Int. J. Dev. Biol. 52: 969-977 (2008)
doi: 10.1387/ijdb.072541nt

THE INTERNATIONAL JOURNAL OF DEVELOPMENTAL BIOLOGY
www.intjdevbiol.com

Evolution of *CUT class* homeobox genes: insights from the genome of the amphioxus, *Branchiostoma floridae*

NAOHITO TAKATORI and HIDETOSHI SAIGA*

Department of Biological Sciences, Graduate School of Science and Engineering, Tokyo Metropolitan University, Hachiohji, Japan

ABSTRACT *CUT class* homeobox genes, including *CUX/CASP, ONECUT, SATB* and *COMPASS* family genes, are known to exhibit diverse features in the homeodomain and the domain architecture. Furthermore, the intron/exon organization of *CUX/CASP* is different between vertebrates and protostomes, and *SATB* genes are only known for vertebrates, whereas *COMPASS* genes have only been found in protostomes. These observations suggest a complex evolutionary history for the *CUT class* homeobox genes, but the evolution of *CUT class* homeobox genes in the lineage to vertebrates remained largely unknown. To obtain clearer insights into this issue, we searched the genome of amphioxus, *Branchiostoma floridae*, a lower chordate, for *CUT class* homeobox genes by extensive BLAST survey and phylogenetic analyses. We found that the genome of *Branchiostoma floridae* encodes each single orthologue of *CUX/CASP, ONECUT,* and *COMPASS,* but not the *SATB* gene, and one atypical *CUT* gene likely specific to this species. In addition, the genomic structure of the amphioxus *CUX/CASP* gene turned out to be protostome-type, but not vertebrate-type. Based on these observations, we propose a model in which *SATB* is suggested to evolve at the expense of *COMPASS* and this change, together with the structural change in *CUX/CASP*, is supposed to take place in the lineage to vertebrates after divergence of the amphioxus and vertebrate ancestors. The present study provides an example of dramatic evolution among homeobox gene groups in the vertebrate lineage and highlights the ancient character of amphioxus, retaining genomic features shared by protostomes.

KEY WORDS: *Cux, COMPASS, Onecut, SATB, chordate*

Introduction

Homeobox genes, encoding transcription factors with the DNA-binding homeodomain, are classified into many groups according to the degree of sequence similarity in the homeodomain (Bürglin, 1994; Bürglin, 1995; Ruddle *et al.,* 1994). CUT class is a group of homeobox genes and/or homeodomain proteins, consisting of CUX/CDP (hereafter referred to as CUX), ONECUT, SATB and COMPASS family genes and/or proteins (Bürglin and Cassata, 2002; Lannoy *et al.,* 1998). Although the members of the CUT family except for COMPASS share the CUT domain, another DNA-binding domain (Harada *et al.,* 1994; Lannoy *et al.,* 1998), they exhibit diverse features in the homeodomain. Furthermore, it has been shown that protostome and deuterostome genomes contain different complements of the CUT class genes (Bürglin and Cassata, 2002), indicating that CUT class genes have gone through a complex evolution.

CUX family proteins contain three CUT domains and a homeodomain (Bürglin and Cassata, 2002) and have been shown to participate in a wide range of cell fate decisions, functioning as an effecter of Notch signaling (Blochlinger *et al.,* 1991; Ellis *et al.,* 2001; Gillingham *et al.,* 2002; Nepveu, 2001; Tufarelli *et al.,* 1998). CASP, an alternative splice form of CUX that shares the N-terminal half with the CUX protein and lacks both the CUT domains and the homeodomain, has been identified in a wide variety of animals from human to *C. elegans* (Bürglin and Cassata, 2002; Lievens *et al.,* 1997). Interestingly, the genomic structure of the CUX/CASP gene differs between protostomes and vertebrates (Bürglin and Cassata, 2002). In human and mouse, the exon encoding the unique C-terminal half of the CASP protein is interposed between the exons which encode the second and third CUT domains of CUX, while in *C. elegans*, the C-terminal exon is positioned in the 3' downstream of the exons encoding the homeodomain (Bürglin and Cassata, 2002; Gillingham *et al.,* 2002).

***Address correspondence to:** Hidetoshi Saiga. Department of Biological Sciences, Graduate School of Science and Engineering, Tokyo Metropolitan University, 1-1 Minamiohsawa, Hachiohji, 192-0397 Tokyo, Japan. Fax: +81-426-77-2572. e-mail: saiga-hidetoshi@tmu.ac.jp

Supplementary Material for this paper (an amino acid alignment) is available at: http://dx.doi.org/10.1387/ijdb.072541nt

Published online: 14 August 2008

0214-6282/2008/$35.00

ONECUT family proteins contain a CUT domain and an atypical homeodomain with phenylalanine at the 48th residue (Fig. 1). ONECUT was initially identified as a gene encoding the transcriptional activator of a hepatocyte nuclear factor, *HNF-3β* (Samadani and Costa, 1996). ONECUT is also expressed in the central nervous system (Landry *et al.,* 1997; Rausa *et al.,* 1997), which appears to be conserved in invertebrates (Nguyen *et al.,* 2000; Sasakura and Makabe, 2001).

SATB family proteins contain two CUT domains, one atypical homeodomain with phenylalanine at the 48th residue like that of ONECUT and a single amino acid insertion between the first and second helices (Fig. 1B) and additionally, a unique domain designated the COMPASS domain (Bürglin and Cassata, 2002). SATB genes have been isolated only from vertebrates and are known to be essential for the development of thymocytes, playing important roles in regulating gene expression through chromatin modification (Yasui *et al.,* 2002).

Finally, COMPASS family proteins also contain the COMPASS domain as well as two atypical homeodomains with a 10 amino acid insertion between the second and third helices but they lack the CUT domain (Fig. 1). Nevertheless, they have been classified into the CUT class, because they are one of only the two gene groups that possess the COMPASS domain (Bürglin and Cassata, 2002). By contrast to SATB, COMPASS genes have been isolated only from protostomes. It has been reported that the COMPASS gene of *Drosophila melanogaster* known as *defective proventriculus* is required for the formation of the proventriculus and the proximodistal patterning of the wing disc (Fuss and Hoch, 1998; Kolzer *et al.,* 2003; Nakagoshi *et al.,* 1998).

As described above, the homeodomains encoded by the CUT class homeobox genes are mostly atypical, displaying diverse features. Also, the gene structures of CUX/CASP are different between vertebrate and protostome genes. Previously, Bürglin and Cassata proposed a model for the evolution of CUT class homeobox genes (Bürglin and Cassata, 2002). Although the model explained very well the emergence of CUT class homeobox genes, there are three points yet to be explained: when dramatic change in CUX/CASP took place, how SATB genes have emerged in vertebrates and why the COMPASS genes are missing from vertebrates. This is largely because at the time the study was conducted, only limited number of sequence resources to survey CUT class genes were available, including *Drosophila melanogaster* and *Caenorhabditis elegans* for protostomes, and several vertebrate species and an ascidian species, *Halocynthia roretzi* for deuterostomes (Bürglin and Cassata, 2002). Recently, the genome resources of non-vertebrate deuterostomes have become available, including ascidian *Ciona intestinalis* (http://genome.jgi-psf.org/Cioin2/Cioin2.home.html), sea urchin *Strongylocentrotus purpuratus* (http://www.ncbi.nlm.nih.gov/genome/guide/sea_urchin/) and amphioxus *Branchiostoma floridae* (http://genome.jgi-psf.org/Brafl1/Brafl1.home.html). In the genome of *C. intestinalis*, however, only ONECUT has been identified as CUT class homeobox genes (Wada *et al.,* 2003) and in the *S. purpuratus* genome, two members, ONECUT and CUX, have been reported (Howard-Ashby *et al.,* 2006), raising a possibility that other members of CUT class homeobox genes may have been lost in these animal species. Therefore, the genome resource of amphioxus is particularly important. Amphioxus is an invertebrate chordate and belongs to the subphylum Cephalochordata. According to the recent molecular phylogenetic studies (Bourlat *et al.,* 2006; Delsuc *et al.,* 2006), amphioxus likely occupies the phylogenetic position at the stem of chordates and is expected to provide informative missing link between protostomes and vertebrates.

In order to obtain clearer insights into the evolution of CUT class homeobox genes in the chordate lineage, we have extensively searched for CUT class homeobox genes in the genome of *B. floridae*. We have found that the amphioxus genome encodes each single orthologue of the CUX, ONECUT and COMPASS, but not SATB family genes, and additionally, a CUT class gene with novel domain architecture. We also found that the genomic structure of the amphioxus CUX/CASP gene was similar to that of *C. elegans* counterpart, suggesting that the genomic structure of this gene changed after the divergence of amphioxus from the lineage to vertebrates. Based on the present findings, we propose a new model for the evolution of CUT class homeobox genes in the chordate lineage and suggest that SATB gene may have arisen from a COMPASS gene in the lineage to vertebrates. The evolutionary scenario for CUT class genes highlights its complex evolutionary change and the ancestral character of the amphioxus genome in the chordate lineage.

Results and Discussion

BLAST Survey of CUT *class homeobox genes in the genome of* Branchiostoma floridae

For search of CUX, ONECUT and SATB family genes, the CUT domain and homeodomain sequences of human, mouse and *Drosophila* CUT class proteins were used as queries, and for COMPASS family genes, the COMPASS domain and homeodomain sequences were used as queries. Presumably due to the high level of polymorphism between haplotypes in the amphioxus genome, BLAST analysis often yielded two highly similar amino acid sequences for each gene. We judged whether these sequences represent haplotypes or paralogues by examining similarity between putative intronic nucleotide sequences and the synteny in the neighborhood regions. A pair of haplotype sequences was handled as a single sequence in the following analyses.

The BLAST search yielded several candidate sequences, of which protein domain architectures were analyzed using an online resource, the Simple Modular Architecture Research Tool (SMART, http://smart.embl-heidelberg.de/) (Letunic *et al.,* 2006; Schultz *et al.,* 1998). A series of the analyses has revealed that the amphioxus genome has three typical CUT class homeobox gene candidates: the first encoding a protein with three CUT domains and one homeodomain, the second encoding a protein with one CUT domain and one homeodomain, and the third encoding a protein with one COMPASS domain and two atypical homeodomains. We could not find a gene that encodes the protein with the domain architecture corresponding to SATB. In addition, we found a putative atypical CUT class gene encoding a novel protein with four CUT domains and one homeodomain. These genes were tentatively designated as *Bf-Cux* (Gene model ID: TAK_gw.74.260.1 and TAK_fgenesh2_pg.scaffold_130000023, on scaffold 74 from 1588925 to 1623355 base pairs and scaffold 130 from 315232 to 340622 base pairs, respectively), *Bf-Onecut* (Gene model ID: TAK_gw.66.125.1 on scaffold

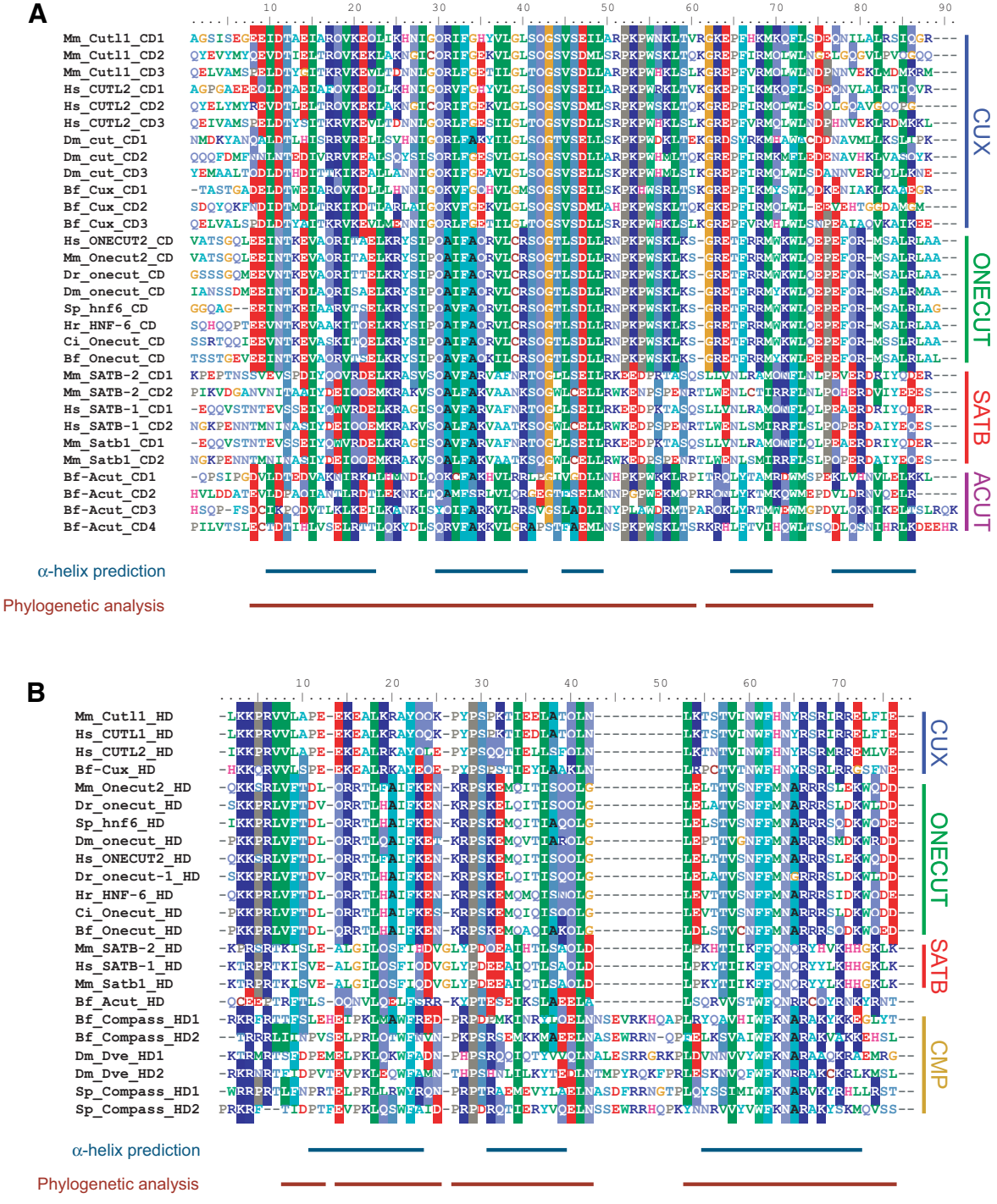

Fig. 1. Alignments of amino acid sequences of CUT domains and homeodomains by CLUSTALW. (A) *Alignment of CUT domains.* **(B)** *Alignment of homeodomains. On the left side, indicated are species names in an abbreviated form, protein names and domain names. As regards the domain names, CD and HD indicate the CUT domain and the homeodomain, respectively, and the appendix numbers indicate the relative position of the domain in the protein. The names of the family of CUT proteins are shown on the right side of the alignment. CMP stands for COMPASS. Below the alignment, predictions of alpha helices are shown by blue lines, which turned out to be consistent with the results by Bürglin and Cassata (Bürglin and Cassata, 2002). Sequence data used for subsequent phylogenetic analyses are shown by brown lines. Species codes are as follows and the same for all figures: Mm, Mus musculus; Hs, Homo sapiens; Dr, Danio rerio; Dm, Drosophila melanogaster; Hr, Halocynthia roretzi; Ci, Ciona intestinalis; Sp, Strongylocentrotus purpuratus; Bf, Branchiostoma floridae. The accession numbers of the sequences used are as follows and the same for all et al.,ures, Mm_Cutl1: AAH90847; Hs_CUTL2: O14519; Dm_cut: NP-524764; Hs_ONECUT2: O95948; Mm_Onecut2: AAI03669; Dr_onecut: AAH90847; Dm_onecut: NP-524842; Sp_hnf6: NP-999824; Hr_HNF-6: BAB15952; Ci_Onecut: BAE06494; Mm_SATB-2_Q546B3; Hs_SATB-1: Q59EQ0; Mm_Satb1: Q91XB1; Dm_Dve: BAA32660.*

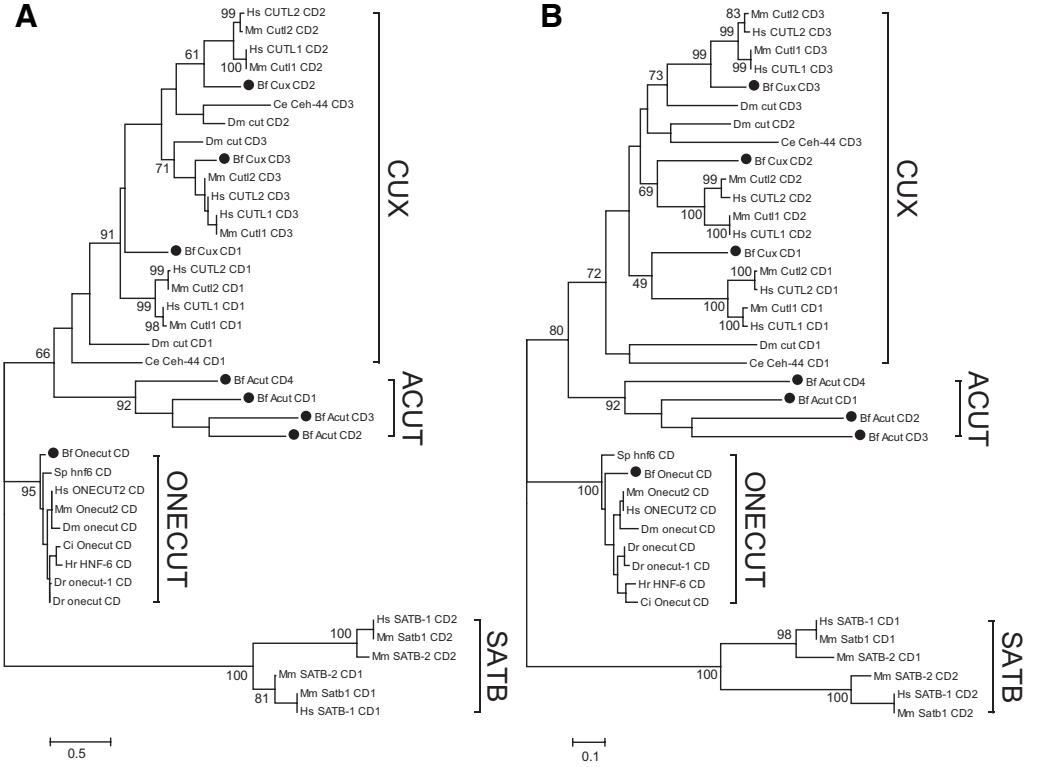

Fig. 2. Phylogenetic treenof CUT class proteins based on CUT domains generated by the maximum likelihood method (A) or neighbor joining method (B). *Numbers at nodes indicate bootstrap values (represented in %) for the percentage of 500 bootstrap pseudoreplications* **(A)** *or the number of times the node was supported out of 1000 pseudoreplications* **(B)**. *Protein names are indicated as shown in Fig. 1. Amphioxus proteins are marked by small black circles. Scale bars indicate the evolutionary distance of 0.5 and 0.1 amino acid substitutions per position for (A,B), respectively.*

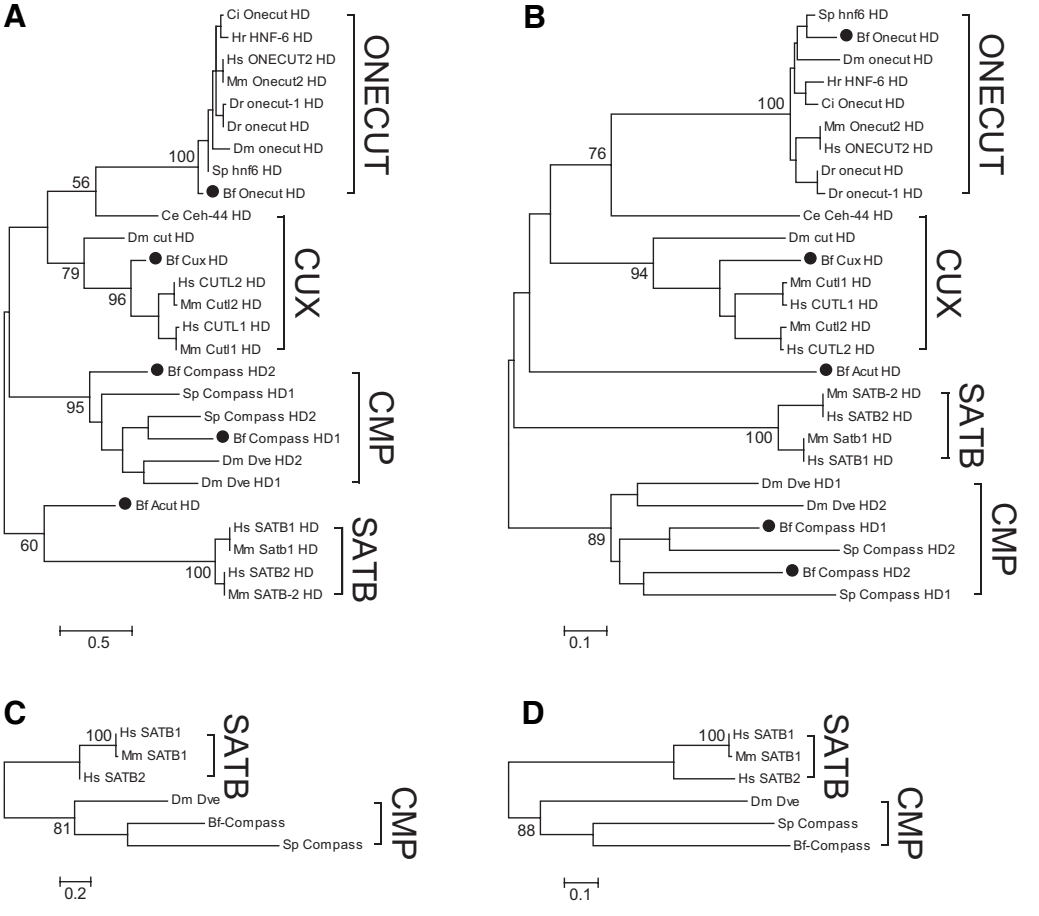

Fig. 3. Phylogenetic tree of CUT class proteins based on homeodomain (A,B) or COMPASS domain sequences (C,D) generated by the maximum likelihood method (A,C) or neighbor joining method (B,D). *Numbers at nodes indicate bootstrap values (represented in %) for the percentage of 500 bootstrap pseudoreplications* **(A,C)** *or the number of times the node was supported out of 1000 pseudoreplications* **(B,D)**. *Protein names are indicated as shown in Fig. 1. Amphioxus proteins are marked with small closed circles. Scale bars indicate the evolutionary distance of 0.5, 0.1, 0.2 and 0.1 amino acid substitutions per position for (A,B,C,D), respectively.*

66, from 273011 to 331608 base pairs), *Bf-Compass* (Gene model ID: KOB_estExt_fgenesh2_pg.C_4820014 and KOB_estExt_fgenesh2_pg.C_630132, on scaffold 482 from 363973 to 374411 base pairs and scaffold 63 from 2482728 to 2493434 base pairs, respectively) and *Bf-Acut* (Gene model ID: TAK_query and TAK_290639, on scaffold 3 from 5910690 to 5916726 base pairs and scaffold 115 from 1458244 to 1476604 base pairs, respectively), respectively.

Phylogenetic analyses of CUT *class homeobox genes and the structure of the* CUX/CASP *gene in the genome of* Branchiostoma floridae

In order to confirm our putative designation of the *B. floridae* CUT class homeobox genes, molecular phylogenetic analyses were performed with the CUT domain, homeodomain and the COMPASS domain by the maximum likelihood (ML) and neighbor-joining (NJ) methods.

First, in both phylogenetic trees drawn with the CUT domain by the ML and NJ methods, the CUT domains of amphioxus, *Drosophila* and vertebrate CUX, except for the first CUT domain of *Drosophila* CUT, formed a clade supported with a high bootstrap value, suggesting that *Bf-Cux* is the amphioxus CUX orthologue (Fig. 3 A,B). Similarly, *Bf-Onecut*, vertebrate ONECUT, HNF-6 and *Drosophila* ONECUT formed a well-supported clade, suggesting that *Bf-Onecut* is the amphioxus ONECUT orthologue (Fig. 3 A,B). The CUT domains of *Bf-Acut* are significantly related to those of CUX genes (Fig. 3 A,B). The trees also suggest that the vertebrate CUX genes increased in number after the divergence of amphioxus and vertebrate ancestors and that the CUT domains of SATB evolved relatively fast.

Second, phylogenetic analyses with the homeodomain also supported the orthology of *Bf-Cux*, *Bf-Onecut* and *Bf-Compass* (Fig. 4 A, B). The homeodomain of *Bf-Onecut* has phenylalanine at the position 48 of the homeodomain, the feature shared with ONECUT and SATB family genes, and that of *Bf-Compass* has a 10 amino acid insertion between the second and third helices, which is characteristic for COMPASS homeodomains as reported previously (Fig. 1B; Bürglin and Cassata, 2002). On the other hand, the homeodomain of *Bf-Acut* did not exhibit strong relationship with any other families of CUT class homeobox genes (Fig. 4 A,B).

Third, analyses with the COMPASS domain were performed. Alignment and the secondary structure prediction as shown in Fig. 2 suggest that *Bf-Compass* does possess a COMPASS domain. Phylogenetic analyses with the COMPASS domain by the ML and NJ methods further support with a high bootstrap value that *Bf-Compass* is the amphioxus COMPASS orthologue (Fig. 4 C,D).

Lastly, we addressed whether the structure of CUX/CASP gene in the amphioxus genome is of vertebrate-type or the *C. elegans*-type. Two gene models were identified in the amphioxus CUX/CASP locus, one corresponding to CUX and the other to CASP, the latter being supported by an EST cluster (data

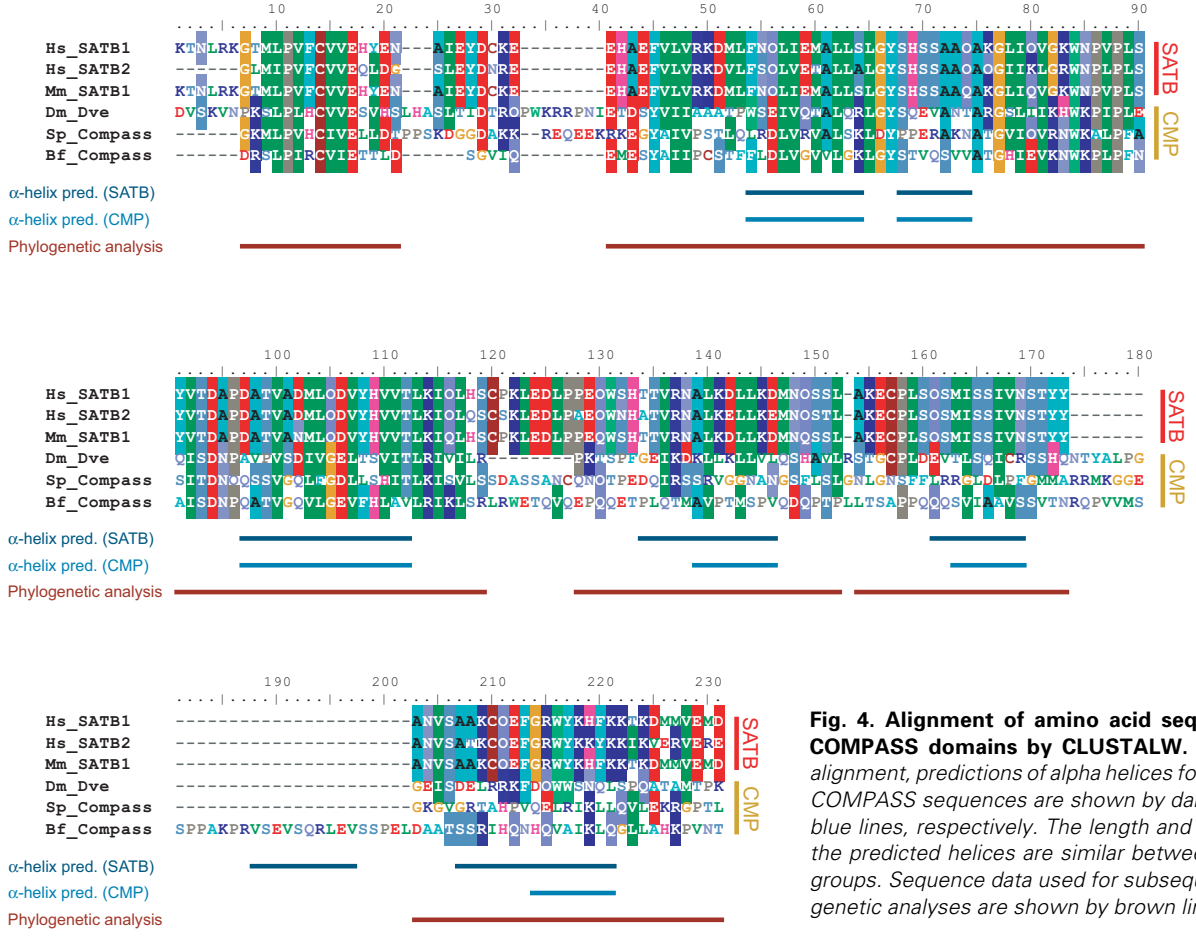

Fig. 4. Alignment of amino acid sequences of COMPASS domains by CLUSTALW. *Below the alignment, predictions of alpha helices for SATB and COMPASS sequences are shown by dark and light blue lines, respectively. The length and position of the predicted helices are similar between the two groups. Sequence data used for subsequent phylogenetic analyses are shown by brown lines.*

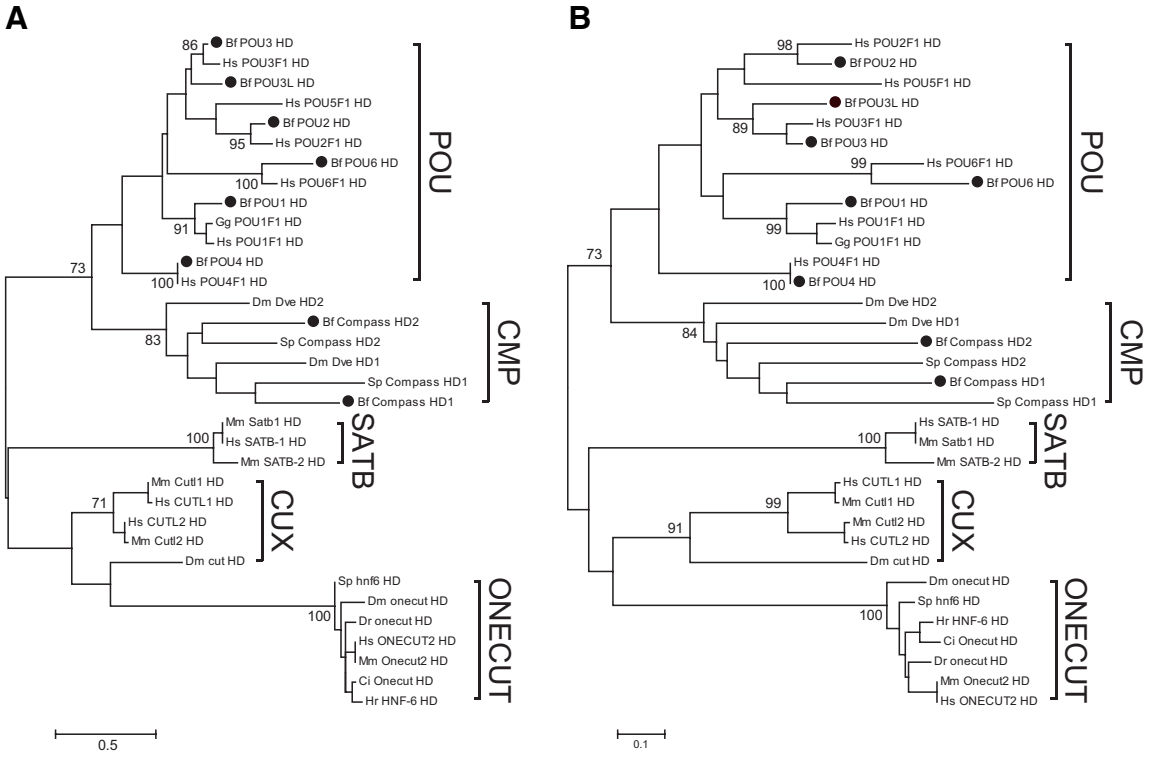

Fig. 5. Phylogenetic tree of CUT and POU class proteins based on homeodomain sequences generated by the maximum likelihood method (A) or neighbor joining method (B). *Numbers at nodes indicate bootstrap values (represented in %) for the percentage of 500 bootstrap pseudoreplications* **(A)** *or the number of times the node was supported out of 1000 pseudoreplications* **(B)**. *Protein names are indicated as shown in Fig. 1. Amphioxus proteins are marked with small closed circles. Scale bars indicate the evolutionary distance of 0.5 and 0.1 amino acid substitutions per position for (A,B), respectively.*

not shown). As schematized in Fig. 6, the genomic structure of the amphioxus CUX/CASP gene turned out to be similar to the *C. elegans* counterpart. This suggests that a dramatic change in the genomic structure of the CUX gene occurred in the lineage to the vertebrates after divergence of the amphioxus and vertebrate ancestors.

The homeodomains of COMPASS exhibit higher affinity to those of POU class protein than the homeodomain of CUX, ONECUT and SATB

Previously, COMPASS family of which protein does not posses the CUT domain, has been associated with the SATB family due to the shared conservation of COMPASS domains (Bürglin and Cassata, 2002). However, the homeodomains of COMPASS and SATB are different in three points; (1) the presence of a ten amino acid insertion between the second and third helices in COMPASS (Fig. 1B), (2) the presence of a single amino acid insertion between the first and second helices in SATB and (3) the presence of phenylalanine instead of standard tryptophan at the position 48 in SATB (Fig. 1B). We addressed to what extent the association of COMPASS family with SATB family is supported by the phylogenetic analysis based on the homeodomain. BLAST analysis using homeodomain sequences of the COMPASS genes revealed that the homeodomain of COMPASS has affinity to the homeodomain of POU class transcription factors rather than CUX, ONECUT and SATB. We next performed phylogenetic analyses with the homeodomain sequences of POU and CUT class transcription factors. The alignment of the homeodomains is shown in Supplemental Fig. 1. The phylogenetic trees drawn by the ML and NJ methods showed that the homeodomains of COMPASS are more closely related to those of POU proteins than those of CUX, ONECUT and SATB (Fig. 5). Phylogenetic analysis using all the amphioxus homeodomain sequences revealed that COMPASS homeodomains show weak but reproducible affinity with the POU homeodomains (data not shown). Considering that homeodomain transcription factors should primarily be classified according to properties and phylogenetic relationships of the homeodomain, it might be better to regard the COMPASS genes as a separate class consisting of themselves.

The COMPASS *gene is also present in the genome of the sea urchin* Strongylocentrotus purpuratus

The identification of a COMPASS gene in the amphioxus genome encouraged us to update the previous model, in which an ancestral COMPASS/CUT gene was proposed and suggested that it might have evolved into COMPASS and SATB in the protostome and deuterostome, respectively (Bürglin and Cassata, 2002). Before proposing a new model, we decided to address whether the COMPASS gene is present in the genome of a sea urchin, *Strongylocentrotus purpuratus* (http://www.ncbi.nlm.nih.gov/genome/guide/sea_urchin/), although COMPASS gene has not been referred to in the recent paper on the comprehensive analysis of sea urchin homeobox genes (Howard-Ashby *et al.,* 2006). In the sea urchin genome database,

we found a gene encoding a protein possessing one putative COMPASS domain and two homeodomains with a ten amino acid insertion between the second and third helices (putative gene: hmm31832 in Sea Urchin Genome Resources; Fig. 2). Phylogenetic analyses using the putative COMPASS domain and the homeodomain sequences indicate that it is the COMPASS but not SATB gene of the sea urchin (Fig. 4C, D). SATB gene was not found in the sea urchin genome. The identification of a COMPASS gene in the amphioxus genome as well as the sea urchin genome prompted us to propose a new model of the evolution of CUT class homeobox genes.

A new model for the evolution of CUT *class homeobox genes in the lineage to vertebrates*

The new model for the evolution of the CUT class homeobox genes in the lineage to the vertebrates is as shown in Fig. 6. The ancestral triploblast animal probably had a complement of CUT homeobox genes constituting of ONECUT, COMPASS and CUX/CASP with *C. elegans*-type genomic structure. COMPASS genes had emerged from a common ancestor with the POU class homeobox genes, acquiring the insertion of COMPASS domain in the N-terminal region and the insertion of ten amino acids in the homeodomain. The CUT class homeobox gene complement has been retained in the protostomes and amphioxus. In the lineage to vertebrates, SATB gene emerged and also the genomic structure of CUX/CASP gene changed after the divergence of the amphioxus and vertebrate ancestors. SATB gene may have arisen from the COMPASS gene by domain shuffling between COMPASS and ONECUT genes and eventually the COMPASS gene was lost from the ancestral vertebrate. Prior to the duplication of the entire gene, its CUT domain increased in number by tandem duplication to form the SATB gene with the homeodomain of ONECUT-type with an amino acid insertion between the first and second helices.

This model requires the smallest number of evolutionary changes of the homeodomain. If we assume the putative ancestral COMPASS/CUT gene (Bürglin and Cassata, 2002) had COMPASS-type homeodomain, it has to lose its insertion, convert the tryptophan at the position 48th residue to phenylalanine and acquire a single amino acid insertion between the first and second helices to form SATB-type homeodomain. This seems less likely compared to the domain shuffling in our model, since it requires two independent conversions of a stringently conserved tryptophan to phenylalanine in the evolution of the ONECUT and SATB genes. On the other hand, if the putative COMPASS/CUT gene had the ONECUT-type homeodomain, the homeodomain should have to change its phenylalanine to tryptophan at the position 48 and gain an insertion between the first and second helices to form the COMPASS-type homeodomain. Compared to our scenario, this scenario also requires an additional amino acid conversion. In support of our scenario, any gene with the genomic structure corresponding to the putative COMPASS/CUT (Bürglin and Cassata, 2002) as well as SATB has not been identified in both protostomes and non-vertebrate deuterostomes. On the other hand, weakness of our scenario may lie in that our molecular phylogenetic analyses did not suggest strong relationship between CUT domains and homeodomains of the ONECUT and SATB families. However, as suggested previously (Bürglin and Cassata, 2002), we feel that this may be due to the fast evolution of the CUT domain and homeodomain in the SATB

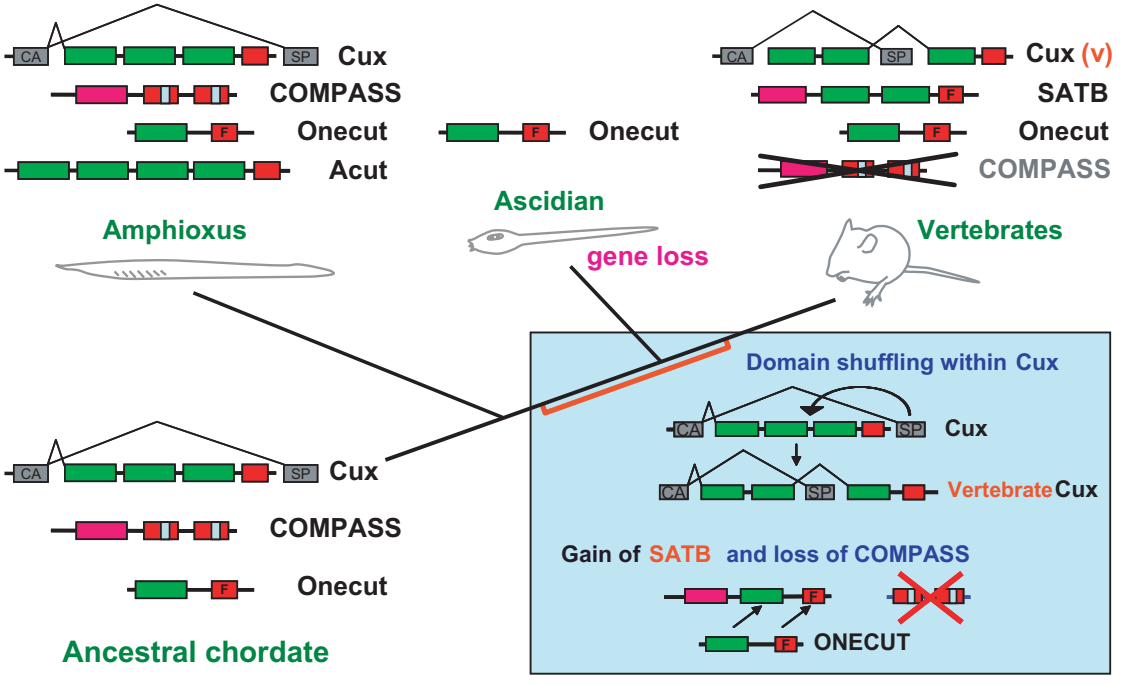

Fig. 6. Model for the evolution of the *CUT class* homeobox genes in the lineage to the vertebrates. *Protein structures are schematically shown with boxes and lines. Green, red, pink and gray boxes indicate the CUT domain, homeodomain, COMPASS domain and the CASP protein exons, respectively. The letter "F" and the light blue box within the red box indicates the substitution of the 48th amino acid to phenylalanine and the ten amino acid insertion between the second and third helixes in the homeodomain, respectively.*

family. Lastly, we have searched the recently released draft genome of *Nematostella vectensis* for CUT class genes and found ONECUT and COMPASS genes but not an SATB gene (data not shown). A part of the N-terminal sequence of the CASP gene flanked by nucleotide sequence gaps was also found, although we were unable to determine whether *N. vectensis* has a CUX gene. The presence of a COMPASS but not SATB in the sea anemone genome supports our hypothesis on the ancestry of the COMPASS. Thus, we propose our scenario as the most likely scenario of the evolution of COMPASS and SATB homeobox genes.

The functional significance of the loss and emergence of COMPASS and SATB genes, respectively, in vertebrates are not clear at the moment. COMPASS gene is known to be involved in the formation of the proventriculus in *Drosophila* (Fuss *et al.,* 1998; Nakagoshi, 2005) and its loss in vertebrates may reflect the evolution of the mechanism of fore gut formation in vertebrates. However, very little is known about the function of COMPASS genes in other invertebrates including amphioxus, and further study of their functions are necessary to pursue this possibility. The emergence of SATB, on the other hand, may represent vertebrate innovations, and in accord with this notion, it has been shown that SATB genes are regulators of the genes essential for the differentiation and activation of T-cells (Alvarez *et al.,* 2000). Closer examination of the evolutionary emergence and the function of SATB genes in relevance to the evolution of acquired immunity is an interesting topic that should be addressed in future studies.

Regarding the evolution of the CUX/CASP genes, the large evolutionary change in the genomic organization of the CUX/CASP gene likely occurred through the domain shuffling within the gene in the lineage to the vertebrates after divergence of amphioxus and vertebrate ancestors. As shown in the phylogenetic trees in Fig. 3, all the three CUT domains of amphioxus and vertebrate CUX form a clade, and BLAST search of the amphioxus genome confirmed that no other genes encode C-terminal half of CASP (data not shown). These observations make a possibility of recruiting the CUT domains or the C-terminal half of CASP from some other genes unlikely and in turn suggest that the repositioning of the C-terminal half of CASP occurred through domain shuffling within the CUX gene (Fig. 6). At present, it is difficult to decide when the change in the genomic structure of CUX/CASP or the gain and loss of SATB and COMPASS took place due to the limited genome sequence resources in the lower chordates, although previous study reported that genes encoding CUX, COMPASS and SATB were not found in the genome of *Ciona intestinalis* (Wada *et al.,* 2003). Future genome sequencing of lower chordates will resolve this issue.

Bf-Acut might have arisen by an amphioxus lineage specific duplication of a CUX gene and subsequent multiplication of CUT domains followed by their diversification. A less likely scenario may be that the ancestral triploblast animals had a gene encoding the CUT transcription factor with four CUT domains, which has been lost in all animals so far examined except for amphioxus. Another point to be noted here is that no EST sequence is found in the genome browser to support the existence of this gene. It remains to be proven whether *Bf-Acut* is a genuine active gene or a pseudogene.

Conclusion

We have identified a full complement of CUT class homeobox genes in the genome of amphioxus, *Branchiostoma floridae*. The amphioxus genome contained one CUX of protostome-type genomic structure, one ONECUT, one COMPASS and one atypical CUT class homeobox gene. The present findings lead us to the new evolutionary scenario of the CUT class homeobox genes in the deuterostomes, which highlights dramatic evolution of a transcription factor gene family in the lineage to vertebrates and also the ancient character of amphioxus, retaining certain protostome traits in the evolution.

Materials and Methods

Search for CUT class sequences genome databases

CUT protein sequences were searched throughout the draft genome database of *Branchiostoma floridae* (http://genome.jgi-psf.org/Brafl1/Brafl1.home.html) with the tblastn algorithm using CUX, SATB, ONECUT and COMPASS sequences of human, mouse, zebrafish, *Ciona intestinalis, Halocynthia roretzi* and *Drosophila melanogaster* out of the NCBI sequence database (http://www.ncbi.nlm.nih.gov/). Also, putative COMPASS protein sequence was searched from the draft genome database of *Strongylocentrotus purpuratus* (http://www.ncbi.nlm.nih.gov/genome/guide/sea_urchin/) using the COMPASS protein sequence of *Drosophila melanogaster.*

Sequence analysis

Sequences were aligned using the CLUSTALW program in the Molecular Evolutionary Genetics Analysis (MEGA) software package (Kumar *et al.,* 2004) (http://www.megasoftware.net/). Alignments were checked and gaps were removed manually before construction of phylogenetic trees. Trees based on neighbor joining method were constructed using the MEGA package. For the construction of maximum likelihood trees, first, the best-fitting model of protein substitution was selected using the ProtTest program (Abascal *et al.,* 2005), second, trees were calculated using the Phyml program (Guindon and Gascuel, 2003; Guindon *et al.,* 2005) and lastly, the resulting unrooted trees were displayed in Newick format using the MEGA package.

Acknowledgements

This work was supported by KAKENHI (Grant-in-Aid for Scientific Research) on Priority Areas «comparative genomics» from the Ministry of Education, Culture, Sports, Science and Technology of Japan and Grant-in-Aid for Scientific Research (B) from JSPS to H.S.

References

ABASCAL, F., ZARDOYA, R. AND POSADA, D. (2005). ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21: 2104-5.

ALVAREZ, J. D. YASUI, D. H. NIIDA, H. JOH, T. LOH, D. Y. (2000) The MAR-binding protein SATB1 orchestrates temporal and spatial expression of multiple genes during T-cell development. *Genes Dev* 14: 521-35

BLOCHLINGER, K., JAN, L. Y. AND JAN, Y. N. (1991). Transformation of sensory organ identity by ectopic expression of Cut in *Drosophila*. *Genes Dev* 14:1124-35.

BOURLAT, S. J., JULIUSDOTTIR, T., LOWE, C. J., FREEMAN, R., ARONOWICZ, J., KIRSCHNER, M., LANDER, E. S., THORNDYKE, M., NAKANO, H., KOHN, A. B. ET AL. (2006). Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. *Nature* 444: 85-8.

BÜRGLIN, T. R. (1994). A comprehensive classification of homeobox genes. In *Guidebook to the Homeobox Genes* (Ed.Duboue, D.). *Oxford University Press, Oxford*, 25-71.

BÜRGLIN, T. R. (1995). The evolution of homeobox genes. In *Biodiversity and Evolution* (Eds. Arai, R., Kato, M. and Doi, Y.). *The National Science Museum Foundation, Tokyo*, 291-336.

BÜRGLIN, T. R. AND CASSATA, G. (2002). Loss and gain of domains during evolution of cut superclass homeobox genes. *Int J Dev Biol* 46: 115-23.

DELSUC, F., BRINKMANN, H., CHOURROUT, D. AND PHILIPPE, H. (2006). Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* 439: 965-8.

ELLIS, T., GAMBARDELLA, L., HORCHER, M., TSCHANZ, S., CAPOL, J., BERTRAM, P., JOCHUM, W., BARRANDON, Y. AND BUSSLINGER, M. (2001). The transcriptional repressor CDP (Cutl1) is essential for epithelial cell differentiation of the lung and the hair follicle. *Genes Dev* 15: 2307-19.

FUSS, B. AND HOCH, M. (1998). *Drosophila* endoderm development requires a novel homeobox gene which is a target of Wingless and Dpp signalling. *Mech Dev* 79: 83-97.

GILLINGHAM, A. K., PFEIFER, A. C. AND MUNRO, S. (2002). CASP, the alternatively spliced product of the gene encoding the CCAAT-displacement protein transcription factor, is a Golgi membrane protein related to giantin. *Mol Biol Cell* 13: 3761-74.

GUINDON, S. AND GASCUEL, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52: 696-704.

GUINDON, S., LETHIEC, F., DUROUX, P. AND GASCUEL, O. (2005). PHYML Online—a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res* 33: W557-9.

HARADA, R., DUFORT, D., DENIS-LAROSE, C. AND NEPVEU, A. (1994). Conserved cut repeats in the human cut homeodomain protein function as DNA binding domains. *J Biol Chem* 269: 2062-7.

HOWARD-ASHBY, M., MATERNA, S. C., BROWN, C. T., CHEN, L., CAMERON, R. A. AND DAVIDSON, E. H. (2006). Identification and characterization of homeobox transcription factor genes in *Strongylocentrotus purpuratus*, and their expression in embryonic development. *Dev Biol* 300: 74-89.

KOLZER, S., FUSS, B., HOCH, M. AND KLEIN, T. (2003). *Defective proventriculus* is required for pattern formation along the proximodistal axis, cell proliferation and formation of veins in the *Drosophila* wing. *Development* 130: 4135-47.

KUMAR, S., TAMURA, K. AND NEI, M. (2004). MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform* 5: 150-63.

LANDRY, C., CLOTMAN, F., HIOKI, T., ODA, H., PICARD, J. J., LEMAIGRE, F. P. AND ROUSSEAU, G. G. (1997). HNF-6 is expressed in endoderm derivatives and nervous system of the mouse embryo and participates to the cross-regulatory network of liver-enriched transcription factors. *Dev Biol* 192: 247-57.

LANNOY, V. J., BÜRGLIN, T. R., ROUSSEAU, G. G. AND LEMAIGRE, F. P. (1998). Isoforms of hepatocyte nuclear factor-6 differ in DNA-binding properties, contain a bifunctional homeodomain, and define the new ONECUT class of homeodomain proteins. *J Biol Chem* 273: 13552-62.

LETUNIC, I., COPLEY, R. R., PILS, B., PINKERT, S., SCHULTZ, J. AND BORK, P. (2006). SMART 5: domains in the context of genomes and networks. *Nucleic Acids Res* 34: D257-60.

LIEVENS, P. M., TUFARELLI, C., DONADY, J. J., STAGG, A. AND NEUFELD, E. J. (1997). CASP, a novel, highly conserved alternative-splicing product of the CDP/cut/cux gene, lacks cut-repeat and homeo DNA-binding domains, and interacts with full-length CDP in vitro. *Gene* 197: 73-81.

NAKAGOSHI, H., HOSHI, M., NABESHIMA, Y. AND MATSUZAKI, F. (1998). A novel homeobox gene mediates the Dpp signal to establish functional specificity within target cells. *Genes Dev* 12: 2724-34.

NEPVEU, A. (2001). Role of the multifunctional CDP/Cut/Cux homeodomain transcription factor in regulating differentiation, cell growth and development. *Gene* 270: 1-15.

NGUYEN, D. N., ROHRBAUGH, M. AND LAI, Z. (2000). The *Drosophila* homolog of Onecut homeodomain proteins is a neural-specific transcriptional activator with a potential role in regulating neural differentiation. *Mech Dev* 97: 57-72.

RAUSA, F., SAMADANI, U., YE, H., LIM, L., FLETCHER, C. F., JENKINS, N. A., COPELAND, N. G. AND COSTA, R. H. (1997). The cut-homeodomain transcriptional activator HNF-6 is coexpressed with its target gene HNF-3 beta in the developing murine liver and pancreas. *Dev Biol* 192: 228-46.

RUDDLE, F. H., BARTELS, J. L., BENTLEY, K. L., KAPPEN, C., MURTHA, M. T. AND PENDLETON, J. W. (1994). Evolution of Hox genes. *Annu Rev Genet* 28: 423-42.

SAMADANI, U. AND COSTA, R. H. (1996). The transcriptional activator hepatocyte nuclear factor 6 regulates liver gene expression. *Mol Cell Biol* 16: 6273-84.

SASAKURA, Y. AND MAKABE, K. W. (2001). A gene encoding a new ONECUT class homeodomain protein in the ascidian *Halocynthia roretzi* functions in the differentiation and specification of neural cells in ascidian embryogenesis. *Mech Dev* 104: 37-48.

SCHULTZ, J., MILPETZ, F., BORK, P. AND PONTING, C. P. (1998). SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A* 95: 5857-64.

TUFARELLI, C., FUJIWARA, Y., ZAPPULLA, D. C. AND NEUFELD, E. J. (1998). Hair defects and pup loss in mice with targeted deletion of the first cut repeat domain of the Cux/CDP homeoprotein gene. *Dev Biol* 200: 69-81.

WADA, S., TOKUOKA, M., SHOGUCHI, E., KOBAYASHI, K., DI GREGORIO, A., SPAGNUOLO, A., BRANNO, M., KOHARA, Y., ROKHSAR, D., LEVINE, M. ET AL. (2003). A genomewide survey of developmentally relevant genes in *Ciona intestinalis*. II. Genes for homeobox transcription factors. *Dev Genes Evol* 213: 222-34.

YASUI, D., MIYANO, M., CAI, S., VARGA-WEISZ, P. AND KOHWI-SHIGEMATSU, T. (2002). SATB1 targets chromatin remodelling to regulate genes over long distances. *Nature* 419: 641-5.