented by the large box (length = 1858 bp). Solid black portions represent coding sequences, hatched regions represent untranslated leader and tailer sequences and open blocks represent each of the three introns. The entire sequence of the first exon of the distal transcript is given with marks to show the extents of the RI32 and KA27 deletions. The ∇1 sequence difference, which begins at nucleotide 448, is also shown.

sequences immediately upstream of each promoter in conjunction with distant larval- and adult-specific enhancer elements (GOLDBERG, POSAKONY and MANIATIS 1983; POSAKONY, FISCHER and MANIATIS 1985; CORBIN and MANIATIS 1989a, 1990).

In natural populations throughout the world, the *Adh* locus is polymorphic for two allozymes, designated Slow and Fast (OAKESHOTT *et al.* 1982). Fast homozygotes generally have a two- to threefold higher ADH activity level than Slow homozygotes. This difference has two causes: Fast homozygotes have a higher concentration of ADH protein than Slow homozygotes and the Fast ADH molecule has a higher catalytic efficiency (see LAURIE-AHLBERG, 1985; LAURIE and STAM 1988; CHOUDHARY and LAURIE 1991). DNA sequencing of several Fast and Slow alleles has shown that the two forms of the protein generally differ by just a single amino acid, a threonine/lysine substitution at residue 192 (KREITMAN 1983). This amino acid is clearly the cause of the catalytic efficiency difference between the allozymes, but the molecular basis of the difference in ADH protein concentration remains an open question. Using site-directed *in vitro* mutagenesis and *P*-element transformation, we have shown that the amino acid replacement has no detectable effect on the concentration of ADH protein, nor does a highly associated silent substitution at nucleotide 1443 (CHOUDHARY and LAURIE 1991). In this report we analyze population survey data in order to develop alternative hypotheses, which will be tested by the same methods.

## MATERIALS AND METHODS

**Fly stocks:** The 11 second chromosome substitution lines described by KREITMAN (1983) were used in two experiments These lines derive from five geographic locations: Florida, France, Africa, Washington and Japan. For experiment I, the original KREITMAN lines were used. For experiment II, each of the KREITMAN second chromosomes was extracted into the isogenic Hochi-R genetic background by the method in LAURIE-AHLBERG *et al.* (1980). Experiments I and II included several additional second chromosome substitution lines that have the Hochi-R genetic background (LAURIE-AHLBERG *et al.* 1980). These second chromosomes derive from populations in Rhode Island, North Carolina, Wisconsin and Kansas. Experiment I also included an analysis of two second chromosomes from a California population, which were provided by J. MCDONALD.

**Polymerase chain reaction:** DNA was extracted from sets of 50 flies to use as template in polymerase chain reactions (PCR). Approximately 2.5 fly equivalents of DNA were used in each reaction. In most cases, reactions were catalyzed by *Taq* DNA polymerase, using conditions recommended by the supplier (Perkin-Elmer Cetus Corp.). In some cases, discussed below, Vent polymerase (New England Biolabs) was used instead of *Taq*. One of the two primers was kinased prior to PCR. The Perkin-Elmer Cetus Thermocycler was set for one minute of denaturation at 94° (3 min on the first round), 1-min annealing at 65° and 2-min extension at 72° for a total of 28 cycles. The PCR product was then digested with λ exonuclease to obtain single-stranded template for sequencing (HIGUCHI and OCHMAN 1989). Template for sequencing each strand was obtained from two separate amplifications. The *Adh* region to be sequenced was amplified in two segments, −145 to 1154 and 984 to 1955, where the numbering is based on KREITMAN's (1983) consensus sequence.

**Sequencing:** Complete *Bgl*II digests of genomic DNA from lines KA27 and NC16 were used to make λEMBL-4 libraries (FRISCHAUF, MURRAY and LERACH 1987). *Adh*-containing clones were identified by plaque hybridization. These clones were provided by C. F. AQUADRO. For each of the two phage clones, a 2.7-kb *Sal*I/*Cla*I fragment (nucleotides −62 to 2665) was subcloned into a plasmid vector for sequencing (pBSM13$^-$; Stratagene, Inc.). The entire 2.7-kb fragment was sequenced for both alleles. Both strands were completely sequenced by the dideoxy chain termination

**TABLE 1**

**Line means from experiment I ($n = 4$)**

| Line | Allozyme | ∇1 | Haplotype | Activity | CRM | RNA |
|---|---|---|---|---|---|---|
| Fl-2s | S | − | CCACTA | 15.1 | 21.9 | 1.08 |
| CA-S1 | S | − | CCACTA | 16.7 | 25.2 | 1.10 |
| WI09 | S | − | CTATCC | 19.5 | 30.8 | 1.10 |
| Wa-s | S | − | CCACTA | 19.6 | 33.7 | 1.11 |
| Fr-s | S | − | CCACTA | 21.1 | 32.8 | 1.37 |
| Ja-s | S | − | CTATCC | 23.0 | 36.9 | 1.17 |
| NC16 | S | + | CCACTA | 33.5 | 51.9 | 1.13 |
| RI32 | F | + | GTCTCC | 34.5 | 29.0 | 0.76 |
| KA12 | F | − | GTCTCC | 34.6 | 27.1 | 1.22 |
| Fl-f | F | + | GTCTCC | 34.7 | 32.0 | 1.13 |
| Wa-f | F | + | GTCTCC | 45.2 | 42.3 | 1.09 |
| Hochi-R | F | + | GTCTCC | 48.4 | 47.0 | 1.12 |
| Fr-f | F | + | GTCTCC | 50.9 | 46.9 | 1.19 |
| Ja-f | F | + | GTCTCC | 51.5 | 48.7 | 1.31 |
| Ca-F1 | F | + | GTCTCC | 52.5 | 49.7 | 1.14 |

Haplotype refers to the nucleotides as positions 1443, 1452, 1490, 1518, 1527 and 1557 [with reference to KREITMAN's (1983) consensus sequence]. Units of activity are nanomoles $NAD^+$ reduced per minute milligram of fly wet weight. Units of CRM are milligrams of Hochi-R standard flies per milligram of fly wet weight. Units of RNA are number of counts in wild-type band per number of counts in standard *fn23* band [see LAURIE and STAM (1988) for further explanation].

**TABLE 2**

**Line means for experiment II ($n = 4$)**

| Line | Allozyme | ∇1 | Haplotype | Activity | CRM |
|---|---|---|---|---|---|
| Fl-2s | S | − | CCACTA | 20.3 | 66.3 |
| KA16 | S | − | CCACTA | 21.6 | 70.9 |
| Wa-s | S | − | CCACTA | 21.9 | 70.4 |
| KA27 | S | − | CTATCC | 22.4 | 78.5 |
| WI09 | S | − | CTATCC | 22.5 | 76.1 |
| RI37 | S | − | CTATCC | 23.1 | 81.9 |
| Af-s | S | − | CCACTA | 24.0 | 80.5 |
| Fr-s | S | − | CCACTA | 26.8 | 92.8 |
| Ja-s | S | − | CTATCC | 30.8 | 110.3 |
| Fl-1s | S | − | CCACTA | 31.5 | 114.7 |
| NC16 | S | + | CCACTA | 40.6 | 124.3 |
| KA12 | F | − | GTCTCC | 41.4 | 72.1 |
| Fl-f | F | − | GTCTCC | 45.8 | 82.4 |
| RI32 | F | + | GTCTCC | 46.7 | 86.1 |
| Wa-f | F | + | GTCTCC | 57.4 | 110.7 |
| Af-f | F | + | GTCTCC | 57.8 | 104.6 |
| WI08 | F | + | GTCTCC | 59.7 | 117.2 |
| Fr-f | F | + | GTCTCC | 60.9 | 118.6 |
| KA13 | F | + | GTCTCC | 62.8 | 121.5 |
| Ja-f | F | + | GTCTCC | 68.5 | 138.2 |

Haplotype refers to nucleotides at positions 1443, 1452, 1490, 1518, 1527 and 1557 [with reference to KREITMAN's (1982) consensus sequence]. Units of activity are nanomoles $NAD^+$ reduced per minute per milligram total protein (multiplied by 0.1). Units of CRM are Hochi-R fly equivalents per milligram total protein (multiplied by 10).

method using oligonucleotide primers located every 200 bases along the sequence. Some features of the KA27 sequence were previously reported (LAURIE *et al.* 1990).

The complete *Adh* gene (−62 to 1858) from lines KA12 and RI32 was sequenced from PCR product templates. We had difficulty sequencing through the run of approximately 10 A residues (beginning at 1698) on template amplified with *Taq* polymerase, evidently because of replication slippage. Vent polymerase was much better and we were able to get clear sequence in this region on one strand for both alleles. Both strands were completely sequenced for both alleles with the following exceptions: a 60-base region surrounding the run of As in KA12, one 32- and one 33-base region surrounding the run of As in RI32, a 34-base region at the 3′ end of KA12, 11 scattered sites in KA12 and 2 individual sites in RI32. In each of these regions the sequence of the other strand was clearly readable and in only three cases did the single stranded sequence show a difference with KREITMAN's (1983) consensus: 1693 and 1698 for KA12 and 1741 for RI32, none of which represents a unique polymorphism.

Parts of the *Adh* gene from several second chromosome substitution lines were also sequenced from PCR products (Tables 1 and 2). In all of these cases, sequence reported is from both strands.

Sequences were aligned with KREITMAN's (1983) consensus sequence using the GAP program of the University of Wisconsin Genetics Computer Group software package (DEVEREAUX, HAEBERLI and SMITHIES 1984). Sequence differences based on those alignments are given in Figure 2. During the course of sequencing, we discovered three errors in the sequences reported by KREITMAN (1983). These errors can be corrected by inserting a T at 1732, deleting a T at 1761 and changing sites 2372 and 2373 from TC to CT. These changes are made in the sequences reported in Figure 2. The only difference it makes is changing the site of a C/G polymorphism from 1740 to 1741.

**Experiment I:** This experiment was a measurement of ADH activity, cross-reacting material (CRM) and RNA levels on several of KREITMAN's lines, two California lines and 5 lines from the original Hochi-R set (Table 1). Methods and some results for the KREITMAN and California lines were previously described by LAURIE and STAM (1988). Here we provide a more complete description of the results of that experiment.

**Experiment II:** In this experiment, the second chromosome lines in Table 2 were assayed for ADH activity and CRM level. Each of these lines had the Hochi-R genetic background except for Wa-s, which was on a $ry^{506}$ genetic background. (The Wa-s background difference probably has little effect because a comparison of ADH activity and CRM levels between lines in which the Wa-f chromosome was on either of the two backgrounds showed no significant difference.) Immediately prior to collecting samples from these lines, each was checked by partial sequencing of PCR products to obtain or verify the sequence information provided in Table 2. Males from each line were crossed to females from an ADH-null stock, $Adh^{fn6}cn$; $ry^{506}$, during each of two time blocks. Within each block, four separate crosses of 5 pairs each were set up in 8-dram vials. From the pooled progeny of those four crosses, two sets of ten 6–8-day-old males were homogenized for the assay of ADH activity and CRM level. This makes a total of four observations per line, which were averaged to give the numbers in Table 2. Activity, CRM and total protein assay procedures are described by CHOUDHARY and LAURIE (1991).

## RESULTS and DISCUSSION

**A population survey reveals associations between ADH expression and certain DNA sequence variants:** AQUADRO *et al.* (1986) previously reported a survey of 49 second chromosome substitution lines for ADH activity variation and restriction fragment

| Reference Sequence | | | W a - S # | F I 1 S # | A f - S # | F r - S # | F I 2 S # | J a - S # | K A 2 7 | N C 1 6 | K A 1 2 | F I - F | F r - F # | W a - F # | A f - F # | J a - F # | R I 3 2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5'Flanking | -3 | C | . | . | . | . | . | . | . | T | . | . | T | T | T | T | T |
| Sequence | -2 | C | . | . | . | . | . | . | . | G | . | . | G | G | G | G | G |
| | -1 | G | . | C | . | . | . | C | . | C | . | C | C | C | C | C | C |
| Adult | 20 | ΔR | . | . | . | . | . | . | . | . | . | . | . | . | . | . | Δ |
| Leader (Exon 1) | 32 | ΔK1 | . | . | . | . | . | . | Δ | . | . | . | . | . | . | . | . |
| | 107 | C | . | . | . | . | A | . | . | A | . | . | A | A | A | A | A |
| | 113 | A | . | . | . | . | G | . | . | G | . | . | G | G | G | G | G |
| | 143 | A | . | . | . | . | . | . | . | . | . | . | . | . | . | G | . |
| | 169 | T | . | . | . | . | . | . | . | . | . | . | . | . | . | G | . |
| | 173 | A | . | . | . | . | . | . | . | . | . | . | . | . | . | G | . |
| | 175 | T | A | . | . | . | A | . | . | A | . | . | A | A | A | A | A |
| Intron 1 | 219 | T | . | . | . | . | . | . | . | . | G | . | . | . | . | . | . |
| (Adult Intron) | 287 | G | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| | 293 | G | . | . | . | . | T | . | . | T | . | . | T | T | T | . | T |
| | 304 | G | . | . | . | . | C | . | . | C | . | . | C | C | C | . | C |
| Hpa I | 327 | ---- | | | | | | | | | | | | | | | |
| | 448 | ∇1 | . | . | . | . | . | . | . | ∇ | . | . | ∇ | ∇ | ∇ | ∇ | ∇ |
| | 486 | T | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . |
| | 488 | T | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . |
| | 491 | ∇K2 | . | . | . | . | . | . | ∇ | . | . | . | . | . | . | . | . |
| | 496 | G | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . |
| | 516 | C | . | . | . | . | . | . | . | . | . | . | G | G | G | . | . |
| | 551 | ∇2 | . | . | . | . | . | . | . | . | . | . | ∇ | ∇ | ∇ | . | . |
| | 586 | G | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . |
| Larval Leader | 713 | C | . | . | . | . | A | . | . | . | . | . | . | . | . | . | . |
| | 769 | G | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . |
| Trans. region | 816 | T | . | . | . | . | G | G | G | . | . | G | G | G | G | G | . |
| of Exon 2 | 896 | A | . | . | . | G | G | . | . | . | . | . | . | . | . | . | . |
| Intron 2 | 925 | C | . | . | . | T | T | . | . | . | T | . | . | . | . | . | . |
| | 979 | C*1 | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . |
| | 1053 | G*2 | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . |
| | 1068 | C | T | T | . | . | . | . | . | . | . | . | . | . | . | . | . |
| Exon 3 | 1082 | C | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . |
| | 1229 | C | T | T | . | . | . | . | T | . | . | . | . | . | . | . | . |
| | 1235 | C | . | . | . | . | . | . | . | . | . | . | . | . | . | A | . |
| | 1244 | C | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . |
| | 1283 | C | A | A | . | . | . | . | . | . | . | . | . | . | . | . | . |
| | 1354 | G | C | C | . | . | . | . | . | . | . | . | . | . | . | . | . |
| | 1362 | G | A | A | . | . | . | . | . | . | . | . | . | . | . | . | . |
| Intron 3 | 1387 | C | . | . | . | . | . | . | . | G | . | . | . | . | . | . | . |
| | 1388 | A | . | . | . | . | . | . | . | . | . | . | . | . | . | G | . |
| | 1399 | G | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . |
| | 1400 | A | T | T | . | . | . | . | . | . | . | . | . | . | . | . | . |
| | 1402 | T | . | . | . | . | . | . | . | G | . | . | . | . | . | . | . |
| | 1405 | T | A | A | . | . | . | . | . | . | . | . | . | . | . | . | . |
| | 1425 | C | A | A | . | . | . | . | . | . | . | . | . | . | . | . | . |
| | 1431 | T | C | C | . | . | . | . | . | . | . | . | . | . | . | . | . |
| | 1443 | C | . | . | . | . | . | . | . | . | G | G | G | G | G | G | G |
| Translated | 1452 | C | . | . | . | . | . | T | T | . | T | T | T | T | T | T | T |
| Region of | 1490 | A*3 | . | . | . | . | . | . | . | . | C | C | C | C | C | C | C |
| Exon 4 | 1518 | C | . | . | . | . | . | T | T | . | T | T | T | T | T | T | T |
| | 1527 | T | . | . | . | . | . | C | C | . | C | C | C | C | C | C | C |
| | 1557 | A | . | . | . | . | . | C | C | . | C | C | C | C | C | C | C |
| | 1596 | G | . | . | A | A | . | A | . | A | . | . | . | . | . | . | . |
| 3'Untranslated | 1693 | A | . | . | . | . | C | C | C | C | C | C | C | C | C | C | C |
| Region | 1698 | ∇3 | . | . | . | -1 | 3 | 4 | 4 | 4 | 4 | 4 | 4 | 5 | 4 | 4 | |
| | 1741 | C | . | . | . | . | . | . | . | G | G | . | G | G | G | . | G |
| | 1908 | A | . | . | . | T | . | . | . | . | | | | | | | |
| | 1920 | G | . | . | . | . | . | . | . | . | | | | | | | |
| | 1921 | C | . | . | . | . | . | . | . | . | | | | | | | |
| | 1925 | G | . | . | . | A | . | . | A | A | | | | | | | |
| | 1937 | C | . | . | T | . | . | . | . | . | | | | | | | |
| | 2066 | C | . | . | . | . | . | . | . | T | | | | | | | |
| | 2081 | ∇4 | . | . | ∇ | . | . | . | . | . | | | | | | | |
| | 2097 | ∇N | . | . | . | . | . | . | . | ∇ | | | | | | | |
| | 2130 | C | . | . | . | . | . | T | . | . | | | | | | | |
| | 2234 | A | . | . | . | . | . | . | C | . | | | | | | | |
| 3'Flanking | 2303 | ∇5 | . | . | 1 | . | . | . | . | -1 | | | | | | -1 | |
| Sequence | 2324 | T | . | . | . | . | . | . | . | A | | | | | | | |
| | 2347 | T | . | . | A | . | . | . | . | . | | | | | | | |
| | 2350 | G | . | . | . | . | . | . | C | . | | | | | | | |
| | 2379 | Δ6 | Δ | Δ | . | . | . | . | . | . | | | | | | | |
| | 2482 | C | . | . | . | . | . | . | . | T | | | | | | | |
| | 2553 | T | . | . | . | . | . | . | . | C | | | | | | | |

ΔR= delete TGTCAGTTGCAGTTCAGCAGACGGGCTAAC
insert A
ΔK1= delete TTCAGCAGACGGGCTAACGAGTACTTGCATCTCTTCAAATT
∇1= delete GTTGGGCATAAATTATAAACATACAAACC
insert TAATATACTAATACTAATACTAATACTAATATAA
∇K2= insert 2A's
∇2= insert AATAAACCATAAACTAGGCTGCTCAGCCGGCGACGGC
∇3= insert or delete A's
∇4= insert GT
∇N= insert GGCGGCTATGGGGTT
∇5= insert or delete T's
Δ6= delete AATCGCCAGATAAGCAATATATTGTGT

∇= place insertions before the base at the number given so that the insertion would begin with that number.

*1= GCC ala becomes GTC val
*2= GCC ala becomes TCC ser
*3= AAG lys becomes ACG thr

#= Sequences from Kreitman (1983)

← S/F allozyme site

FIGURE 2.—Sequence comparisons among 15 *Adh* alleles. The reference sequence is the consensus sequence of KREITMAN (1983). A "·" refers to a match with the consensus at the indicated nucleotide site (numbering begins with the distal transcript start site). All alleles except KA12 and RI32 were sequenced from a *SalI* site at −62 to a *ClaI* site at 2665; all sequence differences within that region are shown in the table. The KA12 and RI32 alleles were sequenced from the *SalI* site at −62 through the 3′ end of the transcript at 1858. Blank cells from 1858 to 2665 for those two alleles represent the lack of sequence information in that region. The table also shows the location of the *HpaI* restriction site at 327.

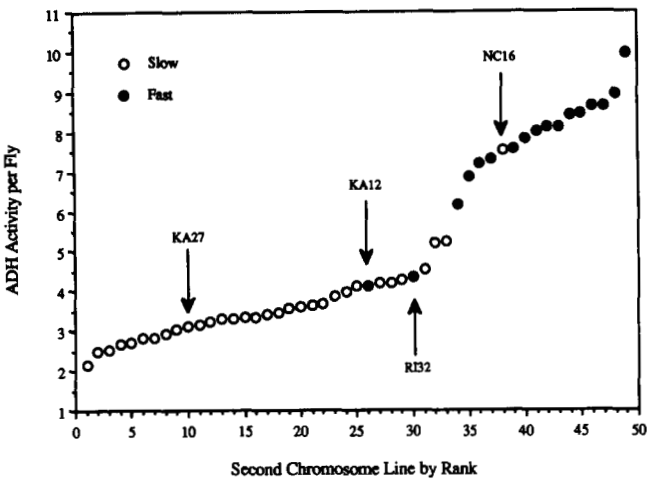


FIGURE 3.—The distribution of ADH activity levels among 49 isogenic second chromosome substitution lines (redrawn from Figure 4 of AQUADRO *et al.* 1986). Each point is the average of 30 observations. Units of ADH activity are nanomoles of $NAD^+$ reduced per minute.

length polymorphism (RFLP) in a 13-kb region containing the *Adh* gene. Figure 3 shows the distribution of the second chromosome effects on ADH activity level. The most striking feature of this plot is that there is clearly a large difference in ADH activity level between lines having the Fast *vs.* Slow allozyme. There is an approximately continuous distribution of activity levels within an allozymic class, but there is a clear break between the two distributions. Three lines have very unusual activity levels for their allozymic type: NC16, RI32 and KA12. This report provides new data concerning the possible molecular bases of the unusual activities in these three lines.

The *Adh* gene of RI32 has been sequenced from 64 bp upstream of the distal start site to the 3′ end of the transcript at +1858 (Figure 2). It has a predicted amino acid sequence identical to each of the Fast alleles previously sequenced by KREITMAN (1983). Only one new polymorphic site is found in this allele, a 29-bp deletion within the first exon of the distal transcriptional unit (Figure 1). This deletion causes a predicted reduction in the size of the 5′ untranslated leader of the distal transcript from 123 to 94 bases. Experiment I shows that the ADH activity, CRM and mRNA levels of RI32 are coordinately reduced to about 60–70% of the levels in a typical Fast line, Hochi-R (Figure 4 and Table 1).

It is tempting to speculate that the RI32 deletion causes the low level of *Adh* expression in this line, but analysis of another line, KA27, suggests that this association may be coincidental. The RFLP study of AQUADRO *et al.* (1986) showed that KA27 also has a small deletion in the same restriction fragment as the deletion in RI32. The sequence of the KA27 *Adh* gene shows that this allele has a different deletion of 41 bp, which is also located in the 5′ untranslated

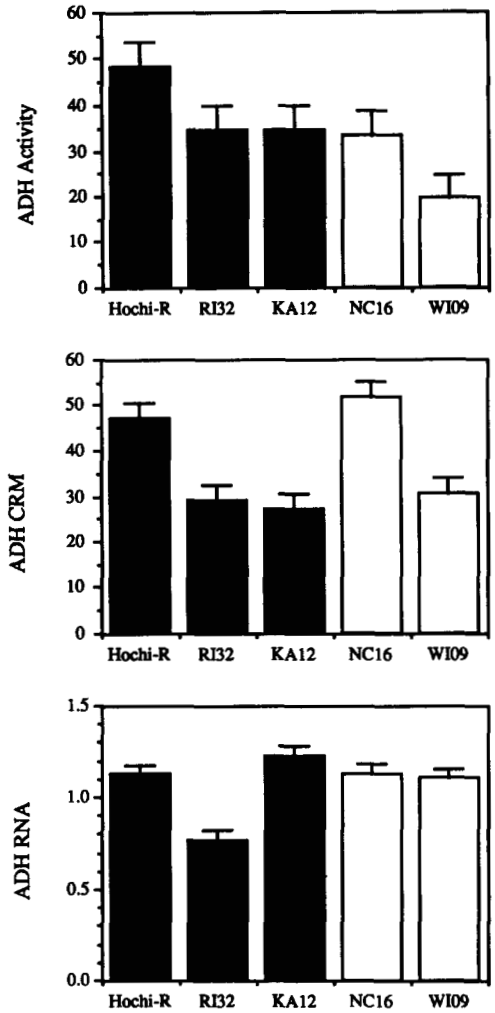


FIGURE 4.—Line means for four isogenic second chromosome substitution lines from experiment I. Each mean is the average of four observations. The error bars represent the least significant difference at the 5% level. Units are defined in Table 1.

leader of the distal transcript (Figure 1; LAURIE *et al.* 1990). Previous analysis of *Adh* expression in the Slow line KA27 showed that activity, CRM and RNA levels are very similar to those of a typical Slow line, Wa-s (THOMSON, JACOBSON and LAURIE 1991; see also Figure 3). A direct analysis of the RI32 deletion by *in vitro* mutagenesis and *P*-element transformation will be needed to determine whether it has any effect on *Adh* expression.

The RFLP study of AQUADRO *et al.* (1986) showed a strong linkage disequilibrium between the allozyme polymorphism and a *Bam*HI polymorphism located about 7 kb upstream (Figure 1). Nearly all Slow chromosomes have the *Bam*HI site and nearly all Fast chromosomes lack it. Only 4 of the 49 chromosomes analyzed has one of the two rare haplotypes. Two of those rare haplotype lines do not have unusual levels of ADH, but the other two are the Slow line NC16 and the Fast line KA12, which appear to have switched activity class with respect to their allozymic type (Figure 3). These results suggested that the difference in

ADH activity between the allozymic classes might not be due entirely to the amino acid replacement, but rather might be due in part to association between the amino acid polymorphism and a regulatory site polymorphism located 5′ of the amino acid substitution. Under this hypothesis, the two rare haplotypes that appear to have switched activity class are interpreted as recombinants between the amino acid site and the putative regulatory site, while the inferred recombination site for the other two rare haplotypes occurs further upstream, between the putative regulatory site and the *Bam*HI site.

**A molecular mapping experiment localized the allozymic differences in ADH activity and CRM to a 2.3-kb *Hpa*I/*Cla*I restriction fragment:** The possibility of association between the allozymic polymorphism and a regulatory site polymorphism in the 5′ flanking region of the *Adh* gene was investigated by a *P*-element transformation experiment (LAURIE-AHLBERG and STAM 1987). In that experiment, an ADH-negative strain was transformed with chimeric *Adh* fragments in which the 5′ flanking regions were exchanged between a pair of Slow and Fast alleles that show the typical difference in ADH activity and CRM levels (Wa-f and Wa-s; LAURIE and STAM 1988). Both the activity and CRM differences clearly map to a 2.3-kb *Hpa*I/*Cla*I restriction fragment that includes all of the *Adh* coding sequence and some intronic and 3′ flanking sequence, but excludes all of the 5′ flanking sequence of the distal transcriptional unit (Figure 1; LAURIE-AHLBERG and STAM 1987). Although this result rejects the hypothesis of a regulatory site polymorphism within the 5′ flanking region, it does not eliminate the possibility that some site (other than the amino acid replacement) within or 3′ of the *Adh* transcriptional unit affects the concentration of ADH protein.

When AQUADRO *et al.* (1986) and LAURIE-AHLBERG and STAM (1987) first formulated the hypothesis of nonrandom association between the amino acid replacement polymorphism and a 5′ flanking region regulatory polymorphism, there was a report in the literature that the higher CRM level of Fast lines appears to be explained by a higher level of ADH-mRNA (ANDERSON and MCDONALD 1983). When the transformation experiment ruled out any major effect of sites in the 5′ flanking region, LAURIE and STAM (1988) reconsidered the possibility of an allozymic difference in ADH-mRNA. Their study of several alleles of each type showed that there is no consistent difference in RNA level between the allozymic classes. They concluded that the allozymic difference in level of ADH protein is due either to a difference in translation rates of the two mRNAs or to a difference in protein stability.

**Sequence comparisons suggested that the amino acid replacement at 1490 and/or the silent substitution at 1443 affect ADH CRM level, but *in vitro* mutagenesis experiments show they do not:** The 2.3-kb *Hpa*I/*Cla*I fragment has been sequenced by KREITMAN (1983) for 11 different *Adh* alleles from 5 geographic locations. These alleles were cloned from stocks that were made isogenic for the second chromosome (which contains the *Adh* gene) and we have analyzed *Adh* expression in these stocks. LAURIE and STAM (1988) measured ADH activity, CRM and RNA levels for four pairs of Slow and Fast KREITMAN stocks that derive from different geographic locations (Fl, Fr, Ja, Wa). The Fast member of each pair had a higher ADH activity and CRM level than the Slow member of the pair. A comparison of the sequences of the four pairs of alleles for the *Hpa*I/*Cla*I fragment shows that only three nucleotide substitutions distinguish all four of the Fast alleles from all four of the Slow alleles (KREITMAN 1983; Figure 2). Therefore, we previously concluded that one or more of these three substitutions is likely to account for the allozymic difference in activity and CRM levels (LAURIE-AHLBERG and STAM 1987). One of those substitutions is, of course, the amino acid replacement at 1490 and the others are nearby third position silent substitutions, one at 1443 and one at 1527. Subsequently, the silent substitution at 1527 was eliminated as a likely candidate when the Slow allele KA27, which has the typical Slow-like low expression, was sequenced and found to have the Fast-typical C at this site (Figure 2). In addition, two other Slow alleles with typical Slow-like low expression have been partially sequenced and they also show the Fast-typical C at 1527 (WI09 and RI37; Table 2).

At this point in the project it appeared likely that the amino acid replacement at 1490 could explain not only the charge difference and the catalytic efficiency difference between the allozymic classes, but that it might also account for the CRM level difference through an effect on protein stability. This hypothesis was tested directly by *in vitro* mutagenesis and the result was very clear: there is no detectable effect of the amino acid replacement on the concentration of ADH protein (CHOUDHARY and LAURIE 1991). That result left a difference in translation rates as the only likely candidate for the mechanism that brings about the difference in protein concentration. It also left the 1443 silent substitution as the most likely cause of that difference, based on the sequence comparisons discussed above. However, when the 1443 substitution was tested directly by *in vitro* mutagenesis, it was found to have no detectable effect on ADH CRM level either (CHOUDHARY and LAURIE 1991).

**Additional sequence data show that ∇1 within the adult intron is strongly associated with the amino acid replacement polymorphism and with ADH**

**CRM level:** The data on lines KA12 and NC16 discussed above first suggested that there might be a sequence difference separate from the amino acid replacement that contributes to the overall difference in ADH activity between the allozymic classes. These two lines were studied in more detail in an attempt to find clues about the nature and location of the second site. Figures 3 and 4 show that these two lines have ADH activity levels intermediate between a typical Slow line, WI09, and a typical Fast line, Hochi-R. Figure 4 also shows that the intermediate activity level of the Fast line KA12 is caused by having an unusually low level of ADH CRM, whereas the intermediate activity level of the Slow line NC16 is caused by having an unusually high level of ADH CRM. Figure 4 also shows that Hochi-R, WI09, KA12 and NC16 have very similar levels of ADH RNA, which is consistent with the conclusion of LAURIE and STAM (1988) that there is no difference in RNA level between the allozymic classes.

A comparison between the *Adh* sequences of KA12, NC16 and the KREITMAN alleles is shown in Figure 2. KA12 was sequenced from −64 through the 3′ end of the transcriptional unit at +1858. The only substitution that is unique to KA12 is a G at 219 within the adult intron. However, KA12 shares one unusual sequence feature with KREITMAN's Fl-f allele: they both lack the ∇1 sequence within the adult intron, whereas all of the other Fast alleles have ∇1. NC16 was sequenced from −64 through the *Adh* gene and its 3′ flanking sequence to a *Cla*I site about 0.8 kb downstream. Several substitutions and one insertion are unique to the NC16 allele. The unique substitutions include two amino acid replacements, one at nucleotide +979 which causes an Ala to Val substitution at amino acid residue 45 and the other at nucleotide +1053 which causes an Ala to Ser substitution at amino acid residue 70. In addition to those unique substitutions, NC16 is also unusual because it is the only sequenced Slow allele that has ∇1. Thus, the presence of ∇1 in NC16 is associated with an increased level of ADH CRM relative to other Slow lines and the absence of ∇1 in KA12 is associated with a decreased level of ADH CRM relative to other Fast lines.

To assess the strength of the associations among ∇1, the amino acid replacement polymorphism and the level of ADH activity, the set of chromosome lines in Figure 3 were each scored for the presence or absence of ∇1. This was done by PCR amplification of *Adh* DNA from the fly stocks, followed by sequencing through the relevant region of the adult intron. Three of the original lines had been lost, but among the remaining 46 lines, all but one of the 14 Fast lines have ∇1 and all but one of the 32 Slow lines lack ∇1. The two exceptional lines are KA12 and NC16,

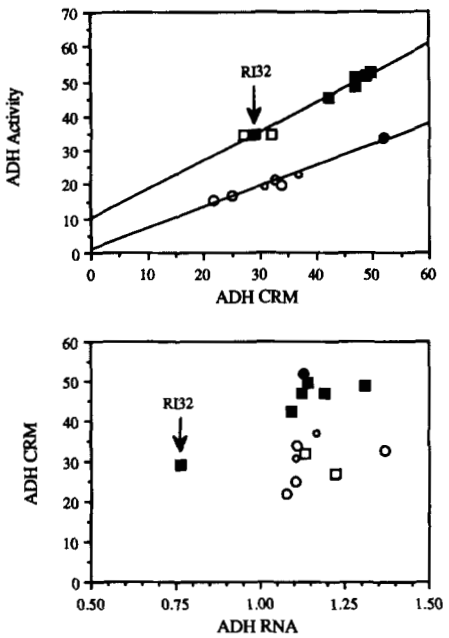


FIGURE 5.—Plots of the line means from experiment I (given in Table 1). Regression lines for activity on CRM were calculated for Slow and Fast alleles separately. "Sa" refers to the haplotype CTATCC and "Sb" refers to the haplotype CCACTA (at nucleotides 1443, 1452, 1490, 1518, 1527 and 1557). The "+" and "−" designations indicate presence or absence of ∇1, respectively. Units are defined in Table 1. $Sa^-$ (o), $Sb^-$ (○), $Sb^+$ (●), $F^+$ (■), $F^-$ (□).

which, along with RI32, are the only lines that appear to have switched activity class with respect to their allozymic type (Figure 3). As discussed above, RI32 is clearly different from the other low activity Fast line, KA12, because it has a low RNA level as well as a low CRM level and because it has a 29-bp deletion in the 5′ untranslated leader of the distal message (see Figures 4 and 5). These results suggest that the ∇1 sequence difference may be the cause of the CRM level difference between the allozymic classes.

The ∇1 sequence difference had previously been ruled out as a likely candidate to explain the CRM level difference between the allozymic classes because of line Fl-f, which lacks ∇1. In experiment I, Fl-f had a substantially higher CRM level than the Fl-2s line with which it was compared. However, Table 1 shows that the Fl-f CRM level is actually lower than some of the Slow lines analyzed in that experiment. These observations were previously interpreted as a geographic location effect. Fl-f is the lowest of the Fast lines, but Fl-2s is also the lowest of the Slow lines. By coincidence, KREITMAN (1983) cloned and sequenced two Slow alleles from the Florida population. The ADH activity and CRM levels of all three of the Florida alleles were analyzed in experiment II (Table 2). The results of this experiment show that the other Florida allele, Fl-1s, actually has a higher CRM level than Fl-f, which indicates that the geographic location interpretation is probably incorrect. Therefore, the unusually low level of ADH CRM in Fl-f also supports

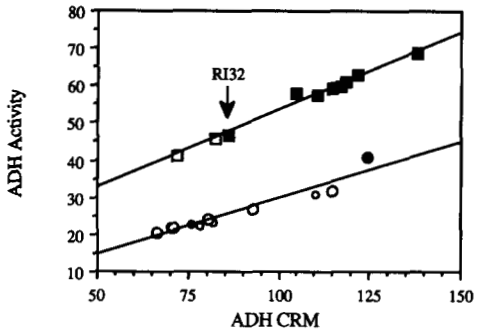


FIGURE 6.—Plot of the line means from experiment II (given in Table 2). Regression lines for activity on CRM were calculated for Slow and Fast alleles separately. "Sa" refers to the haplotype CTATCC and "Sb" refers to the haplotype CCACTA (at nucleotides 1443, 1452, 1490, 1518, 1527 and 1557). The "+" and "−" designations indicate presence or absence of ∇1, respectively. Units are defined in Table 2. $Sa^-$ (o), $Sb^-$ (○), $Sb^+$ (●), $F^+$ (■), $F^-$ (□).

the hypothesis that the ∇1 sequence difference causes the CRM level difference between the allozymic classes.

The *Adh* expression data from experiment I are summarized in Figure 5. This experiment includes several of the KREITMAN lines as well as several lines from the set of 49 in Figure 3 (see also Table 1). Figure 5 shows that all of the lines with ∇1 have higher CRM levels than all of the lines that lack ∇1, with the exception of the deletion line RI32.

Experiment II is essentially a repeat of experiment I, but with the following differences. The second chromosome from each of KREITMAN's stocks was extracted into the Hochi-R genetic background so that all of the stocks analyzed in experiment II would have the same *X* and third chromosomes (with one exception discussed in MATERIALS AND METHODS). Then each of the stocks was checked for the presence or absence of important sequence differences, which include ∇1 within the adult intron and a cluster of 6 sites within the region of exon 4 that includes the amino acid replacement (see Table 2). This information was obtained by direct sequencing of PCR-amplified *Adh* DNA from each fly stock. The only discrepancy between the PCR sequences and the sequences reported by KREITMAN (1983) is that the site at 1527 contains a C in the Ja-s flies, whereas KREITMAN reported a T at this position. This discrepancy does not alter any of the interpretations previously made, because other Slow alleles also have a C at this position.

The results of experiment II (summarized in Table 2 and Figure 6) are essentially the same as experiment I, except that there is some overlap between the high CRM lines having ∇1 and the low CRM lines lacking ∇1. Nevertheless, among the Slow lines, the only one having ∇1 (NC16) has the highest CRM level and, among the Fast lines, the only two that lack ∇1 have the lowest CRM levels (KA12 and Fl-f). These results support the suggestion that ∇1 has an effect on ADH CRM level.

Figure 1 shows that ∇1 is actually a complex sequence difference within the adult intron. The typical Slow allele has a segment of 29 bases beginning at +448, which is replaced in the typical Fast allele by a segment of 34 bases of a completely different sequence. The Fast-typical sequence contains several short direct repeats of the sequence TAATA(C), which resembles a "TATA" box promoter element. It is not clear how such a sequence difference within an intron can affect the level of ADH protein without also affecting the level of ADH RNA. As discussed above, the lack of a difference in RNA level between the allozymic classes and the lack of an effect of the only amino acid difference on ADH protein level suggest that the protein level difference is due to a difference in translational efficiency.

How might the ∇1 sequence difference affect translational efficiency? There are at least three possibilities, which all involve heterogeneity in the structure of the mRNA population. In this context it is important to briefly describe the two procedures that LAURIE and STAM (1988) used to conclude that there is no RNA level difference between the allozymic classes. In the first procedure, total RNA was prepared by urea lysis and pelleting through a CsCl cushion and this preparation was subjected to formaldehyde/agarose electrophoresis and Northern blotting. The flies that provided the RNA were all heterozygous for either a Slow or Fast allele and an null mutant, $Adh^{nLA248}$, which makes a message about 200 bases larger than normal. The amount of RNA provided by the wild type allele was measured relative to the amount provided by the mutant. Using this procedure, message that is unprocessed with respect to the adult intron, which is 654 bp in length, could have been detected, but no RNA of the predicted length was detected on the Northern blots. In the second procedure, total nucleic acids were extracted from adult flies and subjected to an RNase protection assay, which makes use of another *Adh* mutant as an internal control. The probe for this assay spans the last intron of the *Adh* gene and the measured RNA was therefore processed with respect to that intron. Neither of these procedures discriminates between nuclear *vs.* cytoplasmic, polyadenylated *vs.* non-polyadenylated RNA or distal *vs.* proximal RNAs.

Since the ∇1 sequence difference lies within the 5′ flanking region of the proximal transcriptional unit, it is theoretically possible that its phenotypic effect (if it really has one) occurs through regulation of the level of production of the proximal transcript. If this were true, then Fast and Slow lines might have the same total level of processed *Adh* message, but the message populations might differ in the relative

amounts of distal *vs.* proximal forms. If the two types of transcripts have different translation rates, then this could account for the different levels of ADH protein. This explanation is unlikely because RNase protection assays show only trace amounts of the proximal transcript in adults of either Fast or Slow genotypes (FISCHER and MANIATIS 1986; CORBIN and MANIATIS 1989b; THOMSON, JACOBSON and LAURIE 1991) and all of the expression data discussed in this paper so far comes from adult flies.

It is also possible that the ∇1 sequence difference might result in a heterogeneous message population with respect to splicing of the adult intron. However, RNase protection assays using a probe that spans the 3′ acceptor site of the adult intron provide no evidence for such heterogeneity in either Fast or Slow genotypes (FISCHER and MANIATIS 1986; CORBIN and MANIATIS 1989b; Thomson, Jacobson and LAURIE 1991). Furthermore, as discussed above, no significant amount of unprocessed RNA was detected on the Northern blots of LAURIE and STAM (1988).

There is a growing body of evidence that intron splicing is mechanistically coupled with polyadenylation and transport of mRNA into the cytoplasm (BUCHMAN and BERG 1988; HUANG and GORMAN 1990). Also, there is considerable evidence that a 3′-poly(A) tail acts to stimulate translation (MUNROE and JACOBSON 1990). Thus, it is possible that a sequence difference within an intron may indirectly affect translation rate through an effect on polyadenylation or transport. Investigation of this possibility will require careful analysis of possible allozymic differences in message structure at the 3′ end and its distribution between the nucleus and cytoplasm.

Because of the location of the ∇1 sequence difference, it seems likely that its effect, if any, would be different in adults, in which the distal transcript is predominant, than in larvae up to the mid-third instar stage, in which the proximal transcript is predominant. As far as we know, there has been no comparison between allozymes of the ADH CRM level in larvae of the appropriate ages. MARONI *et al.* (1982) analyzed wandering third instar larvae and found that the activity level difference between allozymes is very similar to that in adult flies. However, this comparison is not very informative because wandering third instars have predominantly distal transcript (SAVAKIS, ASHBURNER and WILLIS 1986; FISCHER and MANIATIS 1986).

**A cluster of six polymorphic sites between 1443 and 1557 show strong linkage disequilibrium, but these haplotypes are not associated with ADH CRM level:** The region between 1443 and 1527 contains six polymorphic sites that show strong linkage disequilibrium. One of those sites is the amino acid replacement polymorphism and the others are silent polymorphisms at 1443, 1452, 1518, 1527 and 1557. A total of 11 Fast alleles have been sequenced through this region and they all have a single haplotype (GTCTCC); this includes the low CRM alleles RI32, KA12 and Fl-f (Tables 1 and 2). A total of 12 Slow alleles have been sequenced through this region and two haplotypes have been found, which differ at four of the six sites (CTATCC and CCACTA). Figures 5 and 6 show that there is no suggestion of a difference in ADH CRM level between the two Slow haplotypes.

KREITMAN and HUDSON (1991) have recently pointed out that two Slow alleles, Wa-s and Fl-1s, differ from the other Slow alleles sequenced by KREITMAN by nine silent substitutions within the *Adh* gene as well as a number of other substitutions within the *Adh-dup* locus, which begins less than 150 bases downstream of *Adh*. These two alleles account for a spike of polymorphism among Slow alleles that is greater than expected under a neutral model. This result suggests the possibility that selection may somehow distinguish the two groups of Slow alleles. However, there is no obvious difference in *Adh* expression between the two Slow groups. In terms of CRM level, Wa-s ranks 3 and Fl-1s ranks 10 out of a total of 11 (Table 2).

KREITMAN and HUDSON (1991) also show that, even without alleles Wa-s and Fl-1s, there is an excess of silent polymorphism centered around the amino acid replacement at 1490, which suggests that the amino acid replacement polymorphism or some other site in the immediate vicinity is being maintained by some form of balancing selection. The adult intron, which contains ∇1, does not show an excess of polymorphism. This region shows low levels of interspecific divergence and polymorphism. If ∇1 actually does cause the CRM level difference associated with the allozyme polymorphism, then evidently the region immediately surrounding this site is subject to different evolutionary dynamics than the region containing the amino acid replacement polymorphism, in spite of the strong linkage disequilibrium between the two polymorphisms. There are many reasons why that might be the case, but perhaps it is related to the fact that the 3′ end of this intron contains the promoter and initiation site of the proximal transcript.

To identify any other substitutions that might also account for the ADH CRM level difference, the completely sequenced alleles can be divided into two classes (excluding RI32): The low CRM class includes KA12, Fl-f and all of the Slow alleles except NC16. The high CRM class includes NC16 plus all of the Fast alleles except KA12 and Fl-f. When the sequences of these alleles are compared across the *Hpa*I/*Cla*I fragment, ∇1 is the only site that distinguishes all of the high CRM from all of the low CRM alleles. Even considering the possibility that Ja-s and Fl-1s might

belong to the high CRM class (as suggested by experiment II but not experiment I), there are no other sites that distinguish the two groups.

If an *in vitro* mutagenesis experiment shows that the ∇1 polymorphism does not have an effect on CRM level, then epistasis will have to be considered. It is possible that no single site substitution accounts for the CRM level difference between the allozymic classes. For example, it may be that the 1490 amino acid replacement substitution has an effect on CRM level, but only within the appropriate context of the other five polymorphic sites within the region 1443 to 1527. This hypothesis can be tested by an *in vitro* mutagenesis experiment in which all six substitutions are changed at once. If the sequence differences within the 1443 to 1527 interval actually do affect CRM level, then alternative explanations will be required to account for the unusual CRM levels in the rare haplotype lines NC16, KA12 and Fl-f. The two novel amino acid replacements in NC16 provide obvious candidates to account for the unusual CRM level in that line, but there are no obvious alternatives to ∇1 for the other two lines.

## CONCLUSIONS

A large part of the genetic variation in ADH activity level in natural populations is associated with segregation of an amino acid replacement polymorphism at nucleotide 1490 that generates a difference in electrophoretic mobility. Part of the allozymic difference in activity level is due to a catalytic efficiency difference, which is also caused by the amino acid replacement, and part is due to a difference in the concentration of ADH protein. It is clear that the difference in ADH concentration is not due to the amino acid replacment itself, but rather to one or more polymorphisms that are in linkage disequilibrium with the amino acid replacement polymorphism. The population survey data discussed above suggest that the ∇1 sequence difference within the adult intron may cause the ADH concentration difference. If *in vitro* mutagenesis shows that it does not, multiple site models will have to be considered.

Studies of some other enzyme polymorphisms in *D. melanogaster* suggest that the *Adh* situation may not be uncommon. Analyses of activity variation and RFLP polymorphisms of amylase (LANGLEY *et al.* 1988), glucose-6-phosphate dehydrogenase (MIYASHITA 1990) and esterase-6 (GAME and OAKESHOTT 1990) all show activity differences between allozymic classes and extensive linkage disequilibrium among the allozymic polymorphism and various RFLPs. In these cases it is not yet clear whether any of the activity difference is due to substitutions other than the amino acid replacements that cause the charge differences, but the *Adh* example suggests that it would be worth investigating that question.

We are grateful for the expert technical assistance of ELLEN FLANAGAN, DAVID KEYS, LYNN STAM and JUSTINA WILLIAMS. We are also grateful for helpful comments on the manuscript by C. H. LANGLEY and C. F. AQUADRO. In addition, we thank C. AQUADRO, C. LANGLEY, E. MONTGOMERY and W. QUATTLEBAUM for providing the λ clones and some preliminary sequence information about allele KA27. This work was supported by National Science Foundation grant BSR 861-15632.

*Note added in proof:* The nucleotide sequence data reported in this paper will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession numbers M36580 for KA27, X60793 for NC16, X60791 for KA12 and X60792 for RI32.

## LITERATURE CITED

ANDERSON, S. M., and J. F. MCDONALD, 1983 Biochemical and molecular analysis of naturally occurring *Adh* variants in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **80:** 4798–4802.

AQUADRO, C. F., S. F. DEESE, M. M. BLAND, C. H. LANGLEY and C. C. LAURIE-AHLBERG, 1986 Molecular population genetics of the alcohol dehydrogenase gene region of *Drosophila melanogaster*. Genetics **114:** 1165–1190.

BENYAJATI, C., N. SPOEREL, H. HAYMERLE and M. ASHBURNER, 1983 The messenger RNA for alcohol dehydrogenase in *Drosophila melanogaster* differs in its 5′ end in different developmental stages. Cell **33:** 125–133.

BEWLEY, G. C., 1981 Genetic control of the developmental program of 1-glycerol-3-phosphate dehydrogenase isozymes in *Drosophila melanogaster*: identification of a *cis*-acting temporal element affecting GPDH expression. Dev. Genet. **2:** 113–129.

BUCHMAN, A. R., and P. BERG, 1988 Comparison of intron-dependent and intron-independent gene expression. Mol. Cell. Biol. **8:** 4395–4405.

CHOUDHARY, M., and C. C. LAURIE, 1991 Use of *in vitro* mutagenesis to analyze the molecular basis of the difference in *Adh* expression associated with the allozyme polymorphism in *Drosophila melanogaster*. Genetics **129:** 481–488.

CHOVNICK, A., M. MCCARRON, S. H. CLARK, A. J. HILIKER and C. A. RUSHLOW, 1980 Structural and functional organization of a gene in *Drosophila melanogaster*, pp. 3–23 in *Development and Neurobiology of Drosophila*, edited by O. SIDDIQUI, P. BABU, L. M. HALL and J. C. HALL. Plenum Press, New York.

CORBIN, V., and T. MANIATIS, 1989a The role of specific enhancer-promoter interactions in the *Drosophila Adh* promoter switch. Genes Dev. **3:** 2191–2200.

CORBIN, V., and T. MANIATIS, 1989b Role of transcriptional interference in the *Drosophila melanogaster Adh* promoter switch. Nature **337:** 279–282.

CORBIN, V., and T. MANIATIS, 1990 Identification of *cis*-regulatory elements required for larval expression of the *Drosophila melanogaster* alcohol dehydrogenase gene. Genetics **124:** 637–646.

DEVEREUX, J., P. HAEBERLI and O. SMITHIES, 1984 A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:** 387–395.

DICKINSON, W. J., 1978 Genetic control of enzyme expression in Drosophila: a locus influencing tissue specificity of aldehyde oxidase. J. Exp. Zool. **206:** 333–342.

FISCHER, J. A., and T. MANIATIS, 1986 Regulatory elements involved in *Drosophila Adh* gene expression are conserved in

divergent species and separate elements mediate expression in different tissues. EMBO J. **5:** 1275–1289.

FRISCHAUF, A. M., N. MURRAY and H. LEHRACH, 1987 λ phage vectors–EMBL series. Methods Enzymol. **153:** 103–115.

GAME, A. Y., and J. G. OAKESHOTT, 1990 Associations between restriction site polymorphism and enzyme activity variation for esterase 6 in *Drosophila melanogaster*. Genetics **126:** 1021–1031.

GOLDBERG, D. A., J. W. POSAKONY and T. MANIATIS, 1983 Correct developmental expression of a cloned alcohol dehydrogenase gene transduced into the Drosophila germ line. Cell **34:** 59–73.

HIGUCHI, R. G., and H. OCHMAN, 1989 Production of single-stranded DNA templates by exonuclease digestion following the polymerase chain reaction. Nucleic Acids Res. **17:** 5865.

HUANG, M. T. F., and C. M. GORMAN, 1990 Intervening sequences increase efficiency of RNA 3′ processing and accumulation of cytoplasmic RNA. Nucleic Acids Res. **18:** 937–947.

KREITMAN, M., 1983 Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. Nature **304:** 412–417.

KREITMAN, M., and R. R. HUDSON, 1991 Inferring the evolutionary histories of the *Adh* and *Adh-dup* loci in *Drosophila melanogaster* from patterns of polymorphism and divergence. Genetics **127:** 565–582.

LANGLEY, C. H., A. E. SHRIMPTON, T. YAMAZAKI, N. MIYASHITA, Y. MATSUO and C. F. AQUADRO, 1988 Naturally occurring variation in the restriction map of the *Amy* region of *Drosophila melanogaster*. Genetics **119:** 619–629.

LAURIE, C. C., and L. F. STAM, 1988 Quantitative analysis of RNA produced by Slow and Fast alleles of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **85:** 5161–5165.

LAURIE, C. C., E. M. HEATH, J. W. JACOBSON and M. S. THOMSON, 1990 Genetic basis of the difference in alcohol dehydrogenase expression between *Drosophila melanogaster* and *Drosophila simulans*. Proc. Natl. Acad. Sci. USA **87:** 9674–9678.

LAURIE-AHLBERG, C. C., 1985 Genetic variation affecting the expression of enzyme-coding genes in Drosophila: an evolutionary perspective. Isozymes Curr. Top. Biol. Med. Res. **12:** 33–88.

LAURIE-AHLBERG, C. C., and L. F. STAM, 1987 Use of *P*-element-mediated transformation to identify the molecular basis of naturally occurring variants affecting *Adh* expression in *Drosophila melanogaster*. Genetics **115:** 129–140.

LAURIE-AHLBERG, C. C., G. MARONI, G. C. BEWLEY, J. C. LUCCHESI and B. S. WEIR, 1980 Quantitative genetic variation of enzyme activities in natural populations of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **77:** 1073–1077.

MARONI, G., and C. C. LAURIE-AHLBERG, 1983 Genetic control of *Adh* expression in *Drosophila melanogaster*. Genetics **105:** 921–933.

MARONI, G., C. C. LAURIE-AHLBERG, D. A. ADAMS and A. N. WILTON, 1982 Genetic variation in the expression of *Adh* in *Drosophila melanogaster*. Genetics **101:** 431–446.

MIYASHITA, N. T., 1990 Molecular and phenotypic variation of the *Zw* locus region in *Drosophila melanogaster*. Genetics **125:** 407–419.

MUNROE, D., and A. JACOBSON, 1990 Tales of poly(A): a review. Gene **91:** 151–157.

OAKESHOTT, J. G., J. B. GIBSON, P. R. ANDERSON, W. R. KNIBB, D. G. ANDERSON and G. K. CHAMBERS, 1982 Alcohol dehydrogenase and glycerol-3-phosphate dehydrogenase clines in *Drosophila melanogaster* on different continents. Evolution **36:** 86–96.

POSAKONY, J. W., J. A. FISCHER and T. MANIATIS, 1985 Identification of DNA sequences required for the regulation of *Drosophila* alcohol dehydrogenase gene expression. Cold Spring Harbor Symp. Quant. Biol. **50:** 515–520.

SAVAKIS, C., M. ASHBURNER and J. H. WILLIS, 1986 The expression of the gene coding for alcohol dehydrogenase during the development of *Drosophila melanogaster*. Dev. Biol. **114:** 194–207.

SHAFFER, J. B., and G. C. BEWLEY, 1983 Genetic determination of *sn*-glycerol-3-phosphate dehydrogenase in *Drosophila melanogaster*: a *cis*-linked controlling element. J. Biol. Chem. **258:** 10027–10033.

THOMSON, M. S., J. W. JACOBSON and C. C. LAURIE, 1991 Comparison of alcohol dehydrogenase expression in *Drosophila melanogaster* and *D. simulans*. Mol. Biol. Evol. **8:** 31–48.

Communicating editor: A. G. CLARK