

Evolutionary Rearrangement of the Amylase Genomic Regions Between *Drosophila melanogaster* and *Drosophila pseudoobscura*

S. W. SCHAEFFER, M. J. BERNHARDT, AND W. W. ANDERSON

From the Department of Biology and Institute of Molecular Evolutionary Genetics, Pennsylvania State University, 208 Erwin W. Mueller Laboratory, University Park, PA 16802-5301 (Schaeffer), and Department of Genetics, University of Georgia, Athens, GA 30602 (Bernhardt and Anderson). M. J. Bernhardt is currently at 92 New Wickham Dr., Penfield, NY 14526.

Address correspondence to Stephen W. Schaeffer at the address above, or e-mail: sws4@psu.edu.

Abstract

Two *Drosophila pseudoobscura* genomic clones have sequence similarity to the *Drosophila melanogaster* amylase region that maps to the 53CD region on the *D. melanogaster* cytogenetic map. The two clones with similarity to amylase map to sections 73A and 78C of the *D. pseudoobscura* third chromosome cytogenetic map. The complete sequences of both the 73A and 78C regions were compared to the *D. melanogaster* genome to determine if the coding region for amylase is present in both regions and to determine the evolutionary mechanism responsible for the observed distribution of the amylase gene or genes. The *D. pseudoobscura* 73A and 78C linkage groups are conserved with the *D. melanogaster* 41E and 53CD regions, respectively. The amylase gene, however, has not maintained its conserved linkage between the two species. These data indicate that amylase has moved via a transposition event in the *D. melanogaster* or *D. pseudoobscura* lineage. The predicted genes within the 73A and 78C regions show patterns of molecular evolution in synonymous and nonsynonymous sites that are consistent with previous studies of these two species.

Conserved synteny occurs when two or more homologous genes are located on the same chromosome in two or more species. Conserved linkage occurs when two or more homologous genes are syntenic and are in the same order on the chromosome in two or more species (Ehrlich et al. 1997). The genes within the *Drosophila* genome are organized into six chromosomal elements or syntenic groups that have been conserved during evolution (Ashburner 1989; Muller 1940; Sturtevant and Novitski 1941). The hybridization of conserved DNA probes to the polytene chromosomes of the salivary chromosomes of *Drosophila* have supported the conserved synteny of the six chromosomal elements (Ranz et al. 1997, 2001; Segarra and Aguade 1992; Segarra et al. 1995, 1996; Steinemann 1982; Steinemann et al. 1984), but have suggested that paracentric inversions have provided a powerful mechanism to break up conserved linkage groups (Ranz et al. 2001).

The third chromosome of *Drosophila pseudoobscura* is polymorphic for more than 30 gene arrangements that were generated through a series of overlapping paracentric

inversions. The α -amylase gene maps to the third chromosome of *D. pseudoobscura* (Yardley 1974) and was cloned to test hypotheses about the origin of paracentric inversions and how evolutionary forces have maintained the gene arrangements (Aquadro et al. 1991). The α -amylase region was cloned from *D. pseudoobscura* (Aquadro et al. 1991) using DNA clones that contained the *Drosophila melanogaster* amylase region (Gemmill et al. 1985). The putative amylase clones that were isolated from the *D. pseudoobscura* genomic library mapped to two different locations on the *D. pseudoobscura* third chromosome. The first clone family mapped to section 73A (Aquadro et al. 1991), while the second family of clones mapped to section 78C (Bernhardt 1997).

The 73A region encodes one to three amylase genes (Aquadro et al. 1991; Brown et al. 1990; Popadic et al. 1996) depending on which of the 30 gene arrangements is examined. The common ancestor of *D. pseudoobscura* gene arrangements is postulated to have had three α -amylase genes, and several descendant gene arrangements sub-

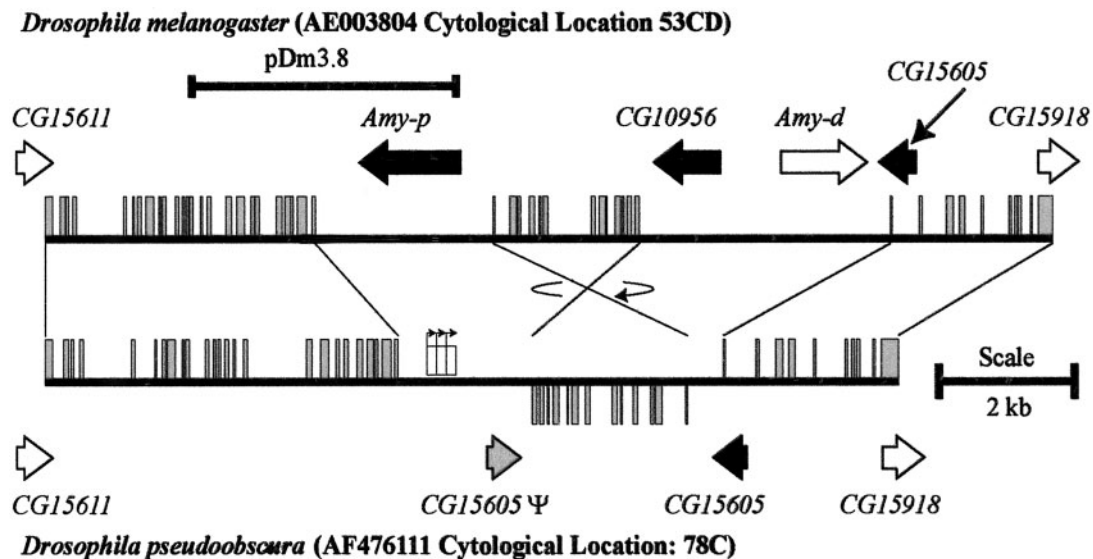


Figure 1. Comparison of the 53CD region of *D. melanogaster* with the 78C region of *D. pseudoobscura*. A horizontal line represents the sequence of each species. The shaded boxes indicate conserved sequence blocks between the two species. Conserved blocks shown above the line indicate that the sequence is on the positive DNA strand and shown below the line indicate that the sequence is on the opposite strand. The lines between the two sequences help to line up the conserved sequences between the two species. Gene regions are shown on the plus strand (black arrow) and minus strand (gray arrow). The central section of the sequence is inverted in *D. pseudoobscura* and is indicated by circular arrows. Three randomly repeated sequences are drawn as blocks with arrows on the *D. pseudoobscura* sequence. The original *D. melanogaster* plasmid clone that was used to isolate the F6 clone is shown above the *D. melanogaster* sequence and indicated by pDM3.8.

sequently lost genes through deletion events (Popadic et al. 1996). Only a single gene, *Amy 1*, is associated with all gene arrangements and is the only gene that produces a functional AMY protein (Hawley et al. 1990). The *D. melanogaster* amylase region is located at section 53CD on the cytogenetic map (Flybase Consortium 1999) and encodes two functional amylase genes, a proximal and distal amylase (Bahn 1967; Boer and Hickey 1986; Levy et al. 1985). Phylogenetic analysis shows that the duplication events of the *D. melanogaster* and *D. pseudoobscura* α -amylase genes occurred after the two species diverged (Figure 1) (Brown et al. 1990). What was not clear is whether the 78C cytogenetic region of *D. pseudoobscura* also encodes a functional α -amylase gene.

There are two broad explanations for why two regions of the third chromosome in *D. pseudoobscura* are similar to α -amylase. First, one of several α -amylase copies was moved to a new location through a paracentric inversion. The α -amylase gene appears to be capable of repeated duplication events and diversification to new functions (Bahn 1967; Boer and Hickey 1986; Brown et al. 1990; Da Lage et al. 1998; Levy et al. 1985; Popadic et al. 1996). A newly discovered divergent amylase gene, *amylase-related*, raises the possibility that the 78C region may harbor a dispersed member of the gene family (Da Lage et al. 1998). In this case we might expect to see a break in conserved linkage groups that flanks a copy of the α -amylase gene at the 73A and 78C cytogenetic locations when the *D. pseudoobscura* and *D. melanogaster*

genomes are compared. The second alternative is that a duplicate α -amylase gene moved via a transposition event. Duplication and transposition of the *exuperantia* locus in the *D. pseudoobscura* lineage led to one copy on the third chromosome (*exu1*) and a second copy on the X chromosome (*exu2*) (Luk et al. 1994). In the case of the α -amylase genes, we might expect to see a duplicate α -amylase gene inserted into a linkage group that is conserved between *D. pseudoobscura* and *D. melanogaster*.

This article presents a nucleotide sequence comparison of the 73A and 78C sections of the *D. pseudoobscura* third chromosome with the conserved linkage groups from the *D. melanogaster* genome. We will show here that the *D. pseudoobscura* 73A and 78C linkage groups are conserved with the *D. melanogaster* 41E and 53CD regions, respectively. The α -amylase gene, however, has not maintained its conserved linkage between the two species. These data suggest that the α -amylase locus has shifted its location via a nonreplicative transposition event in either the *D. pseudoobscura* or *D. melanogaster* lineage.

Materials and Methods

Nucleotide Sequences of the *D. pseudoobscura* 73A and 78C Regions

A 12.6 kb clone that maps to section 78C on the *D. pseudoobscura* cytogenetic map was digested with the *EcoRI*

restriction endonuclease into four fragments 2.9 kb, 3.8 kb, 1.5 kb, and 4.6 kb in length. These *EcoRI* fragments were subcloned into the plasmid sequencing vector pWSK29 (Wang and Kushner 1991) using the methods described in Bernhardt (1997). A total of 97 oligonucleotide primers (18 bp each) were used to sequence across each plasmid subclone in both directions. Three polymerase chain reaction (PCR) products that straddle the three internal *EcoRI* sites were sequenced to confirm that the order of the four *EcoRI* fragments was correct. An additional six primers were used to sequence the PCR fragments. Sequences were determined on an ABI Model 373 DNA Stretch automated sequencer at the University of Georgia or a Beckman CEQ2000 automated sequencer at Penn State. The sequences were assembled with the GCG (Madison, WI) nucleotide sequence analysis software. In addition, a 73 kb contig from *D. pseudoobscura* 78C was obtained (Contig1060; Human Genome Sequencing Center 2003) to determine the degree of conserved linkage of the 78C region with its counterpart in the *D. melanogaster* genome. A 185 kb scaffold sequence for the 73A region was obtained from the first freeze assembly of the *D. pseudoobscura* whole genome shotgun (Contig6731_Contig3011; Human Genome Sequencing Center 2003).

Nucleotide Sequence Analysis

The 12 kb sequence of the 78C clone was submitted to the GenBank (Benson et al. 2002) database (accession no. AF476111). We used BLAST (Altschul et al. 1997) to compare the three *D. pseudoobscura* sequences (185 kb sequence for the 73A region and the 12.6 and 73 kb sequences for the 78C region) with the GenBank database (Benson et al. 2002). All searches of GenBank were limited to the 1.7 million sequence accessions currently available for *D. melanogaster*. The statistical significance of matches was determined based on the *E* value of the alignment score. The *E* value is a measure of the number of sequences in GenBank expected to have an alignment score less than or equal to the observed accession (Altschul et al. 1997). We chose an *E* value cutoff of 5×10^{-9} , which is the *E* value whose expected number of random matches to the 1.7 million *D. melanogaster* sequences is 0.01.

The alignment of sequences was achieved with the MEGALIGN program within the LASERGENE package of DNA sequence analysis software (Madison, WI). Molecular evolutionary sequence analysis was completed with DnaSP 3.50 (Rozas and Rozas 1999) and MEGA 2.0 (Kumar et al. 2001).

Results

Sequence of the 78C Region of *D. pseudoobscura*

The complete sequence of the clone from the *D. pseudoobscura* 78C region is 12,592 bp in length. There is a slight A/T bias in the sequence, but the difference is not statistically significant from a 50:50 ratio of A/T (50.6%) to G/C (49.6%) with a chi-square goodness-of-fit test ($\chi^2 = 1.64$, $df = 1$, $P = .19$).

GenBank Searches

The 12.6 kb sequence from 78C of *D. pseudoobscura* matched eight *D. melanogaster* accessions with an *E* value less than 5×10^{-9} . All of these accessions map to section 53CD on the *D. melanogaster* cytogenetic map. Two accessions, AC099032 and AC007520, were genomic sequences from *D. melanogaster* BAC clones of 163 and 171 kb, respectively. One accession, AE003804, was a 260 kb genomic scaffold assembled in the *Drosophila* genome project (Adams et al. 2000). The other five accessions matched short cDNA or genomic sequences in the region.

The 12.6 kb sequence from 78C may not provide a complete picture of the degree of conserved linkage between *D. melanogaster* and *D. pseudoobscura* because this cloned sequence is relatively small. We used this 12.6 kb sequence to identify a 73.7 kb contig from the *D. pseudoobscura* whole genome shotgun assembly (Contig1060; Human Genome Sequencing Center 2003). Contig1060 matched 40 *D. melanogaster* accessions with an *E* value less than 5×10^{-9} , including the two BAC clones (AC099032 and AC007520) and the genomic scaffold (AE003804) found with the shorter 78C sequence. This BLAST search also found a match to an additional 158 kb BAC clone (AC005713). Thirteen of the matched accessions were between the cytosolic leucine tRNA genes CR30234 and CR30235 in the 78C region and similar tRNA genes of *D. melanogaster* at four locations in the genome. The last 23 matched accessions were to short cDNA or genomic clones in the 78C region. The 260 kb genomic scaffold AE003804 was used in all comparisons of the *D. pseudoobscura* 78C region with the 53CD conserved linkage group of *D. melanogaster*.

The 185 kb sequence (Contig6731_Contig3011) from *D. pseudoobscura* 73A matched 123 GenBank (Benson et al. 2002) accessions within the *D. melanogaster* database (*E* values less than 5×10^{-9}). Six accessions either matched the first 94 kb or the last 91 kb of Contig6731_Contig3011 due to a break in conserved linkage in the sequence. Three *D. melanogaster* accessions—AE003811, AC008343, and AE003810—map to 51E on chromosome 2R in *D. melanogaster* and match sequences in the first 94 kb of Contig6731_Contig3011. The other three *D. melanogaster* accessions—AC007054, AE003785, and AC009847—map to 41D of chromosome 2R in *D. melanogaster* and match sequences in the last 91 kb of Contig6731_Contig3011. We used the *D. melanogaster* accession AE003785 for comparisons to *D. pseudoobscura* Contig6731_Contig3011 because the three α -amylase genes map within the last 91 kb of this genomic scaffold. The other 117 accessions match short cDNAs or genomic sequences.

Comparison of 53CD in *D. melanogaster* with 78C in *D. pseudoobscura*

The cytological section 78C in *D. pseudoobscura* corresponds to cytological position 53CD on the *D. melanogaster* map based on the conserved linkage shared between the two regions. The comparison of the *D. melanogaster* and 12.6 kb *D.*

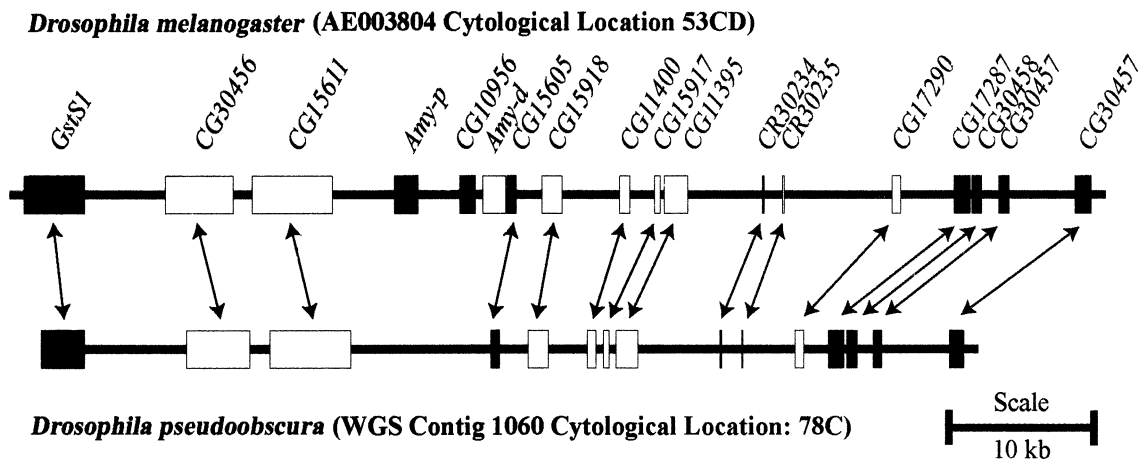


Figure 2. Comparison of the 53CD region of *D. melanogaster* with the 78C region of *D. pseudoobscura*. A horizontal line represents the sequence of each species. The beginnings and ends of the coding sequences are shown for genes on the plus (white boxes) or minus DNA strands (black boxes). The arrows between the two sequences help to line up the conserved genes between the two species. “CG” prefixes indicate predicted protein encoding genes and “CR” prefixes indicate cytosolic tRNA genes that are 100% conserved between the two species.

pseudoobscura sequences found 48 segments with an average similarity of 90% that were distributed uniformly across the region (Figure 1). The sizes of the conserved domains varied from 22 to 131 bp. The 25 and 10 conserved sequences at the 5' and 3' ends of the aligned regions match directly, while the matching central domain matches on the complementary DNA strand. These data suggest that the central region has inverted since the two species diverged.

The *D. melanogaster* sequence has two α -amylase genes, *Amy-p* and *Amy-d*, which are divergently transcribed on the chromosome (Figure 1). In addition, there are four predicted genes located within the *D. melanogaster* region—*CG15611*, *CG10956*, *CG15605*, and *CG15918*. In contrast, the 78C region of *D. pseudoobscura* lacks both α -amylase genes and the predicted *CG10956* gene. *D. pseudoobscura* does have *CG15605* and parts of *CG15611* and *CG15918*, but a larger contig from the whole genome shotgun shows that *CG15611* and *CG15918* are conserved in this region. In addition, the intron-exon structure of these three genes is conserved. The region occupied by *Amy-d* and *CG10956* in *D. melanogaster* has been reduced from 3845 to 537 bp in *D. pseudoobscura*, while the region occupied by *Amy-p* in *D. melanogaster* has not changed significantly in size in *D. pseudoobscura*.

The *CG15605* gene appears to have duplicated in *D. pseudoobscura* lineage based on dot plot analysis and on translated BLAST searches with the 78C region sequence. This second copy of *CG15605* is on the opposite strand from the *D. melanogaster* homologue of the *CG15605* gene, is located near the central microinversion, and appears to be a pseudogene (see below).

One might argue that the 12.6 kb sequence from *D. pseudoobscura* 78C is too small to infer the degree of conserved linkage between the two species. Comparison of Contig1060 from the *D. pseudoobscura* genome project with the *D. melanogaster* 53CD region shows that the conserved linkage extends over at least 73.7 kb (Figure 2).

Comparison of 41E in *D. melanogaster* with 73A in *D. pseudoobscura*

The conservation of genes between 41E in *D. melanogaster* and 73A in *D. pseudoobscura* is shown in Figure 3. Six genes map to the *D. melanogaster* 41E region: *CG7791*, *gp120*, *BEST:CK02137*, *Or42a*, *CG14468*, and *Or42b*. Amylase is absent from 41E of *D. melanogaster*. Five of the six genes found in *D. melanogaster* map to the 73A region of *D. pseudoobscura*: *CG7791*, *gp120*, *BEST:CK02137*, *Or42a*, and *Or42b*. The *CG14468* gene is not detected in *D. pseudoobscura* and the *Or42a* gene has duplicated into a proximal and distal copy, *Or42a-p* and *Or42a-d*. The direction of transcription of these five genes is the same in the two species with the exception of the *Or42a* genes, which have reversed their orientation in one of the species. Two complete α -amylase genes, *Amy 1* and *Amy 2*, and a partial amylase gene, *Amy 3*, are inserted into the *D. pseudoobscura* 73A region between the *BEST:CK02137* and the *Or42a-p* genes, which led to an expansion of the region. A short segment, 613 bp in *D. melanogaster* and 691 bp in *D. pseudoobscura*, is conserved between the two species and is located between *BEST:CK02137* and the *Or42a* genes in *D. melanogaster* and between *Amy 2* and *Amy 3* in *D. pseudoobscura*. All genes had similar intron-exon structure except for *CG7791*, which has two introns in *D. melanogaster* and three introns in *D. pseudoobscura*.

Molecular Evolutionary Comparisons of Genes in *D. melanogaster* and *D. pseudoobscura*

We estimated the frequency of synonymous and non-synonymous changes between the two species for the genes in the two chromosomal regions (Figures 1 and 3) (Li et al. 1985). The complete sequences of the *CG15611* and *CG15918* predicted genes were obtained from Contig1060

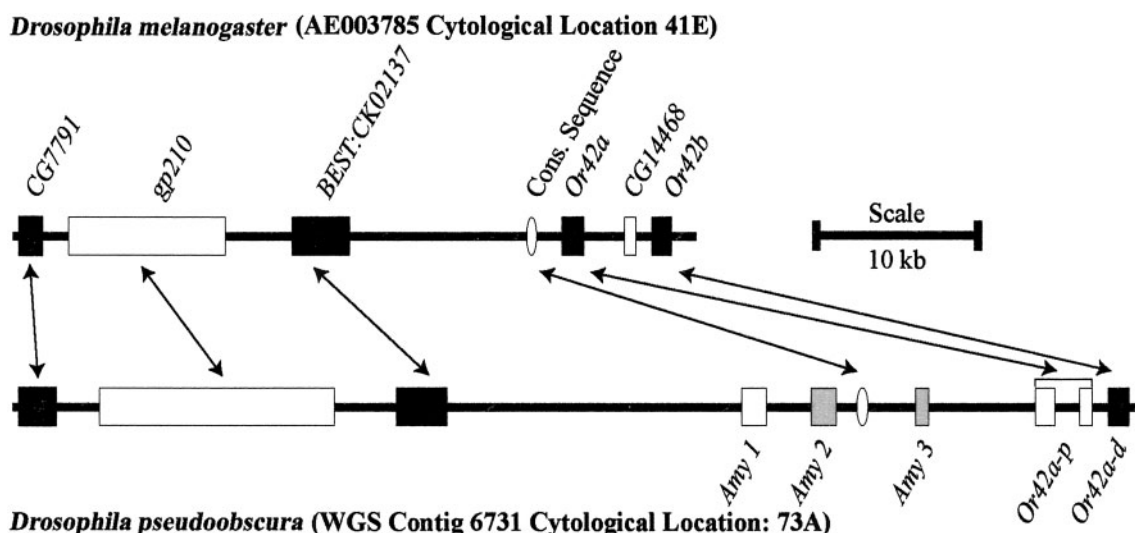


Figure 3. Comparison of the 41D region of *D. melanogaster* with the 73A region of *D. pseudoobscura*. A horizontal line represents the sequence of each species. The beginnings and ends of the coding sequences are shown for genes on the plus (white boxes) or minus DNA strands (black boxes). Two amylase pseudogenes, *Amy 2* and *Amy 3*, are indicated with gray boxes. The arrows between the two sequences help to line up the conserved genes between the two species. “CG” prefixes indicate predicted protein encoding genes.

(Human Genome Sequencing Center 2003). Estimates of synonymous and nonsynonymous changes between *D. melanogaster* and *D. pseudoobscura* amylase genes have been published previously (Brown et al. 1990). Synonymous sites have changed more frequently than nonsynonymous sites in the nine compared genes in the two chromosomal regions. Estimates of synonymous and nonsynonymous changes between *D. melanogaster* and *D. pseudoobscura* have been published previously (Brown et al. 1990). The surprising result is that the number of nonsynonymous substitutions per site is quite high in *CG15605* and *Or42b*, with 35.7% and 48.8%, respectively, of the sites changing between the two species. The frequency of synonymous and nonsynonymous changes between the proximal and distal copies of the *Or42a* gene within *D. pseudoobscura* is less than that observed between *D. melanogaster* and *D. pseudoobscura* for the other genes. Given that *CG15605* was predicted with *in silico* analyses, one might suggest that the high level of change in nonsynonymous sites reflects a comparison of nonfunctional DNA rather than protein coding regions. *CG15605*, however, is likely to produce a functional gene because a partial expressed sequence tag (EST) of *CG15605* has been isolated from a *D. melanogaster* 0–3-day-old testes-specific cDNA library (Flybase Consortium 1999) and has been detected with microarray analysis (Andrews et al. 2000).

A second *CG15605* gene segment was identified in a dot plot analysis and appears to be a pseudogene that we designate *CG15605*ψ (Figure 1). This gene segment is found only in *D. pseudoobscura* and not *D. melanogaster*. Several pieces of evidence suggest that *CG15605*ψ is now a pseudogene. First, the start codon is missing from *CG15605*ψ. Second, the gaps assumed to maximize the similarity of the nucleotide

amino acid sequence alignment lead to frameshift mutations in the amino acid sequence. Finally, the estimates of the frequency of synonymous and nonsynonymous changes show a nearly equivalent amount of change in the two classes of mutation, with a slight bias for synonymous substitutions (Table 1).

Discussion

Rearrangement of the Amylase Region Between *D. melanogaster* and *D. pseudoobscura*

The 78C region of *D. pseudoobscura* and the 53CD section of the *D. melanogaster* genome are conserved linkage groups. The amylase genes, however, are no longer part of this conserved linkage group because the *D. pseudoobscura* amylase genes are now located within section 73A on the *D. pseudoobscura* cytogenetic map (Aquadro et al. 1991). The amylase related (*Amyrel*) gene is not located at the 78C region. The region where the amylase gene is located in *D. pseudoobscura* (73A) and 41E of *D. melanogaster* are conserved linkage groups. The *D. melanogaster* pDm3.8 plasmid clone was able to hybridize to the 73A and 78C regions of *D. pseudoobscura* because the 5' region hybridized to conserved sequences at 78C, while the 3' region that contained amylase hybridized to the amylase coding sequences at 73A.

The comparison of the *D. pseudoobscura* 73A and the 78C regions with the conserved linkage groups of *D. melanogaster* suggest that a transposition event rather than a paracentric inversion was responsible for the relocation of the amylase gene. If an inversion event had moved the amylase gene in *D. pseudoobscura*, then we might expect parts of both the 73A and

Table 1. Synonymous and nonsynonymous substitutions per nucleotide site for genes in the 41E/73A and 53CD/78C regions of *D. melanogaster* and *D. pseudoobscura*

Gene	Bases	Synonymous	Nonsynonymous
41E in <i>D. melanogaster</i> versus 73A in <i>D. pseudoobscura</i>			
<i>CG7791</i>	2115 (2091)	1.936 ± 0.226	0.115 ± 0.009
<i>gp210</i>	5649 (5622)	1.615 ± 0.098	0.205 ± 0.008
<i>BEST:CK02137</i>	1584 (1449)	1.485 ± 0.172	0.256 ± 0.018
<i>Or42a</i> versus <i>Or42a-p</i>	1182 (1152)	2.606 ± 0.740	0.125 ± 0.013
<i>Or42a</i> versus <i>Or42a-d</i>	1182 (1152)	2.060 ± 0.356	0.096 ± 0.011
<i>Or42b</i>	1242 (966)	1.668 ± 0.296	0.488 ± 0.034
53CD in <i>D. melanogaster</i> versus 78C in <i>D. pseudoobscura</i>			
<i>CG15611</i>	1413 (1374)	0.944 ± 0.092	0.096 ± 0.010
<i>CG15605</i>	510 (441)	2.204 ± 0.733	0.357 ± 0.040
<i>CG15918</i>	1233 (1194)	1.457 ± 0.180	0.107 ± 0.011
Paralogous genes within <i>D. pseudoobscura</i>			
<i>Or42a-p</i> versus <i>Or42a-d</i>	1182 (1152)	0.172 ± 0.028	0.052 ± 0.008
<i>CG15605</i> versus <i>CG15605ψ</i>	507 (485)	1.329 ± 0.321	0.982 ± 0.100

78C region to show conserved linkage with 53CD in *D. melanogaster*. This was not the case. The *D. pseudoobscura* 78C region had conserved linkage with the entire length of the *D. melanogaster* 53CD region, except the *D. melanogaster* region had two α -amylase genes and the *D. pseudoobscura* region had none. The *D. pseudoobscura* 73A showed conserved linkage with the *D. melanogaster* 41D region, except the *D. melanogaster* region had no amylase genes and the *D. pseudoobscura* had three amylase genes.

We were unable to directly infer if the transposition occurred in the *D. melanogaster* or *D. pseudoobscura* lineage from the comparison of these two genome regions. The history of the amylase genes is confounded because these loci are subject to frequent gene conversion events, which can mask the relationships among genes (Popadic and Anderson 1995; Zhang et al. 2002). Gene conversion will tend to maintain the similarity of paralogous genes, preventing the detection of the true history of duplicate amylase genes. Thus we cannot infer with confidence the ancestral copy of amylase in either species.

The critical feature that will help us to polarize the transposition event is not comparisons among amylase sequences, but genomic sequences that include the genes that flank amylase. While there are many *Drosophila* amylase sequences in GenBank, we were unable to find amylase sequences that include flanking gene sequences for species in the genus *Drosophila*. The *Anopheles gambiae* genome does not provide any help to resolve this issue either. The *A. gambiae* genome has an amylase homologue, but none of the flanking genes resemble the genes linked to amylase in either *D. pseudoobscura* or *D. melanogaster*. It is tempting to think that the transposition occurred in *D. pseudoobscura* because of the duplication of the *CG15605* predicted gene and subsequent decay of the new gene's coding sequence. This arrangement appears to have reduced the size of the amylase region in *D. pseudoobscura*. This is speculation at best and will require additional sequence data from an outgroup species within either the *willistoni* or *saltans* subgroups to determine the ancestral location of the amylase gene region.

Transposable elements can mediate the movement of genes from one genomic location to another through duplicative or nonduplicative transposition events (Li 1997). The regions that flank amylase in *D. melanogaster* or *D. pseudoobscura* are not associated with any obvious transposable elements. What we do know about the transposition event is that the movement of the amylase gene was associated with small-scale rearrangements at both chromosomal locations. The *D. melanogaster* 53CD/*D. pseudoobscura* 78C region has a 2.2 kb segment that is inverted between the two species. The odorant receptor gene *Or42a* was inverted in the *D. melanogaster* 41E/*D. pseudoobscura* 73A region and duplicated in the *D. pseudoobscura* lineage. The movement of the amylase gene was also coupled with a deletion event of the original gene copy.

It is unclear from the sequence comparison what happened to the predicted gene *CG10956*. If *CG10956* is an essential gene, then we would hypothesize that this gene also transposed to section 73A along with amylase. We found no evidence for the *CG10956* gene at 73A in *D. pseudoobscura*, nor is there any evidence for this gene in the *D. pseudoobscura* genome (Human Genome Sequencing Center 2003). Either this gene is an artifact of the gene prediction analysis, the gene was deleted since *D. melanogaster* and *D. pseudoobscura* shared a common ancestor, or the gene experienced a rapid accumulation of nucleotide changes so that it is no longer similar to the *D. melanogaster* gene.

Future Prospects

Comparison of the 73A and 78C regions of *D. pseudoobscura* with the conserved linkage groups of *D. melanogaster* revealed tantalizing patterns that may be a prelude of what will be revealed when the complete genomes of these two species are compared. It will be of interest to see how many predicted genes are conserved between these two species and what new genes may be found as a result of the comparison. We are curious to know how often paracentric inversions

split predicted genes on the chromosome and if these rearrangements lead to the loss or gain of function. These two regions revealed genes that have duplicated to form new functional genes or pseudogenes. One gene was found to have an additional intron in one species. Comparative genomics between these two model *Drosophila* species will allow careful quantification of the rates of these types of changes. Finally, the *D. melanogaster*–*D. pseudoobscura* comparison may provide valuable insights into how chromosomes evolve in nature through transpositions and inversions.

Acknowledgments

We would like to thank the Human Genome Sequencing Center at the Baylor College of Medicine for access to the draft *Drosophila pseudoobscura* genome sequences, C. F. Aquadro who collaborated on the isolation of the 78C clones, and two anonymous reviewers for comments that improved the manuscript. The work presented here was supported by the National Science Foundation (grant 9726285 to S.W.S. and grant 9729144 to W.W.A.). Any opinions, findings, conclusions, or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

- Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, George RA, Lewis SE, Richards S, Ashburner M, Henderson SN, Sutton GG, Wortman JR, Yandell MD, Zhang Q, Chen LX, Brandon RC, Rogers YH, Blazej RG, Champe M, Pfeiffer BD, Wan KH, Doyle C, Baxter EG, Helt G, Nelson CR, Gabor Miklos GL, Abril JF, Agbayani A, An HJ, Andrews-Pfannkoch C, Baldwin D, Ballew RM, Basu A, Baxendale J, Bayraktaroglu L, Beasley EM, Beeson KY, Benos PV, Berman BP, Bhandari D, Bolshakov S, Borkova D, Botchan MR, Bouck J, Brokstein P, Brottier P, Burtis KC, Busam DA, Butler H, Cadieu E, Center A, Chandra I, Cherry JM, Cawley S, Dahlke C, Davenport LB, Davies P, de Pablos B, Delcher A, Deng Z, Mays AD, Dew I, Dietz SM, Dodson K, Doup LE, Downes M, Dugan-Rocha S, Dunkov BC, Dunn P, Durbin KJ, Evangelista CC, Ferraz C, Ferriera S, Fleischmann W, Fosler C, Gabriellian AE, Garg NS, Gelbart WM, Glasser K, Glodek A, Gong F, Gorrell JH, Gu Z, Guan P, Harris M, Harris NL, Harvey D, Heiman TJ, Hernandez JR, Houck J, Hostin D, Houston KA, Howland TJ, Wei MH, Ibegwam C, Jalali M, Kalush F, Karpen GH, Ke Z, Kennison JA, Ketchum KA, Kimmel BE, Kodira CD, Kraft C, Kravitz S, Kulp D, Lai Z, Lasko P, Lei Y, Levitsky AA, Li J, Li Z, Liang Y, Lin X, Liu X, Mattei B, McIntosh TC, McLeod MP, McPherson D, Merkulov G, Milshina NV, Mobarry C, Morris J, Moshrefi A, Mount SM, Moy M, Murphy B, Murphy L, Muzny DM, Nelson DL, Nelson DR, Nelson KA, Nixon K, Nusskern DR, Pacleb JM, Palazzolo M, Pittman GS, Pan S, Pollard J, Puri V, Reese MG, Reinert K, Remington K, Saunders RD, Scheeler F, Shen H, Shue BC, Siden-Kiamos I, Simpson M, Skupski MP, Smith T, Spier E, Spradling AC, Stapleton M, Strong R, Sun E, Svirskas R, Tector C, Turner R, Venter E, Wang AH, Wang X, Wang ZY, Wassarman DA, Weinstock GM, Weissenbach J, Williams SM, Woodage T, Worley KC, Wu D, Yang S, Yao QA, Ye J, Yeh RF, Zaveri JS, Zhan M, Zhang G, Zhao Q, Zheng L, Zheng XH, Zhong FN, Zhong W, Zhou X, Zhu S, Zhu X, Smith HO, Gibbs RA, Myers EW, Rubin GM, and Venter JC, 2000. The genome sequence of *Drosophila melanogaster*. *Science* 287:2185–2195.
- Altshul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, and Lipman DJ, 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402.
- Andrews J, Bouffard GG, Cheadle C, Lu J, Becker KG, and Oliver B, 2000. Gene discovery using computational and microarray analysis of transcription in the *Drosophila melanogaster* testis. *Genome Res* 10:2030–2043.
- Aquadro CF, Weaver AL, Schaeffer SW, and Anderson WW, 1991. Molecular evolution of inversions in *Drosophila pseudoobscura*: the amylase gene region. *Proc Natl Acad Sci USA* 88:305–309.
- Ashburner M, 1989. *Drosophila*: a laboratory handbook. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Bahn E, 1967. Crossing over in the chromosomal region determining amylase isozymes in *Drosophila melanogaster*. *Hereditas* 58:1–12.
- Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Rapp BA, and Wheeler DL, 2002. GenBank. *Nucleic Acids Res* 30:17–20.
- Bernhardt MJ, 1997. Molecular characterization of a third chromosome clone of *Drosophila pseudoobscura* (Masters thesis). Athens, GA: University of Georgia.
- Boer PH and Hickey DA, 1986. The alpha-amylase gene in *Drosophila melanogaster*: nucleotide sequence, gene structure and expression motifs. *Nucleic Acids Res* 14:8399–8411.
- Brown CJ, Aquadro CF, and Anderson WW, 1990. DNA sequence evolution of the amylase multigene family in *Drosophila pseudoobscura*. *Genetics* 126:131–138.
- Da Lage JL, Renard E, Chartois F, Lemeunier F, and Cariou ML, 1998. *Amyrel*, a paralogous gene of the amylase gene family in *Drosophila melanogaster* and the *Sophophora* subgenus. *Proc Natl Acad Sci USA* 95:6848–6853.
- Ehrlich J, Sankoff D, and Nadeau JH, 1997. Synteny conservation and chromosome rearrangements during mammalian evolution. *Genetics* 147:289–296.
- Flybase Consortium, 1999. The FlyBase database of the *Drosophila* genome projects and community literature. *Nucleic Acids Res* 27:85–88.
- Gemmill RM, Levy JN, and Doane WW, 1985. Molecular cloning of α -amylase from *Drosophila melanogaster*. I. Clone isolation by use of a mouse probe. *Genetics* 110:299–312.
- Hawley SA, Norman RA, Brown CJ, Doane WW, Anderson WW, and Hickey DA, 1990. Amylase gene expression in intraspecific and interspecific somatic transformants of *Drosophila*. *Genome* 33:501–508.
- Human Genome Sequencing Center, 2003. *Drosophila* genome project (visited Feb. 27, 2003) <http://www.hgsc.bcm.tmc.edu/projects/drosophila/index.html>.
- Kumar S, Tamura K, Jakobsen IB, and Nei M, 2001. MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 17:1244–1245.
- Levy JN, Gemmill RM, and Doane WW, 1985. Molecular cloning of α -amylase genes from *Drosophila melanogaster*. II. Clone organization and verification. *Genetics* 110:313–324.
- Li W-H, 1997. *Molecular evolution*. Sunderland, MA: Sinauer Associates.
- Li W-H, Wu C-I, and Luo C-C, 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol Biol Evol* 2:150–174.
- Luk SK-S, Kilpatrick M, Kerr K, and Macdonald PM, 1994. Components acting in localization of *bicoid* mRNA are conserved among *Drosophila* species. *Genetics* 137:521–530.
- Muller HJ, 1940. Bearings of the “*Drosophila*” work on systematics. In: *The new systematics* (Huxley J, ed). Oxford: Clarendon Press; 185–268.
- Popadic A and Anderson WW, 1995. Evidence for gene conversion in the amylase multigene family of *Drosophila pseudoobscura*. *Mol Biol Evol* 12:564–572.
- Popadic A, Norman RA, Doane WW, and Anderson WW, 1996. The evolutionary history of the amylase multigene family in *Drosophila pseudoobscura*. *Mol Biol Evol* 13:883–888.

- Ranz JM, Casals F, and Ruiz A, 2001. How malleable is the eukaryotic genome? Extreme rate of chromosomal rearrangement in the genus *Drosophila*. *Genome Res* 11:230–239.
- Ranz JM, Segarra C, and Ruiz A, 1997. Chromosomal homology and molecular organization of Muller's elements D and E in the *Drosophila repleta* species group. *Genetics* 145:281–295.
- Rozas J and Rozas R, 1999. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* 15:174–175.
- Segarra C and Aguade M, 1992. Molecular organization of the X chromosome in different species of the *obscura* group of *Drosophila*. *Genetics* 130:513–521.
- Segarra C, Lozovskaya ER, Ribo G, Aguade M, and Hartl DL, 1995. P1 clones from *Drosophila melanogaster* as markers to study the chromosomal evolution of Muller's A element in two species of the *obscura* group of *Drosophila*. *Chromosoma* 104:129–136.
- Segarra C, Ribó G, and Aguadé M, 1996. Differentiation of Muller's chromosomal elements D and E in the *obscura* group of *Drosophila*. *Genetics* 144:139–146.
- Steinemann M, 1982. Analysis of chromosomal homologies between two species of the subgenus *Sophophora*: *D. miranda* and *D. melanogaster* using cloned DNA segments. *Chromosoma* 87:77–88.
- Steinemann M, Pinsker W, and Sperlich D, 1984. Chromosome homologies within the *Drosophila obscura* group probed by in situ hybridization. *Chromosoma* 91:46–53.
- Sturtevant AH and Novitski E, 1941. The homologies of the chromosome elements in the genus *Drosophila*. *Genetics* 26:517–541.
- Wang RF and Kushner SR, 1991. Construction of versatile low-copy-number vectors for cloning, sequencing, and gene expression in *Escherichia coli*. *Gene* 100:195–199.
- Yardley DG, 1974. Linkage analysis of the α -amylase locus in *pseudoobscura*. *Drosophila Information Serv* 51:25.
- Zhang Z, Inomata N, Ohba T, Cariou ML, and Yamazaki T, 2002. Codon bias differentiates between the duplicated amylase loci following gene duplication in *Drosophila*. *Genetics* 161:1187–1196.

Received March 24, 2003

Accepted July 31, 2003

Corresponding Editor: Sudhir Kumar