

Highly conserved syntenic blocks at the vertebrate Hox loci and conserved regulatory elements within and outside Hox gene clusters

Alison P. Lee, Esther G. L. Koh, Alice Tay, Sydney Brenner[†], and Byrappa Venkatesh[†]

Institute of Molecular and Cell Biology, 61 Biopolis Drive, Singapore 138673

Contributed by Sydney Brenner, February 22, 2006

Hox genes in vertebrates are clustered, and the organization of the clusters has been highly conserved during evolution. The conservation of Hox clusters has been attributed to enhancers located within and outside the Hox clusters that are essential for the coordinated “temporal” and “spatial” expression patterns of Hox genes in developing embryos. To identify evolutionarily conserved regulatory elements within and outside the Hox clusters, we obtained contiguous sequences for the conserved syntenic blocks from the seven Hox loci in fugu and carried out a systematic search for conserved noncoding sequences (CNS) in the human, mouse, and fugu Hox loci. Our analysis has uncovered unusually large conserved syntenic blocks at the HoxA and HoxD loci. The conserved syntenic blocks at the human and mouse HoxA and HoxD loci span 5.4 Mb and 4 Mb and contain 21 and 19 genes, respectively. The corresponding regions in fugu are 16- and 12-fold smaller. A large number of CNS was identified within the Hox clusters and outside the Hox clusters spread over large regions. The CNS include previously characterized enhancers and overlap with the 5' global control regions of HoxA and HoxD clusters. Most of the CNS are likely to be control regions involved in the regulation of Hox and other genes in these loci. We propose that the regulatory elements spread across large regions on either side of Hox clusters are a major evolutionary constraint that has maintained the exceptionally long syntenic blocks at the HoxA and HoxD loci.

conserved noncoding sequences | conserved synteny | fugu | global control region

Hox genes code for homeodomain-containing transcription factors that determine the anterior-posterior patterning of tissues along the body axis of animals. In vertebrates, Hox genes are organized into tight clusters containing up to 14 paralogs. A remarkable feature of these clusters is that the positions of genes in the cluster are colinear with their spatial and temporal expression pattern along the anterior-posterior axis of the embryo. Hox genes at the 3' end of the cluster (see Fig. 1) express in the anterior segments of the embryo and are turned on relatively early during development, whereas genes at the 5' end express in the posterior segments and at a later stage during development (1). More than two decades after the discovery of Hox clusters, the molecular mechanisms underlying this orchestrated expression pattern of Hox genes and the selective pressure that has maintained the clustering and colinear organization of vertebrate Hox genes still are not well understood.

The colinear organization of Hox genes has been shown to be essential for the temporal progression of expression but not for the spatial expression patterns. Transgenic Hox genes inserted at ectopic locations in the genome retain spatial expression to a large extent but fail to recapitulate their temporal pattern of regulation (1). The mechanism of colinear expression of vertebrate Hox genes may require coordinated alterations in the higher-order chromatin structure from one end of the cluster to the other (2). Such a mechanism could require the clustering and colinear organization of Hox genes to be strictly maintained.

Another factor that may have contributed to the clustering and colinearity of Hox genes is the existence of global enhancers that regulate several genes in a manner independent of their local enhancers but dependent on their location in the cluster. Global control regions (GCRs) have been identified on either side of the HoxD cluster. The 5' HoxD GCR, regulating the expression of *Lnp*, *Evx2*, and posterior *HoxD* genes in developing digits and the central nervous system, is located ≈240 kb upstream of *HoxD13* in the intergenic region between *Atp5g3* and *Lnp* (3). The 3' HoxD GCR mapped downstream of *HoxD1*, called early limb control region, is implicated in the early colinear expression of HoxD genes along the anterior-posterior axis in limb buds. Its exact location and sequence are not known (4). A 5' long-range enhancer located upstream of *HoxA* has been proposed to direct expression of *HoxA13* and four upstream genes (*Evx1*, *Hibadh*, *Tax1bp1*, and *Jazf1*) in the distal limb and genital bud (5). It is not known whether the 3' region of HoxA cluster or the flanking regions of HoxB and HoxC clusters contain GCRs similar to the HoxD cluster.

Comparisons of Hox clusters from phylogenetically distant vertebrates have proved to be effective in discovering conserved regulatory sequences. Several conserved putative regulatory elements have been identified in the HoxA cluster by comparing sequences from mammals, the horn shark, and ray-finned fishes (6–10). Comparisons of subsets of Hox genes from HoxB (11, 12), HoxC (e.g., refs. 13 and 14), and HoxD (e.g., refs. 10 and 15) clusters have also identified a number of putative regulatory elements. However, a clusterwide comparison of HoxB, HoxC, or HoxD clusters has not been carried out. Three blocks of noncoding sequences in the intergenic region of *Lnp* and *Evx2* genes at the HoxD locus have been shown to be highly conserved in mammals, chicken, fugu, and the horn shark, indicating that flanking regions of Hox clusters also contain conserved regulatory elements (16). Subsets of the 5' GCRs of the HoxD and HoxA clusters that are located outside the Hox clusters were found to be conserved in pufferfishes (3, 5). Apart from these elements, no systematic search for conserved regulatory elements outside the Hox clusters has been attempted. We therefore decided to do a systematic search for evolutionary conserved noncoding sequences (CNS) within and outside the vertebrate Hox clusters.

In contrast to a single Hox cluster in *Amphioxus* (a cephalochordate), mammals contain four Hox clusters (HoxA, HoxB, HoxC, and HoxD) that have presumably evolved through two

Conflict of interest statement: No conflicts declared.

Abbreviations: BAC, bacterial artificial chromosome; CNS, conserved noncoding sequences; GCR, global control region; NCBI, National Center for Biological Information; TFBS, transcription factor binding site.

Data deposition: The fugu Hox loci sequences reported in this paper have been deposited in the GenBank database (accession nos. DQ481663–DQ481669).

[†]To whom correspondence may be addressed. E-mail: sbrenner@salk.edu or mcbby@imcb.a-star.edu.sg.

© 2006 by The National Academy of Sciences of the USA

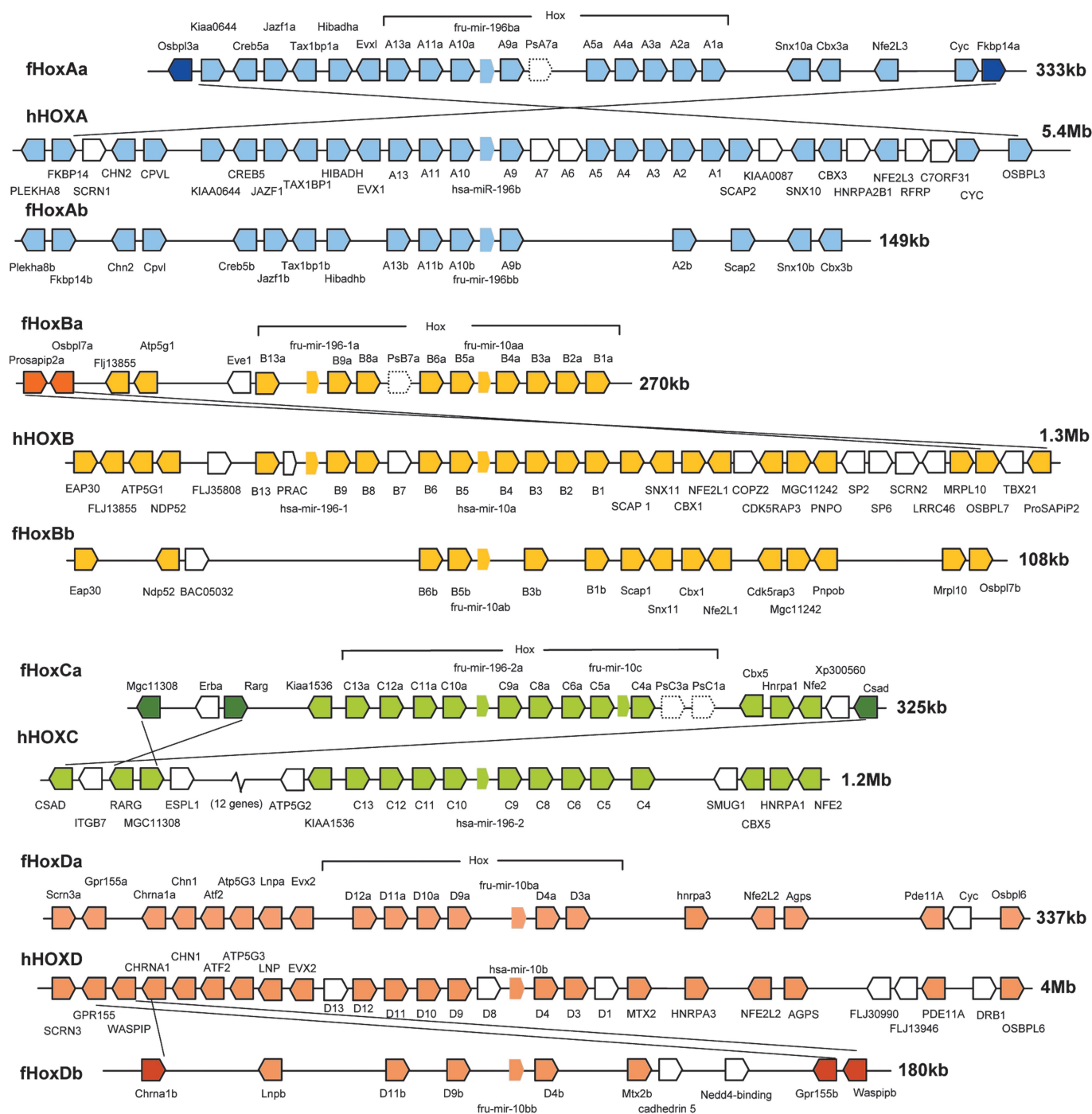


Fig. 1. Conserved syntenic blocks at the *fugu* and human *Hox* loci. *Fugu* has two duplicate loci for human *HoxA*, *HoxB*, and *HoxD* loci. Hexagons represent genes. MicroRNA genes are shown as borderless hexagons and pseudogenes with dotted borders. Genes that are present in only *fugu* or human locus are shown as open hexagons. Dark colored hexagons in the *fugu* (linked with their human orthologs with a diagonal line) represent genes that have undergone rearrangements.

rounds of duplication of a single Hox cluster. Ray-finned fishes, on the other hand, contain up to seven Hox clusters. The additional clusters in fishes have been attributed to a whole genome duplication in the fish lineage (17). To facilitate identification of CNS within and outside Hox clusters, we obtained contiguous sequences corresponding to conserved syntenic blocks from the seven Hox loci in the pufferfish, *Fugu rubripes*, and identified CNS in the human, mouse, and fugu Hox loci. These CNS represent potential regulatory elements associated with Hox gene clusters and non-Hox genes present in the conserved syntenic blocks.

Results

Fugu Hox Loci and Conserved Syntenic Blocks. We carried out an exhaustive search for Hox genes in the FUGU GENOME ASSEMBLY 3.0 and generated contiguous sequences for syntenic blocks that are conserved in human and mouse. Our analysis identified seven fugu Hox loci, HoxAa, HoxAb, HoxBa, HoxBb, HoxCa, HoxDa, and HoxDb, spanning 130–399 kb (Table 1 and Fig. 1). Analysis of an earlier fugu assembly (v2.0) had identified a putative eighth locus, designated HoxAc (18). We investigated the assembly v2.0 scaffolds that were assigned to this locus and

Table 1. Conserved syntenic fragments at the *fugu* and human *Hox* loci

| Fugu Hox cluster | Fugu contiguous length, kb | Syntenic genes | | | Fugu syntenic fragment, kb | Human syntenic fragment, kb | Fold compaction of fugu locus |
|------------------|----------------------------|----------------|----------|--------------|----------------------------|-----------------------------|-------------------------------|
| | | From | To | No. of genes | | | |
| HoxAa | 385.8 | Osbp13a | Fkbp14a | 21 | 332.5 | 5,270.7 | 16 |
| HoxAb | 163.1 | Plekha8b | Cbx3b | 16 | 148.5 | 3,934.2 | 26 |
| HoxBa | 297.9 | Prosapip2a | HoxB1a | 13 | 269.7 | 1,246.5 | 5 |
| HoxBb | 130.2 | Eap30 | Osbp17b | 15 | 107.6 | 1,212.4 | 11 |
| HoxCa | 378.8 | Mgc11308 | Csad | 16 | 325.3 | 1,182.8 | 4 |
| HoxDa | 366.8 | Scrn3a | Osbp16 | 19 | 337.4 | 4,035.7 | 12 |
| HoxDb | 398.5 | Chrna1b | Waspi pb | 8 | 180.0 | 2,783.8 | 15 |

Fugu has two orthologs each for human HoxA (HoxAa and Ab), HoxB (HoxBa and Bb), and HoxD (HoxDa and Db) clusters and one ortholog for HoxC (HoxCa) cluster.

found that most of them correspond to a *Tilapia* bacterial artificial chromosome (BAC) sequence (GenBank accession no. AF533976). These sequences are not present in the third fugu assembly.

Consistent with the small size of the fugu genome (400 Mb), fugu Hox loci are 4- to 26-fold smaller compared with human loci (Table 1). The longest conserved syntenic block is found at fugu HoxDa locus, which spans 337 kb compared with 4.0 Mb of its corresponding region in human, and contains 19 coding genes and a microRNA (miRNA) gene (Fig. 1 and Table 1). The highest number of syntenic genes, however, is found at fugu HoxAa locus (Fig. 1). This locus contains 21 coding genes and a miRNA gene across 333 kb. Its human ortholog covers 5.3 Mb. Conservation of such extensive gene arrangements in the fugu and human genomes suggests that they may contain functional elements that are shared by several genes across the locus. Comparisons between paralogous fugu Hox loci show that the duplicate loci (HoxAb, HoxBb, and HoxDb) are located on shorter syntenic blocks than their original loci, indicating that, besides the loss of a large number of genes after duplication, extensive rearrangements have occurred in the flanking regions of the duplicate Hox clusters. One exception to this trend is HoxBb locus, which has retained more syntenic genes outside the Hox cluster than HoxBa locus (Fig. 1).

Rearrangements at the Fugu Hox Loci. Although the relative positions and orientations of most of the genes are conserved at the fugu and human syntenic blocks, a few fugu genes at the ends of the fugu HoxAa, HoxBa, HoxCa, and HoxDb loci (Fig. 1) appear to have undergone intrachromosomal rearrangements. Because the positions and orientations of paralogs of some of these fugu genes are the same as their human orthologs (e.g., *Fkbp14b* in HoxAb locus, *Osbp17b* in HoxBb locus, and *Gpr155a* in HoxDa locus), we infer that the rearrangements occurred in the fugu lineage after the duplication of the Hox loci in the fish lineage. Parsimonious models that can explain the rearrangements at the fugu HoxAa, HoxBa, HoxCa, and HoxDb loci are shown in Figs. 3 and 4, which are published as supporting information on the PNAS web site. This model involves intrachromosomal pairing of homologous sequences and recombination, resulting in the inversion of the loop region containing the Hox cluster and the non-Hox genes linked to the Hox cluster. In addition to the inversion of the Hox cluster, localized inversions of short segments have occurred in the flanking regions of some of these Hox clusters (Figs. 3 and 4). The maintenance of contiguous segments containing the Hox cluster and the non-Hox genes linked to the cluster despite extensive rearrangements in the flanking regions imply a strong selection against rearrangements within these syntenic blocks.

Conserved Noncoding Sequences in the Hox Loci. The common ancestors of mammals and ray-finned fishes diverged ≈ 450 million years (Myr) ago, and the noncoding sequences conserved in these lineages over such a long evolutionary period are likely to represent functional elements such as RNA genes and cis-regulatory elements. Indeed, many of the noncoding elements conserved in fugu and mammals have been shown to function as tissue-specific enhancers (reviewed in refs. 19 and 20). To identify putative conserved regulatory elements in the Hox loci, we searched for CNS in the fugu, human, and mouse loci by using MLAGAN (21). CNS between human and mouse are typically defined by using 100-bp windows with a minimum 70% identity (22). Considering the much longer divergence period between ray-finned fishes and mammals than that between humans and rodents (≈ 80 Myr), we chose a lower-stringency criteria of 50-bp windows with minimum 60% identity for identifying fugu-mammal CNS. We also identified human-mouse CNS for the same regions by using the higher stringency and compared the fugu-mammal CNS with human-mouse CNS. Approximately 88% of fugu-mammal CNS were found to be part of human-mouse CNS, and another 3% partially overlap human-mouse CNS (Table 3, which is published as supporting information on the PNAS web site), indicating that most of the fugu-mammal CNS identified by using the lower stringency represent evolutionary-conserved sequences.

The VISTA (23) plots of MLAGAN alignments between seven fugu Hox loci and their orthologous human and mouse loci are shown in Figs. 5–11, which are published as supporting information on the PNAS web site. Because the numbers and extent of CNS identified between fugu-human and fugu-mouse alignments were similar, results of only fugu-human CNS will be discussed here as a representative of the fugu-mammal comparisons. Because our objective was to identify conserved regulatory elements, RNA genes in the initial set of CNS were excluded. A final set of 368 CNS comprising 39.3 kb was identified in the seven fugu Hox loci. The location and sizes of all of the 368 CNS are given in Table 3. The longest CNS is 805 bp long (79.6% identity) and is located in the intergenic region of *Lnp* and *Evx2*. Although only 1.5% of the mammalian genome code for proteins, up to 62.5% of the mouse genome has been found to be transcribed (24), indicating that a large proportion of noncoding regions are transcribed. To determine whether any of the CNS at the Hox loci are transcribed, we searched them against the human and mouse ESTs and FANTOM3 cDNAs. One hundred fifty of the 368 CNS (41%), most of them located within the Hox clusters, showed overlap with the ESTs (Table 3, column I). These CNS may represent previously undescribed noncoding exons. Based on comparisons of fugu and human genome sequences by using MEGABLAST (<http://genopole.toulouse.inra.fr/blast/megablast.html>), Woolfe *et al.* (20) have identified 1,389 conserved noncoding elements (CNEs). Only 54 of the CNS

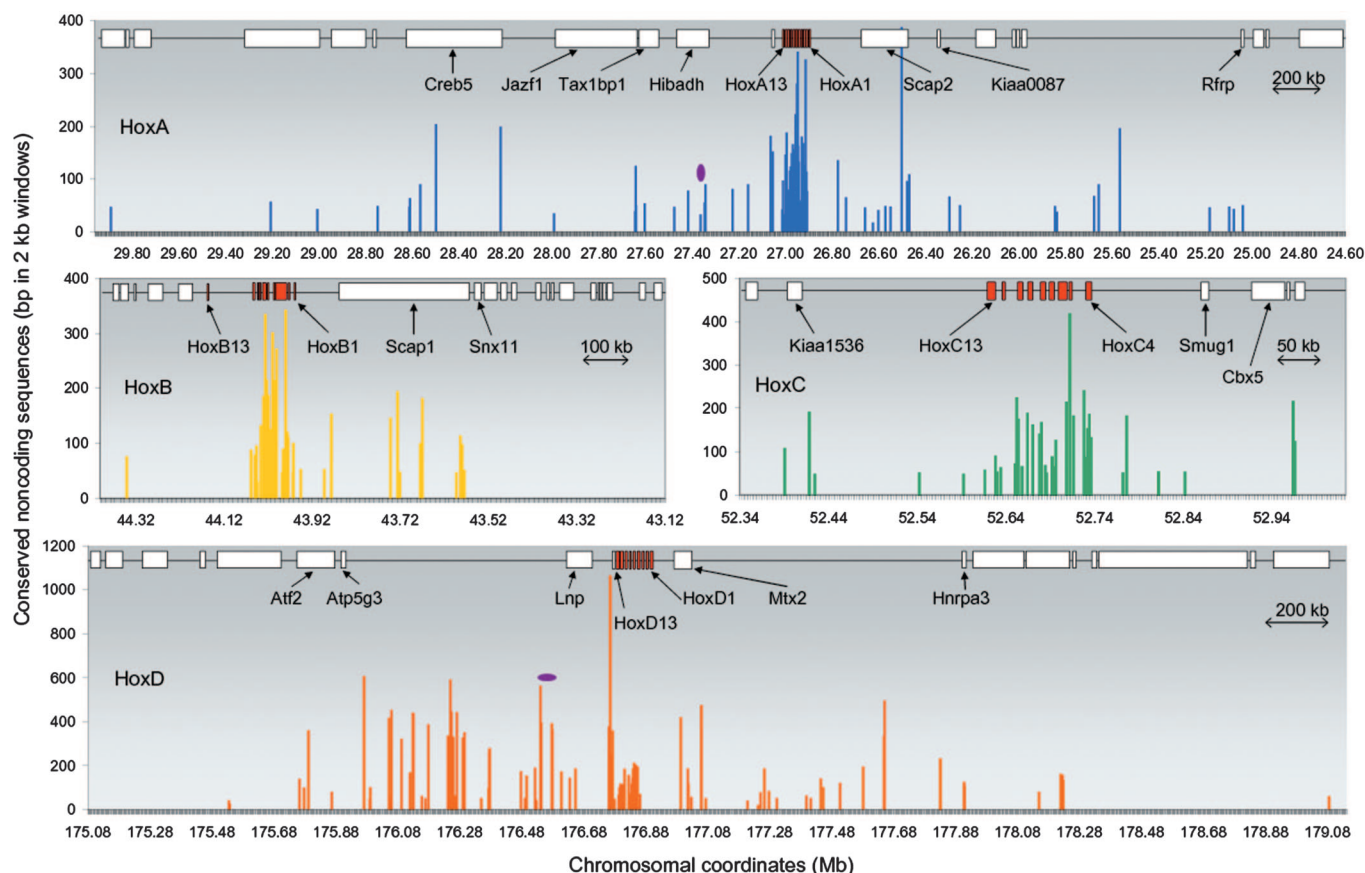


Fig. 2. Profiles of CNS at the human HoxA, HoxB, HoxC, and HoxD loci. x axis represents chromosomal coordinates, and y axis represents CNS. Genes at each locus are shown at the top as red boxes (Hox genes) or open boxes (non-Hox genes) linked with a thin line. Names of some genes are indicated. The previously identified 5' global enhancers at HoxD and HoxA loci (3, 5) are represented by purple oval shapes.

cbil.upenn.edu/tess). Our TFBS analysis identified 12,292 binding sites spread among 368 CNS (data not shown). Of these binding sites, 2,478 were found on sequences that are 100% identical in human, mouse, and fugu. The names and locations of these TFBS are shown in Table 3 (column M). A list of the most abundant of these TFBS is given in Table 5, which is published as supporting information on the PNAS web site. These transcription factors are likely to be involved in the regulation of Hox and non-Hox genes in the Hox loci analyzed by us. Among these TFBS, sites for Cdx-1, USF, E12, IUF-1, Pbx-1, and c-Jun were found to be significantly enriched ($P < 8 \times 10^{-3}$) in these conserved sequences as compared with TFBS predicted on random noncoding sequences in the human genome (see footnote to Table 5). Cdx-1 and USF are known regulators of Hox genes (25, 26), and because our loci contain a large number of Hox genes, many of the binding sites identified for these factors might be functional. The TFBS predicted on the CNS constitute targets for assaying the potential functions of the associated CNS.

Discussion

Chromosomal segments that have escaped rearrangements over a long evolutionary period reflect a strong selective pressure that has preserved the integrity of the cluster of genes located on the segment. Vertebrate Hox clusters typify such conserved chromosomal segments. The relative order and orientation of Hox genes in each cluster have been highly conserved during vertebrate evolution. In the present study, we have identified large syntenic blocks extending far beyond the Hox clusters that have

been conserved in fugu and mammals, indicating that the flanking regions of Hox clusters are also under selective pressure. The syntenic blocks at the human and mouse HoxA and HoxD loci are exceptionally large, spanning 5.4 Mb and 4 Mb, respectively. Genomewide comparisons of fugu and human had indicated that the conserved segments in the two genomes were restricted to short fragments with few syntenic genes on them (27). Nevertheless, longer syntenic regions have been identified at several loci for which contiguous sequences are available (28–30). The longest syntenic block identified so far is at the fugu and human *Shh* locus. The order and orientation of 16 genes spread across 4 Mb in the human *Shh* loci are totally conserved in fugu, although the syntenic block in rodents is restricted to only 12 genes (30). Approximately 25 conserved noncoding elements were identified in this locus, and a large number of them were found to exhibit enhancer activity in transgenic assays. It was proposed that association of these elements with several genes in this locus is responsible for maintaining the large syntenic block (30). The conserved syntenic blocks at the human HoxA and HoxD loci contain 97 and 149 CNS, respectively. These CNS include previously characterized enhancer elements, including a subset of elements associated with the 5' global enhancers of the HoxD and HoxA clusters (3, 5). We propose that most of these CNS are likely to be regulatory elements, and the distribution of these CNS over large regions may be responsible for the maintenance of the unusually large syntenic blocks at these loci.

The distribution of potential regulatory elements over a large region in the mammalian genome poses a major challenge in

testing their function. For example, our analysis has identified 59 putative regulatory elements, including subsets of the HoxD 5' GCR spread across a 741-kb intergenic region of *Atp5g3* and *Lnp*. Functional assay of this control region should ideally include all of the 59 elements, but it is not easy to obtain a mammalian genomic clone containing all these elements. The compact intergenic regions of fugu offer an attractive alternative for testing the function of such clusters of CNS. The intergenic region between *Atp5g3* and *Lnp* in the fugu genome is only 74 kb, and the locus from *Atp5g3* to *HoxD3a* is 131 kb. A fugu BAC containing this region will be invaluable for testing the function of these elements. Although the fugu clone may lack mammal-specific elements, transgenic experiments in zebrafish and mice should provide useful insights into the role of this cluster of CNS in the regulation of Hox and other genes in this locus.

Materials and Methods

Sequences and Annotation. We analyzed the third fugu assembly (www.fugu-sg.org) to identify scaffolds containing Hox gene sequences (Table 6, which is published as supporting information on the PNAS web site) by TBLASTN (www.ncbi.nlm.nih.gov/genome/seq/HsBlast.html) search by using known fugu, zebrafish, and human Hox protein sequences. A contig map for each locus was generated based on cosmid and BAC end-sequence linkage information obtained by searching the scaffold sequences against the cosmid and BAC end sequences. Contiguous sequences extending on both sides of Hox clusters were obtained by filling gaps within and between scaffolds until two or more genes that were unrelated to genes in the orthologous human Hox loci were encountered. Gaps were filled by sequencing cosmid or BAC DNA directly or by sequencing PCR products amplified from genomic DNA. Protein-coding genes were annotated by BLASTX search of the fugu sequence against the NR protein database at the National Center for Biological Information (NCBI) and followed by TBLASTN search of the most similar full-length protein sequence in the NR database against the fugu sequence. For some non-Hox genes with ambiguous exon boundaries, gene structures were predicted by using GENEWISE (www.ebi.ac.uk/Wise2). Genomic sequences and annotation details for the human (v30.35c; NCBI 35) and mouse (v25.33; NCBI m33) Hox cluster loci were extracted from ENSEMBL (www.ensembl.org). Human, mouse, and rat multiple alignments (Hu-

man hg16, mouse mm4, and rat rn3) were obtained from <http://pipeline.lbl.gov/downloads.shtml>. Human-mouse CNS that fulfilled the criteria of $\geq 70\%$ identity across 100-bp windows were extracted, and coordinates were taken with respect to the NCBI 35 (hg17) assembly.

Identification and Analysis of CNS. Repetitive sequences in the human, mouse, and fugu loci were masked by using REPEAT-MASKER REPEAT (Update September 8, 2005; Genetic Information Research Institute, Mountain View, CA). Multiple alignments were generated by using MLAGAN (21) and visualized by using VISTA (23). All MLAGAN alignments and VISTA visualization were carried out on a Compaq (Palo Alto, CA) DEC Alpha server, and all scripts were written in PERL (O'Reilly Media, Sebastopol, CA). CNS between fugu and mammals were identified by using a criterion of $\geq 60\%$ identity across 50 bp windows.

The initial set of fugu-mammal CNS were BLASTX searched against the NCBI NR protein database (cutoff 1×10^{-4}) to identify coding exons that were missed during the annotation. This search resulted in a refinement of the exon annotation of some fugu genes (e.g., *Atf2* and *Hnrpa3*). The revised set of CNS was then BLASTN searched against RFAM (www.sanger.ac.uk/Software/Rfam) and MIRBASE (<http://microrna.sanger.ac.uk/sequences>) to identify and eliminate RNA genes. The fugu-mammal CNS were searched against the fugu-human conserved noncoding elements (26) and human and mouse ESTs and FANTOM3 cDNAs by using the cutoff of 1×10^{-4} and minimum 50% coverage to determine the extent of overlap. Finally, CNS between paralogous fugu loci were compared by using BLASTN.

Identification of TFBS. The CNS were submitted to TESS (www.cbil.upenn.edu/teess) for the identification of putative TFBS. A combined search (by using filtered string-matching and searching against weight matrices) was carried out with vertebrate transcription factors and a maximum allowable string mismatch of 10%. Only string matches with L_a value equal to 2, and matrix matches with $S_c > 0.9$ and $S_m > 0.8$ were retained.

This project was funded by Singapore's Agency for Science, Technology, and Research. A.P.L. is supported by the A*STAR Graduate Scholarship. B.V. is an adjunct staff of the Department of Pediatrics, Yong Loo Lin School of Medicine, National University of Singapore.

1. Kmita, M. & Duboule, D. (2003) *Science* **301**, 331–333.
2. Chambeyron, S., Da Silva, N. R., Lawson, K. A. & Bickmore, W. A. (2005) *Development (Cambridge, U.K.)* **132**, 2215–2223.
3. Spitz, F., Gonzalez, F. & Duboule, D. (2003) *Cell* **113**, 405–417.
4. Zakany, J., Kmita, M. & Duboule, D. (2004) *Science* **304**, 1669–1672.
5. Lehoczy, J. A., Williams, M. E. & Innis, J. W. (2004) *Evol. Dev.* **6**, 423–430.
6. Kim, C.-B., Amemiya, C., Bailey, W., Kawasaki, K., Mezey, J., Miller, W., Minoshima, S., Shimizu, N., Wagner, G. & Ruddle, F. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 1655–1660.
7. Chiu, C.-H., Amemiya, C., Dewar, K., Kim, C.-B., Ruddle, F. H. & Wagner, G. P. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 5492–5497.
8. Santini, S., Boore, J. L. & Meyer, A. (2003) *Genome Res.* **13**, 1111–1122.
9. Chiu, C.-H., Dewar, K., Wagner, G. P., Takahashi, K., Ruddle, F., Ledje, C., Bartsch, P., Scemama, J. L., Stellwag, E., Fried, C., et al. (2004) *Genome Res.* **14**, 11–17.
10. Prohaska, S. J., Fried, C., Flamm, C., Wagner, G. P. & Stadler, P. F. (2004) *Mol. Phylogenet. Evol.* **31**, 581–604.
11. Marshall, H., Studer, M., Popperl, H., Aparicio, S., Kuroiwa, A., Brenner, S. & Krumlauf, R. (1994) *Nature* **370**, 567–571.
12. Hadrys, T., Prince, V., Hunter, M., Baker, R. & Rinkwitz, S. (2004) *J. Exp. Zool.* **302**, 147–164.
13. Geadia, A. M., Coletta, P. L. & Sharpe, P. T. (1996) *Mamm. Genome* **7**, 81–84.
14. Wang, W. C., Anand, S., Powell, D. R., Pawashe, A. B., Amemiya, C. T. & Shashikant, C. S. (2004) *J. Exp. Zool.* **302**, 436–445.
15. Nolte, C., Amores, A., Nagy Kovacs, E., Postlethwait, J. & Featherstone, M. (2003) *Mech. Dev.* **120**, 325–335.
16. Sabarinadh, C., Subramanian, S., Tripathi, A. & Mishra, R. K. (2004) *BMC Genomics* **5**, 75.
17. Amores, A., Force, A., Yan, Y. L., Joly, L., Amemiya, C., Fritz, A., Ho, R. K., Langeland, J., Prince, V., Wang, Y. L., et al. (1998) *Science* **282**, 1711–1714.
18. Amores, A., Suzuki, T., Yan, Y. L., Pomeroy, J., Singer, A., Amemiya, C. & Postlethwait, J. H. (2004) *Genome Res.* **14**, 1–10.
19. Venkatesh, B. & Yap, W. H. (2005) *BioEssays* **27**, 100–107.
20. Woolfe, A., Goodson, M., Goode, D. K., Snell, P., McEwen, G. K., Vavouri, T., Smith, S. F., North, P., Callaway, H., Kelly, K., et al. (2005) *PLoS Biol.* **3**, e7.
21. Brudno, M., Do, C. B., Cooper, G. M., Kim, M. F., Davydov, E., Green, E. D., Sidow, A. & Batzoglou, S. (2003) *Genome Res.* **13**, 721–731.
22. Loots, G. G., Locksley, R. M., Blankespoor, C. M., Wang, Z. E., Miller, W., Rubin, E. M. & Frazer, K. A. (2000) *Science* **288**, 136–140.
23. Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M. & Dubchak, I. (2004) *Nucleic Acids Res.* **32**, W273–W279.
24. Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M. C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., et al. (2005) *Science* **309**, 1559–1563.
25. Haerry, T. E. & Gehring, W. J. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 13884–13889.
26. Zhu, J., Giannola, D. M., Zhang, Y., Rivera, A. J. & Emerson, S. G. (2003) *Blood* **102**, 2420–2427.
27. Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J. M., Dehal, P., Christoffels, A., Rash, S., Hoon, S., Smit, A., et al. (2002) *Science* **297**, 1301–1310.
28. Venkatesh, B., Gilligan, P. & Brenner, S. (2000) *FEBS Lett.* **476**, 3–7.
29. Smith, S. F., Snell, P., Gruetznher, F., Bench, A. J., Haaf, T., Metcalfe, J. A., Green, A. R. & Elgar, G. (2002) *Genome Res.* **12**, 776–784.
30. Goode, D. K., Snell, P., Smith, S. F., Cooke, J. E. & Elgar, G. (2005) *Genomics* **86**, 172–181.