J Mol Evol (1997) 44:23–32

JOURNAL OF MOLECULAR EVOLUTION

Selection and Methionine Accumulation in the Fat Body Protein 2 Gene (FBP2), a Duplicate of the *Drosophila* Alcohol Dehydrogenase (ADH) Gene

Gilles K. Meghlaoui, Michel Veuille

Institut d'Ecologie, Université Pierre et Marie Curie, 4 Place Jussieu, 75252 Paris Cedex 05, France

Received: 12 December 1994 / Accepted: 15 April 1996

Abstract. The *Drosophila* fat body protein 2 gene (*Fbp2*) is an ancient duplication of the alcohol dehydrogenase gene (*Adh*) which encodes a protein that differs substantially from ADH in its methionine content. In *D. melanogaster,* there is one methionine in ADH, while there are 51 (20% of all amino acids) in FBP2. Methionine is involved in 46% of amino acid replacements when *Fbp2* DNA sequences are compared between *D. melanogaster* and *D. pseudoobscura.* Methionine accumulation does not affect conserved residues of the ADH-ADH^r-FBP2 multigene family. The multigene family has evolved by replacement of mildly hydrophobic amino acids by methionine with no apparent reversion. Its short-term evolution was compared between two *Drosophila* species, while its long-term evolution was compared between two genera belonging respectively to acalyptrate and calyptrate Diptera, *Drosophila* and *Sarcophaga.* The pattern of nucleotide substitution was consistent with an independent accumulation of methionines at the *Fbp2* locus in each lineage. Under a steady-state model, the rate of methionine accumulation was constant in the lineage leading to *Drosophila,* and was twice as fast as that in the calyptrate lineage. Substitution rates were consistent with a slight positive selective advantage for each methionine change in about one-half of amino acid sites in *Drosophila.* This shows that selection can potentially account for a large proportion of amino acid replacements in the molecular evolution of proteins.

Key words: Selection — Molecular evolution — *Drosophila melanogaster* — *Drosophila pseudoobscura* — *Sarcophaga peregrina* — Methionine — *Adh* gene — *Fbp2* gene — Gene duplication

Correspondence to: M. Veuille

Introduction

An important problem in molecular evolution is to determine what proportion of selective changes are involved in protein evolution (Gillespie 1991). According to the neutral theory (Kimura 1983), the rate of molecular evolution depends only on the balance between disadvantageous and neutral mutations. The number of advantageous changes would be so small as to be negligible in comparison to neutral changes. As a consequence, the rate of molecular evolution would never exceed the mutation rate, and would be equal in each lineage to the neutral mutation rate. McDonald and Kreitman (1991) showed that this model does not explain the observed difference between polymorphism and divergence in the ratio of replacement vs synonymous substitutions. This suggests that the contribution of Darwinian selection to protein evolution is substantial. An elaboration of the neutral model suggested by Ohta (1973, 1992, 1993a,b) challenges this explanation. The rate of evolution would depend on population size in each lineage, contrary to Kimura's (1983) conception, and differently affect replacements and synonymous substitutions. A major difficulty in assessing the contribution of selection to molecular evolution is our inability to assess the selective consequences of most amino acid changes. The function of ''active sites'' is known in a few proteins, while the consequences of amino acid replacement for minor sites are still a matter of dispute.

We investigated this question by studying systematic biases in amino acid changes in a family of proteins. The FBP2 protein (formerly called P6, we hereafter follow a nomenclature introduced by Lindsley and Zimm 1992) is a product of the *Fbp2* locus, a member of the *Adh* gene family. It is expressed in the *Drosophila* fat body during the third larval instar (Lepesant et al. 1982; Deutsch et al. 1989), along with alcohol dehydrogenase, where it represents up to 1% of soluble proteins. Its amino acid similarity to ADH is low (28%), but the two proteins have retained structural homologies (Rat et al. 1991). The sequence and the expression pattern of this gene are known in *D. melanogaster* and in *Sarcophaga peregrina* (Matsumoto et al. 1985). *Drosophila* belongs to the acalyptrate Diptera, while *Sarcophaga* belongs to the calyptrate Diptera, along with the house fly.

FBP2 shows a very high methionine content. Methionine generally occurs at a low frequency in proteins, its value ranging from 1.6% to 3.0%, depending on whether conceptually translated or mature proteins are considered (Simon and Cserzö 1990). It is coded by only one codon, ATG, which is also the start codon of proteins. If the start codon is excluded, there is no methionine in ADH. There are 51 methionines (20%) in *D. melanogaster* FBP2 (Rat et al. 1991), and 29 (11.3%) in *S. peregrina* (Matsumoto et al. 1985). Most of these methionines are at different positions in the two species. Only 13 of them occupy homologous codons (Rat et al. 1991). Due to their different pattern of occurence we hypothesise that methionine replacements probably involved independent events in the two lineages. Methionine is involved in 46 out of the 112 amino acid changes recorded between the two species. This trend to accumulate methionines suggests the operation of Darwinian selection. In order to determine whether selection for methionine is still continuing or is at rest in *Drosophila*, we cloned and sequenced *Fbp2* in a species close to *D. melanogaster*, *D. pseudoobscura*. We then estimated the rate of change in methionine content over time and over lineages. These data allowed us to address some predictions of molecular evolutionary models.

Materials and Methods

Unless otherwise noted, nucleic acids were handled according to standard protocols (Sambrook et al. 1989). *D. pseudoobscura* nucleic acids were isolated from the JR38 line (Veuille and King 1995).

Probe Labeling. A 322-bp homologous probe was obtained by PCR amplification of *D. pseudoobscura* genomic DNA between primers (5′-GCATGGAGAACGTGGAGATG-3′ and 5′-GTCTTGTCCATGTAGGGCAT-3′) designed from an alignment of *D. melanogaster* and *S. peregrina Fbp2*. They lie in two conserved regions of the gene. Homology of this fragment with *Fbp2* was checked by sequencing. It encompasses a 60-bp intron in *D. melanogaster* and a 55-bp intron in *D. pseudoobscura*. The probe was labeled by random-hexamer priming using [a-32P]dCTP (ICN biochemicals).

cDNA Library Screening. Total RNA was isolated from dissected fat bodies of wandering third-instar larvae. Poly-A+ mRNA was extracted on an aligo-dT column. A cDNA library was constructed by inserting size-fractionated cDNA into an Uni-Zap XR-vector according to supplier's instructions (Stratagene). Between 150,000 and 200,000 recombinant phages were screened. The labeled probe was hybridized to phage DNA fixed to nitrocellulose filters. Hybridization was conducted overnight at 42°C in 50% formamide, 10% Na-dextran sulfate. Filters were washed twice for 5 min in 0.1% SDS 2 × SSC at room temperature and twice for 15 min in 0.1% SDS 0.1 × SSC at 55°C. Filters were autoradiographed, and positive plaques were transformed into SK + pBluescript plasmids using helper phage *f*1 according to supplier's instructions (Stratagene).

Sequencing. Thirty-four positive clones were isolated. The sequence of six clones was determined using the dideoxy-chain termination method. Sequences were aligned to the *Fbp2* cDNA sequence of *D. melanogaster* (Rat et al. 1991) and *S. peregrina* (Matsumoto et al. 1985).

Statistical Analysis of Amino Acid and Nucleotide Sequences. The first amino acid of all polypeptide sequences (a methionine start codon) was discarded in statistical analyses of amino acid replacement. Likewise, nucleotides involved in intron/exon splice junctions were not counted in silent sites. The codon usage table of a given gene during the evolution of the *melanogaster* and *pseudoobscura* lineages was taken as the joint distribution of codons for this gene in the two species.

Results

Structure of the Fbp2 *cDNA in* D. pseudoobscura

The nucleotide sequence at the ends of the cDNA was identical in the six clones chosen for detailed analysis. The complete cDNA is 906 bp long (Fig. 1). An ATG was found at position 63. It initiates a 256-codon ORF. A polyadenylation signal (AATAAA) begins 33 bp downstream of the stop codon (TAG), and a poly-A tail begins 17 bp further downstream.

Nucleotide Sequence Comparison with D. melanogaster

The coding region shows 115 substitutions out of 768 nucleotides (85.0% similarity). They occur unequally in the three codon positions ($\chi^2 = 86.1$; $df = 2$; $P < 0.001$); most occur in the third position. The numbers of transversions (54) and transitions (61) are similar, deviating significantly from the expected 2:1 ratio of transversions/transitions ($\chi^2 = 19.2$, $df = 1$, $P < 0.001$). Within the transition class, T-C and A-G substitutions were equivalent ($\chi^2 = 0.6$; $df = 1$; N.S.). The most frequent transversions were G-C (20 out of 54), but a test for equal frequency in the four kinds of transversions showed no significant difference ($\chi^2 = 6.3$; $df = 3$; N.S.). In the coding region, 145.9 out of the 768 sites are effectively silent according to Li's (1993) definition. There are 76 synonymous substitutions (52.1%) between *D. pseudoobscura* and *D. melanogaster*. Among the remaining

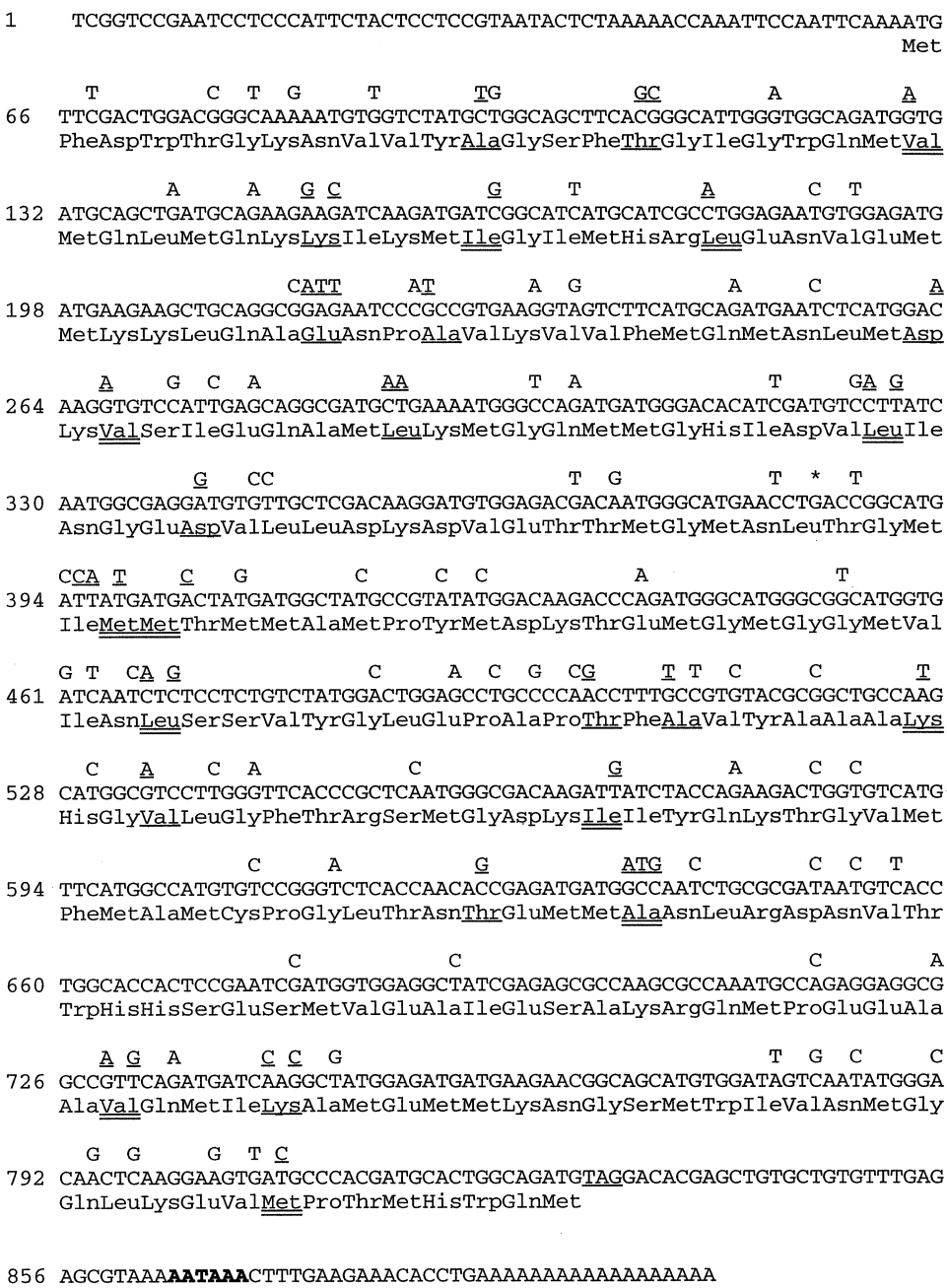

Fig. 1. Sequence of *Fbp2* cDNA in *D. pseudoobscura*. The conceptual amino acid sequence is shown below the nucleotide sequence. It begins at position +63 and ends at +830. A polyadenylation signal is at +866 (*bold letters*). Coding region nucleotide changes with *D. melanogaster* are shown above the nucleotide sequence. Amino acid replacements are *underlined.* Changes involving methionine are *double underlined. Star:* end of the first exon in *D. melanogaster,* corresponding to an intron in *D. pseudoobscura.*

622.1 nonsynonymous sites, the two sequences differ at 39 sites (6.3%). Synonymous substitutions are randomly distributed over the two exons ($\chi^2 = 0.06$; $df = 1$; N.S.), as are nonsynonymous substitutions ($\chi^2 = 0.44$; $df = 1$; N.S.).

Codon usage is highly biased in some *Drosophila* genes (Sharp and Li 1989). A bias toward C- and G-ending codons was observed in *Fbp2:* 73.6% in *D. melanogaster* and 72.6% in *D. pseudoobscura* when excluding methionine from the calculation. The same trend has been observed for *Adh* in *D. melanogaster* (79.2%) and in *D. pseudoobscura* (74.3%).

Amino Acid Comparison with D. melanogaster

In a comparison of FBP2 sequences between *D. melanogaster* and *D. pseudoobscura* 28 amino acid substitutions were observed, 13 of which corresponded to a methionine in one sequence. Ten methionines were present

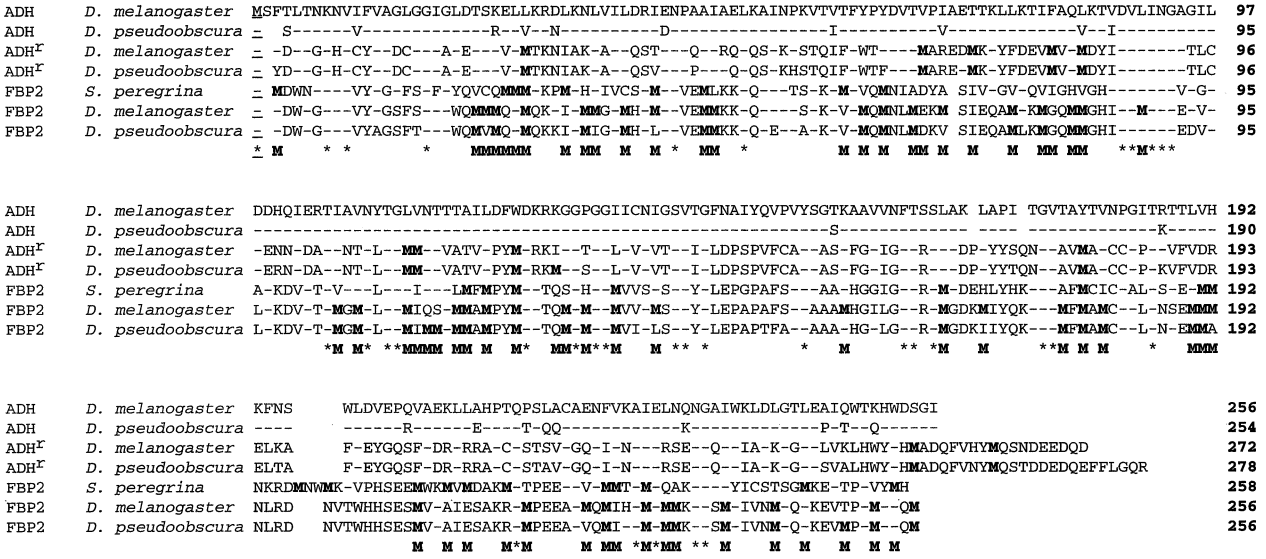

Fig. 2. Alignment of the FBP2-ADH-ADH^r multigene family. *Bold letters:* methionines (excluding start codon). *Dashes:* same amino acids as in *D. melanogaster* ADH. *Gaps:* insertions/deletions. *Bold ''M'' on lower line:* methionine site present in at least one sequence (67 sites). *Stars on lower line:* same amino acids in all sequences (32 sites).

only in the *D. melanogaster* FBP2 sequence and three were present only in *D. pseudoobscura.* Accumulation of methionines between *D. melanogaster* and *D. pseudoobscura* seems to have proceeded independently in the two species, as already noted between *D. melanogaster* and *S. peregrina.*

There is no indication that methionines result from a special class of mutation. Among single-step changes, two occurred in first position (mutation toward A), three in second position (mutation toward T), and two in third position (mutation toward G). Six methionine changes involved at least two nonsynonymous mutation steps; 17.2% of *Fbp2* amino acids (44 out of 256) are methionines in *D. pseudoobscura.*

According to several other parameters, the amino acid structure of the two proteins is similar. The hydropathy profile of FBP2 in *D. pseudoobscura* (not shown) is identical to that formerly published for the two other species (Rat et al. 1991). Its acid/base content (aspartic and glutamic acids/arginine and lysine) is similar and intermediate (24/22) between that of *D. melanogaster* (23/20) and that of *S. peregrina* (21/22). The evolution of the protein has been conservative in respect to the distribution of hydrophobic and hydrophylic residues, and to the electric charge of the protein.

Comparison of Proteins From the ADH-ADHdup-FBP2 Multigene Family

The phyletic diagram obtained from amino acid divergences at nonmethionine positions (Fig. 3) shows that FBP2 diverged from ADH at about the same time as the conceptual protein (ADH^r) encoded by another *Adh* duplicate, *Adh^r* (Schaeffer and Aquadro 1987; Kreitman and Hudson 1991). An alignment of the family of sequences is shown in Fig. 2. ADH and ADH^r from *D. melanogaster* and *D. pseudoobscura* were compared to the three available sequences of FBP2. ADH was taken as the reference. There are 33 strictly conserved sites in this set of sequences. Among these, nine are glycines and four are valines. Glycine is a small nonpolar amino acid that can fit in inner parts of the folded protein structure (Jörnvall et al. 1984). In order to determine whether the conservation of sites was coincidental, χ^2 tests were performed in *D. melanogaster,* taking each protein in turn (ADH, ADH^r or FBP2) and testing for random association of conserved and replaced sites with the other two proteins. The difference was always highly significant whether methionine positions were included ($\chi^2 > 28$, 1 *df*, $P < 0.001$) or excluded ($\chi^2 > 38$, 1 *df*, $P < 0.001$).

Methionines are found in 67 positions in the entire multigene set. There are some methionines in ADH^r for *D. melanogaster* (4.4%) and *D. pseudoobscura* (4.7%). Three methionines are aligned in all ADH^r and all FBP2 sequences. This indicates an accumulation of methionine before divergence between *Adh^r* and *Fbp2.* The three matching methionines between ADH^r and FBP2 suggest either that ADH lost methionines which were present in the common ancestor or that ADH^r is evolutionarily closer to FBP2 than to ADH. Figure 3 clearly shows that methionine fixation was a parallel trend in all FBP2 lineages, since most of them are present at different positions.

Four positions out of 67 depart from a cladistic pattern of methionine accumulation. Two positions show methionines in all ADH^r and FBP2 sequences, except in *S. peregrina* FBP2. Two positions show methionines in FBP2 sequences, except in *D. pseudoobscura.* These

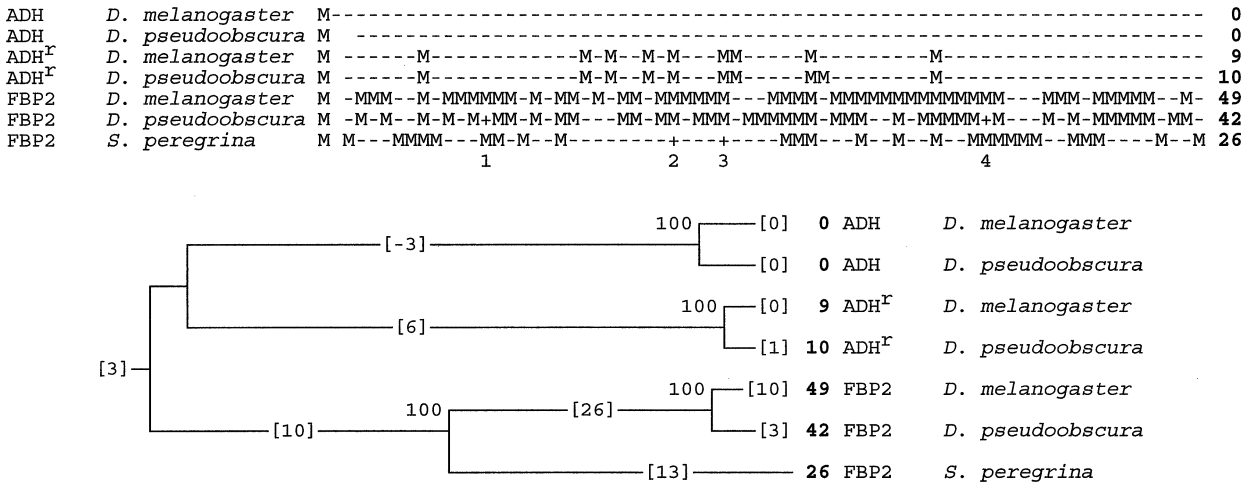

Fig. 3. Patterns of methionine replacement in the FBP2-ADH–ADH^r family. The upper part of the figure shows the alignment of amino acid residues having a methionine in at least one sequence. The *number at the end of each line* is the number of methionines minus both the start codon and those codons that are deleted in at least one aligned sequence. *Dashes* correspond to nonmethionine amino acids. *Numbers 1–4* on the bottom line correspond to ''ambiguous'' cases—that is, to positions where a methionine was either lost in one sequence (''+'') or independently acquired by the others. The lower part of the figure shows the same pattern as projected on an UPGMA tree based on the number of nonmethionine amino acid differences between the aligned proteins. The tree was obtained using the pairwise comparison option of program MEGA (Kumar et al. 1993) for Poisson-corrected differences. Positions where a methionine residue was observed in at least one gene were excluded. *Numbers above nodes* are boostrap values over 100 replicates. *Boxes* show the number of methionine positions separating successive nodes. The *zero* is for ADH. This is why the number of methionines increases from the left (root) to the right (present), except for the ADH branch. The four ambiguous positions are considered to be cases of homoplasy between nonadjacent branches (gain of methionine) in this figure.

four cases may be instances of losses of a methionine in one lineage. Given the large increase in methionine number in all lineages, they may also be instances of parallel gains in methionines in the other genes.

What Kinds of Amino Acids Are Substituted for Methionine?

Methionine is a hydrophobic amino acid. The comparison of *D. melanogaster* and *D. pseudoobscura* sequences shows that in nine cases out of 13, methionine replaced another hydrophobic residue (valine, isoleucine, alanine or leucine, hereafter ''VIAL'' amino acids); it replaced another kind of amino acid (threonine, glycine, serine, lysine) in only four cases. Both the conservation of biochemical properties and the mutation pattern can explain this distribution. Only one mutation step leads to methionine (ATG) from isoleucine codons (ATT, ATA, ATC; 13 residues in *D. melanogaster Fbp2*), and from the frequently used codons for leucine (CTG, six residues out of 11 in *D. melanogaster Fbp2*, none for TTG) and valine (GTG, eight residues out of 19 in *D. melanogaster Fbp2*). However, lysine, a strongly hydrophylic amino acid, also differs from ATG by one step in its mainly used codon (AAG, 14 residues out of 16 in *D. melanogaster Fbp2*); and the same is true for threonine (ACG, four residues out of 12 in *D. melanogaster Fbp2*). The relatively small contribution of lysine suggests that all methionine mutations are not advantageous. The small number of changes involved, and the bias introduced by the fact that most single-step mutations lead to another hydrophobic amino acid, prevent any further conclusions from being drawn. The conservative nature of amino acid replacements as to hydrophylicity was therefore studied at a broader scale of divergence, between FBP2 and ADH in *D. melanogaster*.

Comparisons between highly divergent pairs of sequences cannot be done on a site-by-site basis, since multiple changes may be involved. The net gain or loss in a given class of amino acids was estimated for the protein as a whole. The amount of VIAL amino acids is depleted in FBP2 (57 residues) as compared to ADH (95 residues). The loss in VIAL amino acids thus balances 76% of gains in methionine. Figure 4 compares the proportion of the 20 kinds of amino acids in FBP2 and ADH. Four amino acids are underrepresented in FBP2: threonine, isoleucine, alanine, and leucine, (hereafter ''TIAL'' amino acids). They decrease from 100 in ADH to 50 in FBP2, thus seeming to compensate the gain of 50 methionines. The reduction in isoleucine, alanine, and leucine was expected from their strong hydrophobicity. However, threonine is strongly affected by methionine replacement, while valine, a more hydrophobic amino acid of similar form and size, is only mildly affected.

We also examined the replacement of methionine by amino acids of mild hydrophobicity. We considered the impact of each amino acid on the overall hydropathy of the protein. Conservation in hydropathy can be assessed using Hopp and Woods's (1981) index, which assigns a value to each amino acid. The scale ranges from −3.5 for

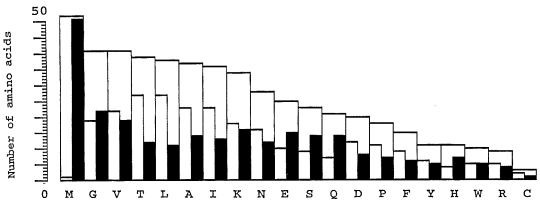

Fig. 4. Number of residues for each kind of amino acid in *D. melanogaster* ADH and FBP2. Amino acids are ranked according to their order in the total distribution (*higher histogram*). *Black rectangles:* FBP2 residues; *empty rectangles:* ADH residues.

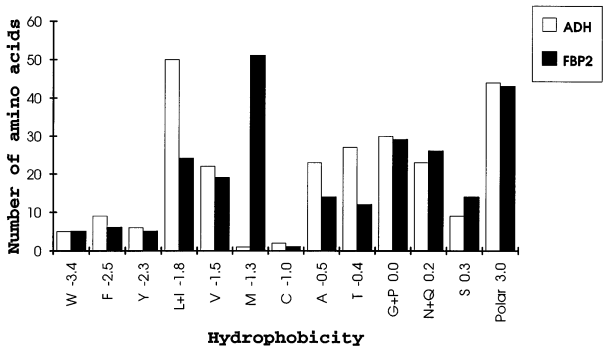

Fig. 5. Proportion of each amino acid from each class of hydropathy in FBP2 and ADH (Hopp and Wood 1981). Values range from extreme hydrophobicity (left) to extreme hydrophylicity (right). Amino acids are shown by their *one-letter code*. Polar amino acids are aspartic acid, glutamic acid, lysine, arginine, and histidine.

extreme hydrophobicity (tryptophane), to +3 for extreme hydrophylicity (lysine, arginine, histidine, aspartic acid, and glutamic acid). The score of methionine is −1.3. The mean and standard deviation of individual amino acid scores over 255 residues are, respectively, −0.23 ± 1.72 for ADH and −0.22 ± 1.67 for FBP2 ($t = 0.915$, N.S.; $F = 1.07$, N.S.). The two values are virtually identical, meaning that methionine accumulation was achieved with a strict conservation of hydrophobicity in the polypeptide. Figure 5 shows the contribution of each amino acid to hydropathy in ADH and FBP2. On this scale, leucine and isoleucine (−1.8) are more hydrophobic than methionine, while alanine (−0.5) and threonine (−0.4) are less. The two groups of amino acids were depleted to a similar extent while the proportion of methionine was increased. The relatively small decrease in valine is probably due to its involvement in four conserved sites. The smaller decrease in alanine than in threonine, two amino acids of similar hydrophobicity (−0.5 vs −0.4), probably results from the fact that a single mutation step can lead from threonine to methionine (ACG to ATG), while two changes, at least, are necessary between alanine and methionine (GCG to ATG). When all six amino acids are considered together (methionine, and valine, threonine, isoleucine, alanine, leucine), their mean contribution to hydrophobicity is remarkably similar in the two genes: −1.19 ± 0.79 in ADH ($N = 123$) and −1.24 ± 0.71 in FBP2 ($N = 120$).

Discussion

The evolution of FBP2 shows a tendency to accumulate one type of amino acid, methionine. Before this study, it was known that the number of methionine residues varied from one—the start codon—in mature ADH to 51 in FBP2 (Rat et al. 1991), while the two proteins are only 256 amino acids long. This suggested that there was a tendency toward either a loss of methionine in ADH or a gain of methionine in FBP2 or both. We confronted these hypotheses by comparing these proteins with related proteins from the ''short-chain dehydrogenase'' family. We counted the number of methionines in nine proteins from this group, including *Bacillus megatherium* glucose dehydrogenase, *Klebsiella aerogenes* ribitol dehydrogenase, *Escherichia coli* glucitol-6-phosphate dehydrogenase, *Pseudomonase pseudoalcaligenes* dihydrodiol dehydrogenase, human 17-β-hydroxysteroid dehydrogenase, mouse adipocyte glucocorticoid controlled gene, *Rhyzobium meliloti* nodulation protein G, *Eubacterium* 27-kD protein, and *Steptomyces coelicolor* Act III gene protein (references in Rat et al. 1991). These proteins show amino acid signatures that are characteristic of the group, and are very divergent for the rest of their sequence. The average number of methionines in these proteins is 6.33 (SD = 1.87), the proportion of methionines among amino acids being 2.35% (57 out of 2419 amino acids). The number of methionines is significantly lower than this in *Drosophila melanogaster* ADH ($G = 10.954$ with Williams correction, 1 *df*, $P < 10^{-3}$), significantly larger in FBP2 ($G > 50$, $P \ll 10^{-3}$), and not significantly different in *Drosophila melanogaster* ADH^r ($G = 0.925$, $P > 0.30$). This suggests that there has been an accumulation of methionines in FBP2 and a loss of methionines in ADH as compared to the ancestral protein. Available data also show that this process is relatively recent since the methionine content differs in the Calyptrate *Sarcophaga peregrina* (29 residues, 11.3%) and in the Acalyptrate *Drosophila melanogaster* (51 residues, 20%). The first purpose of the present study was to check whether methionine accumulation was still active or at rest among drosophilids. The observed proportion of amino acid changes involving methionine between *D. melanogaster* and *D. pseudoobscura* (13 out of 28, 46%) shows that changes in the number of methionine are active, or were active until recently, in FBP2. It is worth noting that in contrast to this, no amino acid replacement involving methionine was observed in ADH between these two species, where this amino acid is virtually lacking, except for one residue at position 119 in species related to *D. mulleri* (Fisher and Maniatis 1985).

Conservation of all the other properties of the polypeptide sequence is equally striking in this gene. The accumulation of one kind of hydrophobic residue over one-fifth of its length could be expected to substantially

change the hydropathy properties of the polypeptide. However, detailed analysis shows that losses of amino acids equally involve residues that are either slightly more hydrophobic (leucine, isoleucine) or slightly less hydrophobic (alanine, threonine) than methionine. This results in hydrophobicity being strictly conserved, both in its mean and in its standard deviation, as compared to ADH. Conservation of other parameters had previously been deduced from the sequence (Rat et al. 1991). The hydropathy profile is the same for all FBP2s, and resembles that of ADH. The subunit binding motif is conserved. The NAD^+ binding site is closer to the consensus of short-chain dehydrogenases in FBP2 than in ADH.

We advance no hypotheses as to why methionines have accumulated in FBP2 and not in ADH. ADH, ADH^r, and FBP2 belong to ''short-chain'' dehydrogenases, a group also comprising polyol dehydrogenases in bacteria and steroid dehydrogenases in mammals (see review in Rat et al. 1991). No substrates are known for ADH^r and FBP2, and we do not know if the common ancestor of these three proteins was an alcohol dehydrogenase. Tamura et al. (1983) consider it a storage protein. This is a possibility, since FBP2 makes up a large fraction of soluble proteins in the last-larval-instar fat body both in *Drosophila* (Lepesant et al. 1978) and *Sarcophaga* (Tamura et al. 1983). There is, however, no evidence of a degeneracy of the protein structure. Conserved parts of the putative polypeptide include the NAD^+ coenzyme-binding site (Rat et al. 1991). Kolker and Trifonov (1995) showed that when proteins are produced from an ancient gene fusion, methionines found within the polypeptide may be remnants of former start sites, a mechanism that would apply to 20% of proteins. This hypothesis cannot apply to FBP2, where methionines make up one out of five residues, and frequently occur in tandems of two or three. We can also rule out that these methionines constitute alternative start codons, first because they occur on the whole length of the protein, and second because the FBP2 protein has been well characterized as a single spot in denaturing gel electrophoresis (Lepesant et al. 1978; Tamura et al. 1983).

Rate of Methionine Fixation in FBP2

The very slow accumulation of methionine over time suggests that this residue contributes little to fitness. This can be substantiated by estimating the average selective gain in methionine. The accumulation of methionines in all FBP2 lineages suggests that the number of methionines in the protein acts as a selected quantitative trait. A reasonable hypothesis is that each new methionine change provides some quantum of selective advantage. The rate of methionine accumulation can be calculated from a model based on the observation that a subset of amino acid positions can be replaced by methionines. It is possible to divide FBP2 residues into three categories

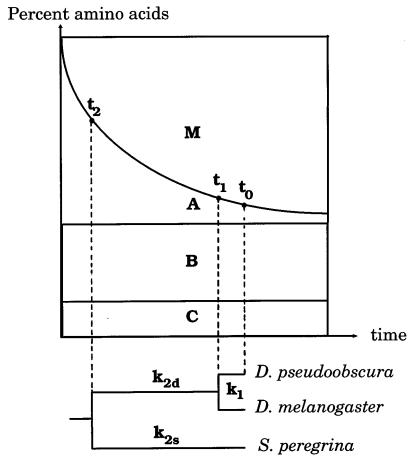

Fig. 6. Graphic model of increase in methionine content over time (see text). *M:* methionine sites. *A:* sites at which a mutation toward methionine is positively selected (''eligible sites''). *B:* sites at which a mutation toward a methionine is negatively selected or neutral. *C:* invariable sites, at which any mutation is negatively selected (including the methionine start codon). t_0: present; t_1: most recent common ancestor (MRCA) between *D. melanogaster* and *D. pseudoobscura;* t_2: MRCA between Acalyptrate and Calyptrate. Rates of substitution k_i are assumed to be constant within lineages. Backward mutations from methionine are assumed to be negligible.

(Fig. 6): Conserved sites are constant in the multigene family; variable sites can change to another amino acid; and eligible sites are those variable sites that can accumulate methionine changes. This allows us to estimate the rate of methionine fixation. Let A_t be the number of eligible sites where no methionine change has occurred at time *t.* Let k_i be the rate of methionine fixation at these sites in lineage *i.* Provided that the selective advantage of each methionine is similar and that there is no backward change, rate k_i can be considered constant. Methionine fixation is an exhaustive process that can be written:

$$A_t = A_0 (1 - k_i)^t \qquad (1)$$

where A_0 is the number of empty sites at the beginning of lineage *i.* Formula (1) cannot be applied in this form, since A_0 is unknown. It cannot be derived from sequence comparison, since methionine replacements can have occurred independently in the same position in two lineages. A derived formula can be used, where methionine replacements are counted only over the eligible sites that have not been replaced by methionines in the other lineage. Let A_i and A_j be the number of eligible sites remaining empty in lineages *i* and *j*, respectively, at time *t.* Let A_{ij} be the number of sites remaining empty in both lineages at time *t.* Assuming that methionine replacements occur independently in the two lineages, k_i can be calculated by comparing A_{ij} to A_j:

$$A_{ij}/A_j = (1 - k_i)^t \qquad (2)$$

Finally, the rate of methionine fixation in lineage *i* is:

$$k_i = 1 - [A_{ij}/A_j]^{1/t} \quad (3)$$

This formula was used after estimating its parameters as follows:

1. The estimate of the number of eligible sites was deduced from the fact that methionine preferentially replaces some kinds of amino acids. It replaces ''VIAL'' amino acids in 76% cases. It also exactly compensates the loss in ''TIAL'' amino acids. This provides an estimate of 125 and 100 eligible sites, respectively, for each hypothesis. This does not imply that FBP2 can accommodate 100 to 125 methionines at a time. These values are at least twice the number of methionines. They only imply that when the protein has accumulated x methionines, any of the remaining eligible sites can provide the next, $x + 1^{th}$, change.

2. The time elapsed since the divergence between *Sarcophaga* and *Drosophila* lineages (hereafter t_2), and since the divergence between *D. melanogaster* and *D. pseudoobscura* lineages (hereafter t_1), is broadly known from the paleontological record. The divergence between acalyptrate and calyptrate flies occurred between 65 and 135 MYA (million years ago) (reviewed by Beverley and Wilson 1984). The divergence between the *melanogaster* group and the *obscura* group probably occurred 20 MYA (Throckmorton, cited in Kreitman 1987). These estimates are necessarily inaccurate, since dipteran fossils are scanty (Grimaldi 1987). We estimated the phyletic distance between genes from the internal standard provided by variable sites in FBP2. Amino acid replacements were counted over sites, excluding all positions where a methionine change had occurred, and all conserved positions of the multigene family. When Poisson corrected for multiple hits, this factor was $t_2/t_1 = 5.04$. Taking the divergence between *D. melanogaster* and *D. pseudoobscura* as 20 MY, the divergence between *Drosophila* and *Sarcophaga* would be about 100 MY. These values may involve some error, but they provide an order of magnitude and are reasonably consistent with paleontological evidence.

The rate of methionine fixation was calculated per eligible site and per 100 MY over lineages. In the acalyptrate lineage, it was based on the mean number of methionines in *D. melanogaster* and *D. pseudoobscura*. The rates calculated under hypotheses H_{100} and H_{125} led to similar results (Table 1). In either case, the rate in the line leading to *Drosophila* since the Calyptrate/Acalyptrate divergence was equal to twice the rate in the line leading to *Sarcophaga* (respectively 0.589 vs 0.297 under hypothesis H_{100}; and 0.403 vs 0.194 under hypothesis H_{125}). The difference was significant in both cases.

The rate of methionine fixation between *D. melanogaster* and *D. pseudoobscura* was calculated from the total number of methionine changes (13) between the two species. We assume that the rate of fixation is the same in the *melanogaster* and the *pseudoobscura* branches. This rate was nearly exactly equal to the rate in the acalyptrate (i.e., *Drosophila*) lineage (0.535 vs 0.589 under hypothesis H_{100}; 0.378 vs 0.403 under hypothesis H_{125}). This rate is not significantly different from the rate in *Sarcophaga* since it is based on small values. It cannot be compared to the acalyptrate rate since the two values are interdependent. They are, however, very close, thus showing that methionine accumulation has undergone no change in rate over time in the *Drosophila* lineage.

Table 1. Rate of methionine fixation per 10^8 years in FBP2[a]

	Hypothesis	
	H_{100}	H_{125}
Sarcophaga	0.297	0.194
lineage	(0.157–0.408)	(0.099–0.275)
Drosophila	0.589	0.403
lineage	(0.450–0.704)	(0.299–0.497)
D. pseudoobscura	0.535	0.378
vs *melanogaster*	(0.272–0.772)	(0.187–0.556)

[a] The rate was calculated per eligible amino acid site for methionine fixation; under hypothesis H_{100} there are 100 eligible sites; under hypothesis H_{125}, there are 125. The rate was calculated according to equation 3, which compensates for the decrease in eligible sites as methionine fixation goes on. The time scale in units of 10^8 years is only intended to give an order of magnitude, since only the relative age of each lineage (see text) is important for comparing species. Confidence limits ($P < 0.05$) are calculated from the binomial distribution on observed methionine sites

Estimation of the Selective Advantage of Methionines in FBP2

The replacement of a subset of FBP2 amino acids by methionines took several million years. This suggests that the mean selective advantage of a methionine change is small. The average advantage can be estimated as follows. Let u_m be the ATG (methionine codon) mutation rate and u_n the synonymous mutation rate in *Fbp2*. Let K_m be the methionine accumulation rate, and K_n the synonymous substitution rate. If synonymous changes are neutral and methionine changes are selective, it can be shown (Kimura 1977) that $K_n = u_n$ and $K_m = 4N_e u_m s$, where N_e is the effective population size and s the mean selective advantage of a methionine change. These formulae combine to give:

$$2N_e s = \tfrac{1}{2} \cdot (K_m/K_n) \cdot (u_n/u_m) \quad (4)$$

In practice, this formula was applied as follows to the divergence between *D. melanogaster* and *D. pseudoobscura:*

1. The u_m/u_n ratio was estimated by performing all possible synonymous substitutions considered as being equally probable in the combined codon usage table of *D. melanogaster* and *D. pseudoobscura* for *Adh* and *Fbp2*. Calculated over the whole sequence, the u_m/u_n ratio is 0.125 in *Adh* and 0.115 in *Fpb2,* and the average

Table 2. Estimates of the parameter $2N_es$, under two hypotheses[a]

	Hypotheses	
	H_{100}	H_{125}
$K_{s(Fbp2)} = 0.9407$	3.694	1.779
$K_{s(lcg)} = 1.6830$	2.064	0.994

Hypothesis H_{100} and H_{125} are as in Table 1. The rate of synonymous substitution per nucleotide, K_s, is either the rate observed in *Fbp2* between *D. melanogaster* and *D. pseudoobscura* (upper line) or the rate observed in a set of "low-constraint genes" (*lcg*) comprised of *Sry*β, *Est5/Est6,* and *Adh-dup* (lower line)

value of 0.120 was retained. This ratio was not applied to the 256 residues of the protein. Our hypothesis is that methionines appear in a subset of the amino acids and do not revert to another amino acid. Therefore, u_n was weighted for the nonmethionine amino acids, while u_m was weighted for the average number of empty (nonmethionine) eligible sites in the two *Drosophila* species—that, is 57.5 under hypothesis H_{100} and 77.5 under hypothesis H_{125}. In this case, the ratio u_m/u_n is 0.0030 under hypothesis H_{100} and 0.045 under hypothesis H_{125}.

2. The rate of methionine fixation was calculated as a rate of decay per amino acid, as calculated in Table 1. In this case, we used the rate for time units of twice 20 million years, corresponding to the divergence between *D. melanogaster* and *D. pseudoobscura.*

3. The rate of synonymous substitution K_s was calculated using Jukes and Cantor's correction according to Kumar et al. (1993) as a per-site synonymous substitution rate between *D. melanogaster* and *D. pseudoobscura.* The rate of synonymous substitution corresponds to a rate of neutral fixation only if synonymous changes are neutral. The codon usage of *Fbp2* and *Adh* (Sharp and Li 1989) is highly biased, and several studies show that synonymous substitutions are probably constrained by selection (Moriyama and Hartl 1993; Carulli et al. 1993; Hartl et al. 1994; Akashi 1995). For this reason, we used two measures of K_s. First, we used its value for *Fbp2*. We calculated K_s for a group of genes with low codon usage: Sry^{β} (Ferrer et al. 1994), *Est-5B/Est-6* (Brady et al. 1990), and Adh^{r} (Schaeffer and Aquadro 1987; Kreitman and Hudson 1991). K_s was then calculated on pooled codons, except for those that were at the intron/exon junction. These genes constitute a lower limit for selection on silent substitution and therefore provide a confidence limit for the estimate of $2N_es$.

Results are shown in Table 2. The value of $2N_es$ is between 2.0 and 3.6, depending on the hypotheses on the saturation in methionine. When the set of low-constraint genes is used as a reference for synonymous substitution, this value is 1 if saturation in methionine is low (H_{125}). According to the neutral model (Kimura 1983), $2N_es = 1$ is the threshold above which an allele can be selected. Below this value, stochastic effects are prevalent, and allele frequencies fluctuate randomly. The selective advantage of methionines, if any, is very small, as expected from the fact that the protein accumulated them very slowly over millions of years. However, the phylogenetic pattern of this accumulation (Fig. 3) shows little evidence of reversion, as if the selective advantage of methionines had remained high enough to exclude random loss. Since any durable decrease of the population size during evolution would also have led to losses of methionines, this also suggests that the effective population size, estimated at around 1–3.25 10^6 between *D. melanogaster* and *D. pseudoobscura* (Berry et al. 1991; Sawyer and Hartl 1992; Riley et al. 1992), has remained high during *Drosophila* evolution.

These calculations are not necessary to show that methionine changes in FBP2 are selective, that they confer a very slight advantage, and that they are lineage dependent: This is apparent from the protein alignment. FBP2 shows that a protein can tolerate selection at 46% of its amino acid replacements. We do not know if these conclusions are valid for proteins where amino acid changes do not involve the same residue. The divergence between *D. melanogaster* and *D. pseudoobscura* FBP2 involved an average of 6.5 selective (i.e., methionine) replacements per lineage in about 20 millions years. This rate could apply to 41,000 proteins of similar size, at the rate of one selective substitution every 300 generations. In other words, selection could be the cause of about half the amino acid substitutions of an organism like Drosophila with a moderate replacement load per generation.

Another question raised by this study is why the rate of methionine fixation is different in *Drosophila* and in *Sarcophaga.* As noted above, the rate of methionine accumulation per time unit can be written $K_m = 4N_eu_ms$. Differences between lineages can result from changes in the selective advantage of a methionine or in the population size. A third possibility is that time units are different since generation time can differ between lineages. No estimates exist for effective population size and generation time in past dipteran lineages. An allozyme based estimate of heterozygosity among calyptrate flies is known from *Musca,* where its value is 0.186 in *M. automnalis,* and 0.115 in *M. domestica* (Krafsur 1993; Black and Krafsur 1985). It is similar to that of *Drosophila melanogaster* (0.132) and *D. pseudoobscura* (0.123) (from a review by Nevo 1978), thus suggesting a similar effective population size. Effective population size is, however, a short-term parameter in evolution. In the *Drosophila* genus, effective size is probably very variable, as suggested by large differences in allozyme heterozygosity varying between 0.252, in *D. affinis,* and 0.025, in *D. orthofasciata* (Nevo 1978). We do not know which of these species, if any, are representative of past lineages. Generation time also differs between *Drosophila* species. In *D. pseudoobscura,* it is about twice as long as in *D. melanogaster.* This may explain why fewer methionines have accumulated in the first species (43)

than in the second (50). Theory predicts that fewer methionines will be present in species with a long generation time and in species with low effective population size. In the first case, the rate of accumulation will be slower. In the second case, methionines will be lost through random drift. Loci that, like *Fbp2*, have measurable low selective effects on molecular changes, are therefore interesting systems for testing predictions of evolutionary theories.

Acknowledgments. We thank Anne-Marie Petitjean and Didier Contamine for technical advice, Matthew Cobb for corrections, and anonymous referees for criticisms that greatly improved the manuscript.

References

Akashi H (1995) Inferring weak selection from patterns of polymorphism and divergence at ''silent'' sites in *Drosophila* DNA. Genetics 139:1067–1076

Berry AJ, Ajioka JW, Kreitman M (1991) Lack of polymorphism on the Drosophila fourth chromosome resulting from selection. Genetics 129:1111–1117

Beverley SM, Wilson AC (1984) Molecular evolution in *Drosophila* and the higher Diptera II. A time scale for fly evolution. J Mol Evol 21:1–13

Black WC, Krafsur ES (1985) Electrophoretic analysis of genetic variability in the house fly (*Musca domestica* L.). Biochem Gen 23: 93–203

Brady JP, Richmond RC, Oakeshott JG (1990) Cloning of the esterase-5 locus from *Drosophila pseudoobscura* and comparison with its homologue in *D. melanogaster*. Mol Biol Evol 7:525–546

Carulli JP, Krane DE, Hartl D, Ochman H (1993) Compositional heterogeneity and patterns of molecular evolution in the *Drosophila* genome. Genetics 134:837–845

Deutsch J, Laval M, Lepesant JA, Maschat F, Pourrain F, Rat L (1989). Laval fat-body gene expression in *Drosophila melanogaster*. Dev Genet 10:220–231

Ferrer P, Crozatier M, Salles C, Vincent A (1994) Interspecific comparison of *Drosophila Serendipity* δ and β: multi-modular structure of these C2H2 zinc finger proteins. J Mol Evol 34:263–273

Fischer JA, Maniatis T (1985) Structure and transcription of the *Drosophila mulleri* alcohol dehydrogenase genes. Nucleic Acids Res 13:6899–6917

Gillespie JH (1991) The causes of molecular evolution. Oxford University Press, New York

Grimaldi DA (1987) Amber fossil Drosophilidae (Diptera), with particular reference to the Hispaniolan taxa. Am Mus Novit 2880:1–23

Hartl DL, Moriyama EN, Sawyer SA (1994) Selection intensity for codon bias. Genetics 138:227–234

Hopp TP, Woods KR (1981) Prediction of protein antigenic determinants from amino acid sequences. Proc Natl Acad Sci USA 78: 3824–3828

Jörnvall HH, von Bahr-Lindström H, Jany KD, Ulmer W, Fröschle M (1984) Extended superfamily of short alcohol-polyol-sugar dehydrogenases: structural similarities between glucose and ribitol dehydrogenases. FEBS Lett 165:190–196

Kimura M (1977) Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. Nature 267:75–276

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge, England

Kolker E, Trifonov EN (1995) Periodic recurrence of methionines: fossils or gene fusion? Proc Natl Acad Sci USA 92:557–560

Krafsur ES (1993) Allozyme variation in stable flies (Diptera: Muscidae). Biochem Genet 31:231–240

Kreitman M (1987) Molecular population genetics. In: Harvey PH, Partridge L (eds) Oxford surveys in evolutionary biology, vol. 4. Oxford University Press, Oxford, pp 38–60

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

Kumar S, Tamura S, Nei M (1993) MEGA: molecular evolutionary genetics analysis, version 1.0. The Pennsylvania State University

Lepesant JA, Kejzlarova-Lepesant J, Garen A (1978) Ecdysone-inducible functions of larval fat bodies in *Drosophila.* Proc Natl Acad Sci USA 75:5570–5574

Lepesant JA, Levine M, Garen A, Kjezlarvoa-Lepesant J, Rat L, Somme-Martin G (1982) Developmentally regulated gene expression in *Drosophila* larval fat bodies. J Mol Appl Genet 1:371–383

Li W-H (1993) Unbiased estimation of the rates of synonymous substitution. J Mol Evol 36:96–99

Lindsley DL, Zimm GG (1992) The genome of Drosophila melanogaster. Academic Press, San Diego, USA

McDonald JH, Kreitman M (1991) Adaptive protein evolution at the *Adh* locus in *Drosophila.* Nature 351:652–654

Matsumoto N, Sekimizu K, Soma GI, Ohmura Y, Andoh T, Nakanisji T, Obinata M, Natori S (1985) Structural analysis of a developmentally regulated 25-kDa protein gene of *Sarcophaga peregrina.* J Biochem 97:1501–1508

Moriyama EN, Hartl D (1993) Codon Usage bias and base composition of nuclear genes in *Drosophila.* Genetics 134:847–858

Nevo E (1978) Genetic variation in natural populations: patterns and theory. Theor Popul Biol 13:121–177

Ohta T (1973) Slightly deleterious mutant substitution in evolution. Nature 246:96–98

Ohta T (1992) The nearly neutral theory of molecular evolution. Annu Rev Ecol Syst 23:263–286

Ohta T (1993a) Amino acid substitution at the *Adh* locus of *Drosophila* is facilitated by small population size. Proc Natl Acad Sci USA 90:4548–4551

Ohta T (1993b) An examination of the generation-time effect on molecular evolution. Proc Natl Acad Sci USA 90:10676–10680

Rat L, Veuille M, Lepesant J-A (1991) *Drosophila* fat body protein *P6* and alcohol dehydrogenase are derived from a common ancestral protein. J Mol Evol 33:94–203

Riley MA, Kaplan SR, Veuille M (1992) Nucleotide polymorphism at the xanthine dehydrogenase locus in *Drosophila pseudoobscura.* Mol Biol Evol 9:56–62

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. 2nd ed. Cold Spring Harbor Laboratories, Cold Spring Harbor, NY

Sawyer SA, Hartl DL (1992) Population genetics of polymorphism and divergence. Genetics 132:1161–1176

Schaeffer SW, Aquadro CF (1987) Nucleotide sequence of the *Adh* gene region of *Drosophila pseudoobscura:* evolutionary change and evidence for an ancient gene duplication. Genetics 117:61–73

Sharp PM, Li W-H (1989) On the rate of DNA sequence evolution in *Drosophila.* J Mol Evol 28:389–402

Simon I, Cserzö M (1990) Rapid evolution of the composition of proteins. Trends Biochem Sci 15:135–136

Tamura H, Tahara T, Kuroiwa A, Obinata M, Natori S (1983) Differential expression of two abundant messenger RNAs during development in *Sarcophaga peregrina.* Dev Biol 99:145–151

Veuille M, King LM (1995) Molecular basis of polymorphism at the esterase-5B locus in *Drosophila pseudoobscura.* Genetics 141: 255–262