Insect Molecular Biology (2003) **12**(4), 345–351

The evolution of HOM-C homeoboxes in the Dipteran family Drosophilidae

R. DeSalle*, M. A. Branham*, P. O'Grady* and J. Gatesy†

**Division of Invertebrates, American Museum of Natural History, 79th Street at Central Park West, New York, NY 10024, USA;* †*Department of Biology, Spieth Hall, University of California, Riverside, CA 92521, USA*

Abstract

Forty-five new Homeotic Complex (HOM-C) homeobox sequences from six species of Drosophilidae (*Drosophila heteroneura*, *D. adiastola*, *Zaprionus vittiger*, *Chymomyza amoena*, *Scaptodrosophila pattersoni* and *Hirtodrosophila pictiventris*) were obtained using a PCR-cloning method. These new homeoboxes are from the *labial*, *proboscipedia*, *Deformed*, *Sex combs reduced*, *fushi tarazu*, *Antennapedia*, *Ultrabithorax*, *abdominal-A* and *Abdominal-B* genes. Phylogenetic signal in the homeobox sequences was assessed and several aspects of sequence evolution were examined. In particular, codon bias was examined and found to exist between the drosophilid species examined here and *Anopheles gambiae* outgroup sequences. In addition, different patterns of codon bias were detected in homeoboxes interrupted with introns when compared to homeoboxes that are uninterrupted.

Keywords: *Drosophila*, homeobox, codon bias, phylogeny, evolution.

Introduction

While homeobox genes have been studied in a wide variety of organisms (de Rosa *et al*., 1999; Banerjee-Basu *et al*., 2000; Ferrier & Holland, 2001) very few studies have focused on homeobox evolution in a group of relatively closely related organisms (Zardoya *et al*., 1996). Here we compare homeobox sequences from seven taxa in the family Drosophilidae selected because their phylogenetic distribution covers a wide range of the diversity in this family. *Drosophila heteroneura* and *D. adiastola* are species in the Hawaiian *Drosophila* clade and members of the genus *Drosophila* (Remsen & O'Grady, 2002). Both of these species are endemic to the Hawaiian archipelago. The remainder of flies examined are from genera representative of the major branching events in the family. *Scaptodrosophila pattersoni* and *Chymomyza amoena* are considered basal flies in the family (Grimaldi, 1990; Remsen & O'Grady, 2002). *Hirtodrosophila pictiventris* and *Zaprionus vittiger* are considered more derived genera closely related to the subgenera *Drosophila* and *Sophophora*, the latter of which contains the intensively studied species *D. melanogaster* (for which all HOM-C homeobox sequences are also available). Phylogenetic relationships of the drosophilid taxa in this study are well known and robust (Tatarenkov *et al*., 2001; Remsen & O'Grady, 2002), as are divergence times for these taxa, which range between 10 and 60 million years ago (Grimaldi, 1990; Russo *et al*., 1995).

The Homeotic Complex (HOM-C) of *Drosophila* is composed of two separate units, the Antennapedia (ANTC) and Bithorax (BXC) complexes. Although both these complexes are located on the right arm of chromosome 3 (3R) in *Drosophila melanogaster*, they are considered genetically unlinked because they are separated by 7.5 Mb. We obtained homeobox sequences from both the ANTC and the BXC complexes for several drosophilid species by amplifying genomic DNA with a single pair of 'HOM-C specific' primers. Resultant PCR products were cloned and 150 clones from each species were sequenced. Nine homeoboxes (*labial*, *proboscipedia*, *Deformed*, *Sex combs reduced*, *fushi tarazu*, *Antennapedia*, *Ultrabithorax*, *abdominal-A* and *Abdominal-B*) were recovered from these species using this approach. Our taxonomic sampling strategy allows us to examine phylogenetic signal in HOM-C as well as test for patterns of codon usage bias in homeoboxes across taxa in the family Drosophilidae. Because several different homeoboxes were obtained for all the species in this study, we also examined codon usage bias as a function of gene structure by comparing those homeoboxes with introns (*lab*, *pb* and *AbdB*) to those that lack these interruptions.

Received 2 July 2002; accepted after revision 3 March 2003. Correspondence: Dr R. DeSalle, Division of Invertebrates, American Museum of Natural History, 79th Street at Central Park West, New York, NY 10024, USA. E-mail: desalle@amnh.org

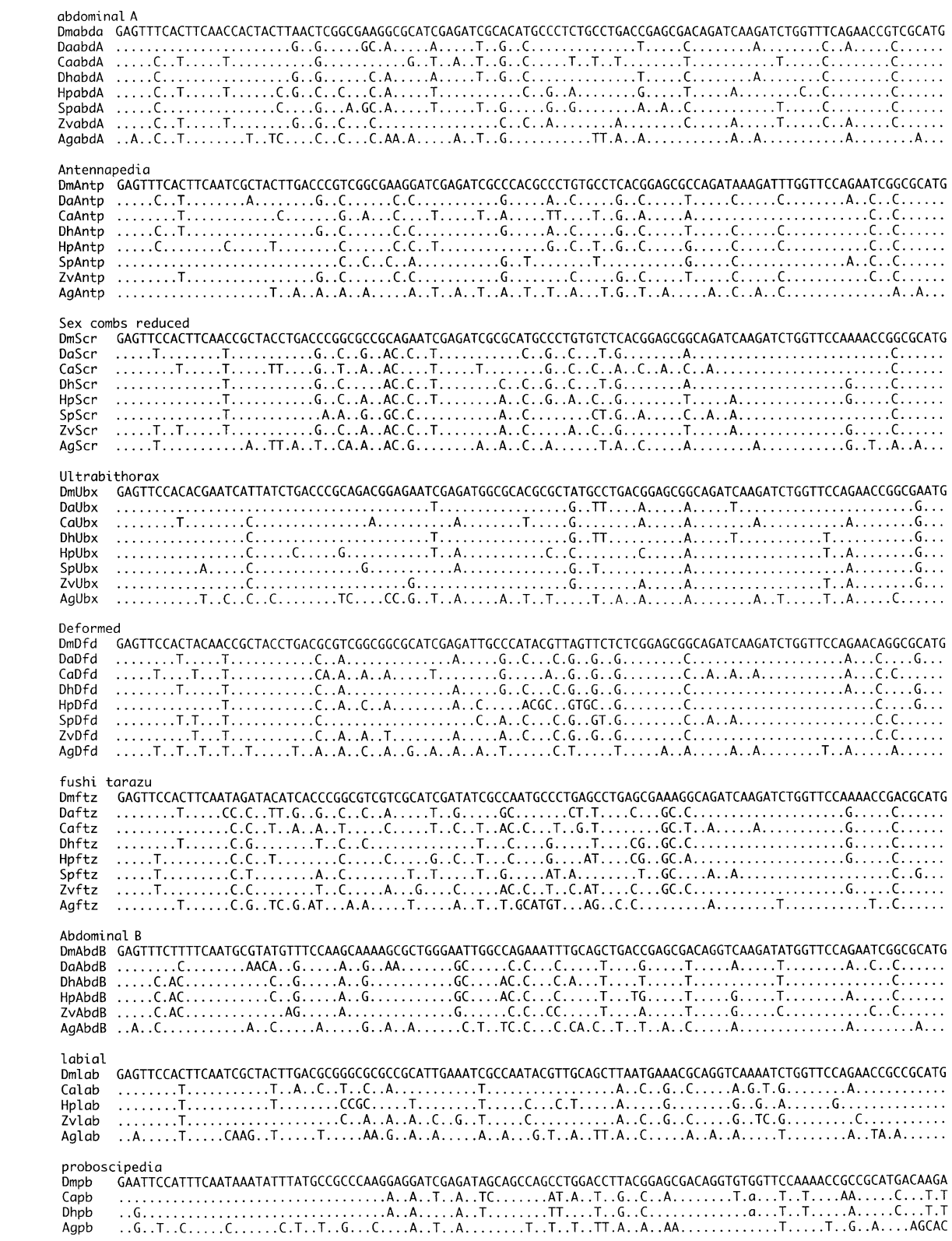

Figure 1. Nucleic acid sequences of 45 new partial *Drosophila* HOM-C homoeboxes. Top sequence for each gene is the *D. melanogaster* sequence from Genbank. Dots in the blocks indicate nucleotide identity to the *D. melanogaster* sequence. Abbreviations: Ag = *Anopheles gambiae*, Dm = *Drosophila melanogaster*, Hp = *Hirtodrosophila pictivientris*; Dh = *D. heteroneura*; Da = *D. adiastola*; Zv = *Zaprionus*; *vittiger*; Sp = *Scaptodrosophila pattersoni*; Ca = *Chymomyza amoena*. Sequences for *A. gambiae* and *D. melanogaster* were obtained from Genbank as detailed in the experimental procedures. All sequences have been deposited in Genbank under accession numbers AY194803 to AY194847.

Results and discussion

Results from PCR and cloning

Of the 900 sequenced clones, about three-quarters contained inserts of the target length 105 bases (or slightly longer to accommodate homeobox products with introns). The majority of the clones were either *Ubx*, *abdA*, *Scr*, *Antp* or *ftz* homologues. Since the amino acid sequences of *Scr* and *Antp* are identical over the region we amplified, diagnosis of clones using amino acids was not possible. Instead, putative orthology was assigned to *Scr* and *Antp* based on nucleotide sequence similarity. Figure 1 shows the nucleotide alignments of the 45 new *Drosophila* homeobox genes and the orthologs from *Anopheles gambiae* and *D. melanogaster*. Amino acid translations of the sequences for these genes were highly invariant except for the *ftz*, *pb*, *AbdB* and *lab* genes (Fig. 2). These sequences, while generally lacking variation at the amino acid level, displayed a great deal of variation in their nucleotide sequences, primarily in third codon positions. Although we looked for polymorphism by sequencing multiple clones of the same homeobox gene in each species, none was discovered. This result is not surprising given that the genomic DNAs used to do the initial PCR amplifications are from strains of flies that have been in culture for decades and most likely have been homogenized for the sequences of these homeoboxes.

Phylogenetic signal in HOM-C homeoboxes

Phylogenetic analysis using the homeobox sequences from the seven drosophilid species and *A. gambiae* as an outgroup is shown in Fig. 3. The topology of this phylogeny is in general agreement with previous studies (DeSalle, 1992; Remsen & DeSalle, 1998; Tatarenkov *et al.*, 2001; Remsen & O'Grady, 2002). Interestingly, support on this tree is spread out somewhat uniformly over the entire tree, suggesting that, in spite of a dearth of amino acid variation, nucleotide sequences from homeoboxes contribute significantly to support on the tree, even across quite different divergence times.

Codon bias

We examined potential patterns of codon bias in these homeobox genes using the program CodonW (http://www.molbiol.ox.ac.uk/cμ/codonW.html; Peden, 1997) and GCUA (McInerney, 2000). The GCUA program was used to generate tables of relative synonymous codon usage (RSCU). The CodonW program was used to perform correspondence analysis to examine if there were statistically significant patterns of codon usage bias. Data were partitioned in two ways: (1) by concatenating the seven homeoboxes for a particular species (by species analysis) and (2) by concatenating the homeoboxes for genes with introns (*lab*, *pb* and *AbdB*) in one group and homeoboxes without introns (*Deformed*, *Sex combs reduced*, *fushi tarazu*, *Antennapedia*, *Ultrabithorax* and *abdominal-A*) as the second group (by gene structure analysis).

Owing to the short length of these sequences, and because of non-random amino acid usage in some genes, not all amino acids and codons were observed in high enough frequency to make statistically significant statements about codon usage bias. A subset of 13 amino acids occur in high enough frequency so they can be characterized for codon usage bias (Phe, Leu, Ile, Tyr, His, Gln, Asn, Lys, Glu, Thr, Ala, Cys and Arg). Figure 4 shows the number of times a codon was used, and the RSCU for comparisons by species and by gene structure. Correspondence analysis revealed several statistically significant codon usage patterns. Interestingly, codon usage among genes was not identical to usage patterns among species, suggesting a difference in the kinds of selection acting at the level of the gene.

One obvious change in codon usage has taken place between the outgroup (*A. gambiae*) and all of the ingroup species. All four significant codon usage bias examples in the 'by species' comparisons involved changes from the *A. gambiae* preferred codon bias to a different codon preferred across all the drosophilids examined. In particular, for Leu – CUG is preferred in drosophilids, for Glu – CAG is preferred in drosophilids, for Thr – ACG is preferred in drosophilids and for Arg – CGC is preferred in drosophilids. The seven drosophilid species we examined appear to be homogeneous in their codon usage in the homeobox genes studied here and in cases where there is a statistically significant shift it is to codons ending in G/C. Besansky (1993) showed a strong G/C bias in third positions of codons in 14 genes of *A. gambiae*. Our results indicate that homeoboxes in the HOM-C complex in *A. gambiae* are actually biased toward A/T in third positions relative to flies in the family Drosophilidae. This result suggests potential differences in transcriptional dynamics of the *Anopheles* homeoboxes relative to the drosophilid homeoboxes.

It is also possible that gene structure may play a role even in genes having very similar cellular functions. The 'by gene structure' comparisons in Fig. 4 show five codons with statistically significant codon usage bias patterns. In particular, for Leu – CUG is used preferentially in the homeoboxes with introns, for Glu – GAG is used preferentially in the homeoboxes with introns, for Ala – GCG is used preferentially in the homeoboxes with introns, for Arg – CGG is used preferentially in the homeoboxes with introns and for Ile – AUC is used preferentially in the homeoboxes without introns. These results suggest a strong bias for four of the five cases to G/C bias in codons in introns. This result is intriguing and may indicate that transcriptional activity or efficiency may differ between genes with and without introns.

```
Abdominal A
ALLAbdA EFHFNHYLTRRRRIEIAHALCLTERQIKIWFQNRRM

Antennapedia
ALLAntp EFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRM

sex combs reduced
ALLscr  EFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRM

Ultrabithorax
ALLUbx  EFHTNHYLTRRRRIEMAHALCLTERQIKIWFQNRRM

Deformed
DmDfd   EFHYNRYLTRRRRIEIAHTLVLSERQIKIWFQNRRM
DaDfd   ................................
CaDfd   ...F............................
DhDfd   ................................
HpDfd   ................QR.C............
SpDfd   ...F............................
ZvDfd   ...F............................
AgDfd   ................................

fushi tarazu
Dmftz   EFHFNRYITRRRRIDIANALSLSERQIKIWFQNRRM
Daftz   .......L......E..H..C.T.........
Caftz   ................H..G............
Dhftz   ................H....T.........
Hpftz   ............M...H..N.T.........
Spftz   ................................
Zvftz   ............V...H..N.T.........
Agftz   .......LN.....E..SM.K.T.........

Abdominal B                          *
DmAbdB  EFLFNAYVSKQKRWELARNLQLTERQVKIWFQNRRM
DaAbdB  ......T.....S...................
DhAbdB  ..H.............................
HpAbdB  ..H................M............
ZvAbdB  ..H....E........................
AgAbdB  ...................N............

Labial                                *
Dmlab  EFHFNRYLTRARRIEIANTLQLNETQVKIWFQNRRM
Calab  ........................RL.......
Hplab  ..........R..................E....
Zvlab  ........................NL...H...
Aglab  .....K............A.H.............

Proboscipedia                        *
Dmpb    EFHFNKYLCRPRRIEIAASLDLTERQVWFQNRRMTR
Capb    ............................H.Q.I.C
Dhpb    ............................H...I.C
Agpb    ..............................KH
```

Figure 2. Amino acid translations for homeoboxes used in this study. *AbdA*, *Antp*, *Scr* and *Ubx* had identical sequences in all taxa studied. For *Dfd*, *ftz*, *AbdB*, *lab* and *pb* we list all species. For these latter five genes the top sequence is for *D. melanogaster*. Dots in the sequences below indicate identity to the *D. melanogaster* sequence on the top line of that block. The asterisk indicates the position in the translated sequence where the intron in *pb*, *lab* and *AbdB* occurs.

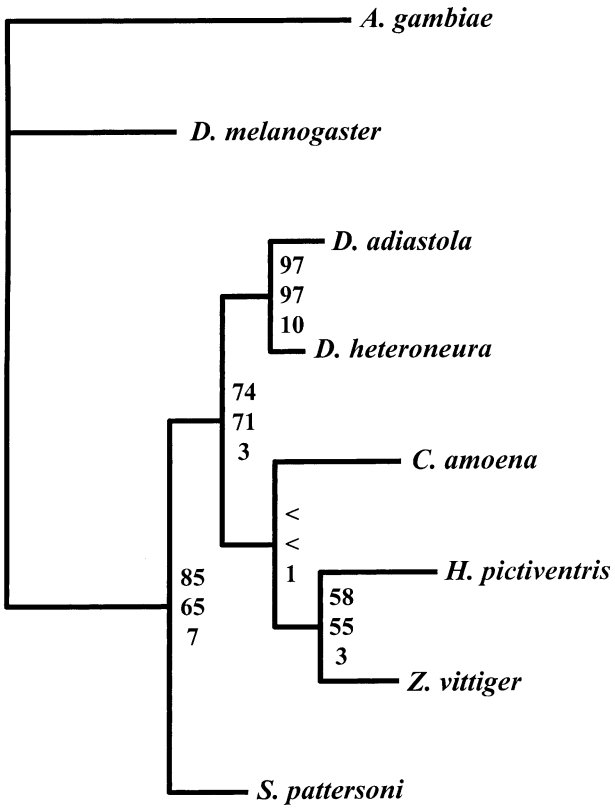

Figure 3. The single most parsimonious phylogenetic tree generated using the concatenated homeobox sequences from this study in combination with sequences for *D. melanogaster* and *A. gambiae* from Genbank (accession numbers for these two species homeoboxes are given in the experimental procedures section). The tree was 494 steps in length and had a CI = 0.6006 and an RI = 0.4167. The numbers on the nodes in the tree represent top: jackknife values (10 000 replications), middle: bootstrap values (10 000 replications), and bottom: Bremer support (Bremer, 1994). A < sign indicates that the bootstrap or jackknife value was less than 50%.

Intron structure and sequence

The point of insertion of introns in *pb*, *lab* and *AbdB* are conserved in all the species we examined (see Fig. 2), as well as in *A. gambiae*, indicating that at the very least such introns are ancestral to the divergence of mosquitoes and fruit flies (see also Cribbs *et al.*, 1992). Several authors (Popodi *et al.*, 1996; Nie *et al.*, 2001; Bastianello *et al.*, 2002) have also surveyed several invertebrate taxa for the presence of *pb*, *lab* and *AbdB* introns. For example, Nie *et al.* (2001) reported that *Tribolium* (Coleoptera) lacked the intron in the *lab* homeobox. Furthermore, Popodi *et al.* (1996) and Bastianello *et al.* (2002) did not detect the intron in *lab*, *pb* and *AbdB* in lower invertebrate and deuterostome homeoboxes. This suggests that the intron in these three homeobox genes may be a Diptera-specific molecular character, although more comprehensive survey work should be done.

The *AbdB*, *lab* and *pb* genes in drosophilids have introns inserted in the 3′ end of the homeobox in the amino acid position indicated in Fig. 2. While these introns are largely unremarkable, one striking characteristic is the short length of the *lab* and *pb* introns compared with other drosophilids. Intron length for the *lab* sequences from *C. amoena*,

By presence of intron

AA		RC1	N1	RC2	N2	
Phe	UUU	0.34	(22)	0.29	(6)	
	UUC	1.66	(109)	1.71	(36)	
Leu	UUA	0.35	(8)	0.86	(7)	
	UUG	1.00	(23)	1.96	(16)	
	CUU	0.22	(5)	1.10	(9)	
	CUC	0.91	(21)	0.24	(2)	
	CUA	0.48	(11)	0.49	(4)	
	CUG*	3.04	(70)	1.35	(11)	int
Ile	AUU	0.75	(47)	1.56	(14)	
	AUC*	1.96	(123)	0.67	(6)	intless
	AUA	0.29	(18)	0.78	(7)	
Tyr	UAU	0.64	(17)	1.14	(8)	
	UAC	1.36	(36)	0.86	(6)	
His	CAU	0.67	(36)	0.93	(7)	
	CAC	1.33	(72)	1.07	(8)	
Gln	CAA	0.62	(30)	0.43	(9)	
	CAG	1.38	(67)	1.57	(33)	
Asn	AAU	0.91	(46)	1.35	(29)	
	AAC	1.09	(55)	0.65	(14)	
Lys	AAA	0.41	(10)	0.80	(12)	
	AAG	1.59	(39)	1.20	(18)	
Glu	GAA	0.29	(20)	0.72	(17)	
	GAG*	1.71	(118)	1.28	(30)	int
Thr	ACU	0.32	(8)	0.46	(3)	
	ACC	1.49	(37)	1.23	(8)	
	ACA	0.57	(14)	0.62	(4)	
	ACG	1.62	(40)	1.69	(11)	
Ala	GCU	0.28	(6)	0.41	(3)	
	GCC	1.10	(24)	1.52	(11)	
	GCA	0.83	(18)	1.38	(10)	
	GCG*	1.79	(39)	0.69	(5)	int
Cys	UGU	0.80	(14)	1.20	(3)	
	UGC	1.20	(21)	0.80	(2)	
Arg	CGU	0.49	(30)	0.45	(6)	
	CGC	2.75	(169)	3.07	(41)	
	CGA	1.12	(69)	0.97	(13)	
	CGG*	1.12	(69)	0.52	(7)	int
	AGA	0.36	(22)	0.83	(11)	
	AGG	0.16	(10)	0.15	(2)	

By Species

AA		RC1	N1	RC2	N2	
	UUU	0.18	(2)	0.64	(8)	
	UUC	1.82	(20)	1.36	(17)	
	UUA	0.00	(0)	1.50	(7)	
	UUG	1.04	(4)	0.64	(3)	
	CUU	0.26	(1)	1.71	(8)	
	CUC	0.78	(3)	0.64	(3)	
	CUA	0.78	(3)	0.86	(4)	
	CUG*	3.13	(12)	0.64	(3)	Dros
	AUU	0.89	(8)	0.93	(9)	
	AUC	1.78	(16)	1.14	(11)	
	AUA	0.33	(3)	0.93	(9)	
	UAU	0.89	(4)	1.40	(7)	
	UAC	1.11	(5)	0.60	(3)	
	CAU	0.40	(3)	1.18	(10)	
	CAC	1.60	(12)	0.82	(7)	
	CAA	0.32	(3)	1.05	(10)	
	CAG	1.68	(16)	0.95	(9)	
	AAU	1.10	(11)	1.13	(13)	
	AAC	0.90	(9)	0.87	(10)	
	AAA	0.20	(1)	1.14	(8)	
	AAG	1.80	(9)	0.86	(6)	
	GAA	0.25	(3)	1.26	(17)	
	GAG*	1.75	(21)	0.74	(10)	Dros
	ACU	0.00	(0)	1.25	(5)	
	ACC	1.87	(7)	1.00	(4)	
	ACA	0.00	(0)	1.50	(6)	
	ACG*	2.13	(8)	0.25	(1)	Dros
	GCU	0.00	(0)	1.65	(7)	
	GCC	1.43	(5)	0.24	(1)	
	GCA	0.86	(3)	1.88	(8)	
	GCG	1.71	(6)	0.24	(1)	
	UGU	0.33	(1)	1.60	(4)	
	UGC	1.67	(5)	0.40	(1)	
	CGU	0.51	(5)	0.48	(5)	
	CGC*	3.56	(35)	1.26	(13)	Dros
	CGA	0.92	(9)	2.32	(24)	
	CGG	0.81	(8)	0.97	(10)	
	AGA	0.20	(2)	0.77	(8)	
	AGG	0.00	(0)	0.19	(2)	

Figure 4. Correspondence analysis results using RSCU values. The data have been partitioned to reflect the three kinds of tests of codon usage bias outlined in the text, by gene structure or by species. RC1 and N1 indicate the RSCU value and the number of codons, respectively, for the upper extreme in the particular correspondence analysis. RC2 and N2 indicate the RSCU value and the number of codons, respectively, for the lower extreme in the particular correspondence analysis. Asterisks (*) indicate statistically significant shifts in codon usage from the correspondence analysis using a chi-square test. The codons that are significant have the particular kind of homeoboxes that show the shift to that particular codon listed in a separate column.

H. pictiventris and *Z. vittiger* are 66, 75 and 68 nucleotides, respectively. In contrast, the *lab* intron length for other species in the subgenus *Drosophila* (*D. repleta* AAK77942, *D. mercatorum* AAK77943, *D. hydei* AAK77944, *D. buzzatii* AAK77945, *D. virilis* AAK77946; J. M. Ranz *et al.* unpublished data) ranges from 250 bases to 630 nucleotides, roughly 4–10 times longer than in three species we examined (above). Likewise, the length of the *pb* intron in *D. heteroneura* and *C. amoena* was 281 and 301 nucleotides, respectively, while the same intron was 480 nucleotides in *D. melanogaster*. While this intron structure information is limited, it does suggest that longer intron sequences in *lab* are a derived characteristic of the *virilis–repleta* radiation (subgenus *Drosophila*) and may coincide with the origin of this group roughly 40 million years ago (Russo *et al.*, 1995). Similarly, the longer *pb* intron in *D. melanogaster* may be characteristic of the subgenus *Sophophora*. Further comparative studies examining the phylogenetic distribution of HOM-C intron length within Drosophilidae will be needed to clarify this issue.

Summary

Our approach to obtain HOM-C homeobox sequences has proven quite efficient. We obtained a full complement of HOM-C homeoboxes for genes uninterrupted by introns. The HOM-C homeoboxes we have obtained from several drosophilid species has yielded some tantalizing results in terms of codon bias, intron structure and phylogenetic signal. Codon bias exists in some among-species and among-gene comparisons we have made. The among-species codon bias is due to a major codon usage shift between the outgroup (*A. gambiae*) and the Drosophilidae species examined in this study. Shifts in codon bias among genes is most significant in homeoboxes with introns, suggesting that intron/exon structure may play a role in codon usage in these flies. The phylogenetic signal provided by the homeoboxes we examined was modest and, surprisingly, distributed mostly within recent divergences.

Experimental procedures

Taxa used and DNA isolation

The following species were used in this study: *Hirtodrosophila pictiventris*, *Drosophila heteroneura* and *D. adiastola* (Hawaiian *Drosophila*), *Chymomyza amoena*, *Zaprionus vittiger*, and *Scaptodrosophila pattersoni*. The Hawaiian *Drosophila* species were obtained from Dr Kenneth Kaneshiro at the University of Hawaii (*D. heteroneura*: Q71G12; *D. adiastola*: Y32). All other species were obtained from the National *Drosophila* Species Stock Center in Tucson, Arizona. DNA was isolated from ten flies from each of the stock centre lines and from single wild caught flies of *D. heteroneura* and *D. adiastola*. Frozen specimens for the taxa in this study have been deposited in the Ambrose Monell Cryo-Collection (AMCC) at the American Museum of Natural History (AMCC102866 – *Z. vittiger*, AMCC102937 – *C. amoena*, AMCC105497 – *S. pattersoni*, AMCC107060 – *D. heteroneura*, and AMCC100171 – *H. pictiventris*). DNA sequences for *Anopheles gambiae* (Ag), *D. melanogaster* (Dm) are from Genbank (NM_057322 = Dm pb; NG_000062 = Dm lab; X14475 = Dm Scr; X00854 = Dm ftz; M20704 = Dm Antp; K01963 = Dm Ubx; X54453 = abdA; X54453 = AbdB; NG_000557 = Dfd; AF269154 = Ag pb; AF269153 = Ag lab; AF080564 = Ag Scr; AF230521 = Ag ftz; AF080565 = Ag Antp; AF080562 = Ag Ubx; AF080566 = Ag abdA; EAA07262 = Ag AbdB; AF269155 = Ag Dfd). Genbank accession numbers for the sequences determined in this study are Dfd: AY194803-AY194808; AbdB: AY194809-AY194812; pb: AY194813-AY194814; lab: AY194815-AY194817; ftz: AY194818-AY194823; abdA: AY194824-AY194829; Antp: AY194830-AY194835; Ubx: AY194836-AY194841; Scr: AY194842-AY194847.

PCR, cloning and sequencing

The following primer pair was designed to amplify as many of the ANT-C and BX-C homeoboxes as possible: HOM5L = 5′caracnytngarytngaraa3′; HOM3R = 5′rttytcyttyttccayttcat3′. PCR products from each species were cloned into the TA cloning vector (Invitrogen). One hundred and fifty clones from each species were picked and DNA was isolated from each. Some clones were cycle sequenced using S35 and a standard protocol (USB). Gels were dried and placed on film. Sequences were read directly from the film, entered into SEQUENCHER and compared to *D. melanogaster* and *A. gambiae* homeobox sequences to determine orthology. Other clones were amplified using colony PCR, sequenced using BigDye (ABI) and run on an ABI 3700. Sequences from this second approach were corrected using the SEQUENCHER software.

Phylogenetic analysis and codon bias

All sequences in the coding portion of the homeoboxes were easily aligned by eye. Introns in the *pb*, *lab* and *Abd-B* genes were aligned using CLUSTALW (Thompson *et al.*, 1994). Phylogenetic analysis was performed with PAUP*, ver 4.0 (Swofford, 1998) using the exhaustive search option. Bootstrap (Felsenstein, 1988) and Jackknife (Farris *et al.*, 1996) analyses, as implemented in PAUP*, ver 4.0 (Swofford, 1998), were used to assess node robustness. Translation of the DNA sequences into amino acid sequences was accomplished using MacClade, ver 3.0 (Maddison & Maddison, 1992). Codon Bias analysis was performed using the programs GCUA (General Codon Usage Analysis, ver 1.0; McInerney, 2000) and CodonW (Peden, 1997). Codon bias analyses were accomplished as described in the text. In assessing codon bias we simply list those codons in the analysis with sufficient data to determine a direction in bias (Phe, Leu, Ile, Tyr, His, Gln, Asn, Lys, Glu, Thr, Ala, Cys and Arg). In some cases there were too few positions in the sequence occupied by a particular amino acid and so the codons for these amino acids were not examined (Met, Val, Asp, Pro, Ser, Gly, and Trp).

Acknowledgements

We thank the Lewis B. and Dorothy Cullman Program for Molecular Systematic Studies at the American Museum of Natural History and an NSF grant (DEB – 0129105) awarded to R.D. and P.O.

References

Banerjee-Basu, S., Ryan, J.F. and Baxevanis, A.D. (2000) The homeodomain resource: a prototype database for a large protein family. *Nucl Acids Res* **28**: 329–330.

Bastianello, A., Ronco, M., Burato, P.A. and Minelli, A. (2002) Hox gene sequences from the geophilomorph centipede *Pachymerium ferrugineum* (C. L. Koch, 1835) (Chilopoda: Geophilomorpha: Geophilidae): implications for the evolution of the Hox class genes of arthropods. *Mol Phyl Evol* **22**: 155–161.

Besansky, N.J. (1993) Codon usage patterns in chromosomal and retrotransposon genes of the mosquito *Anopheles gambiae*. *Insect Mol Biol* **1**: 171–178.

Bremer, K. (1994) Branch support and tree stability. *Cladistics* **10**: 295–304.

Cribbs, D.L., Pultz, M.A., Johnson, D., Mazzulla, M. and Kaufman, T.C. (1992) Structural complexity and evolutionary conservation of the *Drosophila* homeotic gene proboscipedia. *EMBO J* **11**: 1437–1449.

DeSalle, R. (1992) Phylogenetic relationships of flies in the family Drosophilidae. *Mol Phyl Evol* **1**: 31–40.

Farris, J.S., Albert, V.A., Kälersjö, M., Lipscomb, D. and Kluge, A.G. (1996) Parsimony jackknifing outperforms Neighbor-joining. *Cladistics* **12**: 99–124.

Felsenstein, J. (1988) Phylogenies from molecular sequences: inference and reliability. *Ann Rev Genet* **22**: 521–565.

Ferrier, D.E. and Holland, P.W. (2001) Ancient origin of the Hox gene cluster. *Nat Rev Genet* **2**: 33–38.

Grimaldi, D.A. (1990) A phylogenetic, revised classification of genera in the Drosophilidae (Diptera). *Bull Am Mus Nat Hist* **197**: 1–139.

Maddison, W.P. and Maddison, D.R. (1992) *Macclade*, Version 3.0. Sinauer Associates, Sunderland, Massachusetts.

McInerney, J. (2000) *GCUA: General Codon Usage Analysis*, v1.0. The Natural History Museum, London.

Nie, W., Stronach, B., Panganiban, G., Shippy, T., Brown, S. and Denell, R. (2001) Molecular characterization of Tclabial and the 3′ end of the Tribolium homeotic complex. *Dev Genes Evol* **211**: 244–251.

Peden, J. (1997) *CodonW*, Version 1.3. Trinity College, Dublin. Distributed by author at: http://www.molbiol.ox.ac.uk/cu/codonW.html.

Popodi, E., Kissinger, J.C., Andrews, M.E. and Raff, R.A. (1996) Sea urchin hox genes: insights into the ancestral hox cluster. *Mol Biol Evol* **13**: 1078–1086.

Remsen, J. and DeSalle, R. (1998) Character congruence of multiple data partitions and the origin of the Hawaiian Drosophilidae. *Mol Phyl Evol* **9**: 225–235.

Remsen, J. and O'Grady, P. (2002) Phylogeny of Drosophilinae (Diptera: Drosophilidae), with comments on combined analysis and character support. *Mol Phyl Evol* **24**: 249–264.

de Rosa, R., Grenier, J.K., Andreeva, T., Cook, C.E., Adoutte, A., Akam, M., Carroll, S.B. and Balavoine, G. (1999) Hox genes in brachiopods and priapulids and protostome evolution. *Nature* **399**: 772–776.

Russo, C.A.M., Takezaki, N. and Nei, M. (1995) Molecular phylogeny and divergence times of drosophilid species. *Mol Biol Evol* **12**: 391–404.

Swofford, D.L. (1998) *PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods)*, Version 4. Sinauer Associates, Sunderland, MA.

Tatarenkov, A., Zurovcová, M. and Ayala, F.J. (2001) Ddc and amd sequences resolve phylogenetic relationships of *Drosophila*. *Mor Phyl Evol* **20**: 321–325.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucl Acids Res* **22**: 4673–4680.

Zardoya, R., Abouheif, E. and Meyer, A. (1996) Evolutionary analyses of hedgehog and Hoxd-10 genes in fish species closely related to the zebrafish. *Proc Natl Acad Sci USA* **93**: 13036–13041.