Nucleic Acids Research, 1993, Vol. 21, No. 8 ***1687–1688***

A rational nomenclature for vertebrate homeobox (*HOX*) genes

Matthew P.Scott
Departments of Developmental Biology and Genetics, Stanford University School of Medicine, Stanford, CA 94305-5427, USA

Received January 21, 1993; Accepted March 5, 1993

ABSTRACT

Developmental fates along the anterior-posterior axes of animals are controlled by clustered homeotic genes which in vertebrates are called Hox genes. The gene clusters are similar and probably functionally homologous in animals as different as nematodes, flies, and mammals. A new set of names for Hox genes was recently agreed upon by many workers in the field. Remarkably, the order of the Hox genes along the chromosome reflects where they are expressed along the body axis[1]. This simple principle is reflected in the new nomenclature system.

Homeotic genes control cell fates during the development of all animals, as was first revealed by studies of the Drosophila homeotic gene complexes (1,2). Many of these genes contain a homeobox, a 180 bp sequence of DNA which encodes an evolutionarily conserved DNA binding domain, the homeodomain (3, 4). A plethora of mammalian homeobox genes have been reported, among which 38 are located in four clusters. A new nomenclature for the mammalian Hox genes, approved by both the mouse and human nomenclature committees, has recently been announced (5). The main features of the nomenclature are summarized in Figure 1. The new names take advantage of the elegant arrangement of the genes to provide a logical nomenclature system rather than the names given when the genes were discovered. The new system is initially designed only for vertebrate genes, although it is to be hoped that similar systems will be useful, and adopted, for other animals. In order to preserve as much clarity in the literature as possible, it has been agreed by a large number of workers in the field and by the nomenclature committees that homeobox genes not located within the Hox complexes should not be given names containing the word 'Hox'.

There are four clusters of Hox genes (6) now to be known as A, B, C, and D. Based on sequence similarity the genes can be sorted into 13 'paralog' groups, each group having, in most cases, a representative in each complex. The order of paralogs along the chromosome is preserved in the four complexes. The genes within a complex are transcribed in the same direction and are numbered according to their paralog group from 1 at the 3′ end to 13 at the 5′ end. In several cases a representative of a paralog group is absent from a complex, in which case the corresponding gene number is omitted (blank spaces in the figure). The most commonly used old nomenclature is shown in columns marked H, for human, and M, for mouse.

One of the most interesting aspects of Hox genes is the similarity of their organization to Drosophila genes, now thought to represent true homology (7–9). Many Drosophila homeotic genes are located within two clusters called the Antennapedia complex (ANT-C) and the bithorax complex (BX-C), which are thought to have arisen from an ancestral single cluster. The closest correspondences of fly and mammalian genes are indicated in the figure. Due to the remarkable organization of the genes in insects and in mammals, one can also describe them according to their realm of function in the developing animal. The genes required farthest anterior in the animal are at the 3′ end of the complex. Genes with more posterior domains of action are located farther 5′ in the complexes. Thus in the figure the paralog groups are demarcated by 'anterior' and 'posterior'. The conservation of this gene organization for more than 600 million years suggests its importance.

The bracket indicates four fly genes which are all similar to four paralog groups; the precise correspondences, if they exist, have not been determined in these cases. Several paralog groups (10–13) are not represented in Drosophila. The A and D Hox complexes contain an Evx gene at the 5′ end. Evx genes are most similar to Drosophila *even-skipped* gene, which is not located within the clusters of Drosophila homeotic genes. The different type of homeobox sequence of Evx genes has led to the retention of the Evx name. Conversely the Drosophila ANT-C contains four homeobox genes, *zen*, *z2*, *ftz*, and *bcd*, not represented in the mammalian complexes.

The new system should help in following the results in the field, as it allows one to easily remember the relationships between paralogs in different complexes. It is also easy to remember that low numbers mean more anterior function. When the genes are so logically, if mysteriously, arranged, their names should be equally rational.

ACKNOWLEDGEMENTS

I would like to acknowledge the help of the many researchers in the Hox field in arriving at a consensus for the nomenclature, and especially Thomas Bürglin, Eddy De Robertis, Denis Duboule, Peter Gruss, Claudia Kappen, Robb Krumlauf, Atsushi

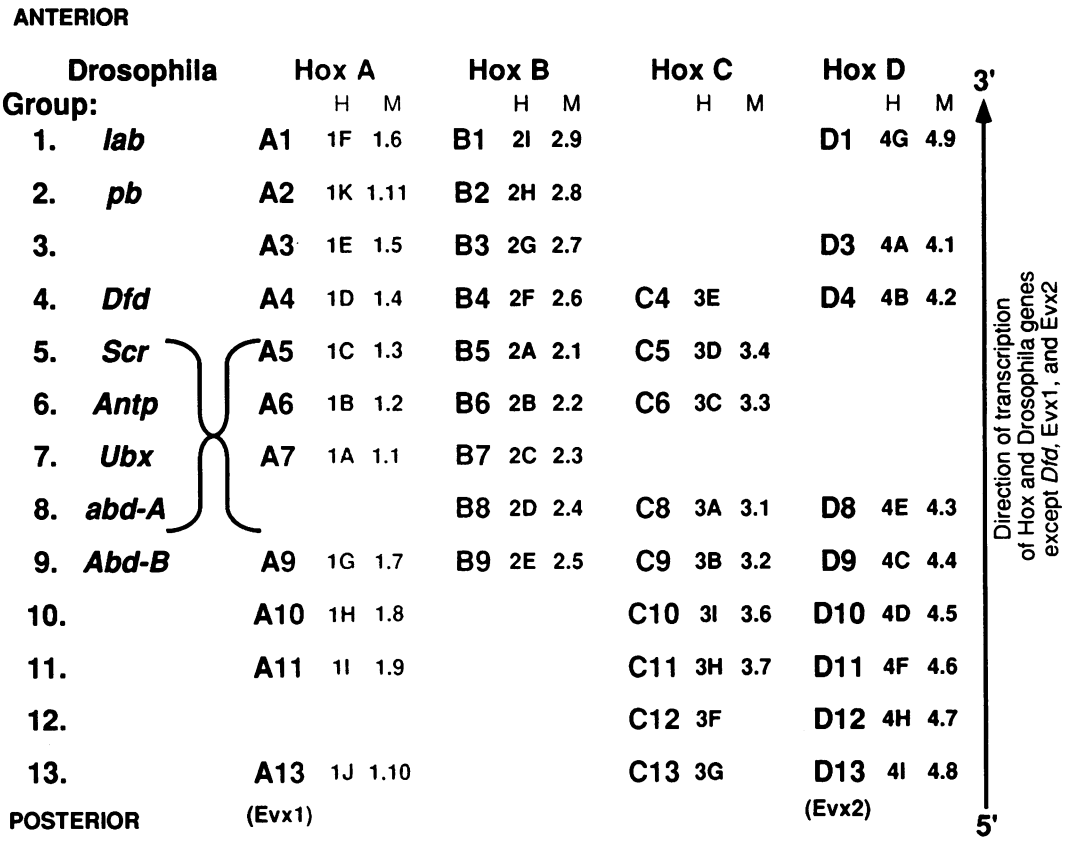

Figure 1.

Kuroiwa, Frank Ruddle, and Cliff Tabin for help with unpublished information and for encouragement.

REFERENCES

1. Lewis, E. B. (1978) Nature *276*, 565–570
2. Kaufman, T. C., M. A. Seeger, and G. Olsen (1990) Adv. Genet. *27*, 309–362
3. Scott, M. P., J. W. Tamkun, and G. W. Hartzell III (1989) BBA Rev. Cancer *989*, 25–48
4. Gehring, W. J. (1992) Trends Biochem Sci *17*, 277–280
5. Scott, M. P. (1992) Cell *71*, 551–553
6. Boncinelli, E., A. Simeone, D. Acampora, and F. Mavilio (1991) Trends Genet *7*, 329–34
7. Duboule, D. and P. Dolle (1989) EMBO J *8*, 1497–1505
8. Graham, A., N. Papalopulu, and R. Krumlauf (1989) Cell *57*, 367–378.
9. McGinnis, W. and R. Krumlauf (1992) Cell *68*, 283–302