Parallel Functional Changes in Independent Testis-Specific Duplicates of Aldehyde dehydrogenase in Drosophila

Mahul Chakraborty*1 and James D. Fry2
1Department of Ecology and Evolutionary Biology, University of California, Irvine
2Department of Biology, University of Rochester
*Corresponding author: E-mail: mchakrab@uci.edu.
Associate editor: John True

Abstract

A large proportion of duplicates, originating from ubiquitously expressed genes, acquire testis-biased expression. Identifying the underlying cause of this observation requires determining whether the duplicates have altered functions relative to the parental genes. Typically, statistical methods are used to test for positive selection, signature of which in protein sequence of duplicates implies functional divergence. When assumptions are violated, however, such tests can lead to false inference of positive selection. More convincing evidence for naturally selected functional changes would be the occurrence of structural changes with similar functional consequences in independent duplicates of the same gene. We investigated two testis-specific duplicates of the broadly expressed enzyme gene Aldehyde dehydrogenase (Aldh) that arose in different Drosophila lineages. The duplicates show a typical pattern of accelerated amino acid substitutions relative to their broadly expressed paralogs, with statistical evidence for positive selection in both cases. Importantly, in both duplicates, width of the entrance to the substrate binding site, known a priori to influence substrate specificity, and otherwise conserved throughout the genus Drosophila, has been reduced, resulting in narrowing of the entrance. Protein structure modeling suggests that the reduction of the size of the enzyme’s substrate entry channel, which is likely to shift substrate specificity toward smaller aldehydes, is accounted for by the positively selected parallel substitutions in one duplicate but not the other. Evolution of the testis-specific duplicates was accompanied by reduction in expression of the ancestral Aldh in males, supporting the hypothesis that the duplicates may have helped resolve intralocus sexual conflict over Aldh function.

Key words: sexual conflict, protein structure evolution, aldehyde dehydrogenase, test of positive selection, sex-specific gene, parallel evolution.

Introduction

A large number of duplicates in animal genomes originating from ubiquitously expressed parental genes have been observed to acquire testis-specific expression patterns (Vinckenbosch et al. 2006; Bai et al. 2007; Gallach et al. 2010; Baker et al. 2012). Identification of the underlying cause of such patterns depends on the question of whether testis-limited duplicates have diverged functionally from their parental counterparts (Gallach and Betran 2011). Although statistical methods are typically employed to detect amino acid sites under positive selection—a proxy for functional divergence—their inferences often lead to misidentification of functional changes in the protein (Wong et al. 2004; Nozawa et al. 2009; Gharib and Robinson-Rechavi 2013). More convincing evidence in this context can come from parallel structural and functional changes in independent duplicates of the same gene, accomplished either by parallel or unique substitutions (Christin et al. 2010; Zhen et al. 2012). Although parallel amino acid substitutions in independent testis-specific duplicates of the same gene have been noted previously, their structural or functional significance has not been investigated (Vemuganti et al. 2007). To gain insight into the process that underlies evolution of testis-specialized duplicates, we investigated whether protein structure and function in independent duplicates of aldehyde dehydrogenase (Aldh) enzyme gene in Drosophila have undergone parallel changes and whether such changes were mediated by parallel substitutions.

The Drosophila aldehyde dehydrogenase encoded by Aldh is a conserved homolog of the vertebrate mitochondrial ALDH (ALDH2). In Drosophila melanogaster, it is capable of catalyzing oxidation of a wide range of aldehydes (Rothacker et al. 1991; Chakraborty and Fry 2011) or acetaldelyde derived from dietary ethanol (Fry and Saweikis 2006). All Drosophila homologs of DmALDH possess mitochondrial leader peptides and share highly conserved sequences, suggesting that the enzyme performs a conserved function within mitochondria. Because most Drosophila species are not resistant to ethanol, protection from aldehydes produced by lipid peroxidation is perhaps the conserved function of Drosophila ALDH. These toxic aldehydes are typically relatively large (six or more carbons), so any new adaptation that requires ALDH enzyme with improved detoxification rate for small aldehydes requires alteration of the conserved specificity of the enzyme for large aldehydes. Such alteration...
of substrate specificity in ALDH, however, is likely to compromise ancestral substrate specificity of the enzyme (Vasiliou et al. 2000).

Substrate specificity of Drosophila ALDH and its conserved homologs in other eukaryotes is controlled by the size and shape of the substrate entry channel (SEC), an intramolecular tunnel within the enzyme that guides the substrate to the active site. Altering the diameter of the channel alters substrate specificity of ALDH predictably: Constricting the entrance or neck of the channel shifts the substrate specificity of the enzyme toward smaller aldehydes (Moore et al. 1998; Sobreira et al. 2011). In D. melanogaster, for example, when SEC diameter is constricted by a naturally occurring Leu479Phe substitution, turnover rate increases for acetaldehyde, a small aldehyde, but decreases for larger aldehydes (Fry et al. 2008; Chakraborty M, Fry JD, unpublished data).

Presumably, fixation of an Aldh allele containing substitutions that would result in altered substrate specificity can be achieved if the allele is a duplicated copy of the Aldh gene. The allele can then evolve independently of the ancestral, conserved copy of the enzyme and thus bypass the deleterious fitness consequences. In Drosophila lineages, as we will show, independent duplications of Aldh have created new gene copies which have diverged considerably from the parental genes in amino acid sequence while acquiring testis-specific expression. We investigated whether the amino acid sequence of the duplicates is consistent with optimization of ALDH for a similar function in both lineages. In particular, we looked for parallel changes in SEC diameter as well as in specific residues.

Results
Throughout the rest of the text, we refer to the diverged duplicates of Aldh as Aldh-dup and conserved duplicates as Aldh. Aldh-dup in the obscura species group was created by a tandem duplication which preserved the exon–intron structure of the autosomal ancestral gene in the duplicate (fig. 1B). Because D. lowei and D. miranda both possess the duplicate, we infer that the duplication event occurred at least 4–5 Ma, before the two species had diverged (fig. 1A) (Gao et al. 2007). We found the occurrence of the second duplication event in the subgenus Drosophila, which consists of Hawaiian Drosophila and species belonging to the virilis-repleta group (fig. 1A). This duplicate, unlike its counterpart in obscura, originated by retroduplication and thus lost the introns and genomic location of the parental Aldh (fig. 1B). The presence of the duplicate in D. virilis, D. mojavensis, and D. grimshawi suggests that the duplication occurred approximately 30 Ma, before the split between the ancestors of the Hawaiian and virilis-repleta groups (Morales-Hojas and Vieira 2012).

Expression of Duplicated and Ancestral Genes
Comparison of expression patterns of Aldh and Aldh-dup in D. miranda and D. pseudoobscura using publicly available RNA sequencing (RNAseq) data indicates that the Aldh-dups have a highly testis-biased expression pattern, whereas expression of Aldhs is female biased (table 1). Sex-biased expression of Aldh-dup appears to be present in the head and thorax of D. pseudoobscura as well, though to a much lesser extent (table 1). In contrast, female-biased expression of DpseAldh appears to be solely due to the higher expression of the gene in ovaries compared with testes (table 1).

As in the obscura group, expression of Aldh-dup and Aldh in the Drosophila subgenus is male- and female-biased, respectively (table 1). Analysis of an independent RNAseq data set suggests that the male-biased expression of DvirAldh-dup is due to higher expression of the gene in the testes (Ahmed Y, personal communication).

The expression pattern of Aldhs in obscura group and Drosophila subgenus differed from expression patterns of their homologs in D. willistoni and D. melanogaster, which lacked Aldh-dup. In these two species, Aldh was expressed in nearly unbiased or testis-biased patterns (table 1). The male-biased expression pattern of Aldh in D. melanogaster appeared to be due solely to the expression differences of Aldh between testes and ovaries, because expression levels of the gene in head and thorax were similar between the sexes.

Tests for Positive Selection
Both duplicates showed accelerated protein evolution compared with their broadly expressed paralogs (fig. 1A). We used the branch site model of Zhang et al. (2005) to test whether this acceleration was the result of positive or relaxed selection. This method tests whether a subset of sites in a specified phylogenetic branch, termed the “foreground” branch, show evidence of positive selection (dN/dS > 1), while allowing some sites to remain under purifying selection (dN/dS < 1). It is therefore more realistic, and less stringent, than testing whether dN/dS > 1 for the entire gene. To test whether a particular duplicate lineage underwent positive selection, we specified the duplicate lineage as the foreground branch, and the remainder of the tree as the background branch (with the other duplicate lineage excluded from the analysis). The model estimates dN/dS = ω for four classes of sites (table 2). Site class 0 denotes codons that are under purifying selection throughout the tree. Site class 1 includes codons that are evolving neutrally (ω = 1) throughout the tree. Classes 2a and 2b comprise codons that are conserved or neutral on the background branches, but are under positive selection on the foreground branches.

When applied to the obscura group duplicates, the test showed that 10% of the sites that were neutral or under purifying selection in Aldh accumulated amino acid substitutions in Aldh-dup under positive selection (P = 10−5, log-likelihood ratio test) (table 2). For the duplicates in the Drosophila subgenus, the corresponding figure is 21% (P = 0.01). The branch site test identified seven sites in the obscura group duplicates, and 13 in the Drosophila subgenus duplicates, as being under positive selection with a high degree of confidence (posterior probability > 0.95, Bayes empirical Bayes analysis). Another 20 and 38 residues, respectively, showed evidence of positive selection in the two groups.
at a somewhat lower level of confidence (0.80 < posterior probability < 0.95).

Functional Changes in the ALDH-dups
To examine whether the substrate specificity of the two ALDH-dups has been altered similarly with respect to the ancestral ALDH, channel diameters of the most recent common ancestors (MRCA) of the ALDH and ALDH-dups of the two groups need to be known. To accomplish that, we reconstructed the diameters of the SEC in the ancestors of the two groups using a maximum-likelihood based statistical method (Paradis et al. 2004). The method, which is implemented in an R package called ape, models evolution of a trait in a given tree following Brownian motion. The estimated ancestral channel mouth diameters for the Drosophila subgenus and the obscura species group were 8.55 Å (95% CI: 7.03–10.71 Å) and 9.42 Å (95% CI: 8.88–9.98 Å), respectively (fig. 3). Furthermore, the estimated channel diameter of ancestral ALDH for the genus Drosophila was 8.73 Å (95% CI: 6.77–11.24 Å), and that for the MRCA of the obscura and melanogaster species groups was 8.73 Å (95% CI: 6.95–11.06 Å) (fig. 3). Hence, the channel diameters of ALDH-dups in the Drosophila subgenus (6.94, 6.5, and 6.2 Å) and obscura species group (5.9, 5.86, 5.89, and 6.28 Å) are smaller than those of the ancestral, as well as extant, ALDHs.

Table 1. Relative (male/female) Expression Level of Aldh and Aldh-dups in Different Species.

<table>
<thead>
<tr>
<th>Group</th>
<th>Species</th>
<th>Gene</th>
<th>Tissue</th>
<th>Fold Difference (log₂(+/−))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obscura species group</td>
<td>Drosophila miranda</td>
<td>Aldh</td>
<td>Testis/Ovary</td>
<td>−3.1667*</td>
</tr>
<tr>
<td></td>
<td>Drosophila miranda</td>
<td>Aldh</td>
<td>Testis/Ovary</td>
<td>11.5447*</td>
</tr>
<tr>
<td></td>
<td>Drosophila pseudoobscura</td>
<td>Aldh</td>
<td>Testis/Ovary</td>
<td>−2.0492*</td>
</tr>
<tr>
<td></td>
<td>Drosophila pseudoobscura</td>
<td>Aldh</td>
<td>Carcass sans gonad</td>
<td>0.006*</td>
</tr>
<tr>
<td></td>
<td>Drosophila pseudoobscura</td>
<td>Aldh</td>
<td>Aldh-dup</td>
<td>8.4064*</td>
</tr>
<tr>
<td></td>
<td>Drosophila pseudoobscura</td>
<td>Aldh</td>
<td>Aldh-dup</td>
<td>1.78*</td>
</tr>
<tr>
<td>Drosophila subgenus</td>
<td>Drosophila virilis</td>
<td>Aldh</td>
<td>Whole body</td>
<td>−0.5992*</td>
</tr>
<tr>
<td></td>
<td>Drosophila virilis</td>
<td>Aldh</td>
<td>Whole body</td>
<td>6.1495*</td>
</tr>
<tr>
<td></td>
<td>Drosophila mojavensis</td>
<td>Aldh</td>
<td>Whole body</td>
<td>−2.3468*</td>
</tr>
<tr>
<td></td>
<td>Drosophila mojavensis</td>
<td>Aldh</td>
<td>Whole body</td>
<td>4.5264*</td>
</tr>
<tr>
<td>Species lacking duplicates</td>
<td>Drosophila melanogaster</td>
<td>Aldh</td>
<td>Testis/Ovary</td>
<td>2.1800*</td>
</tr>
<tr>
<td></td>
<td>Drosophila melanogaster</td>
<td>Aldh</td>
<td>Head</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>Drosophila melanogaster</td>
<td>Aldh</td>
<td>Whole body</td>
<td>0.662</td>
</tr>
<tr>
<td></td>
<td>Drosophila willistoni</td>
<td>Aldh</td>
<td>Abdomen</td>
<td>0.2</td>
</tr>
</tbody>
</table>

*Difference from zero significant at P < 0.01.
Although the change of diameters in the duplicates of the two groups has occurred in the same direction, this change conceivably have resulted from relaxed selection on channel diameter. To test whether the reduced channel diameters in the ALDH-dups are stochastic deviations of the channel diameters of the ancestral enzymes, or the result of adaptive evolution, we asked whether channel diameters of ALDH-dups share an optimum that is different with respect to the optimum channel diameter of ALDH. We modeled the evolution of ALDH channel diameter under the assumption that the evolution of ALDH-dups share an optimum that is different with respect to the optimum channel diameter of ALDH. We modeled the evolution of ALDH channel diameter under the assumption that the evolution of channel mouth diameter is an Ornstein–Uhlenbeck (OU) process: The channel diameter can change randomly in any direction but it is most likely to change toward the optimum channel diameter (Gardiner 1985; Hansen 1997). Inference about the change in optimum is made based on whether the model with a single optimum or two optima fits the observed data better. We used the compa.ou module included in the ape package to compare these two alternative models for ALDH-dups of the both groups. The module allows assignment of a node in the phylogenetic tree as the point at which the character optimum is predicted to change. We assigned the MRCAs of ALDH-dups of the Drosophila subgenus and the obscura species group as the nodes at which the channel diameter optimum might have shifted.

The two optima model fit the data significantly better for both the subgenus Drosophila ($\chi^2 = 18.95, df = 1, P = 7.3 \times 10^{-5}$) and the obscura species group ($\chi^2 = 14.52, df = 1, P = 7.03 \times 10^{-6}$). The estimated optima were 8.7 Å for the ALDHs, compared with 6.56 Å and 5.62 Å for the ALDH-dups in the Drosophila subgenus and obscura group, respectively. Together, these results give evidence that ALDH-dups of both groups had undergone parallel reduction in channel diameters, and that such changes were likely adaptive.

We performed enzyme assays to check whether the substrate specificity of DpseALDH- was shifted toward smaller aldehydes relative to that of DpseALDH, as predicted by in silico analysis. We predicted that the ratio of turnover rates for a small aldehyde (acetalddehyde) to that for a larger aldehyde (hexanal) would be higher in extracts from testes than in extracts from ovaries, because testes in D. pseudoobscura express both ALDH and ALDH-dup, whereas ovaries express only ALDH. The results were consistent with the prediction (fig. 4). No difference in the ratios was observed in D. melanogaster, which lacks ALDH-dup (fig. 4).

Role of Parallel Substitutions

The parallel changes in SEC diameter may have been caused by the same (i.e., parallel) or different amino acid substitutions in the two lineages. In the branch site test, seven sites showed evidence of positive selection (posterior probability $> 0.8$) in both duplicate lineages. Of these, three showed remarkable parallelism, with a residue invariant among the ALDHs in Drosophila and Musca being replaced by the same amino acid in all the duplicate sequences: 218 (Ser $\rightarrow$ Pro), 365 (Val $\rightarrow$ Ile), and 366 (Asn $\rightarrow$ Ser) (coordinates are based on the multiple species alignment in supplementary fig. S1, Supplementary Material online). A fourth, 235 (Glu $\rightarrow$ Pro), showed nearly as striking a pattern, except that the duplicate in D. lowei has alanine rather than proline. In addition, we also noticed that the ALDH-dups in both groups had a single residue (397th deleted (supplementary fig. S1, Supplementary Material online); this could not be included in the branch site test, which ignores indels within a sequence alignment.

To investigate the effects of the parallel substitutions on the SEC diameter, we mutated in silico all four residues (S218, E235, V365, and N366) in DmojALDH and DpseALDH by their corresponding residues (P218, P235, I365, and S366) in the ALDH-dups and also removed the 397th residue of the DmojALDH and DpseALDH. The amino acid substitutions combined with the deletion reduced the channel diameter of DmojALDH to the level of ALDH-dup (fig. 2D–F). The mutations also reduced SEC diameter of DpseALDH, but only slightly (fig. 2A–C). We conclude that the narrowing in SEC diameter in the duplicate lineages is only partly explained by the parallel substitutions.

Discussion

Animal genomes contain numerous genes with testis-biased expression which arose by duplication of broadly expressed genes, and such genes usually show evidence of protein evolution under positive selection following the duplication. Although a general explanation for this pattern is lacking, it suggests that ancestrally, there was a conflict between the ubiquitous function of the gene and its function in the testes, such that the two functions could not be optimized simultaneously. Gene duplication would be expected to mitigate such a conflict, by allowing the novel gene to become optimized for the testis-specific function, without compromising the ubiquitous function. As attractive as this hypothesis is, there is seldom direct evidence on the functional consequences of protein sequence evolution in testis-specific duplicates (Gallach and Betran 2011). Without such evidence, the extent to which replacement substitutions in such duplicates improve a novel function at the expense of an ancestral function is unclear.

We have shown that in the genus Drosophila, two independent testis-specific duplicates of the conserved ortholog of mammalian ALDH2 have arisen, in each case showing evidence for accelerated protein evolution under positive selection. To examine whether the evolution of Drosophila ALDH-dups may have resolved a conflict between the
ancestral function of the ubiquitously expressed ALDH and a testis-specific function, we tested two predictions of the conflict hypothesis. First, we investigated whether the duplicates had experienced structural changes likely to alter substrate specificity, and if so whether functionally similar changes had occurred in the independent duplicates. Consistent with these predictions, both duplicates experienced changes in the SEC diameter, which is known to control substrate specificity in both vertebrate and invertebrate ALDH (Sobreira et al. 2011). Based on protein structure modeling in silico, in each duplicate the width of the channel mouth has been reduced, increasing catalytic activity with smaller aldehydes but decreasing it with larger aldehydes. Results of enzyme activity assays (fig. 4) support this conclusion: in D. melanogaster, a species lacking a duplicate, extracts of male and female abdomens showed similar ratios of activity between a small and large aldehyde, whereas in D. pseudoobscura, a species with a duplicate, male abdomens gave rise to relatively higher activity toward the small aldehyde than female abdomens.

Second, we investigated whether evolution of Aldh-dups resulted in relaxed selection for testis-specific function of Aldh. Consistent with this possibility, in D. melanogaster, which lacks Aldh-dup, expression of Aldh is 2-fold higher in testes than ovaries, but in two obscura group species, which have a duplicate, this difference is reversed (table 1). Although expression data from testes and ovaries are unavailable from the subgenus Drosophila, Aldh expression in whole bodies in D. virilis and D. mojavensis is strongly female-biased, whereas expression in abdomens of D. willistoni, which lacks the duplicate, shows a nonsignificant tendency toward male bias (table 1). These observations are consistent with the hypothesis that the presence of the duplicate relaxes selection for high expression of Aldh in the testes. In species without the duplicate, high expression of Aldh in the testes could be a means of partly compensating for its relative inefficiency with small aldehydes. Such compensation might be costly, however, particularly as Aldh is already strongly expressed in all stages and tissues.

The accelerated evolution and evidence for novel function of the duplicates, together with the apparent retention of the ancestral function in the original Aldh copies, is consistent with three models of gene duplication discussed by Innan and Kondrashov (2010). One of these is the classic neofunctionalization model (Ohno 1970), in which the duplicate is fixed by genetic drift, with the new function evolving later (Innan and Kondrashov 2010). Alternatively, the duplicates may have been fixed by positive selection. One way this could have happened is if the initial duplication event fortuitously resulted in the new copy having testis-biased expression (“modified duplication” model of Innan and Kondrashov), thus potentially relieving a conflict over optimal tissue- and/or sex-specific expression levels. An alternative possibility is that one or more of the substitutions that contribute to the difference in substrate specificity between the Aldhs and Aldh-dups in extant species were segregating prior to the
duplication event. In this case, the duplications could have been beneficial by allowing the complementary variants to coexist in the same genome (“adaptive radiation” model). Although a polymorphism affecting SEC diameter and substrate specificity is present in *D. melanogaster* (Fry et al. 2008; Chakraborty M, Fry JD, unpublished data), we see little prospect of being able to determine whether any of the functionally significant differences between the *Aldh* and *Aldh-dups* derive from polymorphisms present at the time of the duplications.

Regardless of what caused initial fixation of the duplicates, the evidence reviewed above suggests that their preservation and subsequent evolution helped resolve a conflict between optimization for a ubiquitous function and optimization for a testis-specific function. This conflict can be viewed in either of two ways: One possibility is that in species lacking the duplicate, *Aldh* is a potential source of intralocus sexual conflict (IASC), the situation where mutations that are beneficial when in one sex are deleterious when in the other (Bonduriansky and Chenoweth 2008; van Doorn 2009). If this is correct, the evolution of testis-specific duplicates would have resolved, or at least reduced, IASC (Gallach and Betran 2011). On the other hand, it is possible that in species without duplicates, *Aldh* mutations with beneficial effects in the testes do not increase overall fitness in males because they compromise *Aldh* function in other organs and tissues other than the testes. Under this scenario, the conflict would be better viewed as one between optimal ALDH structure in different organs, rather than between the sexes (Stewart et al. 2010). Although we have no evidence to distinguish these possibilities, the IASC hypothesis is testable in principle, because it leads to the prediction that there should be some possible *Aldh* mutations in species lacking the duplicate that increase male fitness at the expense of female fitness.

Presently, we do not know what specific biological function within the testes the duplicates were selected for. In mammals, acetaldehyde is known to compromise male fertility, through mechanisms that remain unclear (Muthusami and Chinnaswamy 2005), so it is possible that ability to detoxify acetaldehyde is at a premium in *Drosophila* testes as well. Acetaldehyde can be produced endogenously (Langevin et al. 2011), and can also be derived from dietary ethanol; in *D. melanogaster*, *Aldh* is essential for detoxifying acetaldehyde from this source (Fry et al. 2004). Owing to the limited knowledge of the natural ecology of most of the *Drosophila* species mentioned here, it remains unclear how much dietary ethanol, acetaldehyde, or other small aldehydes they encounter in nature.

Our in silico mutagenesis study suggests that the parallel narrowing of SEC diameter in the ALDH dupes cannot be
entirely explained by parallel amino acid substitutions. In particular, the four positively selected parallel substitutions plus the parallel deletion were able to account for SEC narrowing in the *Drosophila* subgenus duplicates, but had little effect in the obscura group duplicates (fig. 2). This suggests that other substitutions were involved in the latter case. Notably, in the obscura group, one of the three residues thought to play a major role in substrate specificity variation in chordate ALDHs (Sobreira et al. 2011) has been altered in the ALDH-dups, as have six of the eight identified minor substrate specificity controlling residues (supplementary fig. S1, Supplementary Material online). Although the former site (position 145 in supplementary fig. S1, Supplementary Material online) did not show evidence of positive selection in the branch site test, the latter residues (174 and 358) showed strong evidence of selection (posterior probability = 0.999). Changing these two residues in silico, however, did not narrow SEC diameter, either alone or in combination with the five parallel changes (supplementary table S1, Supplementary Material online). The substitution at the major-substrate specificity controlling site, 145Met → Thr, by itself also had little effect on SEC diameter (supplementary table S1, Supplementary Material online). In combination, however, the three substitutions resulted in full narrowing of the SEC (supplementary table S1, Supplementary Material online). The combination of the 145Met → Thr substitution and the four parallel positively selected substitutions also narrowed SEC diameter (but oddly, this effect was reversed by adding the parallel deletion; supplementary table S1, Supplementary Material online), as did combining 145Met → Thr with a change that had occurred at an adjacent residue, 146Ala → Ser. These results highlight the likely importance of the previously identified major substrate specificity controlling residue, and also give evidence for complex patterns of epistasis between substitutions. It should also be kept in mind that some of the positively selected substitutions, including possibly the parallel changes, were selected for reasons other than their effects on SEC diameter, for example, because of effects on protein stability, NAD+ binding, or oligomerization.

In summary, we have shown that conflicting selection pressures on substrate specificity may underlie the evolution of testis-specific duplicates of *Aldh*. Our results complement previous studies which predicted, primarily based on gene expression pattern (Gallach et al. 2010; Baker et al. 2012), that IASC could be responsible for the evolution of testis-biased expression in many gene duplicates. Because testis-specific proteins are among the most rapidly evolving in *Drosophila* and other animals (Swanson and Vacquier 2002; Nielsen et al. 2005; Haerty et al. 2007), conflicts between testis-specific and conserved functions of proteins may be common.

**Materials and Methods**

**Protein and DNA Sequences**

The duplicates were found by carrying out an unbiased search for aldehyde dehydrogenase genes present in sequenced *Drosophila* genomes. The selection criterion we used for identifying duplicated ALDH1 was that duplicate ALDH should share 70% or more identity with the conserved ALDH2 homolog present in that species. The arbitrary identity cutoff was chosen to ensure that the hit had a high chance of being a functional aldehyde dehydrogenase (70% sequence identity is present between *Drosophila* ALDH and mammalian ALDH2). By aligning with human ALDH2 and *D. melanogaster* ALDH we confirmed that the main catalytic, substrate binding, and cofactor binding residues were present in ALDH-dups.

DNA and protein sequences of ancestral ALDH and its duplicates in *D. pseudoobscura*, *D. persimilis*, *D. virilis*, *D. mojavensis*, and *D. grimshawi* were obtained from FlyBase (Marygold et al. 2013) (table 3). DNA sequences of the ancestral and duplicated *Aldh* of *D. lowei* were obtained from Pseudobase (McGaugh et al. 2012). Sequences of *D. miranda* *Aldh* and *Aldh-dup* were extracted from genomic DNA by aligning the sequences of the two genes extracted from Pseudobase. Protein sequences of ALDH of *D. melanogaster*, *D. sechellia*, *D. simulans*, *D. erecta*, *D. yakuba*, *D. ananalassae*, and *D. willistoni* were obtained from FlyBase. Protein sequences of ALDH of *Glossina morsitans* (GenBank ID ABG77991.1) and *Musca domestica* (RefSeq ID XP_005175573.1) were downloaded from NCBI. Orthologs of *Aldh* and *Aldh-dup* were identified by reciprocal BLAST (Basic Local Alignment Search Tool). Homologous amino acid residues were identified by aligning amino acid sequences of all fly ALDH mentioned here together with that of human ALDH2 in ClustalW.

**Test for Selection**

Nucleotide sequences of the coding region of *Aldh* and *Aldh-dup* were aligned guided by amino acid sequence alignment. The gene tree was created based on this alignment by phylogeny.fr (Dereepe et al. 2008) and reconciled with the

### Table 3. *Aldh* Genes and the Source of Their Corresponding RNAseq Data.

<table>
<thead>
<tr>
<th>FlyBase ID</th>
<th>Gene Name</th>
<th>RNAseq (NCBI accession ID)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GA25309</td>
<td>DpselAldh-dup</td>
<td>SRX102928, SRX102924</td>
</tr>
<tr>
<td>GA17661</td>
<td>DpselAldh</td>
<td></td>
</tr>
<tr>
<td>GL26144</td>
<td>DperAldh-dup</td>
<td></td>
</tr>
<tr>
<td>GL26145</td>
<td>DperAldh</td>
<td></td>
</tr>
<tr>
<td>GJ16294</td>
<td>DviarAldh-dup</td>
<td>SRX247003, SRX247002, SRX054489, SRX054488</td>
</tr>
<tr>
<td>GJ17859</td>
<td>DviarAldh</td>
<td>SRX054487, SRX054486, SRX054485, SRX054484</td>
</tr>
<tr>
<td>GJ18106</td>
<td>DmojaAldh-dup</td>
<td></td>
</tr>
<tr>
<td>GJ17663</td>
<td>DmojaAldh</td>
<td></td>
</tr>
<tr>
<td>GH13227</td>
<td>DgriAldh-dup</td>
<td></td>
</tr>
<tr>
<td>GH13324</td>
<td>DgriAldh</td>
<td></td>
</tr>
<tr>
<td>GK24201</td>
<td>DwilAldh</td>
<td>SRX095362, SRX095361</td>
</tr>
<tr>
<td>CG3725</td>
<td>DmelAldh</td>
<td></td>
</tr>
<tr>
<td>Pseudobase</td>
<td>DmirAldh-dup</td>
<td>SRX106031, SRX106026</td>
</tr>
<tr>
<td>Pseudobase</td>
<td>DmirAldh</td>
<td></td>
</tr>
<tr>
<td>Pseudobase</td>
<td>DlowAldh-dup</td>
<td></td>
</tr>
<tr>
<td>Pseudobase</td>
<td>DlowAldh</td>
<td></td>
</tr>
</tbody>
</table>
known species phylogeny (Powell 1997) by the program Notung (Durand et al. 2005). Lineage-specific $d_\omega/d_s$ was calculated using the “free ratios model” implemented in CODEML program of PAML 4.7 package. Positive selection in duplicates was tested using the “branch sites model” included in the same package (Zhang et al. 2005). In the gene tree provided to CODEML, we specified the branch leading to the duplicates as the foreground branch, which facilitates detection of branch-specific accelerated rate of protein evolution under positive selection. Then, we evaluated the null hypothesis of no selection ($\omega = 1$) against the alternative hypothesis of positive selection ($\omega > 1$) by the log-likelihood ratio test ($\chi^2$ test, df = 1). Analyses were carried out separately for the two duplicates, with all sequences (fig. 1) included except for those of the other duplicate.

Expression Analysis

RNAseq raw reads were downloaded from NCBI short read archive (table 3). The reads were mapped to the genomic sequence of Aldh and Aldh-dup by Tophat (Trapnell et al. 2012). Exon–intron boundaries and untranslated regions were identified by Cufflinks. Expression difference was calculated using the Cuffdiff program (Trapnell et al. 2012). The expression data for DvirAldh and DvirAldh-dup were confirmed from RNAseq data set obtained from an independent experiment (Ahmed Y, unpublished data). Expression data for D. melanogaster were obtained from modENCODE (Celniker et al. 2009).

Protein Structure Modeling and Ancestral Channel Mouth Reconstruction

Protein structure models were constructed by the Populus server (Offman et al. 2008), using the structures of human ALDH2 (PDB ID 1004) and sheep liver ALDH1 (PDB ID 1BXS) as the templates. The top structure model for each enzyme returned by the server was used for further analysis. The SEC was inspected and the surrounding residues were identified using the software Fred Receptor (Omega 2.5.1.4; OpenEye Scientific Software, Santa Fe, NM) (Hawkins et al. 2010). The key residue positions were identified by aligning the models with the structure of human ALDH2 in Pymold (Schrödinger, LLC, NY). The distance between the residues (Thr144 and Asn477) lining the diametrically opposite sides of the channel mouth was measured in Pymold.

Modeling Character Evolution

Maximum-likelihood estimates of ancestral SEC diameters were obtained using the Brownian motion model of character evolution as implemented in the R package ape (Felsenstein 1973; Paradis et al. 2004). For this analysis, we excluded the ALDH-dup lineages and the Glossina moritans and Musca domestica ALDHs. Branch lengths were estimated using ML estimates of synonymous substitution rates provided by PAML. Channel diameters were log-transformed prior to the analysis.

To test whether optimum channel diameter changed in the Drosophila subgenus ALDH-dups, we added the Drosophila subgenus ALDH-dups to the tree used for ancestral character reconstruction. Same procedure was adopted when change of optimum was tested for obscura ALDH-dups but the tests for the change of optimum in ALDH-dups of the two groups were performed separately. The comparison module included in the ape package was used to compare the models with one and two optima. The optimum parameter is designated by the parameter $0$ and the strength of character evolution is designated by a second parameter called $\alpha$. Estimation of both $\alpha$ and $0$ leads to high standard error so $\alpha$ was estimated by recursive function calls over a range of $\alpha$ values until the $\alpha$ that produces the least deviance ($-2\ln L$) was found. This $\alpha$ was then used to estimate $0$. For the two optima model, the MRCA node of the ALDH-dups (supplementary fig. S2, Supplementary Material online) was assigned as the node at which the optimum changes.

Enzyme Assay

To compare ALDH activity of D. pseudoobscura testis and ovary, testis and ovary of the 2- to 4-day-old flies (genome strain MV2-25) were dissected and macerated in grinding buffer (0.25 M sucrose, 5 mM ethylenediaminetetraacetic acid, 15 mM Triton X-100, and 5 mM DTT). Protease inhibitor (Roche Applied Science) was added to the grinding buffer to prevent proteolytic degradation of the enzymes. Next, the unbroken body parts were removed by centrifuging the extracts at 16,000 $g$ for 20 min. The supernatant obtained following centrifugation was used in enzyme activity assays. For D. melanogaster (genome strain), total protein from males and females was extracted following the same procedure as was used for D. pseudoobscura. Protein concentration of the supernatant was measured by Qubit fluorometer following manufacturer’s (Invitrogen, Grand Island, NY) protocol. Enzyme assays were carried out at 25°C in cuvettes of 1-cm path length in 1 ml volume of Na₂HPO₄/NaH₂PO₄ buffer containing 1 mM DTT (pH = 8.5). Pyrazole (0.02 M) was added to the assay buffer to inhibit aldehyde dehydrogenase activity of the enzyme alcohol dehydrogenase (ADH). Acetaldehyde and hexanal were purchased from Sigma-Aldrich (St Louis, MO) and added to the assay mixture at a final concentration of 1 mM ([s] $\gg$ $K_m$) (Rothacker and Ilg 2008). Turnover rates of ALDH were estimated by change in NAD+ concentration as measured by change in absorbance at 340 nm ($E_{340} = 6,220 M^{-1} cm^{-1}$) using Ultrospec spectrophotometer (GE Healthcare Bio-Sciences, Pittsburgh, PA).

Supplementary Material

Supplementary table S1 and figures S1 and S2 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This work was supported by National Science Foundation (0623268) and National Institute of Health (R01AA016178) grants to J.D.F. The authors thank Yasir Ahmed for providing...
unpublished data, and Ryan Bickel for his comments on an early version of this manuscript.

References


doi:10.1093/molbev/msu407

