PNAS
www.pnas.org

Supplementary Information for

The evolution of more: A major role for noncoding regulatory mutations in the evolution of enzyme activity

David W. Loehlin, Jesse R. Ames, Kathy Vaccaro, Sean B. Carroll.

Corresponding author: David W. Loehlin, Sean B. Carroll
Email: dwl1@williams.edu, carrolls@hhmi.org

This PDF file includes:

Supplementary Information Text

Nucleotide-level mapping results of *D. melanogaster* fast/slow.

Fine-scale mapping of the *D. melanogaster* fast and slow alleles uncovered six causative sites (Fig. 5 and S4). These are described in detail below.

5'-flanking region: The source of the difference originating from the 5'-flanking region mapped to the nucleotide on the border with the 5'-UTR region; this border is spanned by a multisite substitution (*Adh-slow*: TCA**C**CG, *Adh-fast*: TCA**T**GC. The underline marks the start of the 5'-UTR region; SI Methods). This sequence binds Initiator, a core promoter transcription factor that positions RNA Polymerase II (1).

Since the multisite substitution spans two regions in our five-region map, we designed two constructs in order to tease apart the potential contribution of the initator and 5'UTR substitutions. One construct recombined 5' of the multisite substitution (construct "X", sequence TCA|**T**GC, "|" is the recombination breakpoint) while the second construct split the multisite substitution by recombining immediately 5' of the presumed transcription start site (construct "Y", sequence TCA**C**|GC). ADH activity of "X" and "Y" was found to differ by 17% ($P < 0.001$, Figure 5), demonstrating that the **C/T** substitution in the Initiator binding motif is a causative contributor to the allelic difference. Also, this substitution is sufficient to explain the entire activity difference arising from the 5'-flanking region (Figure 5). Remarkably, a parallel C/T substitution in the same Initiator site distinguishes *D. yakuba* (TCG**T**CG) from *D. santomea* (TCG**C**CG).

5'-UTR region: This region contributes a 49% difference in activity, the dominant contributor to activity evolution in this lineage (Table 1). Laurie and Stam (2) had implicated an indel polymorphism in the first intron, delta-1, as causative. Our mapping (Fig. 5) confirms the role of delta-1 and reveals the contribution of two additional substitutions: a C/G SNP and the delta-2 indel. All three causative substitutions lie within 100 bp. These are near to, but slightly upstream of, promoter sequences required for larval expression (3), suggesting the presence of a previously undescribed regulatory element.

Coding region: One amino acid substitution occurs here, a lysine to threonine at position 192 (K192T) that has been shown to affect ADH activity (4-5). In our mapping experiment, we tested for possible effects of K192T and three other (noncoding) substitutions in this region (Fig. S4B). None of the noncoding substitutions contributed a detectable effect. Therefore, the K192T substitution is sufficient to explain the entirety of the activity difference from the coding region.

3'-UTR region: We identified a 10% change in activity due to substitution(s) here, but we were unable to further resolve the responsibility of specific site(s). Two substitutions occur in our pair of alleles: a poly-A tract of length 14 in *slow* and length 15 in *fast*, and a complex substitution of sequence CA in *slow* and G- in *fast*. We attempted to separate the effects of the $A_{(14)}/A_{(15)}$ from the CA/G- with a construct that was recombined between the two sites. However, activity of this construct was not significantly different from either neighboring construct (Figure S4D), indicating that we did not have sufficient power to determine the causative site(s). In spite of this, we note that Stam and Laurie (6) also mapped an activity difference with a similar effect size of 1.1-fold to a region containing the 3'-UTR. Their *fast* and *slow* haplotypes carry a different suite of substitutions in this region, but the C/G nucleotide substitution is shared in their study and ours. Additional mapping work would be required to determine if this one substitution or if multiple substitutions are responsible for the activity difference from the 3'-UTR.

Methods.

***Drosophila* strains.** *Adh* loci were PCR amplified and cloned from *D. melanogaster* strains Canton-S (Bloomington Stock Center #64349) and Florida-9 (Bloomington #2374), *D. erecta* (UCSD Stock Center 14021-0224.00), *D. orena* (UCSD 14021-0245.01), *D. yakuba* (UCSD 14021-0261.01), *D. santomea* (UCSD 14021-0271.01), *D. americana* (UCSD 15010-0951.00), *D. virilis* (Bowling Green Stock Center 15010-1051.0, also used in (7)). All transgenes were injected into *D. melanogaster* strain pf86 (7), of genotype y[1] M{vas-int.Dm}ZH-2A w[*]; Adh[fn6] cn[1]; M{3xP3-RFP.attP}ZH-86Fb. This strain carries the Leu479Phe allele of *Aldehyde Dehydrogenase* (8).

Engineering recombinant *Adh* loci. *Adh* locus fragments were PCR amplified and cloned into injection vector pS3aG using steps described below. Site-specific germline transformants were generated using the phiC31 system following the strategy described in (7). One design challenge with such a transgenic approach is that the size of a genetic locus, including all flanking regulatory sequences, can span for hundreds of kilobases in *Drosophila*. We restricted our study to an ~8kb fragment, including the *Adh* transcription unit plus at least 3 kb in either direction. Our constructs did not include a putative larval regulatory element described to be 3-5kb upstream of the promoter (9). It is also possible that additional distant noncoding regulatory elements could contribute to activity divergence. However, initial experiments indicated that most of the divergence in activity was recapitulated by the sequences within the transgenes (Figure 2).

Loci were amplified and cloned as follows. Primers for PCR and sequencing are in Table S1. *D. melanogaster* alleles were amplified using primers S3aAF1/S3aNR1, digested with AscI and NotI-HF (New England Biolabs), and ligated into attB vector pS3aG using the AscI and NotI sites. *D. erecta* and *D. orena Adh* loci were amplified in two overlapping fragments using primers S3aAf2/ere5055R and ere4771F/S3aNR1, A-tailed, ligated into pGem-T-Easy, sequenced, then inserted into pS3aG using the flanking AscI and NotI sites plus a conserved

internal PacI site to join the left and right halves. A similar approach was used for *D. yakuba* and *D. santomea Adh* loci, but using primers S3aAF2/ere5055R and ere4217F/S3aNR1. The *D. americana* locus was amplified using primers Adh_vir_AF1/Adh_ame_lhsR2 and Adh_ame_rhsF2/Adh_vir_NR1 and an internal BamHI site. The *D. virilis* tandem duplicate locus was amplified and sequenced as described in (7).

For the *melanogaster* group species, our constructs include the entirety of *Adhr*, a downstream polycistronic ORF present in ~1% of *Adh* transcripts (in the *D. virilis* group, *Adhr* is located on the other side of *outspread*, ~80kb away). ADHR's function is unclear, other than it does not metabolize ethanol (10), and so it was not quantified here. Each locus was sequenced from multiple clones and a clone that perfectly matched the consensus was used for subsequent work. The *D. melanogaster fast* allele was recalcitrant to Sanger sequencing around the palindromic delta-2 sequence, unless we first denatured the template and primer in 1 microliter of 100 mM Tris, 0.1 mM EDTA, pH 8, at 98 C for 5 min, then placed the mixture on ice, and finally added the sequencing reagents and cycled as normal. Sequences are available in Genbank (MH614199-MH614205; *D. virilis*: KU559568.1).

Our initial design intent was to subdivide the gene at the boundaries of the transcript and translational unit. Subsequently, we became aware that a) the *D. melanogaster Adh* transcriptional start site (TSS) was different in different genome databases and b) causative substitutions occurred right on this boundary. Because the TSS location appeared to be ambiguous, and perhaps because many TSS are biologically variable (11), we chose to not make strong inferences about the relative contribution of causative substitutions in untranscribed versus transcribed regions. When generating recombinant alleles, we used the transcription and translation breakpoints from the following published datasets. For *D. melanogaster* (and *melanogaster* group species), we used the gene coordinates of the adult transcript (Adh-RC, Genbank ID NM_001032095.1) from Flybase Release 5 (12). The TSS of this transcript was called a few nucleotides upstream in Flybase Release 6 (13) and a few nucleotides downstream

in a targeted primer extension assay (1). For *D. virilis* and *D. americana,* we used the transcript boundaries identified by Nurminsky et al. (14).

Recombinant *Adh* alleles were created by PCR and Gibson assembly using the primers in Table S1. Recombination can be done in two directions (e.g., AA and BB can recombine as AB and BA), but for simplicity, we elected to map in only one (arbitrarily chosen) direction. The basic approach was to design primers to the site of the desired breakpoint, amplify the left half of the breakpoint from Species A and the right half from Species B, and assemble these into the pS3aG vector using Gibson Assembly Master Mix or NEBuilder HiFi Master Mix (New England Biolabs). In some cases, to reduce PCR error, we reduced the size of the PCR fragments by assembling into a version of the vector that already contained the appropriate 5' and/or 3' sequence. Amino acid swap alleles were created by incorporating the relevant nucleotide changes into primers for PCR and Gibson assembly. For the *melanogaster* group swaps, we incorporated the "high" amino acids into the "low" allele. Conversely, we swapped the *D. americana* amino acid into the single-copy *D. virilis* locus described below. The transgenic lines of *D. mel slow* allele and "Fast" K192T swap into the *D. mel slow* allele were previously reported in Siddiq et al. (5).

To facilitate recombination mapping between *D. virilis and D. americana*, we modified the locus in two ways. First, the *D. virilis Adh* locus occurs in a nearly-identical tandem duplicate of an 8kb region containing the entire promoter and coding sequence. We previously engineered a single-copy version of the *virilis Adh* locus that matches the right-hand copy of the wild-type tandem duplicate (7). This *virilis* single copy allele was used here for recombination mapping of different subregions of the gene between *virilis* and *americana*. Two aspects of the *virilis* tandem duplicate are noted. First, the left and right copies differ by only two nucleotides in the 3'-noncoding region; these were found to have no apparent effect on ADH activity (7). Second, the tandem duplicate allele differs slightly in sequence from the allele in the reference genome, but we found that this allele was present identically in *D. virilis* strains 15010-1051.00 and 1501-

1051.41 by PCR, cloning, and sequencing. For comparisons in this paper, we used the *virilis* single construct.

The second modification was to the *D. americana* construct. This construct was not a precise match to the source locus due to an unstable microsatellite sequence. Specifically, in a plasmid midi-prep of this construct, we noticed variable sizes of the insert when run on a gel after restriction enzyme digestion. PCR and resequencing indicated that a promoter-proximal repeat of $[CTCTATA]_n$ was variable in size, likely due to slippage as *E. coli* DH5α replicated the plasmid. Because it was not straightforward to either identify the 'correct' length of this repeat (Sanger sequencing of our initial clone suggested that there were 84 repeats) or confidently produce a uniform plasmid preparation of this sequence for injection, we decided to isolate a stable version of this repeat and then control for the possibility that it might influence ADH activity. We isolated a clone that contained a 11-copy repeat $[CTCTATA]_{11}$ that was relatively stable in culture, as determined by restriction digest and gel electrophoresis, and used this as the 'reference' *D. americana* locus. For recombination mapping, we chose the mapping direction such that most constructs would have *D. virilis* sequence at this locus. In addition, we generated constructs with recombination events immediately proximal and distal to this repeat, so as to determine the contribution of other substitutions in the 5'-noncoding region.

ADH enzyme assay. To achieve the large sample sizes required to detect small differences between species and recombinant genotypes, we developed a high throughput implementation of the classic ADH enzyme assay (6). A few experiments presented in this paper (described below) were conducted using early variants of the assay or following a laboratory move. We present these separate experiments as separate plots given the possibility that technical differences (e.g., using a different spectrophotometer or total-protein assay) led to different means. Conversely, data shown in the same plot were all measured concurrently. A description of the typical assay method follows, along with a description of procedural variants conducted in particular experiments.

Flies were cultured under low-crowding conditions on sugar food recipe media at room temperature (20-24˚C) in Madison, WI, as described previously (7). 0-1d old males were collected and aged for 4d in food vials, then transferred to a 1mL titer plate on ice. Each well contained one ball bearing and 300μL of 0.1M sodium phosphate buffer + 0.1% Nonidet P-40, pH 8.6. Groups of four flies (from the same vial) were randomly assigned to wells and transferred under CO_2. After homogenization, the final volume was brought to 800 μL (i.e., 200 μL per fly). Homogenization was performed in a Mini-G vertical ball-bearing grinder (Spex Sample Prep) for five repetitions of (15s at 1000RPM, 1 min on ice) using an aluminum plate insert (Kryo-Tech, Spex Sample Prep) to keep the homogenate cold. Homogenates were then centrifuged for 10 min at 4000G, and a 200 μL aliquot was transferred from the center of the well volume into a 300 μL v-bottom microtiter plate at room temperature, from which the enzyme assay and protein assays were loaded.

The ADH enzyme assay protocol is described in detail in (5 and 7). Enzyme substrate concentrations were chosen to be at least twofold higher than those needed to achieve maximum velocity of the most-active allele *Adh-fast*, the *D. melanogaster* allele from FL-9. We used ethanol as the substrate, whereas other studies (e.g., (6)) have used isopropanol, which magnifies the effect of the K192T amino acid substitution. Total protein was measured from 5 μL of homogenate in 100 μL reagent using the Pierce BCA Kit (Thermo Fisher). Three technical replicate ADH activity measurements and three technical replicate protein measurements, each in separate plates, were performed per sample on a Multiskan GO spectrophotometer (Thermo Fisher).

ADH assay methods varied slightly (described below) as we updated the procedure to be more efficient and moved labs. Qualitatively similar results were observed in pilot studies and in comparison of the same constructs run between method revisions. However, quantitative differences among assay variants are possible, for example as a result of using a more dilute fly homogenate.

Assay Variant 1: Along with Variant 2, this was the preferred method variant and is the method described above. Used in experiments shown in Figures 3, 4, 5, S3, S4C, and S4D.

Assay Variant 2: The enzyme assay was identical to Assay Variant 1, but flies were raised on Bloomington Recipe media at room temperature of 18˚C-21˚C in Williamstown, MA. The spectrophotometer used was a Synergy HT (Bio-Tek). This method was used in Figure 2A and also in Figure 1A for *D. melanogaster fast/slow*, *D. yakuba,* and *D. santomea*.

Assay Variant 3: Same as Assay Variant 1, except that single flies were homogenized in 250 μL of buffer, resulting in a more dilute homogenate. This was an early design of a high-throughput assay using the ball-bearing grinder, and we abandoned it in favor of the more efficient Variant 1 that uses pools of flies. This method was used for the experiment shown in Figure S4B.

Assay Variant 4: This early method is described in detail in (7). The primary differences were that five flies were homogenized in Potter-Elvehjem grinders using 1mL of buffer without detergent. *D. virilis* and *D. americana* flies are larger and thus were homogenized as 2 flies in 1 mL buffer to maintain a similar protein concentration. The protein assay used was the Quant-IT Protein assay (Invitrogen). This method was used for Figures 2B and S1. It was also used in Figure 1A for *D. orena* and *D. erecta* (5 flies per mL) and for *D. virilis* and *D. americana* (2 flies per mL).

Our experiment had the following structure. We set up vials of typically four replicate transgenic lines per construct (exceptions: two lines of *mel* recombinant line E, three of *mel* recombinant line V, three of vir single, three of san). Wildtype species/populations were considered to have one replicate line. For each line, two replicate vials were set up. We collected and measured replicate day batches from each experiment on several separate days. Six day replicates were collected for most experiments, with the following exceptions: Ten day replicates were collected in the Figure 2B *virilis* experiment, eight for the *virilis* mapping experiment (Figures 3, 4, and S3F), and one for the *D. erecta-D. orena* species comparison (Figure S1). For the *D. melanogaster* fine-scale map, in Figure S4D we processed two replicate fly samples per

vial instead of one as a way of increasing sample size and in Figure S4B we processed one to three replicate fly samples per vial (see also description of Assay Variant 3). We analyzed this structure using a mixed-effects model: `relative ADH activity ~ construct` with random effects of `line`, `day`, `vial within day`, and when relevant, `sample within vial`.

To get an approximate sense of the differences in ADH activity we could detect, we used the `power.t.test` function in R. A typical set of data (yak-san recombinants) had standard error of 0.012, df of 18, alpha of 0.05/6 (Bonferroni correction), and power of 0.9, giving a detectable difference of 0.023. This is 4% of the activity of a high-activity genotype (yak) and 8% of the activity of a low-activity genotype (san).

We excluded outlier data in a few extreme cases. Using this site-specific integration method, we had observed that a fraction of transgenic lines (fewer than one in ten) would show extreme activity values associated with ectopic insertion of the construct. The modest number of replicate lines we recovered per genotype did not allow us to treat such genetic variants as random error. Therefore, after the initial batch was run, we graphically examined the ADH activity values of replicate lines for outliers. We detected and excluded three candidate outliers from our analysis. (1) One of four lines of *erecta-orena* recombinant position 1 had extremely low activity. This was tested with PCR and found to be an ectopic insertion, i.e., the insert could be PCR amplified but the junctions between insert and flanking genomic DNA could not. These data were excluded and a new replicate line of that genotype was used in subsequent batches. (2) One line of *virilis-americana* rec1 showed activity greater than twofold higher than other lines with that construct. We could not detect evidence of ectopic insertion using PCR from a genomic DNA sample from this genotype. Later samples set up from a different culture of that line showed activity similar to other lines of this construct. We suspect that this line was contaminated, such as by mating and introgression of a wildtype (non-null) *Adh*, so we excluded all measurements of this line from analysis. (3) A single sample of *melanogaster* recombinant Q showed activity of about three times the grand mean, but other batches from the same vial were

similar to the consensus. We suspect this was due to sample mishandling and excluded it from analysis.

All enzyme assay data were analyzed in R version 3.3.3 (15) with a linear mixed-effects model (`lme`, package `nlme`) and visualized using `ggplot2`. The mixed-effects models had crossed-factor random effects, which were specified in `lme` using modified syntax (7, 16). Residuals analysis suggested that constructs with higher activity had higher variance, therefore we modified the model to weight variance by `construct`. For each separate experiment, we compared models with unweighted and weighted variance. In each case, the weighted variance model showed lower AIC and was used, with the exception of the early mapping experiment presented in Figure S4B which would converge only with unweighted variance. As a check, we also ran equivalent unweighted models using `lmer` (package `lme4`), which does not allow variance weighting. The `lmer` models yielded identical parameter estimates to the unweighted `lme` models. Residuals and random effects were checked for normality and lack of structure when plotted against fitted values and grouping factors.

Our ultimate goal with this statistical analysis was to compare the mean activity of adjacent recombinant constructs. Once a model had been fit, we performed sequential comparisons (i.e., between constructs with recombination at adjacent positions, or between amino acid swap and the wildtype allele). The `lme` model fits did not properly compute degrees of freedom (DF), so we extracted DF for the equivalent unweighted `lmer` model using `lsmeans` (package `lsmeans`), Satterthwaite method. 95% confidence intervals and p-values for multiple comparison t-tests (`consec` contrasts, mvt method) were computed with `lsmeans` on the `lme` model fits, but using the Satterthwaite `df`. In some cases, we wanted to compare non-adjacent constructs. For these, we used P-values from all pairwise comparisons.

Western Blot. Relative Adh protein levels were assessed using Western blot and a commercially produced peptide antibody to the *D. melanogaster* Adh (Adh dG-20, SC-22676, Santa Cruz

Biotech). In our hands, this antibody produced low signal relative to background. Antibody binding to other species' Adh proteins was also plausibly affected by amino acid differences in the peptide region, but we note that the amino acid sequences in this region are identical between the pairs of closely related alleles or species planned for comparison in this study (Figure S2). Still, this antibody appeared to consistently detect differences in protein abundance between pairs of species that were consistent with published data on ADH enzyme activity and our own data on ADH enzyme activity.

Lysate was produced by homogenizing 30 *melanogaster* group males or 15 *virilis* group males. Flies were flash frozen in liquid nitrogen in 2mL tubes containing two ball bearings then thawed on ice. 200 μL of homogenization buffer was added (10 mM Tris-HCl ph 7.5, 1 mM $MgCl_2$, 1% v/v Triton X-100, 0.1% v/v HALT protease inhibitor solution without EDTA (Thermo Fisher), and 1 μL/mL Benzonase nuclease (MilliporeSigma). Flies were then homogenized using a Mini-G vertical ball bearing grinder (Spex Sample Prep), grinding for 15s at 1000RPM followed by chilling the tubes 30s on ice, repeated 7 times. Homogenates were centrifuged for 1 minute at 21000g at 4˚C to pellet the cuticle. The supernatant was then pipetted into a 200μL PCR tube and incubated at room temperature for 5-15 minutes for nuclease activity. To isolate the soluble protein from nucleic acids and floating lipid, lysates were centrifuged at 21000G at 4˚C for 15 minutes, and then the middle aqueous layer was transferred to a new tube. Centrifuging was repeated for an additional 15 minute and 5 minute centrifuge steps to achieve a clear lysate. Lysate protein concentrations were quantified with BCA assay and lysates diluted to equal concentrations. Lysates were run immediately or flash frozen in liquid nitrogen and stored at -20˚C.

Lysate was combined with 4x Laemmli sample buffer (Bio-Rad) with equal μg amounts loaded per sample (50μg-80μg) then heat denatured. Proteins were separated on 12% SDS-PAGE gels (Mini-Protean TGX, Bio-Rad) for 120 min at 100V at room temperature. Proteins were wet transferred overnight onto Immobilon-P (Millipore). Blots were dried, stained with Ponceau-S, cut in half below the 37 kDa molecular weight standard, washed, then rewet with

methanol before blocking with 5% w/v powdered milk in TBSTw (0.1M Tris-buffered saline, pH 7.6, 0.1% Tween-20). Primary antibody incubation at 4˚C overnight used goat-anti-Adh dG-20 (SC-22676, Santa Cruz Biotech) at 1:250 or rabbit-anti-gamma-tubulin T5192 (MilliporeSigma) at 1:2000 in blocking solution. Blots were then washed 5 times for 10 minutes in TBSTw at room temperature. Secondary antibody incubation used donkey-anti-goat-peroxidase SAB3700330 (MilliporeSigma) at 1:4000 and donkey-anti-rabbit-peroxidase SAB3700978 (MilliporeSigma) at 1:4000 in blocking solution for 1-2h at room temperature. Blots were then washed 5 times for 10 minutes in TBSTw at room temperature, Chemiluminescence images were captured using SuperSignal West Pico Plus reagent (Thermo Fisher) on a G-Box Chemi imager (Syngene). Levels for an image were adjusted uniformly in the G-box software, with no other manipulations.

Ethanol Resistance Assay. Flies were cultured on Bloomington recipe media at room temperature (18˚C - 21˚C) in Williamstown, MA. Four lines per construct with three replicate cultures per line were set up in low crowding conditions. 5-12 males (typically 12) were collected from each culture vial under CO_2 and then aged 4d in food vials. A rayon plug (Flugs For Narrow Glass Vials, Genesee Scientific) was pushed to the bottom of a plastic Drosophila vial (Genesee Scientific). In pilot tests, these plugs made scoring flies much easier since drunk flies would tend to crawl into/under looser absorbent materials such as cotton balls or filter paper. For the assay, vials were assigned a numerical ID from a random sequence to blind the observer. 2 mL of a solution consisting of 3% sucrose and 13%-25% ethanol was applied to the plug and flies were then transferred to these assay vials from the food vials by bumping against a rubber pad. Assay vials were then sealed with another rayon plug, flush with the top of the vial, and capped with a piece of Parafilm to reduce evaporation. Vials were then sorted based on the random sequence. After 24h, the number of incapacitated flies per vial was counted. Assay vials were held horizontally and rapped twice on a rubber pad to knock the flies over, after which the

vial sat still. Flies that were dead or failed to right themselves after 20 seconds were scored as incapacitated.

For this assay, we had three replicate vials per line but six ethanol concentrations (13%-23%, in 2% steps). Assuming that ethanol resistance would increase logistically with ethanol concentration, we grouped our ethanol treatments into blocks following a round-robin design. For each day the experiment was replicated, each construct was exposed to three replicates of all six ethanol concentrations and each line was exposed to three concentrations separated by 4% (either (13%, 17%, 21%) or (15%, 19%, 23%)). The particular concentration applied to each vial rotated with each day replicate, such that each vial was subjected to each ethanol concentration after six day replicates. We conducted two rotations of each ethanol concentration block, for a total of twelve day replicates. In addition, we had also conducted a pilot test consisting of two day replicates under similar conditions but using ethanol concentrations from 15%-25%. Since the assay conditions were otherwise similar to those in the round-robin experiment, we combined the data from this pilot with the full experiment. Ultimately, the sample sizes used were: 14 days, 4 lines, 3 cultures per line; 162-168 ethanol treatments per construct, 1537-1833 flies per construct (mean 1734); 328-486 flies per line (mean 433).

Ethanol resistance data were analyzed in R with a generalized linear mixed model (`glmer`, package `lme4`) with `construct` and `ethanol concentration` as fixed effects and `line` and `day` as random effects. We also tested models including `block` (set of ethanol treatments) as a random effect, or replacing `day` with `block`, but these were found to be equivalent by chi-square test. IC_{50} means and standard errors were computed from the model fit using the `dose.p.glmm` function from Venables and Ripley (17).

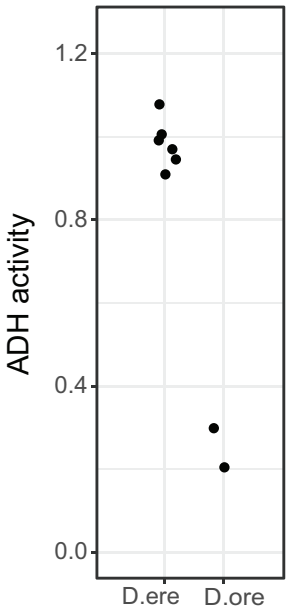


Figure S1. *D. orena* appears to have lower ADH activity than *D. erecta*. Only two replicate samples of *D. orena* males were obtained due to difficulty with culturing the species, but activity of these was ~3.8-fold lower than *D. erecta*. Mean and SD of these data are also presented in Fig. 1. Activity of *D. erecta* flies shown in Fig. 2A is from a separate experiment that used a different assay protocol.

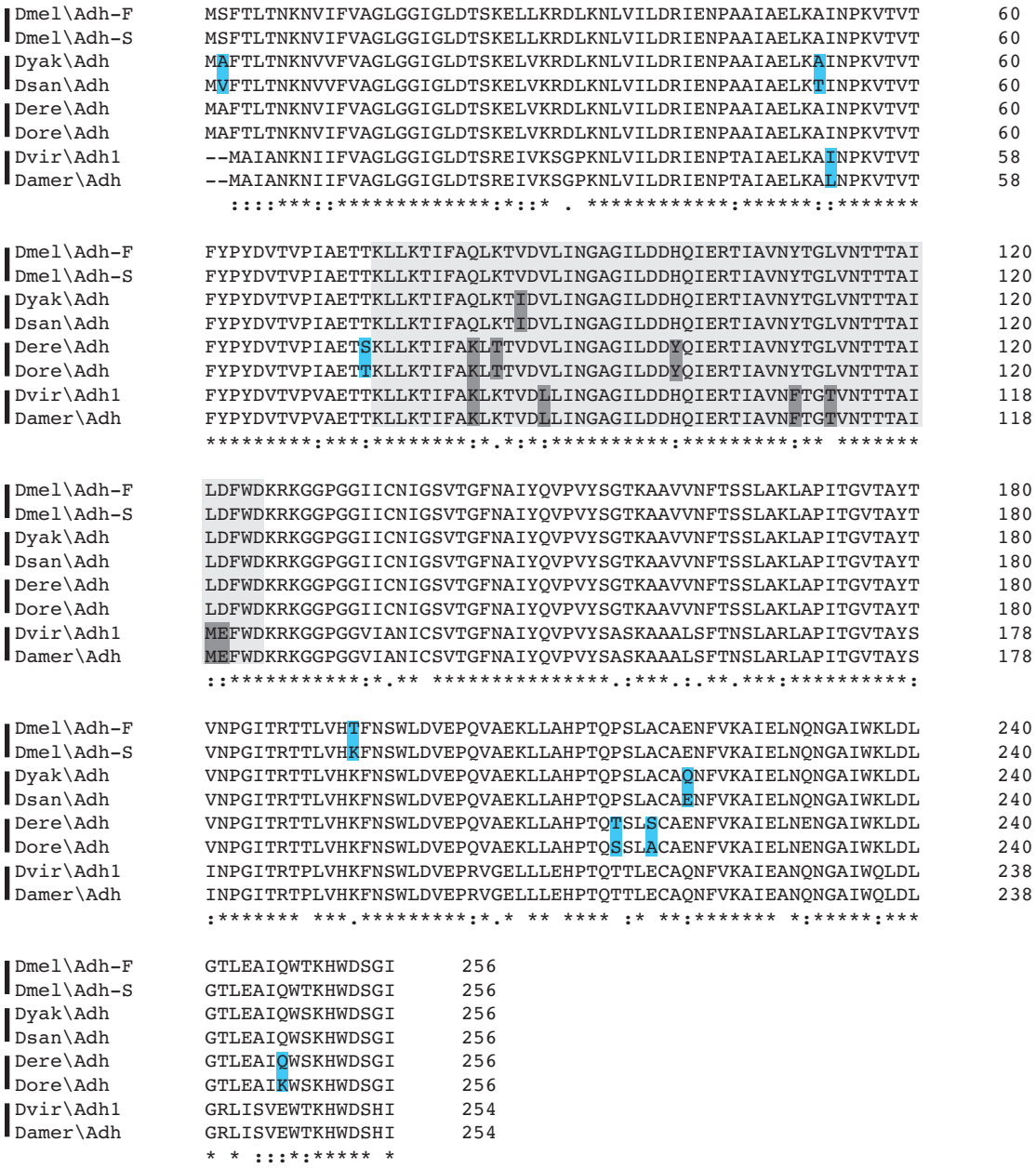


Figure S2. Multiple sequence alignment of ADH protein sequences from the focal species. Amino acid differences within the pairs of species/alleles compared in this study are marked in cyan. Positions 75-125 are highlighted, showing the region to which the peptide-specific antibody Adh dG-20 antibody was raised (Santa Cruz Biotech). Amino acid differences among groups are highlighted in dark gray, and these could potentially affect specificity of the antibody. Low intensity of bands in the *D. erecta* + *D. orena* group (Figure 1, Figure 2) relative to other melanogaster group species could be the result of antibody specificity differences or biological differences in protein levels. The plan of this study was to compare the four pairs of alleles/species, so this difference among pairs was not investigated. In the peptide region, no differences *within pairs* of species occur, suggesting that the observed differences in western blot band intensity within pairs (Figure 1, Figure 2) reveal bona fide differences in protein level. Alignment is by Clustal Omega 1.2.4.

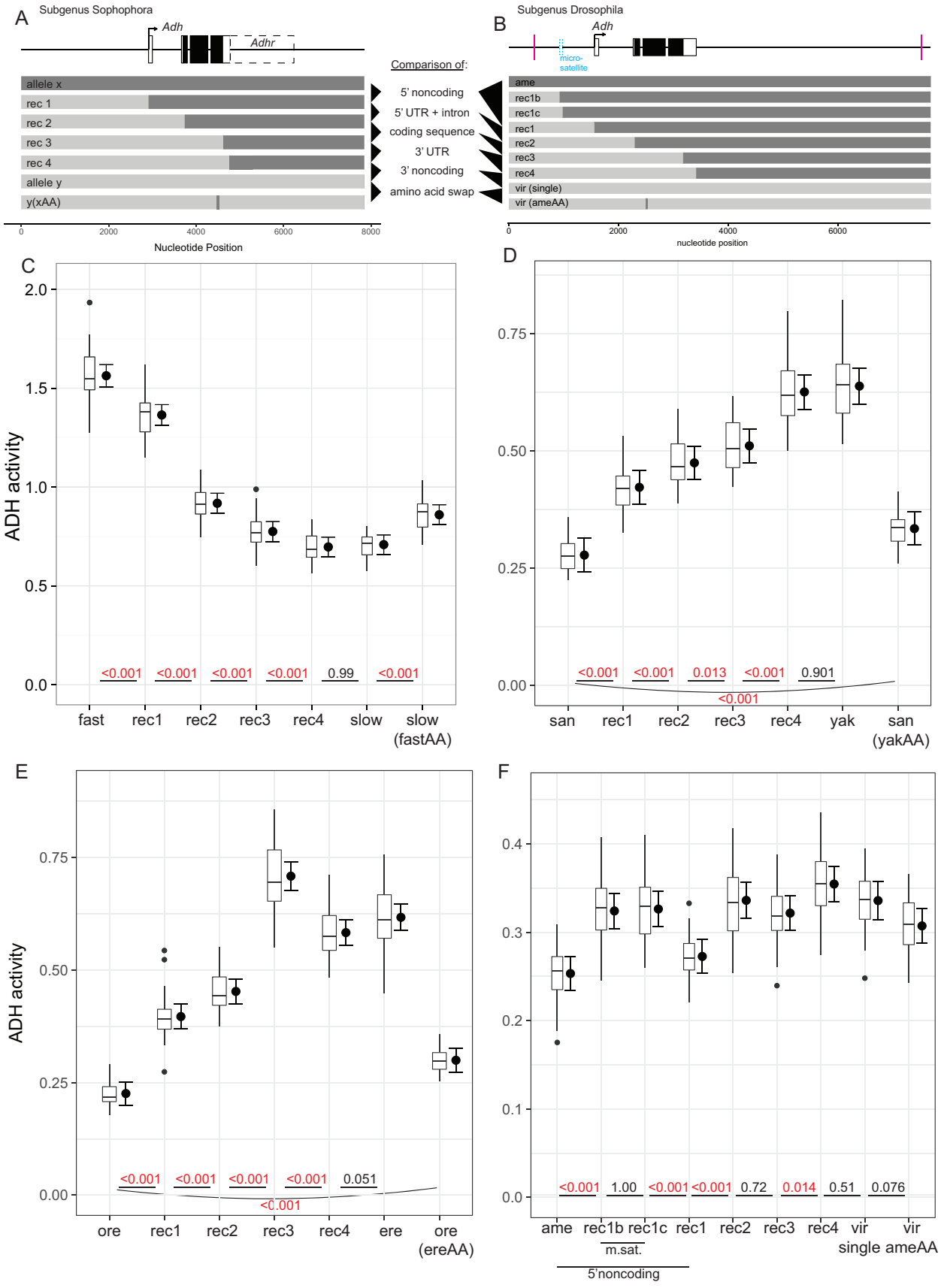

A
Subgenus Sophophora
Adh
Adhr
allele x
rec 1
rec 2
rec 3
rec 4
allele y
y(xAA)
0
2000
4000
6000
8000
Nucleotide Position
Comparison of:
5' noncoding
5' UTR + intron
coding sequence
3' UTR
3' noncoding
amino acid swap
B
Subgenus Drosophila
Adh
micro-satellite
ame
rec1b
rec1c
rec1
rec2
rec3
rec4
vir (single)
vir (ameAA)
0
2000
4000
6000
nucleotide position
C
ADH activity
2.0
1.5
1.0
0.5
0.0
<0.001
<0.001
<0.001
<0.001
0.99
<0.001
fast
rec1
rec2
rec3
rec4
slow
slow (fastAA)
D
0.75
0.50
0.25
0.00
<0.001
<0.001
0.013
<0.001
0.901
<0.001
san
rec1
rec2
rec3
rec4
yak
san (yakAA)
E
ADH activity
0.75
0.50
0.25
0.00
<0.001
<0.001
<0.001
<0.001
0.051
<0.001
ore
rec1
rec2
rec3
rec4
ere
ore (ereAA)
F
0.4
0.3
0.2
0.1
0.0
<0.001
1.00
<0.001
<0.001
0.72
0.014
0.51
0.076
ame
rec1b
rec1c
rec1
rec2
rec3
rec4
vir single
vir ameAA
m.sat.
5'noncoding

Figure S3. Effects of five regions and amino-acid-swaps on ADH enzyme activity for four species pairs. (A) Model of *melanogaster* group *Adh* locus and recombinants used for five-region map. (B) Model of *virilis* group *Adh* locus. Magenta bars show the boundaries of the *D. virilis* tandem duplicate. Cyan box shows the position of the *D. americana* microsatellite. *Adhr* is a separate gene located ~500kb away in this group. (C-F) ADH enzyme activity of transgenic alleles, recombinants, and of swapping the amino acid sequence of one allele into the other allele. Tukey boxplots show distribution of activity measurements and error bars show 95% confidence interval of means. P values and DF are from sequential multiple comparisons between adjacent lines or between an amino-acid swap allele and the unmodified allele (P values for all pairwise multiple comparisons are shown for the same data in Figure 3). Panel C is a subset of the finer-scale mapping experiment presented in Figure 5B. Figure 4 and Table 1 show these data converted into percent differences between adjacent constructs. (F) Mapping within *D. virilis* included additional recombinants rec1b and rec1c to bypass any contribution of the *D. americana* microsatellite, whose length could not be controlled. These recombinants show no effect of the microsatellite, but instead reveal causative and opposing effects from the left and right sides of the 5'-flanking region.

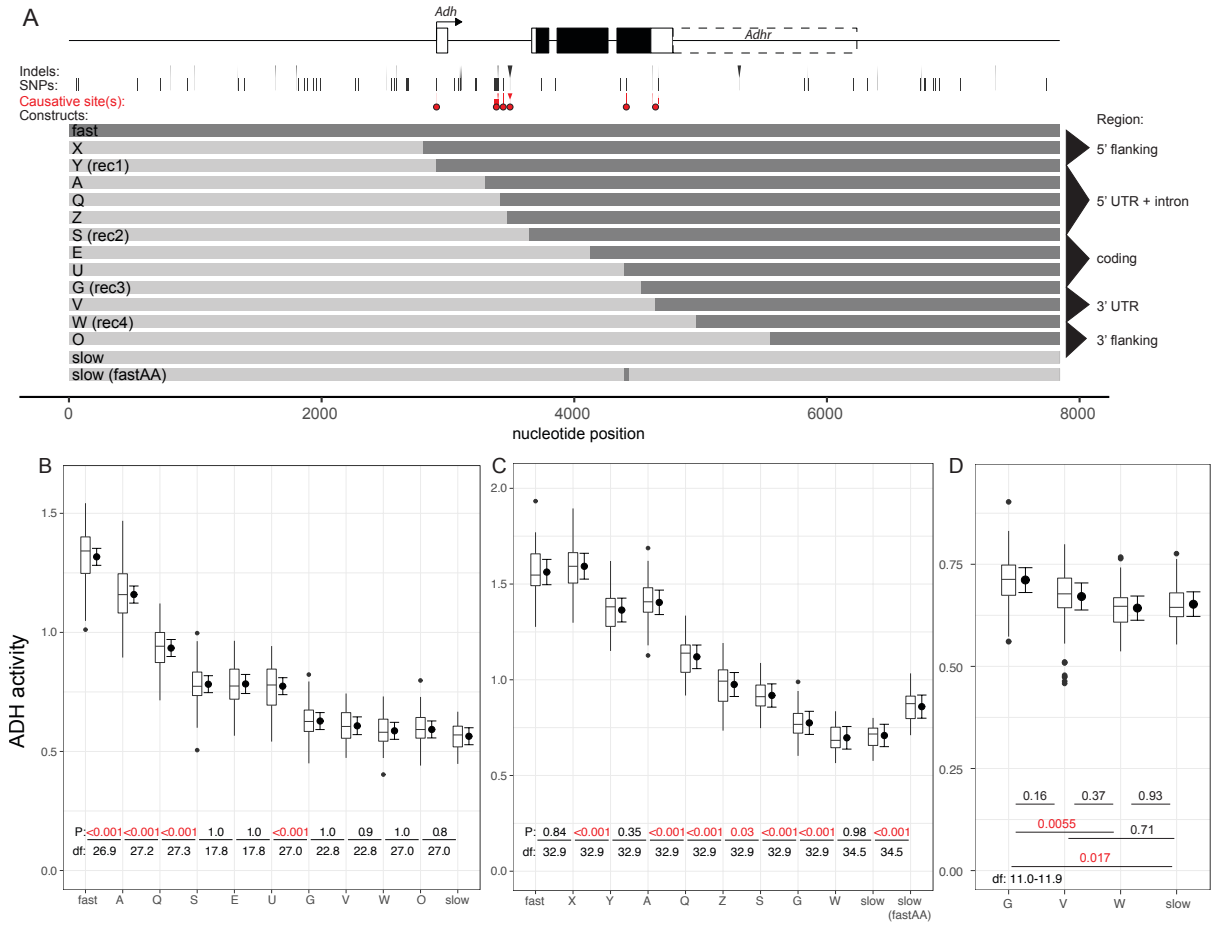
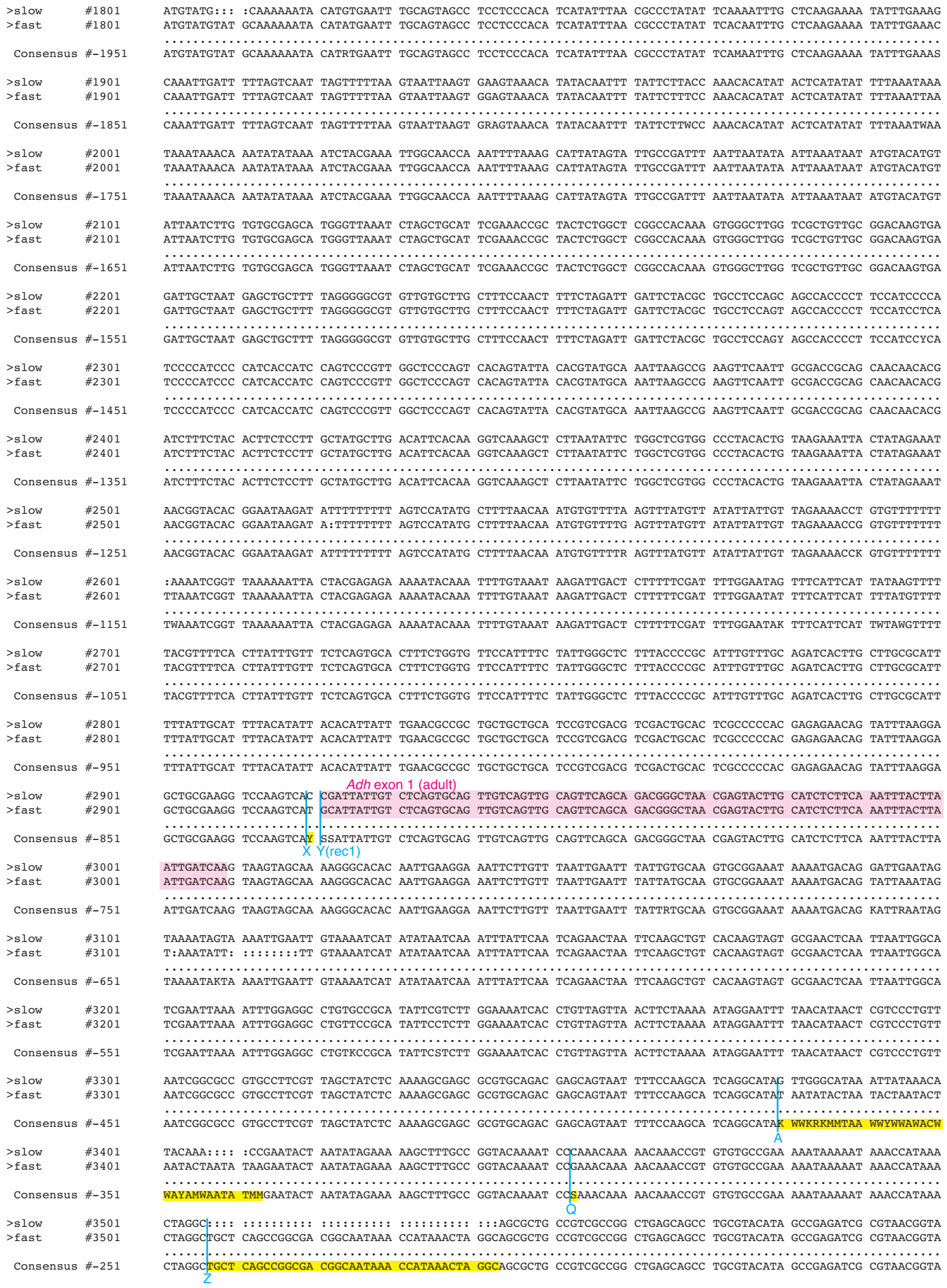


Figure S4. Precision map of *D. melanogaster* alleles. (A) Structure of *D. melanogaster Adh* locus and recombinant constructs used for mapping. The map is presented in different form in Figure S5. Single nucleotide polymorphism (SNP) and indel substitutions are marked below the gene model. Substitutions or groups of substitutions with significant effects are marked below by dots. (B) First round of recombination mapping. This experiment shows nucleotide scale mapping of coding region. Within the coding region, only the K192T amino acid substitution contributes. P values for sequential comparisons are shown. (C) Second round of recombination mapping, focusing on 5' regions. This is the same graph as Figure 5B. (D) A repeat measurement of transgenic lines varying in the 3'-UTR is unable to distinguish between candidate nucleotide substitutions. P values and DF are from sequential comparisons in (B) and (C) and all pairwise comparisons in (D).

```
>slow      #1         GCGAATAATC AAGACTCAGC ACCAATTTTT AGTTTATGGT CTAGTTCTTT ATAGGTTTTG TACTTCTTTT TTTCGCGTTG GCTATTTTGC GATTGAATTC
>fast      #1         GCGAATAATC AAGACTCAGC ACCAATTTTT AGTTTATGGT CTAGTTCTTT ATAGGTTTTG TACCTCTTTT TTTTGCGTTG GCTATTTTGC GATTGAATTC
                      ..........................................................................................................
 Consensus #-3751    GCGAATAATC AAGACTCAGC ACCAATTTTT AGTTTATGGT CTAGTTCTTT ATAGGTTTTG TACYTCTTTT TTTYGCGTTG GCTATTTTGC GATTGAATTC

>slow      #101       ATAAATATGG AATCAAATCT ATAGAGTGGA GAGTGGAACT AACGAGGTGA GAGGTAACAA TATAGTTTTT GGGCAATCAG AAGCAACAAA CAAATATCTG
>fast      #101       ATAAATATGG AATCAAATCT ATAGAGTGGA GAGTGGAACT AACGAGGTGA GAGGTAACAA TATAGTTTTT GGGCAATCAG AAGCAACAAA CAAATATCTG
                      ..........................................................................................................
 Consensus #-3651    ATAAATATGG AATCAAATCT ATAGAGTGGA GAGTGGAACT AACGAGGTGA GAGGTAACAA TATAGTTTTT GGGCAATCAG AAGCAACAAA CAAATATCTG

>slow      #201       CAATAACTCG TTGAATTCGA AACAAAATTA ACTGCATTTA TACTAAATAT ATAATTGCTA TAGGATGAGT TAGCCGTCTT GCGGTTTCCC AAACCCCAAA
>fast      #201       CAATAACTCG TTGAATTCGA AACAAAATTA ACTGCATTTA TACTAAATAT ATAATTGCTA TAGGATGAGT TAGCCGTCTT GCGGTTTCCC AAACCCCAAA
                      ..........................................................................................................
 Consensus #-3551    CAATAACTCG TTGAATTCGA AACAAAATTA ACTGCATTTA TACTAAATAT ATAATTGCTA TAGGATGAGT TAGCCGTCTT GCGGTTTCCC AAACCCCAAA

>slow      #301       AGCAAAGTCA AGCGTGTAGG AAACCTGATC AGATCGCGGG AAAGATTCTC TGCACTCAAT TACGTCAAAC CAGGTTGATT TCCTCCTTTT CGCTGTCGAG
>fast      #301       AGCAAAGTCA AGCGTGTAGG AAACCTGATC AGATCGCGGG AAAGATTCTC TGCACTCAAT TACGTCAAAC CAGGTTGATT TCCTCCTTTT CGCTGTCGAG
                      ..........................................................................................................
 Consensus #-3451    AGCAAAGTCA AGCGTGTAGG AAACCTGATC AGATCGCGGG AAAGATTCTC TGCACTCAAT TACGTCAAAC CAGGTTGATT TCCTCCTTTT CGCTGTCGAG

>slow      #401       AGATTGGCAA ATGGGTCAAA TGGGTCAGGC AGTGGAATAG TAAATTAGAT TATGTTTGCA TCGAGATGCA ATGCAAGCCG CGCCCCAAAT AAATGGAACG
>fast      #401       AGATTGGCAA ATGGGTCAAA TGGGTCAGGC AGTGGAATAG TAAATTAGAT TATGTTTGCA TCGAGATGCA ATGCAAGCCG CGCCCCAAAT AAATGGAACG
                      ..........................................................................................................
 Consensus #-3351    AGATTGGCAA ATGGGTCAAA TGGGTCAGGC AGTGGAATAG TAAATTAGAT TATGTTTGCA TCGAGATGCA ATGCAAGCCG CGCCCCAAAT AAATGGAACG

>slow      #501       TGCGCTAGTA GGGTTCCCCC TTGCCCCTGG TAACCCTTCG TTTCACCACC CGTTTTCCCG CTTTTCCGCT CCCAAACACT AGAGGTAAGC TGCTTAGACC
>fast      #501       TGCGCTAGTA GGGTTCCCCC TTGCCCCTGG TAACCCTTCC TTTCACCACC CGTTTTCCCG CTTTTCCGCT CCCAAACACT AGAGGTAAGC TGCTTAGACC
                      ..........................................................................................................
 Consensus #-3251    TGCGCTAGTA GGGTTCCCCC TTGCCCCTGG TAACCCTTCS TTTCACCACC CGTTTTCCCG CTTTTCCGCT CCCAAACACT AGAGGTAAGC TGCTTAGACC

>slow      #601       CCGGCGTTTA GAAGCCCCAG TTTCGTTTCA CTAGGCAGAC ACACTCGCAG CGGGAAGACA ATGCCACCGC CACCGCCACC GACTTAATCA GCCCGGGAAA
>fast      #601       CCGGCGTTTA GAAGCCCCAG TTTCGTTTCA CTAGGCAGAC ACACTCGCAG CGGGAAGACA ATGCCACCGC CACCGCCACC GACTTAATCA GCCCGGGAAA
                      ..........................................................................................................
 Consensus #-3151    CCGGCGTTTA GAAGCCCCAG TTTCGTTTCA CTAGGCAGAC ACACTCGCAG CGGGAAGACA ATGCCACCGC CACCGCCACC GACTTAATCA GCCCGGGAAA

>slow      #701       CGACATCTCA ATGCTGGCGA GCGTGTACCT ACATATGGAC ATGGGCGTGC GTTGGTGCGG GAGCTGGTGT AAATCGGTTT TGGCAGGTAC GCCGCTGGCG
>fast      #701       CGACATCTCA ATGCTGGCGA GTGTGTACCT ACATATGGAC ATGGGCGTGC GTTGGTGCGG GAGCTGGTGT AAATCGGTTT TGGCAGGTAC GCCGCTGGCG
                      ..........................................................................................................
 Consensus #-3051    CGACATCTCA ATGCTGGCGA GYGTGTACCT ACATATGGAC ATGGGCGTGC GTTGGTGCGG GAGCTGGTGT AAATCGGTTT TGGCAGGTAC GCCGCTGGCG

>slow      #801       TCATTA:CCC CCCAGAGGTT GAATGTCACC GGCGGCATGA CTTGGGGGCC AAGCCGATAA GGCGCACACT GTCCACTGCA CGGTGTACAC TGATAAAAAT
>fast      #801       TCATTACCCC CCCAGAGGTT GAATGTCACC GGCGGCATGA CTTGGGGGCC AAGCCGATAA GGCGCACACT GTCCACTGCA CGGTGTACAC TGATAAAAAT
                      ..........................................................................................................
 Consensus #-2951    TCATTACCCC CCCAGAGGTT GAATGTCACC GGCGGCATGA CTTGGGGGCC AAGCCGATAA GGCGCACACT GTCCACTGCA CGGTGTACAC TGATAAAAAT

>slow      #901       ATATATCAAG ACCAAATATT GTTAAAGATA ATTGATGCGT AAAGGAAATA CACTTGCAAG TTAAAATGTT TTCACCTTAA TGTGTTTTTC TTTT:AATAC
>fast      #901       ATATATCAAG ACCAAATATT GTTAAAGATA ATTGATGTGT AAAGGAAATA CACTTGCAAG TTAAAATGTT TTCACCTTAA TGTGTTTTTC TTTTAAATAC
                      ..........................................................................................................
 Consensus #-2851    ATATATCAAG ACCAAATATT GTTAAAGATA ATTGATGYGT AAAGGAAATA CACTTGCAAG TTAAAATGTT TTCACCTTAA TGTGTTTTTC TTTTAAATAC

>slow      #1001      TCTATTAACT AATATAAATT ATCACCAAAA CAAAACATTA ATTTGGGAAA TGTTATCACC AAAAGCTTTT GCCACTATAG AAAATACAGA TAAATCTAAA
>fast      #1001      TCTATTAACT AATATAAATT ATCACCAAAA CAAAACATTA ATTTGGGAAA TGTTATCACC AAAAGCTTTT GCCACTATAG AAAATACAGA TAAATCTAAA
                      ..........................................................................................................
 Consensus #-2751    TCTATTAACT AATATAAATT ATCACCAAAA CAAAACATTA ATTTGGGAAA TGTTATCACC AAAAGCTTTT GCCACTATAG AAAATACAGA TAAATCTAAA

>slow      #1101      AATAAATTCC TTTGACGCAT GCACGAAATA AGATAAACAA ATTTGATTTT ATTTTCTTAT TTACACAATT CATTTTATTT GCATGCATTT CATTTTGTTC
>fast      #1101      AATAAATTCC TTTGACGCAT GCACGAAATA AGATAAACAA ATTTGATTTT ATTTTCTTAT TTACACAATT CATTTTATTT GCATGCATTT CATTTTGTTC
                      ..........................................................................................................
 Consensus #-2651    AATAAATTCC TTTGACGCAT GCACGAAATA AGATAAACAA ATTTGATTTT ATTTTCTTAT TTACACAATT CATTTTATTT GCATGCATTT CATTTTGTTC

>slow      #1201      AGTGTACCTA ATAAAAACGA TTTCGTTTGC CCCAAGCAGT AAGAAGATGT TAGGCACGTC TGCTGATAAG GAAAACTGTA GCCCCAGACT AGGCCAGACC
>fast      #1201      AGTGTACCTA ATAAAAACGA TTTCGTTTGC CCCAAGCAGT AAGAAGATGT TAGGCACGTC TGCTGATAAG GAAAACTGTA GCCCCAGACT AGGCCAGACC
                      ..........................................................................................................
 Consensus #-2551    AGTGTACCTA ATAAAAACGA TTTCGTTTGC CCCAAGCAGT AAGAAGATGT TAGGCACGTC TGCTGATAAG GAAAACTGTA GCCCCAGACT AGGCCAGACC

>slow      #1301      ATATTAAATT AACGTCTGGA GGCGCGAACA GTCATACGAT TTTTTTTTTT ATATTACTTC GCGGTCAGTT GCCAAGGCAG GAGAGCAACC CGTTCGATTA
>fast      #1301      ATATTAAATT AACGTCTGGA GGCGCGAACA GTCATACGA: TGTTTTTTTT ATATTACTTC GCGGTCAGTT GCCAAGGCAG GAGAGCAACC CGTTTGATTA
                      ..........................................................................................................
 Consensus #-2451    ATATTAAATT AACGTCTGGA GGCGCGAACA GTCATACGAT TKTTTTTTTT ATATTACTTC GCGGTCAGTT GCCAAGGCAG GAGAGCAACC CGTTYGATTA

>slow      #1401      GTGGGTCAAT TTGGAAAATG AGTTATTGAC TCTGGGAAAT TGTTGAGCTG AAAATTTAAT CGGAGCCCGA AAATTTCCAA TCATGCATTC CCCAAGTGAC
>fast      #1401      GTGGGTCAAT TTGGAAAATG AGTTATTGAC TCTGGGAAAT TGTTGAGCTG AAAATTTAAT CGGAGCCCGA AAATTTCCAA TCATGCATTC CCCAAGTGAC
                      ..........................................................................................................
 Consensus #-2351    GTGGGTCAAT TTGGAAAATG AGTTATTGAC TCTGGGAAAT TGTTGAGCTG AAAATTTAAT CGGAGCCCGA AAATTTCCAA TCATGCATTC CCCAAGTGAC

>slow      #1501      CATATATGGA TTAGTGATAA CGCTCGATGC GACCCCCAAA GATTATCAAA AATATTTAAT ATGAATATAT GAAAAAAAGA TTTAACTTTT ATGAATTCTT
>fast      #1501      CATATATGGA TTAGTGATAA CGCTCGATGC GACCCCCAAA GATTATCAAA AATATTTAAT ATGAATATAT GAAAAAAAGA TTTAACTTTT ATGAATTCTT
                      ..........................................................................................................
 Consensus #-2251    CATATATGGA TTAGTGATAA CGCTCGATGC GACCCCCAAA GATTATCAAA AATATTTAAT ATGAATATAT GAAAAAAAGA TTTAACTTTT ATGAATTCTT

>slow      #1601      AAGCGTCCCC AAAGCTGCGG GAGAACTGGG CCATAT:::A CCCGAAATAC ATGTTTATAC TTTAGCAAAT GTATTTTCCA ATTAGGTGAT AGAACTTGTG
>fast      #1601      AAGCGTCCCC AAAGCTGCGG GAGAACTGGG CCATATATGA CCCGAAATAC ATGTTTATAC TTTAGCAAAT GTATTTTCCA ATTAGGTGAT AGAACTTGTG
                      ..........................................................................................................
 Consensus #-2151    AAGCGTCCCC AAAGCTGCGG GAGAACTGGG CCATATATGA CCCGAAATAC ATGTTTATAC TTTAGCAAAT GTATTTTCCA ATTAGGTGAT AGAACTTGTG

>slow      #1701      TGCACACACA CATATAGTTC TATATCAACA AACAGGTTTA AGTTTTATGC AAATTGAAAG CTTATTTCTT CCGCATGCTT ATCTCTTTCC TTCTCATCAT
>fast      #1701      TGCACACACA CATATAGTTC TATATCAACA AACAGGTTTA AGTTTTATGC AAATTGAAAG CTTATTTCTT CCGCATGCTT ATCTCTTTCC TTCTCATCAT
                      ..........................................................................................................
 Consensus #-2051    TGCACACACA CATATAGTTC TATATCAACA AACAGGTTTA AGTTTTATGC AAATTGAAAG CTTATTTCTT CCGCATGCTT ATCTCTTTCC TTCTCATCAT
```

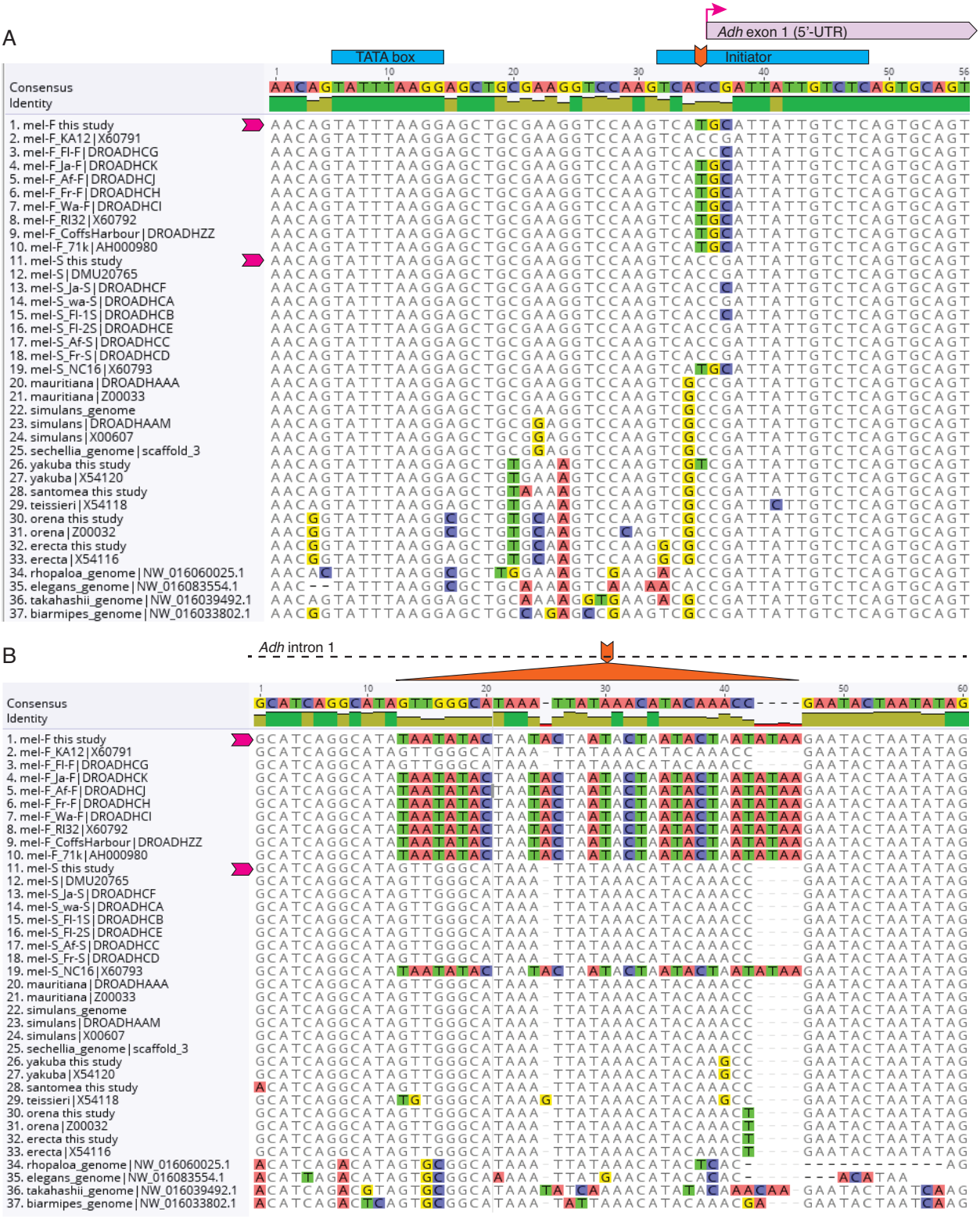

Adh exon 1 (adult)
X
Y(rec1)
A
Q
Z

```
>slow      #3601      GATAATGAAA AGCTCTACGT AACCGAAGCT TCTGCTGTAC GGATCTTCCT ATAAATACGG GGCCGACACG AACTGGAAAC CAACAACTAA CGGAGCCCTC
>fast      #3601      GATAATGAAA AGCTCTACGT AACCGAAGCT TCTGCTGTAC GGATCTTCCT ATAAATACGG GGCCGACACG AACTGGAAAC CAACAACTAA CGGAGCCCTC
                      ..........................................................................................................
 Consensus #-151     GATAATGAAA AGCTCTACGT AACCGAAGCT TCTGCTGTAC GGATCTTCCT ATAAATACGG GGCCGACACG AACTGGAAAC CAACAACTAA CGGAGCCCTC
                                 Adh exon 2                                          Start
>slow      #3701      TTCCAATTGA AACAGATCGA AAGAGCCTGC TAAAGCAAAA AAGAAGTCAC CATGTCGTTT ACTTTGACCA ACAAGAACGT GATTTTCGTT GCCGGTCTGG
>fast      #3701      TTCCAATTGA AACAGATCGA AAGAGCCTGC TAAAGCAAAA AAGAAGTCAC CATGTCGTTT ACTTTGACCA ACAAGAACGT GATTTTCGTG GCCGGTCTGG
                      ..........................................................................................................
 Consensus #-51      TTCCAATTGA AACAGATCGA AAGAGCCTGC TAAAGCAAAA AAGAAGTCAC CATGTCGTTT ACTTTGACCA ACAAGAACGT GATTTTCGTK GCCGGTCTGG
                                                                                                                  S (rec2)
>slow      #3801      GAGGCATTGG TCTGGACACC AGCAAGGAGC TGCTCAAGCG CGATCTGAAG GTAACTATGC GATGCCCACA GGCTCCATGC AGCGATGGAG GTTAATCTTG
>fast      #3801      GAGGCATTGG TCTGGACACC AGCAAGGAGC TGCTCAAGCG CGATCTGAAG GTAACTATGC GATGCCCACA GGCTCCATGC AGCGATGGAG GTTAATCTCG
                      ..........................................................................................................
 Consensus #50       GAGGCATTGG TCTGGACACC AGCAAGGAGC TGCTCAAGCG CGATCTGAAG GTAACTATGC GATGCCCACA GGCTCCATGC AGCGATGGAG GTTAATCTYG
                                 Adh exon 3
>slow      #3901      TGTATTCAAT CCTAGAACCT GGTGATCCTC GACCGCATTG AGAACCCGGC TGCCATTGCC GAGCTGAAGG CAATCAATCC AAAGGTGACC GTCACCTTCT
>fast      #3901      TGTATTCAAT CCTAGAACCT GGTGATCCTC GACCGCATTG AGAACCCGGC TGCCATTGCC GAGCTGAAGG CAATCAATCC AAAGGTGACC GTCACCTTCT
                      ..........................................................................................................
 Consensus #150      TGTATTCAAT CCTAGAACCT GGTGATCCTC GACCGCATTG AGAACCCGGC TGCCATTGCC GAGCTGAAGG CAATCAATCC AAAGGTGACC GTCACCTTCT

>slow      #4001      ACCCCTATGA TGTGACCGTG CCCATTGCCG AGACCACCAA GCTGCTGAAG ACCATCTTCG CCCAGCTGAA GACCGTCGAT GTCCTGATCA ACGGAGCTGG
>fast      #4001      ACCCCTATGA TGTGACCGTG CCCATTGCCG AGACCACCAA GCTGCTGAAG ACCATCTTCG CCCAGCTGAA GACCGTCGAT GTCCTGATCA ACGGAGCTGG
                      ..........................................................................................................
 Consensus #250      ACCCCTATGA TGTGACCGTG CCCATTGCCG AGACCACCAA GCTGCTGAAG ACCATCTTCG CCCAGCTGAA GACCGTCGAT GTCCTGATCA ACGGAGCTGG

>slow      #4101      TATCCTGGAC GATCACCAGA TCGAGCGCAC CATTGCCGTC AACTACACTG GCCTGGTCAA CACCACGACG GCCATTCTGG ACTTCTGGGA CAAGCGCAAG
>fast      #4101      TATCCTGGAC GATCACCAGA TCGAGCGCAC CATTGCCGTC AACTACACTG GCCTGGTCAA CACCACGACG GCCATTCTGG ACTTCTGGGA CAAGCGCAAG
                      ..........................................................................................................
 Consensus #350      TATCCTGGAC GATCACCAGA TCGAGCGCAC CATTGCCGTC AACTACACTG GCCTGGTCAA CACCACGACG GCCATTCTGG ACTTCTGGGA CAAGCGCAAG

>slow      #4201      GGCGGTCCCG GTGGTATCAT CTGCAACATT GGATCCGTCA CTGGATTCAA TGCCATCTAC CAGGTGCCCG TCTACTCCGG CACCAAGGCC GCCGTGGTCA
>fast      #4201      GGCGGTCCCG GTGGTATCAT CTGCAACATT GGATCCGTCA CTGGATTCAA TGCCATCTAC CAGGTGCCCG TCTACTCCGG CACCAAGGCC GCCGTGGTCA
                      ..........................................................................................................
 Consensus #450      GGCGGTCCCG GTGGTATCAT CTGCAACATT GGATCCGTCA CTGGATTCAA TGCCATCTAC CAGGTGCCCG TCTACTCCGG CACCAAGGCC GCCGTGGTCA
                                                                                                                  Adh exon 4
>slow      #4301      ACTTCACCAG CTCCCTGGCG GTAAGTTGAT CAAAGGAAAC GCAAAGTTTT CAAGAAAAAA CAAAACTAAT TTGATTTATA ACACCTTTAG AAACTGGCCC
>fast      #4301      ACTTCACCAG CTCCCTGGCG GTAAGTTGAT CAAAGGAAAC GCAAAGTTTT CAAGAAAAAA CAAAACTAAT TTGATTTATA ACACCTTTAG AAACTGGCCC
                      ..........................................................................................................
 Consensus #550      ACTTCACCAG CTCCCTGGCG GTAAGTTGAT CAAAGGAAAC GCAAAGTTTT CAAGAAAAAA CAAAACTAAT TTGATTTATA ACACCTTTAG AAACTGGCCC
                                                                                  K192T
>slow      #4401      CCATTACCGG CGTGACGGCT TACACTGTGA ACCCCGGCAT CACCCGCACC ACCCTGGTGC ACAAGTTCAA CTCCTGGTTG GATGTTGAGC CTCAGGTTGC
>fast      #4401      CCATTACCGG CGTGACGGCT TACACTGTGA ACCCCGGCAT CACCCGCACC ACCCTGGTGC ACACGTTCAA CTCCTGGTTG GATGTTGAGC CTCAGGTTGC
                      ..........................................................................................................
 Consensus #650      CCATTACCGG CGTGACSGCT TACACTGTGA ACCCCGGCAT CACCCGCACC ACCCTGGTGC ACAMGTTCAA CTCCTGGTTG GATGTTGAGC CTCAGGTTGC
                                E                                                                 U
>slow      #4501      CGAGAAGCTC CTGGCTCATC CCACCCAGCC CTCGTTGGCC TGCGCCGAGA ACTTCGTCAA GGCTATCGAG CTGAACCAGA ACGGAGCCAT CTGGAAACTG
>fast      #4501      CGAGAAGCTC CTGGCTCATC CCACCCAGCC CTCGTTGGCC TGCGCCGAGA ACTTCGTCAA GGCTATCGAG CTGAACCAGA ACGGAGCCAT CTGGAAACTG
                      ..........................................................................................................
 Consensus #750      CGAGAAGCTC CTGGCTCATC CCACCCAGCC CTCGTTGGCC TGCGCCGAGA ACTTCGTCAA GGCTATCGAG CTGAACCAGA ACGGAGCCAT CTGGAAACTG
                                                                                  Stop
>slow      #4601      GACTTGGGCA CCCTGGAGGC CATCCAGTGG ACCAAGCACT GGGACTCCGG CATCTAAGAA GTGATACTCC CAAAAAAAAA AAAAACATA ACATTAGTTC
>fast      #4601      GACTTGGGCA CCCTGGAGGC CATCCAGTGG ACCAAGCACT GGGACTCCGG CATCTAAGAA GTGATACTCC CAAAAAAAAA AAAAACATA ACATTAGTTC
                      ..........................................................................................................
 Consensus #850      GACTTGGGCA CCCTGGAGGC CATCCAGTGG ACCAAGCACT GGGACTCCGG CATCTAAGAA GTGATACTCC CAAAAAAAAA AAAAACATA ACATTAGTTC
                                                                                                                  G (rec3)
>slow      #4701      ATAGGGTTCT GCGAACCACA AAGATATTCA CGCAAGGCAA TAAGGCTGAT TCGATGCACA CTCACATTCT TCTCCTAATA CGATAATAAA ACTTTCCATG
>fast      #4701      ATAGGGTTCT GCGAACCAG: AAGATATTCA CGCAAGGCAA TAAGGCTGAT TCGATGCACA CTCACATTCT TCTCCTAATA CGATAATAAA ACTTTCCATG
                      ..........................................................................................................
 Consensus #950      ATAGGGTTCT GCGAACCASA AAGATATTCA CGCAAGGCAA TAAGGCTGAT TCGATGCACA CTCACATTCT TCTCCTAATA CGATAATAAA ACTTTCCATG
                                V
>slow      #4801      AAAAATATGG AAAAATATAT GAAAATTGAG AAATCCAAAA AACTGATAAA CGCTCTACTT AATTAAAATA GATAAATGGG AGCGGCAGGA ATGGCGGAGC
>fast      #4801      AAAAATATGG AAAAATATAT GAAAATTGAG AAATCCAAAA AACTGATAAA CGCTCTACTT AATTAAAATA GATAAATGGG AGCGGCAGGA ATGGCGGAGC
                      ..........................................................................................................
 Consensus #1050     AAAAATATGG AAAAATATAT GAAAATTGAG AAATCCAAAA AACTGATAAA CGCTCTACTT AATTAAAATA GATAAATGGG AGCGGCAGGA ATGGCGGAGC

>slow      #4901      ATGGCCAAGT TCCTCCGCCA ATCAGTCGTA AAACAGAAGT CGTGGAAAGC GGATAGAAAG AATGTTCGAT TTGACGGGCA AGCATGTCTG CTATGTGGCG
>fast      #4901      ATGGCCAAGT TCCTCCGCCA ATCAGTCGTA AAACAGAAGT CGTGGAAAGC GGATAGAAAG AATGTTCGAT TTGACGGGCA AGCATGTCTG CTATGTGGCG
                      ..........................................................................................................
 Consensus #1150     ATGGCCAAGT TCCTCCGCCA ATCAGTCGTA AAACAGAAGT CGTGGAAAGC GGATAGAAAG AATGTTCGAT TTGACGGGCA AGCATGTCTG CTATGTGGCG

>slow      #5001      GATTGCGGAG GAATTGCACT GGAGACCAGC AAGGTTCTCA TGACCAAGAA TATAGCGGTG AGTGAGCGGG AAGCTCGGTT TCTGTCCAGA TCGAACTCAA
>fast      #5001      GATTGCGGAG GAATTGCACT GGAGACCAGC AAGGTTCTCA TGACCAAGAA TATAGCGGTG AGTGAGCGGG AAGCTCGGTT TCTGTCCAGA TCGAACTCAA
                      ..........................................................................................................
 Consensus #1250     GATTGCGGAG GAATTGCACT GGAGACCAGC AAGGTTCTCA TGACCAAGAA TATAGCGGTG AGTGAGCGGG AAGCTCGGTT TCTGTCCAGA TCGAACTCAA

>slow      #5101      AACTAGTCCA GCCAGTCGCT GTCGAAACTA ATTAAGTTAA TGAGTTTTTC ATGTTAGTTT CGCGCTGAGC AACAATTAAG TTTATGTTTC AGTTCGGCTT
>fast      #5101      AACTAGTCCA GCCAGTCGCT GTCGAAACTA ATTAAGTTAA TGAGTTTTTC ATGTTAGTTT CGCGCTGAGC AACAATTAAG TTTATGTTTC AGTTCGGCTT
                      ..........................................................................................................
 Consensus #1350     AACTAGTCCA GCCAGTCGCT GTCGAAACTA ATTAAGTTAA TGAGTTTTTC ATGTTAGTTT CGCGCTGAGC AACAATTAAG TTTATGTTTC AGTTCGGCTT

>slow      #5201      AGATTTCGCT GAAGGACTTG CCACTTTCAA TCAATACTTT AGAACAAAAT CAAAACTCAT TCTAATAGCT TGGTGTTCAT CTTTTTTTTT AATGATAAGC
>fast      #5201      AGATTTCGCT GAAGGACTTG CCACTTTCAA TCAATACTTT AGAACAAAAT CAAAACTCAT TCTAATAGCT TGGTGTTCAT CTTTTTTTTT AATGATAAGC
                      ..........................................................................................................
 Consensus #1450     AGATTTCGCT GAAGGACTTG CCACTTTCAA TCAATACTTT AGAACAAAAT CAAAACTCAT TCTAATAGCT TGGTGTTCAT CTTTTTTTTT AATGATAAGC
```

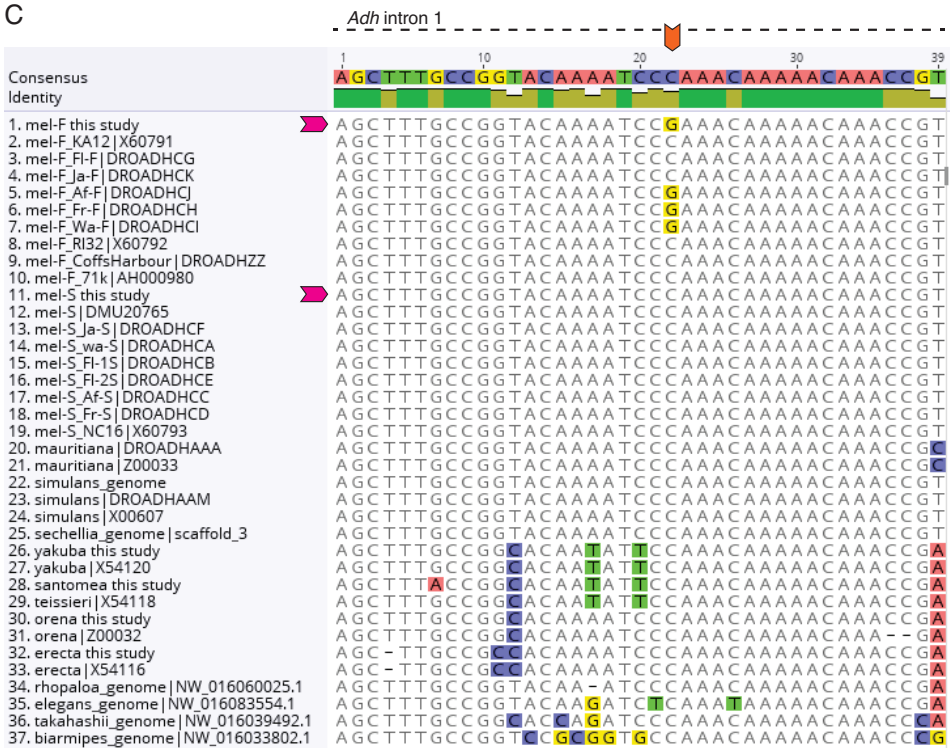
```
>slow      #5301      ATTTTGTCGT TTATACTTTT TATATTTCGA TATTAAACCA CCTATGAAGT CTATTTTAA: :::::::::: :::::::::: ::::::ATAT TTGTATTCTT
>fast      #5301      ATTTTGTCGT TTATACTTTT TATATTTCGA TATTAAACCA CCTATGAAGT CTATTTTAAT CGCCAGATAA GCAATATATT GTGTAAATAT TTGTATTCTT
                      ..................................................................................................
 Consensus #1550      ATTTTGTCGT TTATACTTTT TATATTTCGA TATTAAACCA CCTATGAAGT CTATTTTAAT CGCCAGATAA GCAATATATT GTGTAAATAT TTGTATTCTT
                                                                                    W (rec4)
>slow      #5401      TATCAGGAAA TTCAGGGAGA CGGGGAAGTT ACTATCTACT AAAAGCCAAA CAATTTCTTA CAGTTTTACT CTCTCTACTC TAGAAACTGG CCATTTTACA
>fast      #5401      TATCAGGAAA TTCAGGGAGA CGGGGAAGTT ACTATCTACT AAAAGCCAAA CAATTTCTTA CAGTTTTACT CTCTCTACTC TAGAAACTGG CCATTTTACA
                      ..................................................................................................
 Consensus #1650      TATCAGGAAA TTCAGGGAGA CGGGGAAGTT ACTATCTACT AAAAGCCAAA CAATTTCTTA CAGTTTTACT CTCTCTACTC TAGAAACTGG CCATTTTACA

>slow      #5501      GAGTACGGAA AATCCCCAGG CCATCGCTCA GTTGCAGTCG ATAAAGCCGA GTACCCAAAT ATTTTTCTGG ACCTACGACG TGACCATGGC AAGGGAAGAT
>fast      #5501      GAGTACGGAA AATCCCCAGG CCATCGCTCA GTTGCAGTCG ATAAAGCCGA GTACCCAAAT ATTTTTCTGG ACCTACGACG TGACCATGGC AAGGGAAGAT
                      ..................................................................................................
 Consensus #1750      GAGTACGGAA AATCCCCAGG CCATCGCTCA GTTGCAGTCG ATAAAGCCGA GTACCCAAAT ATTTTTCTGG ACCTACGACG TGACCATGGC AAGGGAAGAT

>slow      #5601      ATGAAGAAGT ACTTCGATGA GGTGATGGTC CAAATGGACT ACATCGATGT CCTGATCAAT GGTGCTACGC TGTGCGATGA AAATAACATT GATGCCACCA
>fast      #5601      ATGAAGAAGT ACTTCGATGA GGTGATGGTC CAAATGGACT ACATCGATGT CCTGATCAAT GGTGCTACGC TGTGCGATGA AAATAACATT GATGCCACCA
                      ..................................................................................................
 Consensus #1850      ATGAAGAAGT ACTTCGATGA GGTGATGGTC CAAATGGACT ACATCGATGT CCTGATCAAT GGTGCTACGC TGTGCGATGA AAATAACATT GATGCCACCA

>slow      #5701      TCAATACAAA TCTAACGGGA ATGATGAACA CTGTGGCCAC AGTGTTACCC TATATGGACA GAAAAATGGG AGGAACTGGT GGGCTAATTG TGAACGTCAC
>fast      #5701      TCAATACAAA TCTAACGGGA ATGATGAACA CTGTGGCCAC AGTGTTACCC TATATGGACA GAAAAATGGG AGGAACTGGT GGGCTAATTG TGAACGTCAC
                      ..................................................................................................
 Consensus #1950      TCAATACAAA TCTAACGGGA ATGATGAACA CTGTGGCCAC AGTGTTACCC TATATGGACA GAAAAATGGG AGGAACTGGT GGGCTAATTG TGAACGTCAC

>slow      #5801      TTCGGTCATT GGATTGGACC CTTCGCCGGT TTTCTGCGCA TATAGTGCAT CCAAATTCGG TGTAATTGGA TTTACCAGAA GTCTAGCGGT GAGTTGAATC
>fast      #5801      TTCGGTCATT GGATTGGACC CTTCGCCGGT TTTCTGCGCA TATAGTGCAT CCAAATTCGG TGTAATTGGA TTTACCAGAA GTCTAGCGGT GAGTTGAATA
                      ..................................................................................................
 Consensus #2050      TTCGGTCATT GGATTGGACC CTTCGCCGGT TTTCTGCGCA TATAGTGCAT CCAAATTCGG TGTAATTGGA TTTACCAGAA GTCTAGCGGT GAGTTGAATM
                                                                                                                        O
>slow      #5901      CAATTTTATG CGGATAAATT CTTTAATTTA TTGCTTGCAG GACCCTCTTT ACTATTCCCA AAACGGGGTA GCTGTGATGG CGGTTTGTTG TGGTCCTACA
>fast      #5901      CGATCTTATG CGGATAAATT CA:TAATTTT TTGGTTTCAG GACCCTCTTT ACTATTCCCA AAACGGGGTA GCTGTGATGG CGGTTTGTTG TGGTCCTACA
                      ..................................................................................................
 Consensus #2150      CRATYTTATG CGGATAAATT CWTTAATTTW TTGSTTKCAG GACCCTCTTT ACTATTCCCA AAACGGGGTA GCTGTGATGG CGGTTTGTTG TGGTCCTACA

>slow      #6001      AGGGTCTTTG TGGACCGGGA ACTGAAAGCG TTTTTGGAAT ACGGACAATC CTTTGCCGAT CGCCTGCGGC GAGCGCCCTG CCAATCGACA TCGGTTTGTG
>fast      #6001      AGGGTCTTTG TGGACCGGGA ACTGAAAGCG TTTTTGGAAT ACGGACAATC CTTTGCCGAT CGCCTGCGGC GAGCGCCCTG CCAATCGACA TCGGTTTGTG
                      ..................................................................................................
 Consensus #2250      AGGGTCTTTG TGGACCGGGA ACTGAAAGCG TTTTTGGAAT ACGGACAATC CTTTGCCGAT CGCCTGCGGC GAGCGCCCTG CCAATCGACA TCGGTTTGTG

>slow      #6101      GTCAGAATAT TGTCAATGCC ATCGAGAGAT CGGAGAATGG TCAGATATGG ATTGCGGATA AGGGTGGACT CGAGTTGGTC AAATTGCATT GGTACTGGCA
>fast      #6101      GTCAGAATAT TGTCAATGCC ATCGAGAGAT CGGAGAATGG TCAGATATGG ATTGCGGATA AGGGTGGACT CGAGTTGGTC AAATTGCATT GGTACTGGCA
                      ..................................................................................................
 Consensus #2350      GTCAGAATAT TGTCAATGCC ATCGAGAGAT CGGAGAATGG TCAGATATGG ATTGCGGATA AGGGTGGACT CGAGTTGGTC AAATTGCATT GGTACTGGCA

>slow      #6201      CATGGCCGAC CAGTTCGTGC ACTATATGCA GAGCAATGAT GAAGAGGATC AAGATTGAAT TCGAATCAAA TAAAATAATG CTTTACACAA AAAGTAGGCA
>fast      #6201      CATGGCCGAC CAGTTCGTGC ACTATATGCA GAGCAATGAT GAAGAGGATC AAGATTGAAT TCGAATCAAA TAAAATAATG CTTTACGCAA AAAGTAGGCA
                      ..................................................................................................
 Consensus #2450      CATGGCCGAC CAGTTCGTGC ACTATATGCA GAGCAATGAT GAAGAGGATC AAGATTGAAT TCGAATCAAA TAAAATAATG CTTTACRCAA AAAGTAGGCA

>slow      #6301      ATTCATTTTC CTATGATAAT AGATATGGGT CATCTATGGG GTGTGAAAGA GTAATGACAA AATTTGGTGT GCCCAAAAGT ATGCAGCGAA TGTTGATGGG
>fast      #6301      ATTCATTTTC CTATGATAAT AGATATGGGT CATCTATGGG GTGTGAAAGA GTAATGACAA AATTTGGTGT GCCCAAAAGT ATGCAGCGAA TGTTGATGGG
                      ..................................................................................................
 Consensus #2550      ATTCATTTTC CTATGATAAT AGATATGGGT CATCTATGGG GTGTGAAAGA GTAATGACAA AATTTGGTGT GCCCAAAAGT ATGCAGCGAA TGTTGATGGG

>slow      #6401      AGGTATAATT AGATGTGCTT AATTATGATG GGGTTACGTT ATGCATGTTG TGGGAATGTG AACTATACTG TTTTTTTTTT TTGACATCAG TCGACTATAT
>fast      #6401      AGCTATAATT AGATGTGCTT AATTATGATG GGGTTACGTT ATGCATGTTG TGGGAATGTG AACTATACTG TTTTTTTTTT ::GACATCAG TCGACTATAT
                      ..................................................................................................
 Consensus #2650      AGSTATAATT AGATGTGCTT AATTATGATG GGGTTACGTT ATGCATGTTG TGGGAATGTG AACTATACTG TTTTTTTTTT TTGACATCAG TCGACTATAT

>slow      #6501      GTTCTCCTAA CAACCTTTGA TCTAAATTGT AAGTGCAAAG TCAAGTCCAA CTACTCAAGG CTTTTTCAAC AACAATATAC GTAGACTTTA CGTGCAACAT
>fast      #6501      GTTCTCCTAA CAACCTTTGA TCTAAATTGT AAGTGCAAAG TCAAGTCCAA CTACTCAAGG CTTTTTCAAC AACAATATAC GTAGACTTTA CGTGCAACAT
                      ..................................................................................................
 Consensus #2750      GTTCTCCTAA CAACCTTTGA TCTAAATTGT AAGTGCAAAG TCAAGTCCAA CTACTCAAGG CTTTTTCAAC AACAATATAC GTAGACTTTA CGTGCAACAT

>slow      #6601      TTCTAGAACG TCTCTAGTGC ATCTCAACTA TGCGAATGCA TGTCGAAATT TTGCTAGAGC AATGCCGTGC ATTGAACTTT GGCACATTTC CAAACGCACA
>fast      #6601      TTCTAGAACG TCTCTAGTGC ATCTCAACTA TGCGAATGCA TGTCGAAATT TTGCTAGAGC AATGCCGTGC ATTGAACTTT GGCACATTTC CAAACGCACA
                      ..................................................................................................
 Consensus #2850      TTCTAGAACG TCTCTAGTGC ATCTCAACTA TGCGAATGCA TGTCGAAATT TTGCTAGAGC AATGCCGTGC ATTGAACTTT GGCACATTTC CAAACGCACA

>slow      #6701      GGCAGACCCA GTCAGCCATA GGAAACCCAA TCGGACGACC GAGAATCGAA TGCGAAAACA AACAAATCTG CCGACACATT GCCCAGGAAT ATTGGCACTG
>fast      #6701      GGCAGACCCA GTCAGCCATA GGAAACCCAA TCGGACGACC GAGAATCGAA TGCGAAAACA AACAAATCTG CCGACACATT GCCCAGGAAT ATTGGCACTG
                      ..................................................................................................
 Consensus #2950      GGCAGACCCA GTCAGCCATA GGAAACCCAA TCGGACGACC GAGAATCGAA TGCGAAAACA AACAAATCTG CCGACACATT GCCCAGGAAT ATTGGCACTG

>slow      #6801      AGAGCCTACA GATACACCAG GAGAAAAACA TAAAAATAAA TGCTGGGATA ATTTTGGTGT TTTCGGTACT AATCTAACGC AATAGCCCTT CTACGGTCGA
>fast      #6801      AGAGCCTACA GATACACCAG GAAAAAAA:A :AAAAATAAA TGCTGGGATA GTTTTGGTAT TTTCGGTACT AATCTAACGC AATAGCCCTT CTACGGTCGA
                      ..................................................................................................
 Consensus #3050      AGAGCCTACA GATACACCAG GARAAAAACA TAAAAATAAA TGCTGGGATA RTTTTGGTRT TTTCGGTACT AATCTAACGC AATAGCCCTT CTACGGTCGA

>slow      #6901      GTAAACTGAA AAACTCTTAA TGTTTTTATG TTGCATTCAT GAAGTAGTAT ACATTTTCTA CTTCCAATAT TGCCTTGGCT TTTTCTCATC GTGTAGCTAT
>fast      #6901      GTAAACTGAA AAACTCTTAA TATTTTTATG TTGCATTCAT TAAGTAGTAT ACATTTTCTA CTTCCAATCT TGCCTTGGCT TTTTCTCATC GTGTAGCTAT
                      ..................................................................................................
 Consensus #3150      GTAAACTGAA AAACTCTTAA TRTTTTTATG TTGCATTCAT KAAGTAGTAT ACATTTTCTA CTTCCAATMT TGCCTTGGCT TTTTCTCATC GTGTAGCTAT
```

```
>slow      #7001      ATCTCAGCTA TATGGTGGGG TGAGTTGTGG GCACGCCGCC TTGGCATACT TCCAATTCGA GTGCGTTGTC TTCCTTCTAG TCGGGGGCCA CTCGGTTCGT
>fast      #7001      ATCTCAGCTA TATGGTGGGG TGAGTTGTGG GCACGCCGCC TTGGCATACT TCCAATTCGA GTGCGTTGTC TTCCTTCTAG TCGGGGGCCA CTCGGTTCGT
                      ..........................................................................................................
 Consensus #3250      ATCTCAGCTA TATGGTGGGG TGAGTTGTGG GCACGCCGCC TTGGCATACT TCCAATTCGA GTGCGTTGTC TTCCTTCTAG TCGGGGGCCA CTCGGTTCGT

>slow      #7101      AAAAAATCAA TAGCCAGCAG ACCAAAAAGA GCCACAATAC CATCGGACAT TGAGGACCCA ACGCAACGGA ATGGAACCAA CGGAATCGCA TGGAGTGGAA
>fast      #7101      AAAAAATCAA TAGCCAGCAG ACCAAAAAGA GCCACAATAT CATCGGACAT TGAGGACCCA ACGCAACGGA ATGGAACCAA CGGAATCGCA TGGAGTGGAA
                      ..........................................................................................................
 Consensus #3350      AAAAAATCAA TAGCCAGCAG ACCAAAAAGA GCCACAATAY CATCGGACAT TGAGGACCCA ACGCAACGGA ATGGAACCAA CGGAATCGCA TGGAGTGGAA

>slow      #7201      TCTGGCCAGG CCATATCAAA GCATTTTTCC AGAGTTCCCA CTCCCTGCCG TCCGCCGGCG TTGCAAGTTT GCTTTCCTTT CGATGGCGCA ACTTAATTCC
>fast      #7201      TCTGGCCAGG CCATATCAAA GCATTTTTCC AGAGTTCCCA CTCCCTGCCG TCCGCCGGCG TTGCAAGTTT GCTTTCCTTT CGATGGCGCA ACTTAATTCC
                      ..........................................................................................................
 Consensus #3450      TCTGGCCAGG CCATATCAAA GCATTTTTCC AGAGTTCCCA CTCCCTGCCG TCCGCCGGCG TTGCAAGTTT GCTTTCCTTT CGATGGCGCA ACTTAATTCC

>slow      #7301      AAAAATCGAT TAGTCACTGT AAGCGAACCA TGCAATTGTC CAGCATTTTT CATTTATAAC TCCTTTTTTT CGTTTAATTT TTGCTGGTTT CCTCCCCTTT
>fast      #7301      AAAAATCGAT TAGTCACTGT AAGCGAACCA TGCAATTGTC CAGCATTTTT CATTTATAAC TCCTTTTTTT CGTTTAATTT TTGCTGGTTT CCTCCCCTTT
                      ..........................................................................................................
 Consensus #3550      AAAAATCGAT TAGTCACTGT AAGCGAACCA TGCAATTGTC CAGCATTTTT CATTTATAAC TCCTTTTTTT CGTTTAATTT TTGCTGGTTT CCTCCCCTTT

>slow      #7401      TCCACTTTTT TTTTT:GGCA CACGCTTGAT TAATTTGGTC AGCTGCCGGT TGGCAAGGAG GTGGGGGGGA GGACATTTAA AAATGCAGGA AGTAATCGAT
>fast      #7401      TCCACTTTTT TTTTTTGGCA CACGCTTGAT TAATTTGGTC AGCTGCCGGT TGGCAAGGAG GTGGGGGGGA GGACATTTAA AAATGCAGGA AGTAATCGAT
                      ..........................................................................................................
 Consensus #3650      TCCACTTTTT TTTTTTGGCA CACGCTTGAT TAATTTGGTC AGCTGCCGGT TGGCAAGGAG GTGGGGGGGA GGACATTTAA AAATGCAGGA AGTAATCGAT

>slow      #7501      TGGCTAGGGC ACAATAGCCA AAATGGGAGA GGCAGTGCAA TCGCAAAAAA TATTGGGGTA TGTTCGCCTG AAGGAAATGC CGGAAAATAT TGAAAATTAC
>fast      #7501      TGGCTAGGGC ACAATAGCCA AAATGGGAGA GGCAGTGCAA TCGCAAAAAA TATTGGGGTA TGTTCGCCTG AAGGAAATGC CGGAAAATAT TGAAAATTAC
                      ..........................................................................................................
 Consensus #3750      TGGCTAGGGC ACAATAGCCA AAATGGGAGA GGCAGTGCAA TCGCAAAAAA TATTGGGGTA TGTTCGCCTG AAGGAAATGC CGGAAAATAT TGAAAATTAC

>slow      #7601      GATACCAAGA TACATTGCAC GAAATAATAA ATATTCTGGG CCGGTAATTG ACCAATTTTG TTTGTCCATT TGATTGTAGG TGTACTCGTA AATGGTAGGT
>fast      #7601      GATACCAAGA TACATTGCAC GAAATAATAA ATATTCTGGG CCGGTAATTG ACCAATTTTG TTTGTCCATT TGATTGTAGG TGTACTCGTA AATGGTAGGT
                      ..........................................................................................................
 Consensus #3850      GATACCAAGA TACATTGCAC GAAATAATAA ATATTCTGGG CCGGTAATTG ACCAATTTTG TTTGTCCATT TGATTGTAGG TGTACTCGTA AATGGTAGGT

>slow      #7701      TATATTAAAT GACTAATAAT TATATTGAAT GCAACTATTC ATACTAAGAG TATCTAGGCT TCGACTAGTT CTAGTATTTT CCCCAACACT GTAGATCGAA
>fast      #7701      TATATTAAAT GACTAATAAT TATATTGAAT GCAACTATTC ATACTAAGAG TATCTAGGCT TCGACTAGTT CTAGTATTTT CCCCAACACT GTAGATCGAA
                      ..........................................................................................................
 Consensus #3950      TATATTAAAT GACTAATAAT TATATTGAAT GCAACTATTC ATACTAAGAG TATCTAGGCT TCGACTAGTT CTAGTATTTT CCCCAACACT GTAGATCGAA

>slow      #7801      CAAAAACATG CGACCGACCG GTTGCCGGCA GTAGGTGGAG ATTAATCAAC AAAAAAAAAT TTGCGGAAAA ATGGAGGAAA AAATTACTGA TGCTTTTTAA
>fast      #7801      CAAAAACATG CGACCGACCA GTTGCCGGCA GTAGGTGGAG ATTAATCAAC AAAAAAAAAT TTGCGGAAAA ATGGAGGAAA AAATTACTGA TGCTTTTTAA
                      ..........................................................................................................
 Consensus #4050      CAAAAACATG CGACCGACCR GTTGCCGGCA GTAGGTGGAG ATTAATCAAC AAAAAAAAAT TTGCGGAAAA ATGGAGGAAA AAATTACTGA TGCTTTTTAA

>slow      #7901      ATGGCAATAC GTGCCTGGTC ACCA
>fast      #7901      ATGGCAATAC GTGCCTGGTC ACCA
                      ........................
 Consensus #4150      ATGGCAATAC GTGCCTGGTC ACCA
```

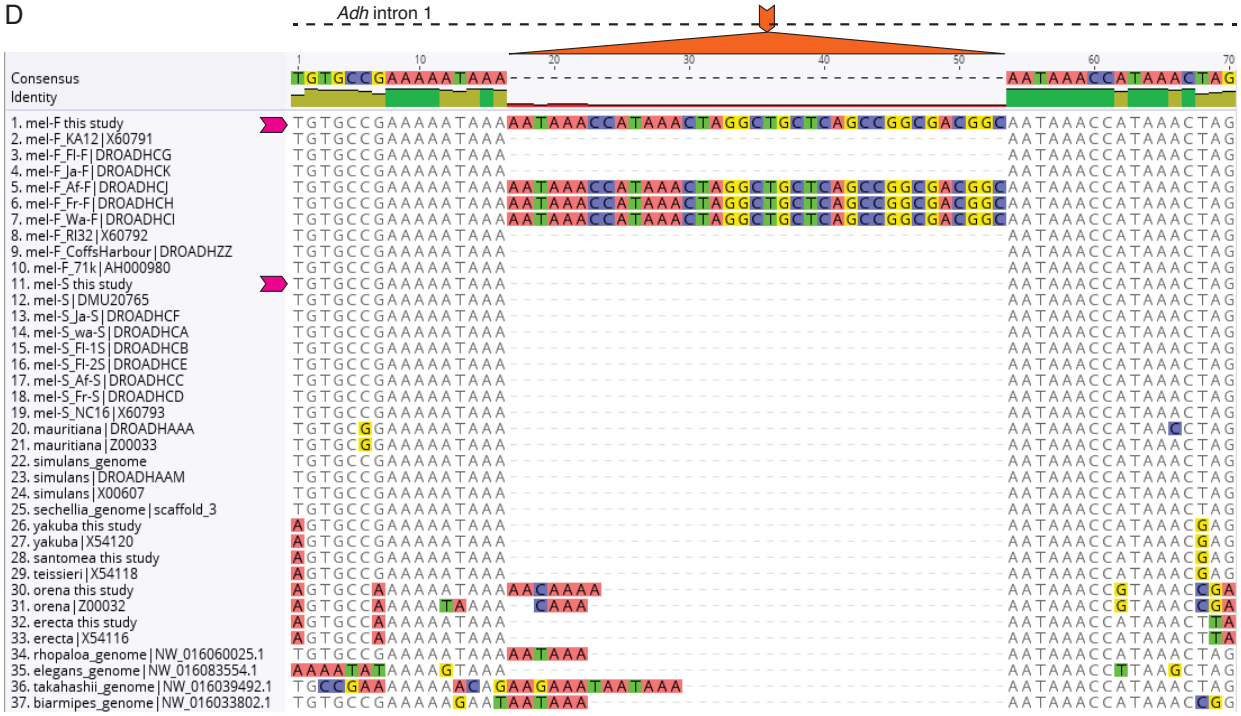
Figure S5. Pairwise sequence alignment of *D. melanogaster slow* and *fast* alleles, showing recombinant mapping positions. This is an alternate view of Figure S4A. Cyan lines mark the right-most boundary of recombination breakpoints. Capital letters denote the recombinant strain name. *Adh* adult transcript structure is outlined in magenta. The "fast/slow" K192T amino acid change is marked in green. Yellow highlights causative substitutions; light yellow highlights substitutions in the 3'-UTR where the specific causative substitution was not resolved (Figure S4 and SI text).


A
Adh exon 1 (5'-UTR)
TATA box
Initiator
Consensus
Identity
1. mel-F this study
2. mel-F_KA12|X60791
3. mel-F_FI-F|DROADHCG
4. mel-F_Ja-F|DROADHCK
5. mel-F_Af-F|DROADHCJ
6. mel-F_Fr-F|DROADHCH
7. mel-F_Wa-F|DROADHCI
8. mel-F_RI32|X60792
9. mel-F_CoffsHarbour|DROADHZZ
10. mel-F_71k|AH000980
11. mel-S this study
12. mel-S|DMU20765
13. mel-S_Ja-S|DROADHCF
14. mel-S_wa-S|DROADHCA
15. mel-S_FI-1S|DROADHCB
16. mel-S_FI-2S|DROADHCE
17. mel-S_Af-S|DROADHCC
18. mel-S_Fr-S|DROADHCD
19. mel-S_NC16|X60793
20. mauritiana|DROADHAAA
21. mauritiana|Z00033
22. simulans_genome
23. simulans|DROADHAAM
24. simulans|X00607
25. sechellia_genome|scaffold_3
26. yakuba this study
27. yakuba|X54120
28. santomea this study
29. teissieri|X54118
30. orena this study
31. orena|Z00032
32. erecta this study
33. erecta|X54116
34. rhopaloa_genome|NW_016060025.1
35. elegans_genome|NW_016083554.1
36. takahashii_genome|NW_016039492.1
37. biarmipes_genome|NW_016033802.1
B
Adh intron 1
Consensus
Identity
1. mel-F this study
2. mel-F_KA12|X60791
3. mel-F_FI-F|DROADHCG
4. mel-F_Ja-F|DROADHCK
5. mel-F_Af-F|DROADHCJ
6. mel-F_Fr-F|DROADHCH
7. mel-F_Wa-F|DROADHCI
8. mel-F_RI32|X60792
9. mel-F_CoffsHarbour|DROADHZZ
10. mel-F_71k|AH000980
11. mel-S this study
12. mel-S|DMU20765
13. mel-S_Ja-S|DROADHCF
14. mel-S_wa-S|DROADHCA
15. mel-S_FI-1S|DROADHCB
16. mel-S_FI-2S|DROADHCE
17. mel-S_Af-S|DROADHCC
18. mel-S_Fr-S|DROADHCD
19. mel-S_NC16|X60793
20. mauritiana|DROADHAAA
21. mauritiana|Z00033
22. simulans_genome
23. simulans|DROADHAAM
24. simulans|X00607
25. sechellia_genome|scaffold_3
26. yakuba this study
27. yakuba|X54120
28. santomea this study
29. teissieri|X54118
30. orena this study
31. orena|Z00032
32. erecta this study
33. erecta|X54116
34. rhopaloa_genome|NW_016060025.1
35. elegans_genome|NW_016083554.1
36. takahashii_genome|NW_016039492.1
37. biarmipes_genome|NW_016033802.1

C


Adh intron 1
Consensus
Identity
1. mel-F this study
2. mel-F_KA12|X60791
3. mel-F_FI-F|DROADHCG
4. mel-F_Ja-F|DROADHCK
5. mel-F_Af-F|DROADHCJ
6. mel-F_Fr-F|DROADHCH
7. mel-F_Wa-F|DROADHCI
8. mel-F_RI32|X60792
9. mel-F_CoffsHarbour|DROADHZZ
10. mel-F_71k|AH000980
11. mel-S this study
12. mel-S|DMU20765
13. mel-S_Ja-S|DROADHCF
14. mel-S_wa-S|DROADHCA
15. mel-S_FI-1S|DROADHCB
16. mel-S_FI-2S|DROADHCE
17. mel-S_Af-S|DROADHCC
18. mel-S_Fr-S|DROADHCD
19. mel-S_NC16|X60793
20. mauritiana|DROADHAAA
21. mauritiana|Z00033
22. simulans_genome
23. simulans|DROADHAAM
24. simulans|X00607
25. sechellia_genome|scaffold_3
26. yakuba this study
27. yakuba|X54120
28. santomea this study
29. teissieri|X54118
30. orena this study
31. orena|Z00032
32. erecta this study
33. erecta|X54116
34. rhopaloa_genome|NW_016060025.1
35. elegans_genome|NW_016083554.1
36. takahashii_genome|NW_016039492.1
37. biarmipes_genome|NW_016033802.1

D


Adh intron 1
Consensus
Identity
1. mel-F this study
2. mel-F_KA12|X60791
3. mel-F_FI-F|DROADHCG
4. mel-F_Ja-F|DROADHCK
5. mel-F_Af-F|DROADHCJ
6. mel-F_Fr-F|DROADHCH
7. mel-F_Wa-F|DROADHCI
8. mel-F_RI32|X60792
9. mel-F_CoffsHarbour|DROADHZZ
10. mel-F_71k|AH000980
11. mel-S this study
12. mel-S|DMU20765
13. mel-S_Ja-S|DROADHCF
14. mel-S_wa-S|DROADHCA
15. mel-S_FI-1S|DROADHCB
16. mel-S_FI-2S|DROADHCE
17. mel-S_Af-S|DROADHCC
18. mel-S_Fr-S|DROADHCD
19. mel-S_NC16|X60793
20. mauritiana|DROADHAAA
21. mauritiana|Z00033
22. simulans_genome
23. simulans|DROADHAAM
24. simulans|X00607
25. sechellia_genome|scaffold_3
26. yakuba this study
27. yakuba|X54120
28. santomea this study
29. teissieri|X54118
30. orena this study
31. orena|Z00032
32. erecta this study
33. erecta|X54116
34. rhopaloa_genome|NW_016060025.1
35. elegans_genome|NW_016083554.1
36. takahashii_genome|NW_016039492.1
37. biarmipes_genome|NW_016033802.1

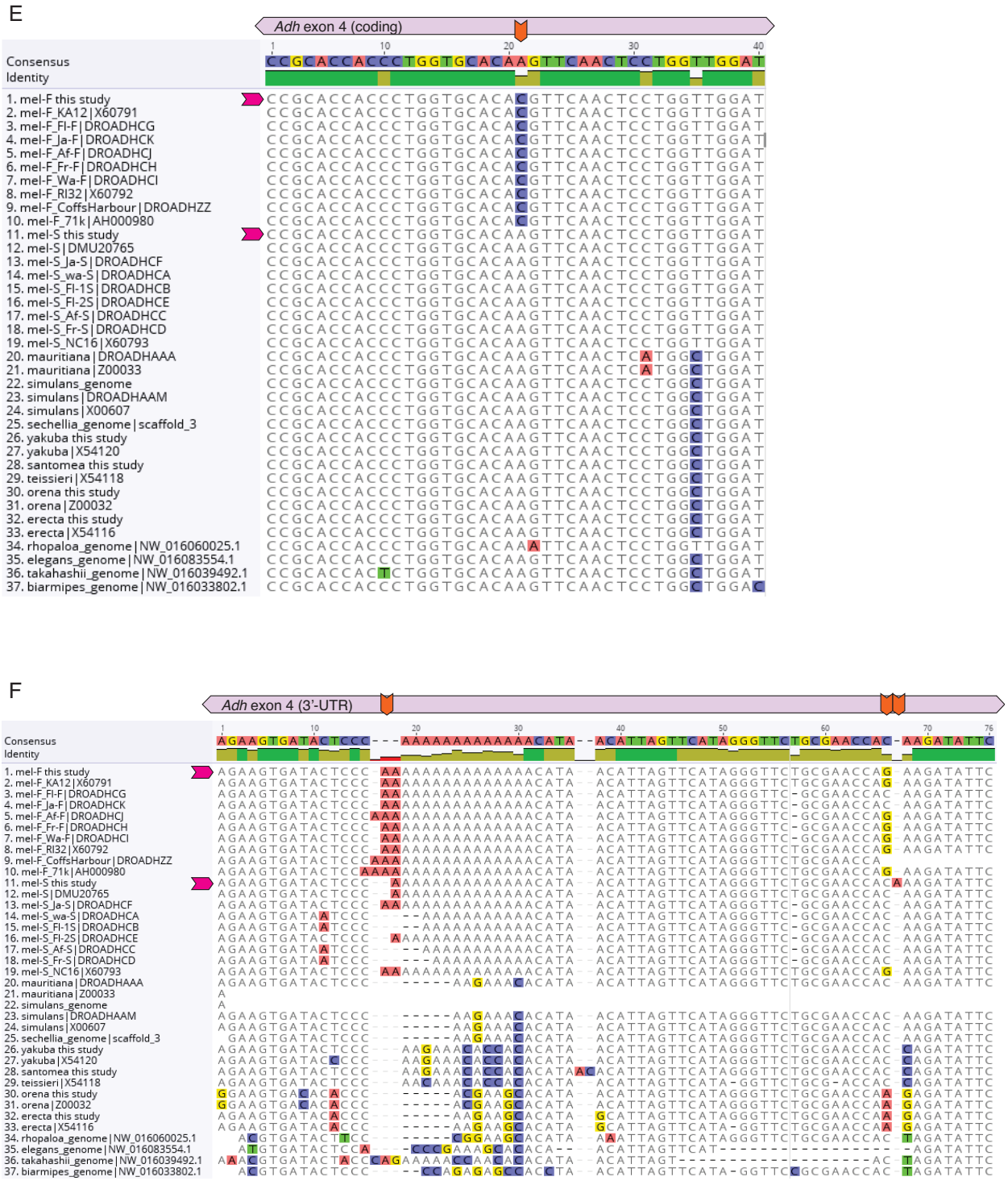


Figure S6. Evolution of causative sites between *D. melanogaster* alleles. For each site, the higher-activity allele (determined in Figure 5 and Figure S4) appears to be derived. The *melanogaster* haplotypes are named "F" or "S" based on amino acid 192 (panel E). Note that linkage disequilibrium among causative sites is partial, such that some *melanogaster* "F" haplotypes have low-activity alleles at other sites and vice versa. Orange arrowheads mark the causative site(s). Magenta arrowheads mark the two *melanogaster* alleles from this study where causative sites were mapped. Initiator and TATA box positions (panel A) were empirically

determined in reference (18). *Adh* sequences were retrieved from Genbank (accessions given after |) or Flybase (species labeled "genome") and aligned using Geneious Prime.

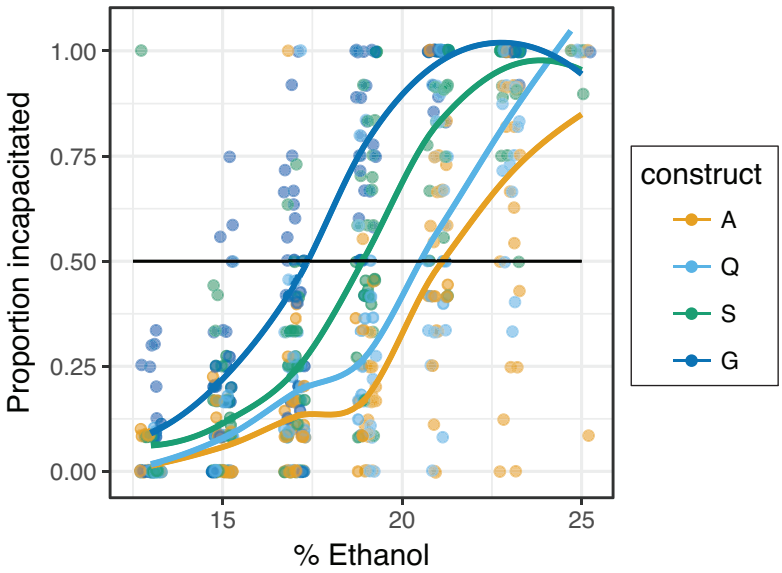


Figure S7. Flies show a positive dose-response relationship to ethanol toxicity, and this is affected by *Adh* genotype. Flies with recombinant *D. melanogaster Adh* alleles were exposed to sucrose solutions containing a series of ethanol concentrations. The proportion of flies in each vial that were incapacitated after 24h of ethanol exposure is shown by points (with x-axis jitter) and summarized by a LOESS curve. These data are presented as 50% incapacitating concentration (IC_{50}) in Figure 6.

Table S1. Oligonucleotide primers.

primer name	species	purpose	sequence
>Adh_ame_lhsR2	ame	amplify locus	TGAAACTTATCTCACACATTTCGC
>Adh_ame_rhsF2	ame	amplify locus	GAGGAATAAGACCAAGGGAAACA
>Adh_CDS_F1	mel	sequencing	ATGTCGTTTACTTTGACCAACAAGA
>Adh_CDS_F2	mel	sequencing	AAGCAAAAAAGAAGTCACCATGTC
>Adh_clone_S3aAF1	mel	amplify locus	GGGCGAATTCGCCggcgcgccGCGAATAATCAAGACTCAGCACCA
>Adh_clone_S3aAF2	yak san ere ore	amplify locus	GGGCGAATTCGCCggcgcgcCTCGCACGCGAATAATCAAG
>Adh_clone_S3aNR1	mel yak san ere ore	amplify locus	TATGATCTAGAGTCgcggccgcTGGTGACCAGGCACGTATTG
>Adh_exon2_R1	mel	sequencing	TTCTCAATGCGGTCGAGGAT
>Adh_exon2_R2	mel	sequencing	CAGGTTCTAGGATTGAATACACGA
>Adh_exon3_R3	mel	sequencing	GAGCGTTTATCAGTTTTTTGGATTT
>Adh_exon3_R4	mel	sequencing	GGAAAGTTTTATTATCGTATTAGGAGAAG
>Adh_vir_AF1	vir ame	amplify locus	TTggcgcgcCTTATCAGTAAATTTACGAGTGGTTTGT
>Adh_vir_il1F	vir	amplify locus	GAAATTCCCTTTTTTGCTGGTC
>Adh_vir_il1R	vir	amplify locus	GACCAGCAAAAAAGGGAATTTC
>Adh_vir_il2F	vir	amplify locus	TTTGCTTAAAATAAGAGGCGAGG
>Adh_vir_il2R	vir	amplify locus	CGCCTCTTATTTTAAGCAAATCTC
>Adh_vir_il3F	vir	amplify locus	GCAAACCGTAAGTCCATTGTCTAC
>Adh_vir_il3R	vir	amplify locus	GTAGACAATGGACTTACGGTTTGC
>Adh_vir_ir1F	vir	amplify locus	TAGCCGAGACATTGCTGTTGAG
>Adh_vir_ir1R	vir	amplify locus	CTCAACAGCAATGTCTCGGCTA
>Adh_vir_ir2F	vir	amplify locus	TACATCACCCACGCTTGAAATG
>Adh_vir_ir2R	vir	amplify locus	CATTTCAAGCGTGGGTGATGTA
>Adh_vir_NR1	vir ame	amplify locus	ATGTTAgcggccgCTGTTTTGCTGTCTGAATTTTGTG
>Adh_vir_s3aAF1	vir ame	amplify locus	CCCGGGCGAATTCGCCggcgcgccTTATCAGTAAATTTACGAGTGGTTTGT
>Adh_vir_s3aNR1	vir ame	amplify locus	CTGATTATGATCTAGAGTCgcggccgcTGTTTTGCTGTCTGAATTTTGTG
>Adhseq_-0278F	mel	sequencing	TGCCGGTACAAAATCCCAAA
>Adhseq_-0408R	mel	sequencing	GCCGATTAACAGGGACGAGT
>Adhseq_-0429F	mel	sequencing	TACTCGTCCCTGTTAATCGGC
>Adhseq_-0773F	mel yak san	sequencing	TCAGTGCAGTTGTCAGTTGC
>Adhseq_-0909R	mel	sequencing	CGCAAGCAAGTGATCTGCAA
>Adhseq_-1294F	mel yak san	sequencing	ACACTTCTCCTTGCTATGCTTG
>Adhseq_-1419R	mel yak san	sequencing	TGGAGGCAGCGTAGAATCAA
>Adhseq_-1857F	mel yak san	sequencing	TCCTCCCACATCATATTTAACGC
>Adhseq_-1915R	mel yak	sequencing	AGCATGCGGAAGAAATAAGCT

	san		
>Adhseq_-2309F	mel yak san	sequencing	GCAACCCGTTCGATTAGTGG
>Adhseq_-2391R	mel	sequencing	GGTCTGGCCTAGTCTGGG
>Adhseq_-2834F	mel	sequencing	ATAAGGCGCACACTGTCCACT
>Adhseq_-2900R	mel yak san	sequencing	CGTACCTGCCAAAACCGATT
>Adhseq_-3285F	mel	sequencing	GGCAAATGGGTCAAATGGGT
>Adhseq_-3409R	mel	sequencing	GCAAGACGGCTAACTCATCC
>Adhseq_+0105R	mel	sequencing	AGTTACCTTCAGATCGCGCT
>Adhseq_+0224F	mel yak san	sequencing	CAATCCAAAGGTGACCGTCA
>Adhseq_+0585R	mel	sequencing	CCTTTGATCAACTTACCGCCA
>Adhseq_+0728F	mel	sequencing	TGGATGTTGAGCCTCAGGTT
>Adhseq_+1209F	mel yak san	sequencing	AATGTTCGATTTGACGGGCA
>Adhseq_+1587F	mel	sequencing	CACCTATGAAGTCTATTTTAATCGCC
>Adhseq_+1611R	mel	sequencing	GCGATTAAAATAGACTTCATAGGTGG
>Adhseq_+2092R	mel	sequencing	TATATGCGCAGAAAACCGGC
>Adhseq_+2219F	mel yak san	sequencing	CTGTGATGGCGGTTTGTTGT
>AdhSeq_+2381R	mel	sequencing	TCCGATCTCTCGATGGCATT
>Adhseq_+2807F	mel	sequencing	GTCAAAACCAGGTTAGGGAGG
>Adhseq_+2937R	mel	sequencing	CTTAATGTACACCCCAACAGCCTA
>Adhseq_+3206F	mel	sequencing	CGAATGCATGTCGAAATTTTGC
>Adhseq_+3600R	mel	sequencing	CCACAACTCACCCCACCATA
>Adhseq_+3710F	mel yak san	sequencing	TCATCGGACATTGAGGACCC
>Adhseq_+4108R	mel yak san	sequencing	GCACTGCCTCTCCCATTTTG
>Adhseq_+4203F	mel	sequencing	ATTCTGGGCCGGTAATTGAC
>AdhSeq-2228R	mel	sequencing	TTGGGGGTCGCATCGAG
>AdhSeq-2274R	mel	sequencing	CTTGGGGAATGCATGATTGG
>AdhSeq-2483F	mel	sequencing	AGATGTTAGGCACGTCTGCTGAT
>AdhSeq-2626F	mel	sequencing	TAAATTCCTTTGACGCATGCAC
>ame_bam_R1	ame	recombination	AAATGTTGACCAATGTCGGCTAGGATCC
>ame_S27_F	ame	sequencing	CGTCCAAATGTTACTGCACTCG
>ame_S28_R	ame	sequencing	CCGACCAAAAATAAAATATTCCC
>as0.1F	ame	sequencing	CCACATGAATTCCGAACGAGTC
>as0.2R	ame	sequencing	TTAAGACAGAAATCAAACGGCTCT
>as0.3F	ame	sequencing	TTTTGTTCCACTACCATATTGAACC
>as0.4F	ame	sequencing	ACATTGCTGTTAAGACTTTTTAATTGAG
>as0.5R	ame	sequencing	TCTGCCGAGTAGGGCGTTC
>eoAAswap_cF1	ere ore	recombination	CCATTGCCGAGACCAGCAAGCTGCTGAAG ACC
>eoAAswap_lR1	ere ore	recombination	GGTCTTCAGCAGCTTGCTGGTCTCGGCAAT GG

>eost1_R1	ere ore	recombination	GCTAGCCGGTCTGCTGAACTGCAACTGAC
>eost2_F1	ere ore	recombination	GGCATTCACCTTGACCAACAAGAACGTCATTTTCG
>eost2_R1	ere ore	recombination	CGAAAATGACGTTCTTGTTGGTCAAGGTGAATGCC
>eost3_F1	ere ore	recombination	TGGAGCAAGCACTGGGACTCCGGCATCTAA
>eost3_R1	ere ore	recombination	TTAGATGCCGGAGTCCCAGTGCTTGCTCCA
>eost4_F1	ere ore	recombination	AGCTCTCCATGAGAAGTATGGAAAAACTATATGAAAATTGAGA
>eost4_R1	ere ore	recombination	TCTCAATTTTCATATAGTTTTTCCATACTTCTCATGGAGAGCT
>eostU_nheF1	ere ore	recombination	GTCAGTTGCAGTTCAGCAGACCGGCTAGC
>eostU_pacR1	ere ore	recombination	TCCAGCCACTCCCATTTATCTATTTTAATTAAGTAGAGGG
>eostU_xhoF1	ere ore	recombination	CATTATTGTATTTAGTTCATATCAGCACTAAAACTAAACTCGAG
>ere0164F	ere ore	sequencing	CGAAGTGAGAGGTAACAATATTTGG
>ere0419F	ere ore yak san	sequencing	AGATTGCCAAATGGGTCAAA
>ere0446R	ere ore	sequencing	CTGACCCATTTGACCCATTT
>ere0637F	ere ore	sequencing	TACATTAGCGGTAAGCTGCTTAGAC
>ere0771R	ere ore	sequencing	TACTCGCCAGCATTGAGATG
>ere1006F	ere ore	sequencing	TGCGTAGGCTGAACACACTT
>ere1115F	ere ore	sequencing	TCGGGAACATTATCACCAAAAG
>ere1211R	ere ore	sequencing	GTTTCTTGTTTGCTTCGCACC
>ere1345R	ere ore yak san	sequencing	AACGTCTTCTTGCTGCTTGG
>ere1350R	ere ore	sequencing	TCCTTATCAGCAAACGTGCCT
>ere1433F	ere ore	sequencing	TGGAAGCGCGAACAGTCATA
>ere1539F	ere ore	sequencing	TTATTGATTGGGGGAAAATGTTG
>ere1576F	ere ore	sequencing	CACTCCCCAAGTGACCGTAT
>ere1610R	ere ore	sequencing	TAGTCCAAATATGGTATACGGTCACTT
>ere1773F	ere ore	sequencing	TTCCAATTAGGTGAAACAATTCGTA
>ere1877R	ere ore	sequencing	TGATGAGAAGGAAAGAGATAAGC
>ere1915R	ere ore yak san	sequencing	TGCAAGGGCAATAGCACATA
>ere1998R	ere ore	sequencing	ACGCCTTAGAAACTAATTGGCAA
>ere2249F	ere ore	sequencing	GGGTTAAGTCGAGCTGCATT
>ere2250R	ere ore	sequencing	CCATGTTCGCATAGCAGATTCA
>ere2285F	ere ore	sequencing	GCTGCATTTGAAACCGCTACT
>ere2374F	ere ore	sequencing	CAACATTTCTAGATTGATTCTACTACGC
>ere2594F	ere ore	sequencing	GCGTTCTGACTCGGGGT
>ere2597R	ere ore	sequencing	GTGTAGGACCCCGAGTCAGA
>ere2814F	ere ore	sequencing	TCTCAGTGCACTTCCTGGTG
>ere2960R	ere ore	sequencing	GGGCGAGTGCAGTTGACG
>ere3006R	ere ore	sequencing	TTGCACAGCTCCTTAAATACCG

>ere3022R	ere ore	sequencing	GACAACTGCACTGAGACAATAATCG
>ere3170R	ere ore	sequencing	TGCGCTGTCTAACACTGGTT
>ere3209F	ere ore	sequencing	AAAGCGGCACAAGCAGTGT
>ere3264F	ere ore	sequencing	CCGTGCCGAATATTTGACTT
>ere3392R	ere ore	sequencing	TGGAAAATTACTGCTCGTCTGC
>ere3603R	ere ore	sequencing	CATTACGCGATCTCGGCTAT
>ere3741F	ere ore	sequencing	CAACATGGCATTCACCTTGA
>ere3765F	ere ore	sequencing	CAAGAACGTCATTTTCGTGGC
>ere3853R	ere ore	sequencing	GCCTACTTACCTTCAGATCACGC
>ere4032F	ere ore	sequencing	TGCCCATTGCCGAGACC
>ere4091R	ere ore yak san	sequencing	AGCTCCGTTGATCAGGACAT
>ere4217F	ere ore yak san	amplify locus, sequencing	CAACATTGGATCCGTGACTG
>ere4251F	ere ore	sequencing	TACCAGGTGCCCGTCTACTC
>ere4388R	ere ore	sequencing	CAGTTTCTAGAGGTGTCATTGAGGA
>ere4433R	ere ore	sequencing	GGTGGTGCGGGTGATGC
>ere4599R	ere ore yak san	sequencing	CCAGGTCCAGTTTCCAGATG
>ere4711F	ere ore yak san	amplify locus, sequencing	ATTCACGCAAGGCAATAAGG
>ere4983F	ere ore	sequencing	CTGCTATGTGGCGGATTGC
>ere5055R	ere ore yak san	amplify locus, sequencing	CACCGCTATGTTCTTGGTCA
>ere5221F	ere ore	sequencing	GTTCAAGGACTTTGCACTTTCG
>ere5366R	ere ore	sequencing	TGGTGGTTTAATAACGAATCTTCTATG
>ere5461F	ere ore	sequencing	CACGGGGGAGTTGCTATTTA
>ere5515R	ere ore	sequencing	AAGGGCAAAACTGTTAAAAATGG
>ere5561F	ere ore yak san	sequencing	GCACAGCTGCAGTCGATAAA
>ere5715F	ere ore	sequencing	ATGGAGCCACGTTGTGCG
>ere5809R	ere ore	sequencing	TCCTCCCATTTTTCTGTCCA
>ere5884R	ere ore yak san	sequencing	TGCACTATAGGCGCAGAAAA
>ere5889F	ere ore	sequencing	AGTTCGGTGTGATTGGCTTT
>ere5908R	ere ore yak san	sequencing	AAAGCCAATCACACCGAACT
>ere6107F	ere ore	sequencing	CCTTTGCCGACCGCCTG
>ere6225R	ere ore	sequencing	TGCAACTTGACCAACTCCAG
>ere6270F	ere ore	sequencing	GCAATGATGAAGACGACCAA
>ere6407R	ere ore	sequencing	TGCATACTTTTGGGCACGC
>ere6408F	ere ore yak san	sequencing	CCCAAAAGTATGCAGCGAAT
>ere6472F	ere ore	sequencing	GTTCTATATGATTCCGTGTTGTGGT
>ere6596F	ere ore	sequencing	GTCAAGTCCAACTACCCAAGGC
>ere6622R	ere ore	sequencing	CGGCAGACAACGAATATTGA
>ere6747R	ere ore	sequencing	TATGGCTGACTGGCTGACTG

>ere6803F	ere ore yak san	sequencing	ACATTGCCCAGGAATATTGG
>ere6910F	ere ore	sequencing	CAAGAGCCCTAGCTGGAAAA
>ere6914R	ere ore	sequencing	TCTTGCGTTGGATTAGTACCGA
>ere6927R	ere ore	sequencing	GCTGAGCTGAAAGCTTTTCCA
>ere7070R	ere ore	sequencing	CACAACTCACCCCACCATATAGC
>ere7090F	ere ore	sequencing	CAGCTATATGGTGGGGTGAGTTG
>ere7170F	ere ore	sequencing	AGCCAGCAGACCAAAAAGAG
>ere7260R	ere ore yak san	sequencing	GATTCCACTCCATGCGATTC
>ere7284R	ere ore	sequencing	GGAGTGGGTACTCTGGAAAAATG
>ere7298F	ere ore	sequencing	ACTCCCTGCCATCTGCTGC
>ere7320R	ere ore	sequencing	GGAAAGCAAACTTGCAACGG
>ere7520R	ere ore yak san	sequencing	TTAATCAAGCGTGTGCCAAA
>ere7587F	ere ore yak san	sequencing	AGGGCACAATAGCCAAAATG
>ere7834R	ere ore	sequencing	ACTTACCATTTACGAGTACACCTACAAC
>melst_-0183F	mel	recombination	CCTGCGTACATAGCCGAGATC
>melst_-0183R	mel	recombination	GATCTCGGCTATGTACGCAGG
>melst_-0301F	mel	recombination	GAATACTAATATAGAAAAAGCTTTGCCG
>melst_-0301R	mel	recombination	CGGCAAAGCTTTTTCTATATTAGTATTC
>melst_-0408F	mel	recombination	ACTCGTCCCTGTTAATCGGC
>ore6486R	ore	sequencing	GTAAATACTACCACAACACGGAAT
>oreAAswap_cR1	ere ore	recombination	GATGGCACCGTTCTCGTTCAGCTCAATGGC CTTGACAAAGTTCTCGGCGCAGGaCAACGA GgTCTGGGTGGGATG
>oreAAswap_rF1	ere ore	recombination	CCATTGAGCTGAACGAGAACGGTGCCATCT GGAAACTGGACTTGGGCACCCTGGAGGCC ATCcAGTGGAGCAAGC
>sanAAswap1F	yak san	recombination	GAAGTCATCATGGCGTTTACCTTGACCAAC AAG
>sanAAswap1R	yak san	recombination	CTTGTTGGTCAAGGTAAACGCCATGATGAC TTC
>sanAAswap2F	yak san	recombination	TCGCCGAGCTGAAGGCAATCAATCCAAAG
>sanAAswap2R	yak san	recombination	CTTTGGATTGATTGCCTTCAGCTCGGCGA
>sanAAswap3F	yak san	recombination	CGTTGGCCTGCGCCCAGAACTTTGTGAAG
>sanAAswap3R	yak san	recombination	CTTCACAAAGTTCTGGGCGCAGGCCAACG
>va_S01_R	vir ame	sequencing	ACTGATGACTTTTGGTTTAAGTATTTCTA
>va_S02_R	vir ame	sequencing	CAGCCGAACGGCATTGATT
>va_S03_F	vir ame	sequencing	TACACAAGAGCATTACATTACCTAGACAC
>va_S04_F	vir ame	sequencing	ATAGCTGCGCATGCATGATA
>va_S05_R	vir ame	sequencing	CGTGTGCTTTGTGTGTTGGTAGA
>va_S06_F	vir ame	sequencing	CGGCAAGGGAACACTAAAAA
>va_S07_R	vir ame	sequencing	GACAATCTCCCGACTGGTGT
>va_S08_F	vir ame	sequencing	TTGCCAAGCTGAAGACTGTG
>va_S09_R	vir ame	sequencing	GATCGTCCAAAATGCCAGC

>va_S10_F	vir ame	sequencing	TGGGCGAGCTACTTCTTGAG
>va_S11_R	vir ame	sequencing	GCCTCAATGGCCTTAACAAA
>va_S12_F	vir ame	sequencing	AAATGTGGTTGGTTGCTTTTCA
>va_S13_R	vir ame	sequencing	TGGAGTAATCGAATTTGCAACG
>va_S14_F	vir ame	sequencing	AATAAAATCTTCCTTTTGCAGGTTAC
>va_S15_R	vir ame	sequencing	AACCTGCAAAAGGAAGATTTTATTAC
>va_S16_F	vir ame	sequencing	GAGCATTTATTATTCAAGCAAGAT
>va_S17_R	vir ame	sequencing	TATCTTGCTTGAATAATAAATGCTC
>va_S18_F	vir ame	sequencing	TAGACATAGGTCAACATTTCCGATT
>va_S19_R	vir ame	sequencing	AATCGGAAATGTTGACCTATGTCT
>va_S20_F	vir ame	sequencing	CCAATCAAAAGTTGCGTGTG
>va_S21_R	vir ame	sequencing	CACACGCAACTTTTGATTGG
>va_S22_F	vir ame	sequencing	AAATGGCGTATACCGCAAAA
>va_S23_R	vir ame	sequencing	CAGCAGAAGCCAGCAAACG
>va_S24_R	vir ame	sequencing	GCTGTGCTTGTTGTTCTTGC
>va_S25_F	vir ame	sequencing	GATACGGATGGATTTGCTACGA
>va_S30_F	vir ame	sequencing	GCCGACGAGTTGCCAATG
>va_S31_R	vir ame	sequencing	GACTCAATCCGCTTGTTCTGTG
>va_S32_F	vir ame	sequencing	GTCACACCTCTGGGATGTCAAT
>vasti_1F	vir ame	recombination	CTTGTTATTATTAATAGTTATTGGTTTCATTTT
>vasti_1R	vir ame	recombination	AAAATGAAACCAATAACTATTAATAATAACAAG
>vasti_2F	vir ame	recombination	TTAATCAATGCCGTTCGGCTGAGACAGT
>vasti_2R	vir ame	recombination	ACTGTCTCAGCCGAACGGCATTGATTAA
>vasti_4F	vir ame	recombination	AGGATGGCGATCGCTAACAAGAACATTATC
>vasti_4R	vir ame	recombination	GATAATGTTCTTGTTAGCGATCGCCATCCT
>vasti_5F	vir ame	recombination	ACCATTCAAATCCGAGCAAGGAGCATCTG
>vasti_5R	vir ame	recombination	CAGATGCTCCTTGCTCGGATTTGAATGGT
>vasti_a3F	vir ame	recombination	AGTGTCAATCAGCACAGTGTAATCAGGC
>vasti_a6F	vir ame	recombination	AATCTGATTTAATGAATTAATGAAGTTTGTATTACATATAAGCTGATTG
>vasti_v3R	vir ame	recombination	acactgtgctgattgacactGTTCTCTATGCGAGTGGCCCCTTAAATACT
>vasti_v6R	vir ame	recombination	atacaaacttcatTAATTCATTAAATCAGATTTTTATTTTTTGGTTATATTCAGAATTCC
>vir_S26_F	vir	sequencing	ACTAGAAGTCCGTGTCAGGAATCTAA
>vir_S29_R	vir	sequencing	GTACACTGATTCATAAGAGACATACATCC
>virAAswap-F1	vir ame	recombination	GCTGAAGGCACTCAACCCAAAGGTG
>virAAswap-R1	vir ame	recombination	CACCTTTGGGTTGAGTGCCTTCAGC
>YSsti1F	yak san	recombination	CGATTATTGTCTCAGTGCAGTTGTCAGTTGCAG
>YSsti1R	yak san	recombination	CTGCAACTGACAACTGCACTGAGACAATAATCG
>YSsti2F	yak san	recombination	AGAGCCTGCTCAAGCAAAAAGAAGTCATCATGG

>YSsti2R	yak san	recombination	CCATGATGACTTCTTTTTGCTTGAGCAGGC TCT
>YSsti3F	yak san	recombination	CAGTGGTCCAAGCACTGGGACTCC
>YSsti3R	yak san	recombination	GGAGTCCCAGTGCTTGGACCACTG
>YSsti4F	yak san	recombination	GAGAAATCCAAAATAGTGATAAGCCGCCT GCC
>YSsti4R	yak san	recombination	GGCAGGCGGCTTATCACTATTTTGGATTTC TC
>ysstiU_nheF1	yak san	recombination	GGAGCGTAGGTTAAATACCCTGGCTAGCA AG
>YSstiU_pacR1	yak san	recombination	CCAGCCACTCCCACTTATCTATTTTAATTA ATTTAGCC

SI References

1. Hansen SK, Tjian R. (1995) TAFs and TFIIA Mediate Differential Utilization of the Tandem *Adh* Promoters. *Cell* 82: 565-575

2. Laurie CC, Stam L. (1994) The Effect of an Intronic Polymorphism on Alcohol Dehydrogenase Expression in *Drosophila melanogaster*. *Genetics* 138: 379-385

3. Heberlein U, England B, Tjian R. (1985) Characterization of *Drosophila* Transcription Factors That Activate the Tandem Promoters of the Alcohol Dehydrogenase Gene. *Cell* 41: 965-977

4. Choudhary M, Laurie CC. (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

5. Siddiq MA, Loehlin DW, Montooth KL, Thornton JW. (2017b) Experimental test and refutation of a classic case of molecular adaptation in *Drosophila melanogaster*. *Nature Ecology and Evolution* 1, 0025 DOI: 10.1038/s41559-016-0025

6. Stam LF, Laurie CC. (1996) Molecular Dissection of a Major Gene Effect on a Quantitative Trait: The Level of Alcohol Dehydrogenase Expression in *Drosophila melanogaster*. *Genetics* 144: 1559-1564

7. Loehlin DW, Carroll SB. (2016) Expression of tandem gene duplicates is often greater than twofold. *Proc. Natl Acad. Sci. USA* 113, 5988–5992 DOI: 10.1073/pnas.1605886113

8. Chakraborty M, Fry JD. (2016) Evidence that Environmental Heterogeneity Maintains a Detoxifying Enzyme Polymorphism in *Drosophila melanogaster*. *Current Biology* 26: 219-223 DOI: 10.1016/j.cub.2015.11.049

9. Corbin V, Maniatis T. (1990) Identification of cis-Regulatory Elements Required for Larval Expression of the *Drosophila melanogaster* Alcohol Dehydrogenase Gene. *Genetics* 124: 637-646

10. Ashburner M. (1998) Speculations on the subject of alcohol dehydrogenase and its properties in *Drosophila* and other flies. *Bioessays* 20: 949-954

11. Vo ngoc L, Wang YL, Kassavetis GA, Kadonaga JT. (2017) The punctilious RNA polymerase II core promoter. *Genes and Development* 31: 1289-1301 DOI: 10.1101/gad.303149.117

12. Hoskins RA, Carlson JW, Kennedy C, Acevedo D, Evans-Holm M, Frise E, Wan KH, Park S, Mendez-Lago M, Rossi F, Villasante A, Dimitri P, Karpen GH, Celniker SE. (2007) Sequence finishing and mapping of *Drosophila melanogaster* heterochromatin. *Science* 316: 1625-1628

13. Gramates LS, Marygold SJ, dos Santos G, Urbano J-M, Antonazzo G, Matthews BB, Rey AJ, Tabone CJ, Crosby MA, Emmert DB, Falls K, Goodman JL, Hu Y, Ponting L, Schroeder AJ, Strelets VB, Thurmond J, Zhou P and the FlyBase Consortium. (2017) FlyBase at 25: looking to the future. *Nucleic Acids Research* 45(D1):D663-D671 DOI: 10.1093/nar/gkw1016

14. Nurminsky DI, Moriyama EN, Lozovskaya ER, Hartl DL. (1996) Molecular phylogeny and genome evolution in the *Drosophila virilis* species group: Duplications of the alcohol dehydrogenase gene. *Mol Biol Evol* 13(1):132–149.

15. R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

16. Galecki A, Burzykowski T. (2013) Linear Mixed-Effects Models Using R: A Step-By-Step Approach (Springer, New York), pp 478–480.

17. Venables WN, Ripley BD. (2002) Modern Applied Statistics with S. Fourth Edition. (Springer, New York), p. 193

18. Hansen SK, Tjian R. (1995) TAFs and TFIIA Mediate Differential Utilization of the Tandem *Adh* Promoters. *Cell* 82: 565-575