Perspective

A Cluster of *Antennapedia*-Class Homeobox Genes in a Nonsegmented Animal

CYNTHIA KENYON AND BRUCE WANG

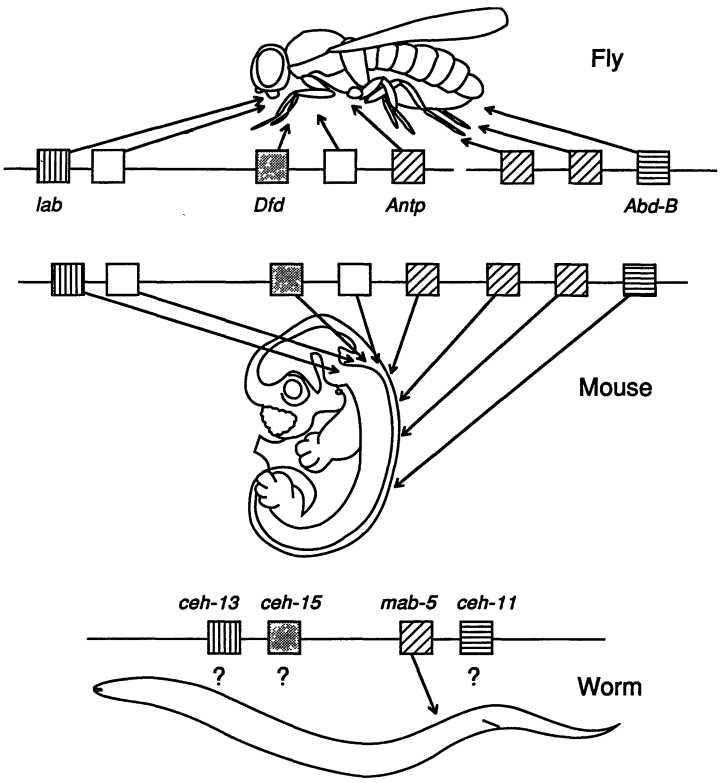
THE *ANTENNAPEDIA*-CLASS HOMEOBOX (HOM) GENES, which give different body regions different identities in insects and possibly vertebrates, have provided a molecular entry point into understanding the evolution of body pattern (*1*). These genes are located within evolutionarily conserved clusters in which, remarkably, the order of individual HOM genes is the same as the order in which they are expressed along the anterior-posterior body axis. The striking conservation of these genes among organisms suggests that the genes used for anteroposterior pattern formation in insects and vertebrates are ancient and that they evolved from a single homeotic gene complex (HOM-C) in a common primitive ancestor.

The bodies of developing insects and vertebrates have a segmental morphology. Furthermore, in both insects and vertebrates, the boundaries of expression and function of HOM genes often coincide with segment (or parasegment) boundaries. This has led to the idea that the function of HOM-C genes is to diversify the different body segments of segmented animals. Organisms that are not segmented, such as the nematode *Caenorhabditis elegans*, have been considered unlikely candidates for HOM-C–based pattern formation. For this reason, it was not too surprising when, on the basis of Southern hybridization analysis, *C. elegans* first appeared to lack HOM-C genes.

Recent findings in several laboratories now force us to reexamine the relation between segmentation and HOM-C–based pattern formation, because *C. elegans* does appear to generate body pattern by means of HOM-C genes after all. Many types of homeobox genes have now been identified in *C. elegans* (*2*). Four homeobox genes in the *Antennapedia* class, *ceh-13* (*3*), *ceh-15* (*4*), *mab-5* (*5*), and *ceh-11* (*2, 3, 6*), lie near one another on the physical map of *C. elegans* (*7*). It is not known whether there are additional homeobox genes in the vicinity of these four. However, it seems significant that each of the four homeobox genes exhibits sequence similarity to a gene positioned in the same relative order in the fly and vertebrate clusters (Fig. 1). When compared to *Drosophila* HOM-C genes, *ceh-13* most closely resembles *labial; ceh-15* most closely resembles *Deformed; mab-5* most closely resembles *Antennapedia*; and *ceh-11* exhibits similarity to both *Antennapedia* and *Abdominal-B*. The homeoboxes are arranged in two closely linked pairs separated by 200 to 300 kb.

Are these genes involved in pattern formation? So far we know the function of only one *C. elegans* HOM-C gene, *mab-5* (*8*), although the mutant phenotype and map position of another gene, *egl-5*, makes this gene an excellent candidate for another member of the complex (*9*). In *mab-5*$^-$ mutants, lineally unrelated sensory organs, muscles, and epidermal structures that characterize a posterior body region are missing. Many of the affected cells undergo posterior-to-anterior homeotic transformations. This suggests that, as in *Drosophila*, *C. elegans* HOM-C genes give cells in different positions along the anteroposterior body axis different identities. The discovery of this *C. elegans* homeobox gene complex is significant because it suggests that nematodes as well as insects and vertebrates evolved from a single ancestor that used HOM-C genes in anteroposterior pattern formation.

Because HOM-C genes are generally thought to function in segmented body regions to specify segmental identity, it is important to examine *C. elegans* closely for any hint of segmental morphology that might be the target of *C. elegans* HOM-C function. *Caenorhabditis elegans* does not have an overt segmental morphology (repeated bulges and grooves). Furthermore, the embryonic ventral nerve cord does not contain consistently repeated cell groups, and a number of unique neural and mesodermal cells are located along the sides of the body. However, the epidermis of the newly hatched animal does consist of repeated sets of cells. Some of these cells, six pairs of V cells, located laterally, and twelve P cells, located ventrally, divide to generate repeated cell groups. In general, the fates of V and P cell descendants are correlated with their positions in the lineage. For example, the Pn.p cells (the posterior




```
Antp     ERKRGRQTYTR YQTLELEKEF HFNRYLTRRR RIEIAHALCL TERQIKIWFQ NRRMKWKKEN
mab-5    •S••T••••S• S••••••••• •YHK••••K• •Q••SET•H• ••••V••••• ••••H•••A

lab      TNNS••TNF•N K•LT•••••• ••••••••A• •••••NT•Q• N•T•V••••• •••••Q••RV
ceh-13   •NGTN•TNF•T H•LT•••••• •TAK•VN•T• •T•••SN•K• Q•A•V••••• •••••E••RE

Dfd      •P••Q•TA••• H•I••••••• •Y••••••••  ••••••T•V• S••••••••• ••••••••D•
ceh-15   GE••Q•TA••• N•V••••••• •THK••••K• •••V••S•M• ••••V••••• •••••H••••

AbdB     SVRKK•KP•SK F••••••••• L••A•VSKQK •W•L•RN•Q• ••••V••••• •••••N••NS
ceh-11   SS•K•••••Q• •••SV••AK• QQSS•VSKKQ •E•LRLQTQ• •D••••••••  •••••A•••K
```

Fig. 1. Genomic organization and expression of homeotic genes in the fly, mouse, and worm. Shading in boxes represents homeobox sequences conserved between specific homeotic genes. Adapted from (*12*) by S. Salser. (Bottom) Comparison of *C. elegans* and *Drosophila* homeobox sequences. Dots indicate residues identical to those in *Antennapedia*. Residues in boxes show additional identities in the genes from *C. elegans* to specific *Drosophila* homologs. Abbreviations: *lab*, *labial*; *Dfd*, *Deformed*; *Antp*, *Antennapedia*; *Abd-B*, *Abdominal-B*.

Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94143.

daughters of the P cells) become epidermal cells whereas their sisters, the Pn.a cells, give rise to neurons. Thus P and V cell divisions generate repeated groups of cells that might, at some level, be compared to segments.

Some of the cells affected by *mab-5* are unique cells in the posterior body region and are not members of repeat units. However, because it is possible to draw an analogy between the repeated V and P cell lineages and body segments, it is informative to ask whether *mab-5* also generates diversity between V or P cell descendants located in different body regions.

In fact, *mab-5* does allow descendants of V and P cells that are located posteriorly to adopt fates that differ from their anterior homologs. The fates of cells located in one position of the P lineage, the Pn.aaap cells, illustrate this point. Cells in the Pn.aaap position of the P(1-10) lineages become motor neurons, but their homologs, the P(11,12).aaap cells, which are located in the posterior *mab-5* domain, undergo programmed cell death. We know that *mab-5* is required for this alternative fate because in *mab-5*$^-$ mutants all these cells adopt the fates of their anterior homologs and become motor neurons. The effect of *mab-5* activity differs for different descendants of P and V cells. For example, the function of *mab-5* in another set of P-derived homologs, the Pn.p cells, is to allow cells in the posterior of the male to generate mating structures instead of behaving like their anterior homologs and becoming simple epidermal cells.

The genes that give different cells within a single V or P cell lineage different developmental potentials may be analogous to genes such as the segment polarity genes in *Drosophila* (*10*), which give different cells within a single segment different developmental potentials. The *C. elegans* lineage genes allow different cells in a single V or P lineage to respond differently to *mab-5* activity, just as segment polarity genes allow cells in different geographical regions of a single segment to respond differently to *Drosophila* homeotic gene activity. The consequence is that each repeat unit itself is highly patterned (because of genes like the segment polarity genes in flies or lineage diversification genes in worms) and each repeat unit is unique (because of the homeotic genes).

The idea that the primary role of HOM-C genes is to function in register with segmentation mechanisms to make different segmental units different from one another has arisen because the boundaries of HOM gene expression and function in insects and vertebrates often coincide with segmental boundaries. From *C. elegans*, we can infer that there is no necessary relation between HOM-C gene function and segmentation, however liberal the definition. For the many cells affected by *mab-5* that are not members of repeat units, this is self-evident. However, one can also ask whether there is a fundamental underlying segmental structure to the expression and function of *mab-5* that can be inferred from its role in V and P cell diversification. If HOM-C genes have such a function in *C. elegans*, then we would expect the boundary of *mab-5* function along the body axis to coincide with the boundary of a single repeat unit, that is, a single V or P cell lineage. This is not the case. Instead, the boundary falls at different positions along the body axis for different sets of V and P cell descendants. For example, in the Pn.p homologs, the boundary of *mab-5* function is near the middle of the animal (between P6 and P7) whereas, for Pn.aaap homologs, the boundary is far more posterior and falls between P10 and P11. Thus even if one accepts the tenuous proposition that the repetitive V and P lineages are equivalent to segments, then *mab-5* still creates anteroposterior diversity without respecting any type of segment boundary. Thus this *C. elegans* HOM-C gene is targeted to a specific body region but not to a specific repeat unit or set of repeat units.

It is possible that the original function of HOM-C genes was simply to create cell diversity along the anteroposterior body axis, rather than to diversify body segments per se. The aspect of HOM-C genes that may have been conserved during evolution is their ability to respond to coarse positioning mechanisms that determine their general domains of expression. Other less well conserved mechanisms may set the precise boundaries of HOM gene expression and function. In some cases these boundaries will be made to coincide with segment boundaries and in other cases they will not. Consistent with this viewpoint is the fact that, even in flies and vertebrates, the domains of HOM gene expression do not always coincide with segment boundaries. Because flies and vertebrates are thought to have evolved segments independently of one another (*11*), it is quite possible that the primitive metazoan in which the homeotic gene cluster first arose was not segmented at all.

REFERENCES AND NOTES

1. L. Reid, *Cell* **63**, 875 (1990).
2. T. R. Burglin *et al.*, *Nature* **341**, 239 (1989).
3. D. Schaller *et al.*, *Nucleic Acids Res.* **18**, 2033 (1990).
4. A. Kamb *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 4372 (1989); B. Wang and C. Kenyon, *C. elegans* Data Base, MRC, Cambridge, England.
5. M. Costa *et al.*, *Cell* **55**, 747 (1988).
6. N. C. Hawkins and J. D. McGhee, *Nucleic Acids Res.* **18**, 6101 (1990).
7. A. Coulson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 7821 (1986).
8. C. Kenyon, *Cell* **46**, 477 (1986).
9. A. Chisholm, *Development* **111**, 921 (1991).
10. P. W. Ingham, *Nature* **335**, 25 (1988).
11. L. H. Hyman, *The Invertebrates: Platyhelminthes and Rhynchocoela* (McGraw-Hill, New York, 1951), vol. 2; R. B. Clark, *Dynamics in Metazoan Evolution* (Oxford University Press, Oxford, U.K., 1964).
12. E. M. De Robertis *et al.*, *Sci. Am.* **263**, 46 (July 1990).

13. We thank members of the Kenyon lab as well as J. Hodgkin for comments on the manuscript.