# A genomewide survey of basic helix–loop–helix factors in *Drosophila*

**Adrian W. Moore, Sandra Barbel, Lily Yeh Jan, and Yuh Nung Jan***

Howard Hughes Medical Institute, Departments of Physiology and Biochemistry, University of California, San Francisco, CA 94143-0725

Contributed by Yuh Nung Jan, June 29, 2000

**The basic helix–loop–helix (bHLH) transcription factors play important roles in the specification of tissue type during the development of animals. We have used the information contained in the recently published genomic sequence of *Drosophila melanogaster* to identify 12 additional bHLH proteins. By sequence analysis we have assigned these proteins to families defined by Atonal, Hairy-Enhancer of Split, Hand, p48, Mesp, MYC/USF, and the bHLH-Per, Arnt, Sim (PAS) domain. In addition, one single protein represents a unique family of bHLH proteins. mRNA *in situ* analysis demonstrates that the genes encoding these proteins are expressed in several tissue types but are particularly concentrated in the developing nervous system and mesoderm.**

The basic helix–loop–helix (bHLH) proteins comprise an evolutionarily ancient, important group of transcription factors in animals, plants, and fungi. Their functions range from control of cellular proliferation to tissue differentiation. They are united by conservation solely within the bHLH domain (1) and act as dimers binding the E-box site CANNTG to regulate transcription (2, 3). The bHLH domain consists of the basic domain, a run of approximately 15 aa with a high number of basic residues, followed by two amphipathic $\alpha$-helices separated by a loop region of variable length. Interaction between the helix regions of two different proteins leads to intermolecular dimerization, and the basic region of each partner binds to half of the E-box sequence (4–6).

In animals, bHLH proteins often are used in cascades to specify tissue identity. Furthermore, closely related families of bHLH proteins as defined by their level of identity in the bHLH domain tend to have functions in a similar tissue type. For example, in *Drosophila*, Twist is required for mesoderm specification. It then acts alongside (the MyoD family ortholog) Nautilus (Nau) in myogenesis. Similarly, in vertebrates the initial specification and division of the mesoderm involves the activity of the mTwist family members; at later stages, the MyoD family acts at different steps of the myogenic pathway (7).

bHLH protein cascades are also well studied in neurogenesis in both *Drosophila* and vertebrates. The Achaete–Scute (AS) proteins [Achaete (Ac), Scute (Sc), Lethal of Scute (L'sc), and Asense (Ase)] and Atonal-related proteins [Atonal (Ato), Amos, Cato, and Tap] function at several stages of this process in *Drosophila*. Initially, the proneural proteins (8) Ac, Sc, and L'sc in the central nervous system (CNS) (9) and Ato, Amos, and Sc in the peripheral nervous system define the field of neural competence (10–12). Later, Tap (13) is required for specifying neural type and Ase (14) and Cato (15) are required for neural differentiation. In vertebrates, orthologs of these genes are active in several neurogenic cascades (16). For example, the AS ortholog Mash1 initiates a cascade of bHLH proteins in the mammalian olfactory neuron lineage followed by the Ato orthologs Neurogenin 1 (Ngn1) and NeuroD.

The Ato-related AS and MyoD families bind to DNA as heterodimers with the ubiquitously expressed Daughterless (Da) in *Drosophila* or Da orthologs such as E47 or E12 in vertebrates (17) to positively regulate transcription. On the other hand, the Hairy-Enhancer of split (HES) proteins act during neurogenesis as antagonizers of Ato-related and AS function. They act to repress transcription via interaction with the non-bHLH Groucho protein (18).

Genes of the MYC family of bHLH proteins contain a leucine zipper (zip) immediately C terminal to the bHLH domain. MYC forms a heterodimeric transcription factor with the related MAX or MAD proteins to regulate cell proliferation (19). The mammalian upstream transcription factor (USF) proteins also have a bHLH-zip structure and are proposed to act to antagonize the proliferation function of MYC (20).

A further group of bHLH proteins contains a Per, Arnt, Sim (PAS) domain C-terminal to the bHLH domain. This functions to mediate protein–protein interaction. The bHLH-PAS proteins form heterodimeric transcription factors with roles in development [e.g., Tango (21)], toxin metabolism [e.g., Rst(1)JH (22)], and the regulation of circadian rhythm-expressed genes (23).

The recent publication of the entire heterochromatic sequence of *Drosophila melanogaster* (24) presents an opportunity to analyze those members of the bHLH family as yet unstudied. bHLH transcription factors act in a cell-autonomous manner. Hence, the *in situ* data (Table 1 and Fig. 3) along with the sequence analysis (Figs. 1 and 2) presented in this paper give clues about the developmental roles of these newly identified proteins in *Drosophila*. The proteins identified in this study also may facilitate the search for their orthologs in vertebrate species. Finally, by using a genomewide survey we can ask whether all tissue types in the developing *Drosophila* embryo have bHLH cascades associated with their genesis.

## Materials and Methods

**Database Search.** We used the BLAST search (25) and PATTERN search programs provided by The National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov) and Berkley *Drosophila* Genome Project (http://www.fruitfly.org) to search the recently published sequence and predicted protein information from the entire *Drosophila* genome (24). As query sequence, we used the amino acid sequences of the bHLH domains of several known bHLH proteins and sequences based on the predicted degenerate sequence for all bHLH domains (26). We restricted our search to those proteins with an intact basic domain.

**Tree Building and Sequence Lineups.** We constructed a database consisting of the amino acid sequences of the bHLH domains of the 12 proteins reported in this study and the closest related proteins to these (identified by BLAST search). In addition, we added all known *Drosophila* bHLH-containing proteins and representative members of all of the different bHLH domain-

Abbreviations: bHLH, basic helix–loop–helix; CNS, central nervous system; HES, Hairy-Enhancer of split; PAS, Per, Arnt, Sim; AS, Achaete–Scute; VNC, ventral nerve cord.

*To whom reprint requests should be addressed. E-mail ynjan@itsa.ucsf.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Article published online before print: *Proc. Natl. Acad. Sci. USA*, 10.1073/pnas.170301897. Article and publication date are at www.pnas.org/cgi/doi/10.1073/pnas.170301897

**Table 1. Summary of gene expression using the expression code of Hartenstein and Jan (48)**

| Gadfly no. | Name | mRNA expression |
|---|---|---|
| CG8667 | Mistr | **N7+ring gland**:9 |
| CG5545 | Doli | **N1**:9 |
| CG10066 | Fer1 | **Epidermis**:15 |
| CG5952 | Fer2 | **N5**:11 |
| CG6913 | Fer3 | **M**:11 |
| CG18144 | Dm Hand | **N5**:13 **M1**:10 **DU2**:10 |
| CG10446 | Side | **N5**:12 |
| CG5927 | Her | **Ubiquitous** |
| CG12952 | Sage | **SH3**:10 |
| CG6211 | Gce | **GO**:9 |
| CG11450 | Shout | **SM1**:5 **AS**:5 |
| CG17592 | Dm Usf | **N7+proventriculus**:14 |

Bold type indicates expression pattern code; normal type indicates the stage at which expression is initiated.

containing families (27). The sequences were aligned and the tree was constructed by using the CLUSTALX program (28). Gaps in the sequence alignment were ignored in the tree-building process. To increase the predictive sequence available to the CLUSTALX program, the genes of the AS complex (AS-C) were not included in the data set for tree drawing. Ten trees were constructed by using different randomizations of sequence input order and bootstrapped 1,000 times to give an indication of the predictive strength of each node.

***In Situ* Analysis.** Analysis of mRNA distribution in *Drosophila* embryos was carried out as described in ref. 29. As hybridization probes, we used digoxigenin-labeled PCR product amplified by using primers designed to the coding region of each of these genes and either *Drosophila* genomic DNA or embryo-derived cDNA as template.

## Results

We have identified 12 additional proteins of the bHLH class and estimate from surveying the genome that the total number of bHLH genes in *Drosophila* with an intact basic domain is 46. We used the neighbor-joining method to classify these 12 proteins into families (28) (Fig. 1). The publication of the genomic sequence of *Drosophila* included an initial annotation of the sequence (24). This study identified 69 putative proteins with a HLH dimerization domain (30) that are listed at the Genome Annotation Database of *Drosophila* (GadFly) web site (http://www.fruitfly.org/annot/). Eleven of the proteins identified in this study are a subset of the HLH proteins identified; CG18144 was not identified in the initial genome annotation analysis.

In two cases, CG5952/Fer2 and CG17592/Dm Usf, the predicted protein sequence contained inserted sequences that disrupted the basic helix1–loop–helix2 domains. Using embryonic cDNA as a template, products spanning the region of doubt were amplified by PCR. Subsequent sequencing of these products demonstrated that the predicted sequence of CG17592/Dm Usf is correct. The sequence of CG5952/Fer2 presented in this paper has been altered from that predicted from the genomic sequence.

**CG8667 (Mistr) and CG5545 (Doli), Part of the Ato-Related Family, Are Expressed in the Developing Nervous System.** CG5545 is closely related to the vertebrate Beta 3 protein, a repressor molecule (31) (96% sequence identity in the bHLH domain), and the Olig proteins involved in oligodendritic precursor formation (32, 33) (Fig. 2*a*). We suggest that this protein should be named Doli (*Drosophila* Olig family). CG8667 has closest sequence identity to the vertebrate Mist1 protein (34), a negative regulatory factor of MyoD activity (78% identical over the entire bHLH domain and 92% identical in the basic domain alone). We propose that this protein should be named Mistr (Mist 1-related protein).

As with the other proteins of the Ato-related family, the genes

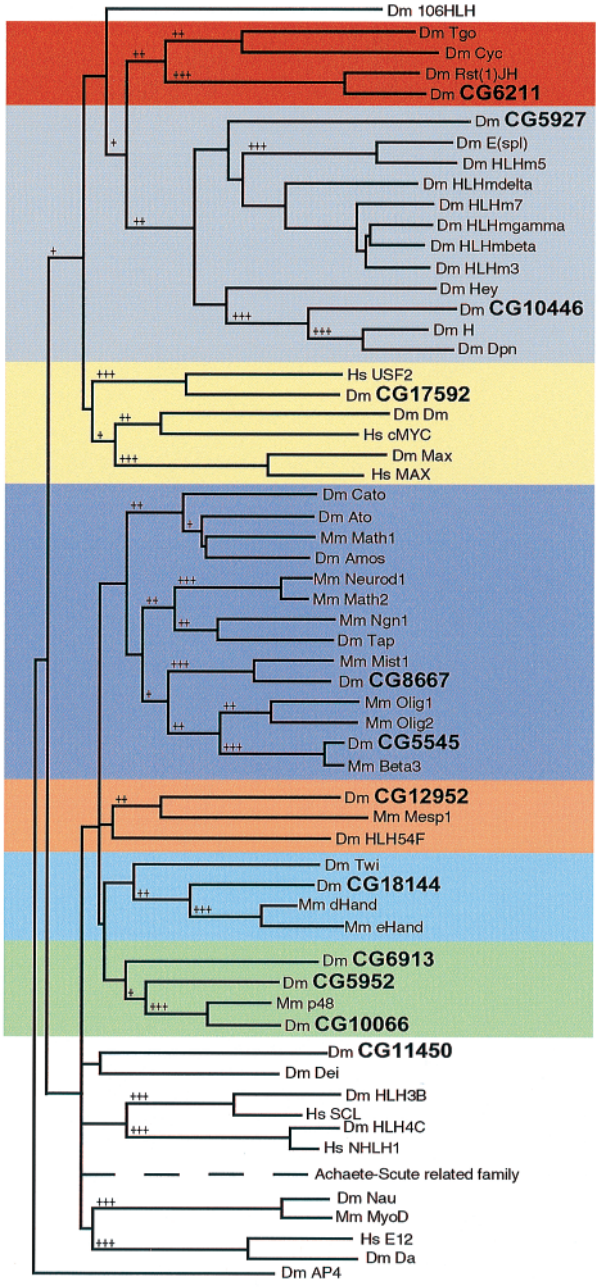


**Fig. 1.** Neighbor-joining plot of bHLH domains. A representative neighbor-joining plot was constructed by using the bHLH domain. To simplify the tree, nodes with a bootstrap value of less than 5% have been removed and the lower branch has been increased in length accordingly. Nodes with a bootstrap value of less than 50% are unmarked, those at 50–75% are marked +, 75–95% are marked ++, and 95–100% are marked +++. Note that the majority of lower (right-hand side) nodes in the tree are of very low predictive value and, hence, are unlikely to give a true indication of the phylogenetic relationship between sequences. On the other hand, upper nodes in the tree (left-hand side) are of a high predictive value. The approximate position of the genes of the AS-C has been indicted on the plot by comparison with trees drawn with the same data set, with the genes of the AS-C included (dashed line). A colored background marks subfamilies of bHLH proteins: bHLH-PAS (red), HES (gray), bHLH-zip (yellow), Atonal-related (dark blue), Mesp-related (orange), Hand family (light blue), and p48-related (green).

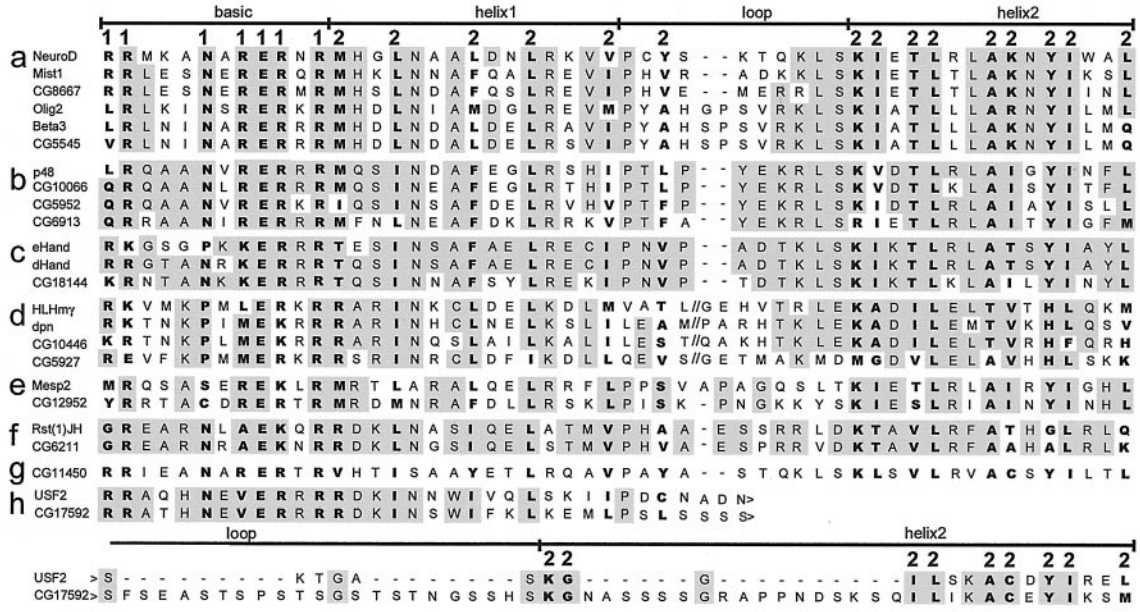


**Fig. 2.** Amino acid sequence lineups of bHLH families. Residues with predicted structural function are in bold (5); 1, DNA binding; 2, helix formation. Residues with amino acid identity between the family members illustrated are shaded. (*a*) Atonal-related. (*b*) p48-related. (*c*) Hand-related. (*d*) HES family; omission of residues in the loop region of this family is marked with two slashes (//). (e) Mesp-related. (*f*) bHLH-PAS. (*g*) CG11450/Shout. (*f*) USF family (note: the loop and the second helix domains of CG17592 are extended with respect to USF2 and contain multiple serine residues).

encoding these proteins are expressed in the developing *Drosophila* nervous system. *CG5545/doli* is expressed first in a subset of cells in both the ventral nerve cord (VNC) and the procephalic region at stage 9. The number of cells in these regions expressing the gene increases to a peak at stage 11 (Fig. 3. 3*a*). By stage 14, levels of expression have fallen such that *CG5545/doli* is expressed only in a few cells per hemisegment on the ventral surface of the VNC.

There is a strong maternal contribution of *CG8667/mistr* mRNA. Zygotic transcription is initiated at stage 14. It is expressed in bilateral domains in the cephalic region, which, as development proceeds, fuse into a U shape forming part of the ring gland (Fig. 3*b*). Concomitant expression of *CG8667/mistr* also begins in the CNS. By stage 17, *CG8667/mistr* is in clusters of cells at the anterior and posterior of the VNC and bilaterally in two lateral cells per hemisegment in the VNC.

**CG10066 (Fer1), CG5952 (Fer2), and CG6913 (Fer3) Are Related to Mammalian p48.** Three new bHLH proteins are most closely related to the bHLH domain of the p48 subunit of PTF1, a pancreatic, exocrine cell-specific transcription factor in the mouse (35), and represent a new bHLH family in *Drosophila*. We name these proteins Fer for 48 related. CG10066/Fer1 is 88%, CG5952/Fer2 is 76%, and CG6913/Fer3 is 62% identical to p48 in the bHLH region (Fig. 2*b*).

*CG10066/fer1* is expressed in the epidermis at the stage when it begins to secrete cuticle and, therefore, may share a common function with p48 in active exocrine cells. It is first transcribed in the epidermal pads adjacent to the posterior spiracles at stage 15. The expression of this gene quickly spreads over the entire epidermal surface of the embryo and is strongest in epidermis underlying the forming denticle belts (Fig. 3*c*).

*CG5952/fer2* shows a strong maternal contribution of mRNA in the early embryo. Zygotic expression of this gene begins at stage 10 in an anterior-to-posterior wave in the VNC and the brain. As development proceeds, the number of *CG5952/fer2*-positive cells increases, so that by stage 12, the expression domain forms a bilateral, dorsal–posterior, crescent-shaped structure (Fig. 3*d*).

*CG6913/fer3* is expressed at stage 11 in part of the posterior midgut primordia and stage 12 in part of the anterior midgut primordia. At later stages, we have detected expression in several unidentified cells scattered throughout the embryo.

**CG10446 (Side) and CG5927 (Her) Are in the HES Family.** CG10446 is most closely related to Dpn (76% identity in the basic and 62% in the entire bHLH domain). We name this protein Side (similar to Deadpan). CG5927 is most closely related to the proteins of the Enhancer of split [E(spl)] complex, such as HLHmγ (76% identity in the basic and 51% identity in the entire bHLH domain) (Fig. 2*d*). We name CG5927 Her (HES-related). Hairy, Dpn, and the proteins of the E(spl) complex have WRPW at the very C terminus to mediate interaction with Groucho. CG5927/Her and CG10446/Side also end in this motif.

There is a strong maternal contribution of *CG10446/side* mRNA. Zygotic transcription of the gene begins at stage 12 in a subset of cells in the CNS (Fig. 3*e*). *CG5927/her* has a low level of maternal mRNA contribution and then is expressed ubiquitously throughout embryogenesis.

**CG12952 (Sage) Is Distantly Related to the Mesp Family and Expressed in the Salivary Gland.** CG12952 represents a protein with little sequence similarity to other known proteins. In the neighbor-joining tree, it is placed in the same family as the vertebrate Mesp proteins, which are necessary for mesoderm segmentation initiation (53% identity in the bHLH domains) (Fig. 2*b*) (36). *CG12952* has a strong maternal mRNA contribution in early embryogenesis. Its zygotic expression begins in the salivary gland anlage at stage 10 and persists until stage 15 (Fig. 3*f*). We name CG12592 Sage (salivary gland-expressed bHLH).

**CG17592 (Dm Usf) Is the Ortholog of the Mammalian USF Proteins.** CG17592 is the single *Drosophila* sequence homologue of the vertebrate USF proteins that are involved in cell proliferation control (92% identical in the basic domain) (Fig. 2*h*) (20). We term this protein Dm Usf. Both vertebrate and *Drosophila* USF

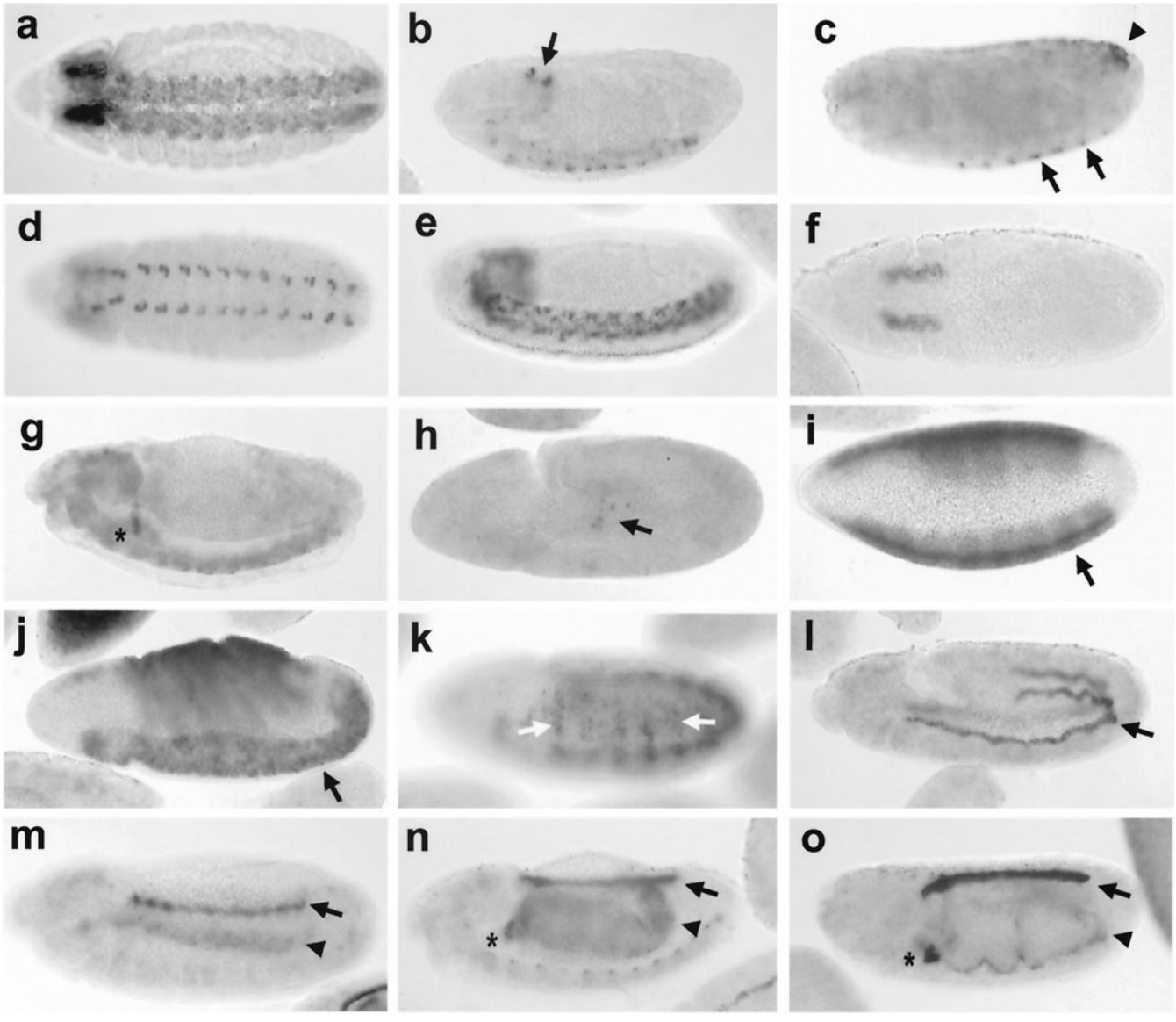


**Fig. 3.** mRNA expression in *Drosophila* embryos. Except where indicated, ventral is down and anterior is to the left. (*a*) *CG5545/doli* widespread in the CNS at stage 12. (*b*) *CG8667/mistr* in the VNC and parts of the ring gland (arrow) at stage 16. (*c*) *CG10066/fer1*, over the whole surface of the embryo but concentrated in two posterior dorsal pads (arrowhead) and around the forming denticle belts (arrows) at stage 16. (*d*) *CG5952/fer2* stg12 expression in the VNC and brain at stage 12. (*e*) *CG10046/side* in the CNS at stage 12. (*f*) *CG12952/sage* dorsal view in the salivary glands at stage 15. (*g*) *CG17592/Dm usf* in the VNC and proventriculus (*) at stage 15. (*h*) *CG6211/gce* in a subset of germ cells (arrow) stage 9. (*i–k*) *CG11450/shout* at stage 5 (*i*), stage 8 (*j*), and stage 11 (*k*). At the cellular blastoderm stage, *CG11450/shout* transcript is expressed first in the dorsal and then in the ventral blastoderm (*i*, arrow). This expression becomes hemisegmentally repeated in the dorsal and ventral blastoderm but is not present in the procephalic region or telson and is not present in the lateral regions of the embryo. During gastrulation, all of the *CG11450/shout*-expressing cells on the ventral surface of the embryo move through the ventral furrow (*j*, arrows). It then is expressed in myoblast cells as they migrate underneath the body wall of the embryo (*k*, arrows). *CG18144/Dm hand* expression at stage 9 (*l*), stage 13 (*m*), stage 14 (*n*), and stage 16 (*o*); *Dm hand* expression is seen first at stage 10 of embryo development. It occurs in bilateral stripes in the ventral mesoderm (*l*, arrow). These stripes represent the precursors of the forming dorsal vessel. *Dm hand* continues to be expressed in this structure, including the latest stage we have investigated. As germ-band retraction proceeds, a second bilateral strip of *Dm hand* expression becomes apparent in the ventral mesoderm (*m*, arrowhead). These cells are the precursors of the circular visceral musculature. *Dm hand* continues to be expressed in the cells of the circular visceral musculature of the gut as they migrate over the gut surface (*m–o*). Additionally, there is a high level of *Dm hand* expression in the external portion of the proventriculus (*n–o*, *). At late stage 13, *Dm hand* expression comes on in the VNC and brain. In the VNC, *Dm hand* is expressed in every hemisegment in a small group of cells at the lateral–dorsal part of the hemisegment adjacent to the perineurium (*n*).

are bHLH-zip proteins. Dm Usf has a loop and a second helix region, high in serines, which is greatly diverged from that of mouse and human and, hence, may have lost its ability to dimerize (Fig. 2*h*). There is a weak maternal contribution of *Dm usf* mRNA. At stage 7, *Dm usf* is expressed in bilateral domains in the ventral cephalic furrow. In later stages (15 onward) of development, *Dm usf* expression is confined to the proventriculus and a subset of cells in the CNS (Fig. 3*g*). This specific expression pattern differs from the ubiquitous *USF* expression pattern reported in vertebrates.

**CG6211 (Gce) Is Closely Related to the bHLH-PAS Rst(1)JH Protein.** CG6211 is closely related to the Rst(1)JH protein, a bHLH-PAS protein (78% identity in the bHLH, 68% in the PAS-A, and 86%

in the PAS-B domains) (Fig. 2*f*). Rst(1)JH originally was isolated in a screen to find *Drosophila* resistant to the Juvenile Hormone Analog insecticide Methoprene. *CG6211* transcript is expressed strongly as a maternally supplied message and then later in a subset of the germ cells of the developing embryo (Fig. 3*h*). We suggest that this protein should be named Gce (germ cell-expressed bHLH-PAS).

**CG11450 (shout) Is Expressed During Mesoderm Formation and in Myoblasts.** CG11450 represents a member of a new bHLH family (Figs. 1 and 2*g*). It is expressed first in the dorsal and ventral cellular blastoderm. In the ventral region of the embryo, the gene is expressed continually in the presumptive mesoderm throughout gastrulation and then in a segmented pattern in the ventral mesoderm layer at the extended germ-band stage. It is expressed in the myoblast cells that then migrate dorsally from this layer (Fig. 3 *i–k*). The expression pattern of *CG11450* overlaps with that of the bHLH transcription factor *twist*, suggesting that it may be playing a role in the same mesoderm specification and myogenic pathways; therefore, we term this gene *shout* after "Twist and Shout" by Lennon and McCartney (1963).

**CG18144 (Dm Hand) Is the *Drosophila* Ortholog of the Vertebrate Hand Proteins.** CG18144 (Dm Hand) represents the single *Drosophila* ortholog of the vertebrate dHand (76% identity) and eHand (69% identity in the bHLH domain) proteins involved in heart formation (37) (Fig. 2*c*). *Dm hand* expression begins at stage 10 of embryonic development in bilateral stripes in the ventral mesoderm. It continues to be expressed in two tissues derived from this mesoderm, the dorsal vessel (heart) and the circular visceral musculature. In addition, at stage 13 *Dm hand* mRNA appears in a small subset of cells in the CNS (Fig. 3 *l–o*).

## Discussion

In this study we have identified 12 additional bHLH proteins and examined the expression pattern of the genes that encode them. The majority of the proteins we identified are members of previously identified families. In these cases, the available knowledge about the biology of known members of these families, together with the expression pattern of these newly identified genes, allows one to make predictions concerning their function. For example, all members of the HES proteins mediate transcription repression via their interaction with Groucho. CG10446/Side and CG5952/Her have the WRPW domain required for this interaction, implying that they are highly likely to act via the same mechanism. *CG10446/side* is expressed solely in the CNS at a stage at which cell differentiation is occurring. We hypothesize that it may play a role in antagonizing the function of transcription factors involved in the later stages of CNS differentiation.

However, sequence homology between species does not always imply functional homology. For example, CG8667/Mistr is a *Drosophila* sequence ortholog of the mammalian Mist1 protein. It is expressed solely in the developing nervous system, whereas *Mist1* is expressed not in the nervous system but in gut, pancreas, submandibular gland, lung, and skeletal muscle (34). In this case, differences in expression pattern of the genes encoding these proteins argue against any conservation of developmental role.

Two developmental processes in which bHLH gene function has been extensively studied in both *Drosophila* and other model systems are neurogenesis and mesoderm specification/myogenesis. We found potential new components in both of these processes in *Drosophila*.

Some of the genes we uncovered were expressed solely in the nervous system: *CG8667/mistr*, *CG5545/doli*, *CG10446/side*, and *CG5952/her*. Although several bHLH orthologs in *Drosophila* and vertebrates are known to control the formation and differentiation of neurons, similar bHLH factors in the determination of glia have not been identified. The *HLH3B* bHLH gene is expressed solely in the *Drosophila* midline glia (A.W.M., unpublished data); however, this gene is the sequence homologue of mammalian hematopoietic stem cell leukemia factor, which has not been described in gliogenesis in mammals (38). In this study, CG5545/Doli is particularly interesting because it is related to the Olig proteins. The *Olig* genes are expressed in the zone from which oligodendrocyte precursors arise and then in the precursors themselves throughout the CNS in mice and rats (32, 33). Ectopic expression of these genes promotes the expression of some oligodendritic precursor markers. From the mRNA *in situ* analysis we have carried out for *CG5545/doli* it is not clear in which cell types in the nervous system this gene is expressed; however, it may have a conserved role in gliogenesis if further experiments demonstrate its expression in a glial cell type.

*CG11450/shout* is the founding member of a new bHLH family. It is expressed in the ventral-most cells of the blastoderm-stage embryo that are fated to become mesoderm. This expression domain overlaps with that of *twist* (7, 39). *twist* and *CG11450/shout* continue to be expressed in the presumptive mesoderm during gastrulation. At the extended germ-band stage, both *twist* and *CG11450/shout* are expressed in alternating high and low levels along the length of the mesoderm. These alternating expression levels of *twist* are required for the specification of muscle derived from this tissue (39). The pattern of *CG11450/shout* expression in the ventral mesoderm implies that it could have a similar role to *twist* in specification of mesoderm derivatives. In *Drosophila*, Twist activates Snail and other downstream, mesoderm-specific regulators such as Tinman, Bagpipe, and Mef2 (7); all of these proteins have orthologs in vertebrates implicated in mesoderm development. Hence, CG11450/Shout represents a good candidate for both sequence and function conservation across species.

In both *Drosophila* and vertebrates, the heart (dorsal vessel in *Drosophila*) assembles at the midline from bilaterally symmetrical ventral mesoderm precursors. In vertebrates, the homeodomain-containing transcription factor *Nkx2.5* is the earliest known marker for the cardiac lineage. The *d* and *eHand* genes additionally are expressed in the early cardiac progenitors. The *Mef2C* transcription factor is expressed in the heart myoblasts. In *Nkx2.5*, *dHand*, and *Mef2C* null mice, the early heart tube forms but it fails to undergo looping and express some heart-specific molecules (40–42), implying that these three genes may lie in the same pathway. *tinman* (*tin*) is the *Drosophila* ortholog of *Nkx2.5*. It is expressed first uniformly in the presumptive mesoderm before gastrulation and in the ventral mesoderm after this process. It is required for the formation of the heart and circular visceral musculature from this mesoderm (43). The *mef2* gene in *Drosophila* is a direct transcription target of Tin and is required for the correct differentiation of the heart myoblasts (44–46). In this study, we identified *Drosophila hand* and showed that it is expressed in the early heart mesoderm, mirroring the situation in vertebrates. All these data imply that *tin*, *mef2*, and *Dm hand* may lie in a pathway in *Drosophila* homologous to that proposed in vertebrates. In addition, we also detected *Dm hand* expression in the forming circular visceral musculature of the *Drosophila* embryo. Because *tin* and *mef2* also are required for this structure, the same interaction may be conserved in this tissue.

It has been proposed that all regions of the body plan are specified by a *hox* gene code (47). Do all tissue types in the embryo have a bHLH component to their specification? The answer seems to be no. The expression of factors such as Da, which forms heterodimers with tissue-specific bHLH genes such as those of the Ato-related or MyoD families, is ubiquitous. However, some tissue types such as the fat body (fly

equivalent of the liver) do not express any other tissue-specific bHLH genes with which this gene can act. In fact, Da may not function in all tissues in which it is expressed; loss of Da function leads not to general embryological defects but specific ones in the muscle, gut, and nervous systems. Therefore, the evidence suggests that bHLH proteins are not involved in the development of all tissues but are concentrated in the genesis of a subset of tissue types wherein often several bHLH proteins act in a cascade.

We thank Sarah Goulding and members of the Jan lab for useful discussions and Sarah Meadows and Mike Rothenberg for reagents. A.W.M. is supported by a Wellcome Prize Traveling Research Fellowship. L.Y.J. and Y.N.J. are Howard Hughes Investigators.

1. Murre, C., McCaw, P. S. & Baltimore, D. (1989) *Cell* **56,** 777–783.
2. Blackwell, T. K. & Weintraub, H. (1990) *Science* **250,** 1104–1110.
3. Blackwell, T. K., Kretzner, L., Blackwood, E. M., Eisenman, R. N. & Weintraub, H. (1990) *Science* **250,** 1149–1151.
4. Ma, P. C., Rould, M. A., Weintraub, H. & Pabo, C. O. (1994) *Cell* **77,** 451–459.
5. Ellenberger, T., Fass, D., Arnaud, M. & Harrison, S. C. (1994) *Genes Dev.* **8,** 970–980.
6. Dang, C. V., Dolde, C., Gillison, M. L. & Kato, G. J. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 599–602.
7. Baylies, M. K., Bate, M. & Ruiz Gomez, M. (1998) *Cell* **93,** 921–927.
8. Ghysen, A., Dambly-Chaudiere, C., Jan, L. Y. & Jan, Y. N. (1993) *Genes Dev.* **7,** 723–733.
9. Modolell, J. & Campuzano, S. (1998) *Int. J. Dev. Biol.* **42,** 275–282.
10. Goulding, S. E., zur Lage, P. & Jarman, A. P. (2000) *Neuron* **25,** 69–78.
11. Huang, M. L., Hsu, C. H. & Chien, C. T. (2000) *Neuron* **25,** 57–67.
12. Jarman, A. P., Grau, Y., Jan, L. Y. & Jan, Y. N. (1993) *Cell* **73,** 1307–1321.
13. Bush, A., Hiromi, Y. & Cole, M. (1996) *Dev. Biol.* **180,** 759–772.
14. Jarman, A. P., Brand, M., Jan, L. Y. & Jan, Y. N. (1993) *Development (Cambridge, U.K.)* **119,** 19–29.
15. Goulding, S. E., White, N. M. & Jarman, A. P. (2000) *Dev. Biol.* **221,** 120–131.
16. Lee, J. E. (1997) *Curr. Opin. Neurobiol.* **7,** 13–20.
17. Murre, C., McCaw, P. S., Vaessin, H., Caudy, M., Jan, L. Y., Jan, Y. N., Cabrera, C. V., Buskin, J. N., Hauschka, S. D., Lassar, A. B., *et al.* (1989) *Cell* **58,** 537–544.
18. Fisher, A. & Caudy, M. (1998) *BioEssays* **20,** 298–306.
19. Bouchard, C., Staller, P. & Eilers, M. (1998) *Trends Cell Biol.* **8,** 202–206.
20. Gregor, P. D., Sawadogo, M. & Roeder, R. G. (1990) *Genes Dev.* **4,** 1730–1740.
21. Crews, S. T. & Fan, C. M. (1999) *Curr. Opin. Genet. Dev.* **9,** 580–587.
22. Ashok, M., Turner, C. & Wilson, T. G. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 2761–2766.
23. Dunlap, J. C. (1999) *Cell* **96,** 271–290.
24. Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., *et al.* (2000) *Science* **287,** 2185–2195.
25. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997) *Nucleic Acids Res.* **25,** 3389–3402.
26. Atchley, W. R., Terhalle, W. & Dress, A. (1999) *J. Mol. Evol* **48,** 501–516.
27. Atchley, W. R. & Fitch, W. M. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 5172–5176.
28. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997) *Nucleic Acids Res.* **25,** 4876–4882.
29. Tautz, D. & Pfeifle, C. (1989) *Chromosoma* **98,** 81–85.
30. Rubin, G. M., Yandell, M. D., Wortman, J. R., Gabor Miklos, G. L., Nelson, C. R., Hariharan, I. K., Fortini, M. E., Li, P. W., Apweiler, R., Fleischmann, W., *et al.* (2000) *Science* **287,** 2204–2215.
31. Peyton, M., Stellrecht, C. M., Naya, F. J., Huang, H. P., Samora, P. J. & Tsai, M. J. (1996) *Mol. Cell. Biol.* **16,** 626–633.
32. Zhou, Q., Wang, S. & Anderson, D. J. (2000) *Neuron* **25,** 331–343.
33. Lu, Q. R., Yuk, D., Alberta, J. A., Zhu, Z., Pawlitzky, I., Chan, J., McMahon, A. P., Stiles, C. D. & Rowitch, D. H. (2000) *Neuron* **25,** 317–329.
34. Lemercier, C., To, R. Q., Swanson, B. J., Lyons, G. E. & Konieczny, S. F. (1997) *Dev. Biol.* **182,** 101–113.
35. Krapp, A., Knofler, M., Frutiger, S., Hughes, G. J., Hagenbuchle, O. & Wellauer, P. K. (1996) *EMBO J.* **15,** 4317–4329.
36. Saga, Y., Hata, N., Koseki, H. & Taketo, M. M. (1997) *Genes Dev.* **11,** 1827–1839.
37. Srivastava, D., Cserjesi, P. & Olson, E. N. (1995) *Science* **270,** 1995–1999.
38. Varterasian, M., Lipkowitz, S., Karsch-Mizrachi, I., Paterson, B. & Kirsch, I. (1993) *Cell Growth Differ.* **4,** 885–889.
39. Baylies, M. K. & Bate, M. (1996) *Science* **272,** 1481–1484.
40. Lyons, I., Parsons, L. M., Hartley, L., Li, R., Andrews, J. E., Robb, L. & Harvey, R. P. (1995) *Genes Dev.* **9,** 1654–1666.
41. Srivastava, D., Thomas, T., Lin, Q., Kirby, M. L., Brown, D. & Olson, E. N. (1997) *Nat. Genet.* **16,** 154–160.
42. Lin, Q., Schwarz, J., Bucana, C. & Olson, E. N. (1997) *Science* **276,** 1404–1407.
43. Bodmer, R. (1993) *Development (Cambridge, U.K.)* **118,** 719–729.
44. Bour, B. A., O'Brien, M. A., Lockwood, W. L., Goldstein, E. S., Bodmer, R., Taghert, P. H., Abmayr, S. M. & Nguyen, H. T. (1995) *Genes Dev.* **9,** 730–741.
45. Lilly, B., Zhao, B., Ranganayakulu, G., Paterson, B. M., Schulz, R. A. & Olson, E. N. (1995) *Science* **267,** 688–693.
46. Gajewski, K., Kim, Y., Lee, Y. M., Olson, E. N. & Schulz, R. A. (1997) *EMBO J.* **16,** 515–522.
47. Gellon, G. & McGinnis, W. (1998) *BioEssays* **20,** 116–125.
48. Hartenstein, V. & Jan, Y.-N. (1992) *Roux's Arch. Dev. Biol.* **201,** 194–220.