A New Split of the Hox Gene Complex in Drosophila: Relocation and Evolution of the Gene labial

Bárbara Negre, José María Ranz, Ferran Casals, Mario Cáceres, and Alfredo Ruiz

Department de Genètica i de Microbiologia, Universitat Autònoma de Barcelona, Bellaterra (Barcelona) Spain

Hox genes encode transcription factors involved in the specification of segment identity in the early metazoan embryo. These genes are usually clustered and arranged in the same order as they are expressed along the anteroposterior body axis. This conserved genomic organization has suggested the existence of functional constraints acting on the genome organization. Partial disassembly of the Hox gene complex (HOM-C) in Caenorhabditis elegans and in two different Drosophila lineages, however, calls into question whether this cluster organization is absolutely required for proper function. Here we report a new split of the HOM-C discovered in the species of the Drosophila repleta group, which relocated the most anterior gene of the complex, lab, to a distant chromosomal site near the two most posterior Hox genes, abd-A and Abd-B. To investigate the evolutionary consequences of natural rearrangements of the Hox gene complex, the gene lab has been cloned and sequenced in D. buzzatii, a member of the D. repleta group with the split, and in D. virilis, a member of a different species group without the split. The results show that the structure of lab in D. buzzatii is intact and place the breakpoint at least 8 kb from its transcription start site. The nucleotide sequence evolution of lab in the genus Drosophila has been investigated by means of maximum likelihood methods. No significant variation has been observed among lineages in the rate of nucleotide substitution or in the nonsynonymous/synonymous substitution ratio. Seemingly, the relocation of lab has not induced a change in evolution rate or degree of functional constraint. Nevertheless, further work is needed to ascertain whether the lab-pb split has had any effects on gene expression.

Introduction

Homeotic (Hox) genes encode transcription factors that specify the future cellular identity along the anteroposterior axis in the early metazoan embryo by activating or repressing their target genes (Gehring and Hiromi 1986). In the genome, Hox genes are found clustered and arranged in the same order as they are expressed along the anteroposterior body axis (Lewis 1978; McGinnis and Krumlauf 1992). This colinearity appeared before the separation between vertebrates and invertebrates, and it is widely conserved across the bilaterian phylogeny, which has suggested the existence of functional constraints acting on their genome organization (Duboule and Morata 1994). Partial disassembly of Hox complexes in Caenorhabditis elegans (Ruvkun and Hobert 1998) and Drosophila (vonAllmen et al. 1996; Adams et al. 2000; Ranz, Casals, and Ruiz 2001), together with reported experimental breaks of the Drosophila complexes (Hazelrigg and Kaufman 1982; Struhl 1984; Tiong, Whittle and Gribbin 1987), calls into question whether this organization is absolutely required for proper function.

A single Hox gene cluster (HOM-C) is postulated to have existed in the ancestor of the arthropods (Hughes and Kaufman 2002), although most sequences are partial and, usually, there are no data about the chromosomal localization of the genes (Finnerty and Martindale 1998; deRosa et al. 1999; Cook et al. 2001). The 10 Hox genes included in this ancestral complex are (in anteroposterior order): labial (lab), proboscipedia (pb), Hox3, Deformed (Dfd), Sex comb reduced (Scr), fushi tarazu (ftz), Antennapedia (Antp), Ultrabithorax (Ubx), abdominal-A (abd-A), and Abdominal-B (Abd-B). This was probably also the organization in the ancestor of the Drosophila genus, although in insects Hox3 and ftz have lost their Homeotic function (Hughes and Kaufman 2002). Two rearrangements of this ancestral Hox complex are known to have occurred during the evolution of the Drosophila genus (fig. 1). In the lineage leading to D. melanogaster a split between Antp and Ubx separated the genes in two complexes: the Antennapedia complex (ANT-C) (Kaufman, Seeger, and Olsen 1990) and the Bithorax complex (BX-C) (Lewis 1978; Duncan 1987). These two complexes are located on the right arm of D. melanogaster chromosome 3, separated by approximately 9.6 Mb of euchromatic sequences. The ANT-C spans ~ 400 kb (Adams et al. 2000) and comprises five Hox genes (lab, pb, Dfd, Scr, and Antp), which are expressed in those segments that will become the head and the anterior thorax. The BX-C spans ~ 350 kb (Martin et al. 1995; Adams et al. 2000) and contains three Hox genes (Ubx, abd-A, and Abd-B), which are expressed in the posterior thorax and abdomen. A different arrangement of the Hox genes was found in D. virilis, a species of the Drosophila subgenus (vonAllmen et al. 1996). In this species Ubx mapped by in situ hybridization with Antp instead of with abd-A, indicating that in the lineage leading to this species a disruption of the Hox complex occurred between the genes Ubx and abd-A (fig. 1).

The diversity in Hox gene arrangement observed among Drosophila species along with the reported flexibility of its genome (Stone 1962; Hartl and Lozovskaya 1994; Powell 1997; Ranz, Casals, and Ruiz 2001) prompted us to investigate the organization of Hox genes in D. buzzatii and D. repleta, two representative species of the D. repleta species group of the Drosophila subgenus. The in situ hybridization to the salivary gland chromosomes of the Antp, ftz, pb, Ubx, abd-A, and Abd-B genes (Ranz, Segarra,
Fig. 1.—Phylogenetic reconstruction of Hox gene organization in the genus Drosophila. All known information of Hox gene localization has been plotted in a phylogenetic tree to establish when each reorganization took place. The evolutionary relationships among species and divergence times are from different sources (Powell 1997; Ranz, Segarra, and Ruiz 1997; Kwiatkowski and Ayala 1999). Hox genes are shown as triangles when the orientation (5' → 3') is known; they are otherwise shown as rectangles. Hox genes expressed in the anterior and posterior parts of the embryo appear in red and blue gradients, respectively. The genes ftz, zen1, and Pxd are also included (Wharton 1942; leCalvez 1953; vonAllmen et al. 1996; Powell 1997; Adams et al. 2000); the gene zen2 found in the Hox complex of D. melanogaster (Adams et al. 2000) is not represented. Unless a connecting line appears, the order of the genes has been assumed as in D. melanogaster but has not yet been demonstrated. Arrows indicate known splits in the Hox complex. Vertical lines indicate chromosomal discontinuities in the Hox complex. When known, map position of genes is shown. Information for D. canalinea and D. immigrans: this work; D. melanogaster: Adams et al. 2000; D. pseudoobscura: Terol, Perez-Alonso, and Frutos 1991, Randazzo et al. 1994; and D. virilis: Maier, Preiss, and Powell 1993, vonAllmen et al. 1996, and this work. For all the D. repleta group species, a representative organization obtained from D. buzzatii (Ranz, Segarra and Ruiz 1997; this work), D. hydei (Maier, Preiss, and Powell 1990; Maier, Sperlich, and Powell 1993; this work), D. mercatorum (this work), and D. repleta (Ranz, Segarra, and Ruiz 1997; Ranz, Casals, and Ruiz 2001; this work) is shown.
and Ruiz 1997; Ranz, Casals, and Ruiz 2001; Ranz et al. 2003) indicated that both species lack the Antp-Ubx split present in D. melanogaster and share the split between Ubx and abd-A with D. virilis (fig. 1). However, the location of the lab gene was not yet known in those species. Here, we have mapped the lab gene in different Drosophila species and show that a new split has separated this gene from the other anterior Hox genes and relocated it near abd-A and Abd-B, which direct the posterior development of the fly. To determine if the lab-pb split has altered in any way the structure or evolution of lab, we have cloned and sequenced this gene in D. buzzatii (which presents the split) and also partially in D. virilis (which lacks the split). The results show that the molecular structure of lab is conserved between D. buzzatii, D. melanogaster, and D. virilis, and that the breakpoint must be located at least 8 kb upstream of this gene. In addition, they indicate that lab is a relatively fast evolving gene, although there was no evidence for positive or relaxed selection either before or after the split.

### Materials and Methods

#### Drosophila Stocks

The following species and stocks were used (code numbers of the Tucson Drosophila Species Stock Center are given in parenthesis): D. buzzatii (st-1 and j-19; see Cáceres, Puig, and Ruiz 2001), D. canalinea (15050–1201.0), D. hydei (Carboneras, Spain), D. immigrans (15111–1731.0), D. mercatorum (Comarapa, Bolivia), D. repleta (15084–1611.2), and D. virilis (Tokyo, Japan). All stocks are fixed for the standard chromosomal arrangements except line j-19 of D. buzzatii, which is homozygous for inversion 2j.

#### In Situ Hybridization to Salivary Gland Chromosomes

Clones used as probes to map by in situ hybridization the Antp, Ubx, abd-A, and Abd-B genes have been reported elsewhere (Ranz, Segarra, and Ruiz 1997; Ranz, Casals, and Ruiz 2001). To localize the lab gene in D. buzzatii, D. hydei, D. mercatorum, D. repleta, and D. virilis, homologous probes were generated in each species by polymerase chain reaction (PCR) amplification of a 725–1,172-bp segment encompassing exons 2–3 of the gene and including the homeobox (table 1), followed by cloning and sequencing of the amplified products. The clone of D. virilis was used to map lab in D. immigrans and D. canalinea.

Preparations of slides, probe labeling, and hybridization to salivary gland chromosomes were performed as previously described (Ranz, Segarra, and Ruiz 1997). Homologous and heterologous hybridizations were performed at 37°C and 25°C, respectively. Micrographs were taken by phase contrast with a Nikon Optiphot-1 microscope and a Nikon H-III photomicrographic system at 600× magnification using EKTAR-25 Kodak film and a blue filter. The chromosome map of D. repleta (Wharton 1942) was used to localize the in situ hybridization signals in all D. repleta group species and D. canalinea; the cytological maps of D. immigrans (leCalvez 1953) or D. virilis (Gubenko and Evgen’ev 1984) were used in the remaining cases.

#### Southern Hybridization and Library Screening

Genomic DNA extraction was performed as described in Piñol et al. (1988). DNA was digested with different restriction enzymes and transferred from the agarose gels to nylon membranes by capillarity transfer under neutral conditions (Sambrook, Fritsch, and Maniatis 1989). Probes were labeled with digoxigenin-11-dUTP by random primer with the DIG DNA Labeling and Detection Kit (Roche). Hybridization was performed at 42°C for 16 h using standard hybridization buffer with formamide (50%). Filters were washed twice for 15 min in 2× SSC 0.1% sodium dodecyl sulfate (SDS) at room temperature and twice for 20 min in 0.1× SSC 0.1% SDS at 68°C.

A lambda genomic library derived from D. buzzatii line j-19 (Cáceres, Puig, and Ruiz 2001) was amplified following Sambrook, Fritsch, and Maniatis (1989) and used for cloning of lab. Screening was done by plaque hybridization under the same conditions as for the Southern hybridization. Two probes were used to clone lab in D. buzzatii, one corresponding to exons 2–3 and the other to exon 1, separated in D. melanogaster by a large intron of nearly 14 kb. The probe containing exons 2 and 3 was the same 1,122 bp fragment used in the in situ hybridization (see above). The exon 1 probe consisted in a 1,285-bp fragment amplified by PCR using specific primers (table 1). Sequencing of the PCR product confirmed its homology with exon 1 of D. melanogaster lab.

### Table 1

<table>
<thead>
<tr>
<th>Gene Region</th>
<th>Species</th>
<th>Forward (5′–3′)</th>
<th>Reverse (5′–3′)</th>
<th>Product Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exons 2–3 (homebox)</td>
<td>D. buzzatii</td>
<td>TCCAGCACCAGAAAACATACCTG</td>
<td>TTTACGCGCTTCTGCTTCAT</td>
<td>1,122 bp</td>
</tr>
<tr>
<td></td>
<td>D. hydei</td>
<td></td>
<td></td>
<td>954 bp</td>
</tr>
<tr>
<td></td>
<td>D. mercatorum</td>
<td></td>
<td></td>
<td>725 bp</td>
</tr>
<tr>
<td></td>
<td>D. repleta</td>
<td></td>
<td></td>
<td>727 bp</td>
</tr>
<tr>
<td></td>
<td>D. virilis</td>
<td></td>
<td></td>
<td>1,172 bp</td>
</tr>
<tr>
<td>Exon 1</td>
<td>D. buzzatii</td>
<td>ATGATGGACGTAAGCAGCATG</td>
<td>CATTCCCTTGAGTTGCATCCA</td>
<td>1,285 bp</td>
</tr>
<tr>
<td></td>
<td>D. virilis</td>
<td></td>
<td></td>
<td>1,261 bp</td>
</tr>
<tr>
<td>cDNA</td>
<td>D. buzzatii</td>
<td>CATTCCGTCATCAACAACATCTCGT</td>
<td>TTTACGCGCTTCTGCTTCAT</td>
<td>570 bp</td>
</tr>
<tr>
<td></td>
<td>D. virilis</td>
<td></td>
<td></td>
<td>548 bp</td>
</tr>
<tr>
<td>Intron 1 fragment</td>
<td>D. buzzatii</td>
<td>CCTGGAAGCCATATAATAGGT</td>
<td>GCTGATTAATTGATGGATG</td>
<td>~3.5 kb</td>
</tr>
<tr>
<td>5′ UTR</td>
<td>D. virilis</td>
<td>TGCTCCGCCAGCAATCGAG</td>
<td>AGGCAGTGCGTGGCATGA</td>
<td>818 bp</td>
</tr>
</tbody>
</table>
Polymerase Chain Reaction Amplification

Primers (table 1) were designed with Primer Designer (1.01—1990 Scientific and Educational Software). Primers used to amplify the lab segment comprising exons 2–3 (homeobox) in all species were designed according to the D. melanogaster lab sequence. Exon 1 of D. buzzatii and D. virilis were amplified using the regions conserved between Labial proteins of D. melanogaster (AAD19811.1) and Tribolium castaneum (AAF64147.1). To amplify the cDNA of D. buzzatii and D. virilis, forward primers were designed in each species exon 1 sequence. Primers located in the transcription initiator (Inr) sequence of D. buzzatii and the beginning of the coding region of D. virilis were used to amplify the 5′ UTR in the latter species.

Polymerase chain reaction was carried out in a volume of 50 µl, including 100–200 ng of genomic DNA or cDNA or 1 µl of phage lysate, 20 µM of each primer, 200 µM dNTPs, 1.5 mM MgCl2, and 1 unit Taq DNA polymerase. Temperature cycling conditions were 30 rounds of 30 s at 94°C, 30 s at the annealing temperature, and 60 s at 72°C, with annealing temperature varying between 55° and 61°C depending on the primer pair used.

RNA Isolation and Reverse Transcriptase-Polymerase Chain Reaction

Total RNA from 0–12-h-old embryos from D. buzzatii (line j-19) and D. virilis was isolated with TRIZOL (GIBCO) following the manufacturer’s instructions. Glassware was washed with DEPC-water to eliminate RNases.

cDNA was obtained by random primer with the First Strand cDNA Synthesis Kit for reverse transcriptase polymerase chain reaction (RT-PCR; AMV-Roche). Amplification conditions for RT-PCR were as described for PCR.

DNA Sequencing and Sequence Analysis

Polymerase chain reaction products and restriction fragments from lambda phages were cloned into pGEM-T (Promega) and Bluescript II SK (Stratagene) and sequenced with universal primers. In some cases internal specific primers were designed for primer walking. Sequencing was performed in an ALF express DNA automated sequencer (Pharmacia Biotech) and an ABI 373 A (PerkinElmer) automated sequencer. Nucleotide sequences were analyzed with the Wisconsin Package (Genetics Computer Group). Similarity searches were done with algorithms Fasta, BlastN, and BlastX. ClustalW (Thompson, Higgins, and Gibson 1994) was used to perform multiple alignments. AVID was used to align long sequences (>3 kb) and mVISTA to visualize the alignments (Mayor et al. 2000; Dubchak et al. 2000). Untranslated region (UTR) functional elements were searched with UTRscan (Pesole and Liuni 1999).

Maximum likelihood estimates of the number of synonymous and nonsynonymous substitutions per site (dS and dN, respectively) were obtained with the program codeml of the PAML package (Yang 1997). For hypothesis testing, the base model assumed a rooted tree with constant nucleotide substitution rate (global clock) and a single ω = dN/dS ratio. Codon equilibrium frequencies πc were estimated from the nucleotide frequencies of the three codon sites (F3X4 option of codeml; Yang 1997). The base model included 13 parameters: the transition/transversion rate ratio (κ), the nonsynonymous/synonymous substitution ratio (ω), two relative branch lengths (r), and nine nucleotide frequencies.

Results and Discussion

Chromosomal Localization of labial in D. buzzatii and Other Species in the Drosophila Subgenus

To map labial in D. buzzatii, a fragment encompassing exons 2–3 and including the homeobox was obtained by PCR and used for in situ hybridization to the salivary gland chromosomes. Surprisingly, the lab probe hybridized to the dense polytene band G4g-5a of chromosome 2 (fig. 2) separated by about 70 bands from the site (F1c-e) where the other anterior genes of the Hox complex (pb, ftz, Antp) had been previously localized (Ranz, Casals, and Ruiz 2001). Furthermore, lab cytological position was almost indistinguishable from that of abd-A and Abd-B (Ranz, Casals, and Ruiz 2001). The localization of lab in D. buzzatii chromosomes was later corroborated by hybridization of other probes of the gene (see Materials and Methods) and indicates a split of the Hox gene complex between lab and pb. To date this split and reconstruct the evolutionary history of the Hox gene complex in the Drosophila subgenus, the same gene fragment was amplified by PCR and cloned in D. hydei, D. mercatorum, D. repleta, and D. virilis. The resulting clones were sequenced and used for
in situ hybridization to the chromosomes of those four species, as well as those of D. canalinea and D. immigrans (fig. 1). D. buzzatii, D. hydei, D. mercatorum, and D. repleta are representative species of four subgroups of the D. repleta group (Wasserman 1992), and in all of them lab was localized at band G4g-5a of chromosome 2. Likewise, in D. canalinea, which belongs to a closely related species group (Wasserman 1992), Antp and lab mapped to different chromosomal sites. However, in D. virilis, a phylogenetically more distant species, lab mapped to 24E, the same constricted region of chromosome 2 where the genes Antp, zen, and Ubx have been previously localized (vonAllmen et al. 1996). Finally, five Hox genes (abd-A, Abd-B, Antp, lab, and Ubx) were mapped by in situ hybridization to the chromosomes of D. immigrans, an even more distantly related species. In this species, lab, Antp, and Ubx mapped to the same chromosomal site, whereas abd-A and Abd-B mapped to another location of the same chromosome (fig. 1). Therefore, we conclude that the virilis-immigrans Hox gene organization (including the split between Ubx and abd-A) appears to be ancestral within the Drosophila subgenus. The newly discovered lab-pb split occurred before the radiation of the D. repleta group, 15–22 MYA, but after the divergence between the repleta and virilis species groups, 20–30 MYA (Spicer 1988; Russo, Takezaki, and Nei 1995). To our knowledge, the split lab-pb is unique among the eukaryotes and provides the opportunity to investigate the evolutionary consequences of natural Hox gene complex rearrangements.

### Table 2

Structure of the labial Gene in Three Drosophila Species and Tribolium castaneum

<table>
<thead>
<tr>
<th>Gene Region</th>
<th>D. buzzatii</th>
<th>D. virilis</th>
<th>D. melanogaster</th>
<th>T. castaneum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Promoter</td>
<td>Inr + DPE</td>
<td>Inr + DPE</td>
<td>Inr + DPE</td>
<td>TATA</td>
</tr>
<tr>
<td>5' UTR</td>
<td>757 bp</td>
<td>738 bp</td>
<td>239 bp</td>
<td>181 bp</td>
</tr>
<tr>
<td>Exon 1</td>
<td>1,300 bp</td>
<td>1,276 bp</td>
<td>1,216 bp</td>
<td>630 bp</td>
</tr>
<tr>
<td>Intron 1</td>
<td>~19 kb</td>
<td>ND</td>
<td>14 kb</td>
<td>11.5 kb</td>
</tr>
<tr>
<td>Exon 2</td>
<td>380 bp</td>
<td>380 bp</td>
<td>416 bp</td>
<td>431 bp</td>
</tr>
<tr>
<td>Intron 2</td>
<td>687 bp</td>
<td>737 bp</td>
<td>245 bp</td>
<td>—</td>
</tr>
<tr>
<td>Exon 3</td>
<td>288 bp</td>
<td>&gt;50 bp</td>
<td>258 bp</td>
<td>—</td>
</tr>
<tr>
<td>3' UTR</td>
<td>619–831 bp</td>
<td>ND</td>
<td>647 bp</td>
<td>144 bp</td>
</tr>
</tbody>
</table>

**Note:** ND: not determined.

a Data from this work.
b Accession number NG_000062.
c Data from Nie et al. (2001).
d Inr: transcription initiator; DPE: downstream promoter element.

Molecular Structure of labial in D. buzzatii and D. virilis

To determine if the lab-pb split has affected in any way the molecular structure of lab, this gene was cloned and sequenced in D. buzzatii and partially in D. virilis. The gene lab of D. melanogaster (Dm lab) has three exons and two introns, intron 1 being rather long (table 2 and fig. 3). Cloning of lab in D. buzzatii (Db lab) was carried out by screening a lambda library of D. buzzatii genomic DNA with two probes corresponding to the two distantly located exon 1 and exon 2–3 regions (see Materials and Methods). Three positive phages coming from these two regions were subcloned and partially sequenced (fig. 3): λj-19/21, including exons 2 and 3, and λj-19/68 and λj-19/89, including exon 1. Because those phages did not contain the whole gene, another phage was obtained (λj-19/77), and the remaining gap (3.5 kb) was closed by PCR amplification (table 1 and fig. 3). Overall, 12,069 bp of Db lab were sequenced, corresponding to the entire coding sequence (CDS), intron 2, the 5' UTR, the 3' UTR, and 6,241 bp from the upstream regulatory region (table 2 and fig. 3). Intron 1, with an estimated size of ~19 kb, was only partially sequenced.

In the analysis of the genomic sequence of Db lab (12,069 bp) no other conserved coding regions were found apart from those corresponding to the lab gene. However, two transposable element insertions were identified: a new copy of the ISBu2 element (Cáceres, Puig, and Ruiz 2001) of 738 bp and an ISBu3 element of 993 bp, inserted 5,500 bp and 1,663 bp, respectively, upstream from the Db lab transcription start site (fig. 3). Both elements are closely related to the ISBu1 element of D. buzzatii (Cáceres, Puig, and Ruiz 2001), and all three belong to the mini-me class described by Wilder and Hollocher (2001).

In D. virilis three genomic fragments of the lab gene (Dv lab) where amplified by PCR and sequenced (table 2): a 1,172-bp fragment corresponding to exons 2 and 3 and intron 2, a 1,268-bp fragments corresponding to exon 1, and a 1,818-bp fragment containing the 5' UTR. In addition, a 548-bp cDNA fragment comprising the area from the end of exon 1 to the beginning of exon 3 was also cloned and sequenced. Overall, 3,184 bp were sequenced, with only intron 1, part of exon 3, and the 3' UTR remaining.

The molecular structure of Db lab and Dv lab is similar to that of Dm lab, comprising three exons and two introns (table 2 and fig. 3). Therefore, the structure of the lab transcription unit has not been altered by the lab-pb split in D. buzzatii and the exact breakpoint must be located at least 8 kb upstream from the lab transcription start point. In contrast, Db lab is 32% larger than Dm lab, mostly because of changes in noncoding regions. The total length of the CDS is about 4% larger (1,968 bp versus 1,890 bp), whereas intron 1 is ~5 kb longer and intron 2 and the 5' UTR are more than twice as large in D. buzzatii (table 2). In general, the D. virilis gene is more like that of D. buzzatii than that of D. melanogaster (table 2), as expected from their phylogenetic relationships (fig. 1). Interestingly, the
lab structure differs widely between Drosophila and Tribolium castaneum (table 2). In the latter species, the CDS is 1,061 bp long with two exons only, and intron 1, the 5′ UTR, and the 3′ UTR are also significantly shorter (Nie et al. 2001).

Expression of labial and Identification of Regulatory Sequences

To verify the predicted exon structure of Db lab, total RNA from 0–12-h-old embryos was amplified by RT-PCR. Primers were designed on exon 1 and exon 3 (table 1 and fig. 3), allowing us to examine both exon junctions at the same time. The amplified cDNA fragment had the expected size of 570 bp, and its sequence confirmed precisely the predicted exon-intron junctions. In certain D. melanogaster lines, an alternative splicing of lab intron 1 that produces two mRNAs with 18 nucleotides of difference has been described (Mlodzik, Fjose, and Gehring 1988). This alternative splicing is caused by a small duplication at the end of intron 1 that provides a new functional splice acceptor site and gives rise to the longer form of the mRNA (FlyBase 2002). The analysis of the genomic and cDNA sequences of Db lab indicates that in D. buzzatii there is no alternative splicing and only the short form of mRNA is present. In D. virilis, the sequenced cDNA fragment also corresponds to the short mRNA. To identify the end of the 3′ UTR, polyadenylation signals were searched along the sequence 3′ from the stop codon. Four possible polyA signals were identified between 623 and 825 bp from the stop codon of Db lab. No conserved sequences were found between Dm lab and Db lab on the 3′ UTR.

The open reading frame and translation start site for Db lab were predicted by identification of consensus sequences and homology with Dm lab. To find the promoter and transcription start sites, 1,308 bp of genomic DNA sequence from the 5′ region of Db lab were aligned with 630 bp from the 5′ region of Dm lab (X13104 X12834) including the 5′ UTR and 400 additional nucleotides upstream (fig. 4). The most conserved region between the two sequences corresponds to the transcription start site described for Dm lab (Mlodzik, Fjose, and Gehring 1988; Diederich et al. 1989). In Db lab, we find in this conserved region the sequence TCAGTC that fits well the proposed consensus TCA\(_{+1}\)G/TTT/C of the initiator (Inr) of Drosophila genes, where A\(_{+1}\) is the transcription start site (Arkhipova 1995). No other sequence similar to the Inr consensus was found in the 5′ upstream region of Db lab. An additional alignment including the 738-bp upstream sequence of Dv lab and using the ClustalW algorithm (Thompson, Higgins, and Gibson 1994) with a gap extension penalty of 0.05 indicated also a high identity around both the transcription (Inr) and the translation (Met) start sites between the three Drosophila species (fig. 4). There is a high homology between the Inr and position +32 in the three species (fig. 4), with the sequence CACG located between positions +29 and +32 being fully conserved. This agrees with the characterization of this sequence as the Downstream Promoter Element (DPE) in Dm lab (Kutach and Kadonaga 2000). Because no TATA box was found upstream of the transcription start site, we conclude that Db lab and Dv lab present a TATA-less promoter with Downstream Promoter Element (DPE), as does Dm lab and all other homeotic
genes described in Drosophila (Nie et al. 2001). Analysis with the UTRscan software (Pesole and Liuni 1999) of the 5' UTRs of the three Drosophila species identified internal ribosome entry site (IRES) elements in all of them. These IRES elements promote cap-independent translation in different genes (Hellen and Sarnow 2001). In the Hox genes *Ubx* and *Antp*, they regulate translation during development (Ye et al. 1997).

Given the time elapsed since the separation of the Sophophora and Drosophila subgenera (40–60 Myr), no significant nucleotide identity should be expected in noncoding regions in the absence of selective pressure to conserve the function (O'Neil and Belote 1992). Thus, conserved structures may be interpreted as functionally important for gene expression. The entire sequence for *Db lab* (12,069 bp; fig. 3) was aligned with *Dm lab* (NG_000062) with the AVID algorithm and visualized with the VISTA software (Mayor et al. 2000; Dubchak et al. 2000). The resulting graph (fig. 5) highlights those segments with a nucleotide identity >50% between *Db lab* and *Dm lab*. Conserved regions both in coding and noncoding sequences are shown relative to their position in *Db lab*. There is a high variability in conservation along the coding sequence. Several peaks of conserved sequence appear along exon 1, which could be a consequence of the folding pattern of the protein. Two peaks of conserved sequence at the end of exon 2 and the beginning of exon 3 correspond to the homeobox. In the noncoding regions, two prominent peaks of conserved sequence show up, one segment 180-bp long at the end of intron 1 with an identity of 82.8%, and a second segment 108 bp long in the 5' region with 84.3% identity. In addition, downstream of this second conserved segment there is a ~750-bp region with 50%–75% identity. A Blast search with those conserved sequences from *D. buzzatii* shows similarity only with the same sequence in *D. melanogaster*.

We have compared the position of those noncoding conserved sequences with the regulatory elements described for *Dm lab* (Chouinard and Kaufman 1991). The conserved regions upstream of *lab* are inside the 2.4-kb fragment that promotes the initiation of procephalon expression. The small fragment with 50%–65% identity located 4,800 bp upstream of *Db lab* corresponds to the 1.2-kb *Dm lab* fragment that drives expression in the anterior midgut via a decapentaplegic (DPP)-regulated enhancer (*lab550* – Marty et al. 2001). Finally, the 180-bp conserved segment at the end of intron 1 corresponds to a fragment not tested by Chouinard and Kaufman (1991), which might contain the regulatory element required for embryonic peripheral nervous system (PNS) expression.

### Nucleotide Sequence Comparisons

The coding sequences of *Db lab* and *Dv lab* were aligned and compared to that of *Dm lab* (NM_057265). Table 3 shows the number of synonymous and nonsynonymous substitutions per site ($d_s$ and $d_a$) for pairwise comparisons between the three Drosophila species. These numbers have been estimated using maximum likelihood methods (Yang and Bielawski 2000; Yang and Swanson...
2002) for the entire coding sequence (CDS) as well as for three separate (non-overlapping) regions: exon 1, exon 2 + 3 (excluding the homeobox) and the homeobox. The ratio between the number of nonsynonymous and synonymous substitutions (ω = dN/dS) provides information on the type of selection acting on amino acid sequences. The low average ω values for the entire gene (0.0696–0.0910) suggest a relatively high degree of functional constraint, but there is wide variation among regions (see below).

To investigate if the molecular evolution of lab had been affected by the lab-pb split in the D. repleta species group, three hypotheses related to the molecular evolution of lab were tested by comparing the likelihood of different models. First, we tested the constancy of the evolutionary rates by comparing the base model (which assumes a molecular clock and a single ω ratio for all lineages) with a model assuming no clock. The results were nonsignificant for the entire CDS and the three separate coding regions (table 4). Second, we tested for variation in the ω ratio among lineages by comparing the base “one ratio” model with a “free-ratio” model, involving as many ω ratios as branches in the tree (four). The results again were nonsignificant both for the entire CDS and for the three separate coding regions (table 4). Therefore, the evolutionary rate of lab has not accelerated or decelerated significantly after its separation from the anterior Hox gene complex. Likewise, the ω ratio is not different in the phylogenetic branch leading to D. buzzatii than in the rest of Drosophila, and there is no evidence that positive selection or relaxation of selection has occurred after the relocation of lab by the lab-pb split.

Finally, we tested for differences among the three separate coding regions, fitting a series of fixed-sites models as described by Yang and Swanson (2002). A detailed list of the models and their results is given in table 5. We were particularly interested in the among-region variation in ω ratio. To test the hypothesis of no variation, model D (which assumes different rs, κ, and ω) was compared with model B (which assumes different rs). The log-likelihood difference, 2(lD − lB) = 71.65, is highly significant (df = 19–15 = 4; P < 0.001) indicating differences either in the transition/transversion ratio κ or the ratio ω. A modified D model (designated D′) with a fixed κ value was compared to model B, and the resulting log-likelihood difference, 2(lD′ − lB) = 64.36, showed that the variation in the ω ratio was indeed significant (df = 16–15 = 1; P < 0.001). Similar tests can be conducted comparing model E (which assumes different rs, κ, and ω, and rs) with model C (which assumes different rs and rs). The difference in log-likelihood, 2(lE − lC) = 85.06, is highly significant (df = 37–33 = 4; P < 0.001), indicating again significant differences either in the transition/transversion ratio κ or the ratio ω. If a modified E model

![Fig. 5.—VISTA plot showing similarity between the nucleotide sequences of the lab gene of D. melanogaster and D. buzzatii. Conserved regions are shown relative to their position in Db lab (horizontal axis), and their percent identities (50%–100%) are indicated on the vertical axes. The location of coding exons (black rectangles) and 5' and 3' UTRs (empty rectangles) are shown above the profile, where the thin arrow indicates the direction of transcription and gray boxes represent transposable elements. Peaks with an identity of at least 70% over 100 bp are shaded.](image-url)
Table 4
Likelihood Ratio Tests of Variation in the Rate of Molecular Evolution and the \( \omega \) Ratio of the Gene labial in the Genus Drosophila

<table>
<thead>
<tr>
<th>Region(^a)</th>
<th>Number of Codons</th>
<th>No Clock (df = 1)</th>
<th>Lineage-specific ( \omega ) ratio (df = 3)</th>
<th>( \omega ) Ratio in Base Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDS</td>
<td>695</td>
<td>0.98</td>
<td>2.43</td>
<td>0.0971</td>
</tr>
<tr>
<td>Exon 1</td>
<td>449</td>
<td>3.66</td>
<td>3.55</td>
<td>0.0766</td>
</tr>
<tr>
<td>Exon 2 + 3</td>
<td>186</td>
<td>1.19</td>
<td>3.17</td>
<td>0.3374</td>
</tr>
<tr>
<td>Homebox</td>
<td>60</td>
<td>0.28</td>
<td>0.00</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

\( a \) Exon 2 + 3 does not include homebox.

Amino Acid Sequence Comparisons

Seven different proteins encoded by lab orthologous genes were aligned with the ClustalW algorithm (fig. 6). These include the complete Labial sequences from \( D. buzzatii \) (present work), \( D. melanogaster \) (Mlodzik, Fjose, and Gehring 1988; Diederich et al. 1989), \( T. castaneum \) (Nie et al. 2001), and \( H. sapiens \) (Hong et al. 1995) and partial Labial sequences from \( D. viridis \) (present work), Anopheles gambiae (Powers et al. 2000) and Thermobia domestica (Peterson et al. 1999). Complete sequences were aligned with default parameters. To avoid artifacts, partial sequences were aligned one by one with \( D. buzzatii \) increasing gap extension penalty and included in the global alignment afterwards. The most intriguing observation in this comparison is the much larger size of the Drosophila Labial proteins (629–655 aa) compared to those of \( T. castaneum \) (353 aa) or \( H. sapiens \) (335 aa). Whereas Labial has almost duplicated its size in Drosophila in relation to the other species, other Hox genes do not show this variation. For example, the gene \( abd-A \) presents a similar size in \( T. castaneum \) and \( D. melanogaster \) (Shippy, Brown, and Denell 1998) and it is only slightly smaller in \( A. gambiae \) (Devenport, Blass, and Eggleson 2000).

The most conserved region between all Labial proteins is the 65-aa homeodomain, located at the C-terminal region, with an identity between 83% and 100%. Outside the homeodomain there are some other conserved regions. The YKWM motif, a variant of the core of the hexapeptide motif, YPWM, involved in the interaction with the cofactor EXD (Gehring, Affolter, and Bürglin 1993; Rauskolb and Wieschaus 1994, Johnson, Parker, and Krasnow 1995), presents an identity between 60% and 100% through 15 aa. There is also a short region around the YTNLD motif with an identity between 44% and 100%. Finally the N-terminal region presents between 50% and 100% identity. The conservation of these amino acid sequences across long phylogenetic distances suggests that they suffer functional constraints and are probably involved in highly conserved interactions, such as the YKWM domain (Johnson, Parker, and Krasnow 1995). These results are fully consistent with the previous nucleotide comparison between Drosophila species. The homeodomain is the most conserved region of the protein and shows a high level of nucleotide sequence evolution constraint, whereas all other conserved domains are located in exon 1, which is more constrained than exons 2 and 3 (excluding the homeobox).

Outside those conserved domains there is a very low conservation, which is surprising when compared with other proteins. The Labial proteins from \( D. buzzatii \) and \( D. melanogaster \) have an overall identity of only 62.8%, much lower than the 95% identity found by Hooper et al. (1992) between the Antp proteins from \( D. viridis \) and \( D. melanogaster \) (with the same divergence time). There are very few direct comparisons of homologous genes between \( D. melanogaster \) and \( D. buzzatii \), but \( D. melanogaster \) and \( D. viridis \) homologous proteins usually have an identity between 80% and 90%, with values ranging from 36% to 97% (O’Neil and Belote 1992). Interestingly, all regions

Table 5
Log-Likelihood Values and Parameter Estimates Under Fixed-Sites Models

<table>
<thead>
<tr>
<th>Model</th>
<th>( p )</th>
<th>( l )</th>
<th>( r_2 )</th>
<th>( r_3 )</th>
<th>( \kappa )</th>
<th>( \omega )</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (homogeneous model)</td>
<td>13</td>
<td>-5147.69</td>
<td>1</td>
<td>1</td>
<td>1.6552</td>
<td>0.0971</td>
</tr>
<tr>
<td>B (different ( rs ))</td>
<td>15</td>
<td>-5125.17</td>
<td>1.569</td>
<td>0.387</td>
<td>1.6433</td>
<td>0.0862</td>
</tr>
<tr>
<td>C (different ( rs ) and ( \pi s ))</td>
<td>33</td>
<td>-5087.45</td>
<td>1.398</td>
<td>0.277</td>
<td>1.6547</td>
<td>0.0813</td>
</tr>
<tr>
<td>D (different ( rs ), ( \kappa ), and ( \omega ))</td>
<td>19</td>
<td>-5089.34</td>
<td>0.942</td>
<td>0.983</td>
<td>( \kappa_1 = 1.4036; \kappa_2 = 0.0853; \kappa_3 = 2.5631 )</td>
<td>( \omega_1 = 0.2857; \omega_2 = 0.7441; \omega_3 = 0.0001 )</td>
</tr>
<tr>
<td>D’ (( \kappa ) fixed to 1.643)</td>
<td>16</td>
<td>-5092.99</td>
<td>0.888</td>
<td>0.966</td>
<td>( \kappa = 1.643 )</td>
<td>( \omega_1 = 0.0883; \omega_2 = 0.2718; \omega_3 = 0.0001 )</td>
</tr>
<tr>
<td>E (different ( rs ), ( \kappa ), and ( \omega ), and ( \pi s ))</td>
<td>37</td>
<td>-5044.93</td>
<td>0.726</td>
<td>0.613</td>
<td>( \kappa_1 = 1.4864; \kappa_2 = 0.0809; \kappa_3 = 2.3396 )</td>
<td>( \omega_1 = 0.3255; \omega_2 = 1.1217; \omega_3 = 0.0001 )</td>
</tr>
<tr>
<td>E’ (( \kappa ) fixed to 1.655)</td>
<td>34</td>
<td>-5046.94</td>
<td>0.703</td>
<td>0.670</td>
<td>( \kappa = 1.655 )</td>
<td>( \omega_1 = 0.0824; \omega_2 = 0.3107; \omega_3 = 0.0001 )</td>
</tr>
<tr>
<td>F (separate analysis)</td>
<td>39</td>
<td>-5042.39</td>
<td>1.4978</td>
<td>2.3044; ( \kappa_3 = 1.0623 )</td>
<td>( \omega_1 = 0.0766; \omega_2 = 0.3374; \omega_3 = 0.0001 )</td>
<td></td>
</tr>
</tbody>
</table>

\( \text{Note.} - \) \( p \): number of parameters of the model; \( l \): log-likelihood value; \( r_2 \) and \( r_3 \): relative branch lengths.
FIG. 6.—Multiple alignment of proteins coded by lab orthologs from different organisms. D. buzzatii (Db) and D. virilis (Dv) from present work. GenBank accession numbers: D. melanogaster (Dm, CAB57787), A. gambiae (Ag, AAF91398.1 and AAAB01008961), Tribolium castaneum (Tc, AAF64147.1 and AAF64148.1), Thermobia domestica (Td, AAD50360.1), Homo sapiens (Hs, HoxA1, NP_005513.1). Dv, Ag, and Td are partial sequences. Amino acids in lowercase correspond to degenerate primers. Alignments were performed with ClustalW with default parameters, except for Ag and Td, which have been aligned individually with Db, increasing the gap extension penalty and included in the global alignment.
conserved between Dm lab and Dm lab are also more or less conserved in all other sequences studied, including the human gene HoxA1. The domain conservation suggests that the protein functionality relies basically on a few domains, leaving the rest of the sequence with more freedom to change. This fact could explain why minigenes carrying the chicken Hoxb1 gene are able to rescue null mutants for the lab gene in D. melanogaster (Lutz et al. 1996). Thus, Labial proteins seem to be built on highly conserved blocks joined by rapidly evolving sequences. For example, the spacer or “gene specific region” (Peterson et al. 1999) localized between the YKWM motif and the homeodomain is much more variable, both in sequence and in length, in the lab gene than in other Hox genes even between Drosophila species.

Concluding Remarks

Although Hox gene colinearity and cluster conservation represents a paradigm in Evo-Devo, in the genus Drosophila the Hox cluster has been split at least three times. The lab-pb split described here is the most recent example. Even after moving to a new location, labial appears to evolve consistently, showing no alteration of its molecular evolutionary patterns. A paracentric inversion as well as a gene transposition event might, in principle, account for the lab-pb disruption and the relocation of lab near the genes abd-A and Abd-B observed in the replet group of Drosophila. However, the former possibility seems more likely given the extensive reorganization of the Drosophila genome produced by chromosomal inversions (Powell 1997; Ranz, Casals, and Ruiz 2001). The molecular characterization of the chromosomal regions 5' of lab and 3' of pb in D. buzzatii should help to determine which was the molecular mechanism that caused such a split and the precise location of the breakpoint between lab and pb. This study would also elucidate the exact physical relationship, if any, between lab and the genes abd-A and Abd-B. Furthermore, although the rearrangement has not altered the lab coding region, and although some of its regulatory sequences seem to be in place, whether the relocation of lab had any consequences upon the level or spatiotemporal pattern of expression of this gene remains an open question, one that we are already investigating.

Acknowledgments

We thank S. Celniker, J. E. Hooper, and E. Sánchez-Herrero for probes; the National Drosophila Species Resource Center (Bowling Green, Ohio, USA) for fly strains; and J. González and E. Betrán for advice and helpful comments on the manuscript. This work was supported by grant PB98-0900-C02-01 from the Dirección General de Enseñanza Superior e Investigación Científica (MEC, Spain) awarded to A.R. and by a doctoral FI/DGR fellowship from the Generalitat de Catalunya awarded to B.N.

Literature Cited


William R. Jeffery, Associate Editor

Accepted July 14, 2003