Evidence for at least six Hox clusters in the Japanese lamprey (*Lethenteron japonicum*)

Tarang K. Mehta[a,b,1], Vydianathan Ravi[a,1], Shinichi Yamasaki[c], Alison P. Lee[a], Michelle M. Lian[a], Boon-Hui Tay[a], Sumanty Tohari[a], Seiji Yanai[d], Alice Tay[a], Sydney Brenner[a,c,2], and Byrappa Venkatesh[a,b,2]

[a]Institute of Molecular and Cell Biology, Agency for Science, Technology and Research, Biopolis, Singapore 138673; [b]Department of Paediatrics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 119228; [c]Okinawa Institute of Science and Technology Graduate University, Onna-son, Okinawa 904-0495, Japan; and [d]Department of Environmental Sciences, Ishikawa Prefectural University, Nonoichi, Ishikawa 921-8836, Japan

Contributed by Sydney Brenner, August 22, 2013 (sent for review July 2, 2013)

Cyclostomes, comprising jawless vertebrates such as lampreys and hagfishes, are the sister group of living jawed vertebrates (gnathostomes) and hence an important group for understanding the origin and diversity of vertebrates. In vertebrates and other metazoans, Hox genes determine cell fate along the anteroposterior axis of embryos and are implicated in driving morphological diversity. Invertebrates contain a single Hox cluster (either intact or fragmented), whereas elephant shark, coelacanth, and tetrapods contain four Hox clusters owing to two rounds of whole-genome duplication ("1R" and "2R") during early vertebrate evolution. By contrast, most teleost fishes contain up to eight Hox clusters because of an additional "teleost-specific" genome duplication event. By sequencing bacterial artificial chromosome (BAC) clones and the whole genome, here we provide evidence for at least six Hox clusters in the Japanese lamprey (*Lethenteron japonicum*). This suggests that the lamprey lineage has experienced an additional genome duplication after 1R and 2R. The relative age of lamprey and human paralogs supports this hypothesis. Compared with gnathostome Hox clusters, lamprey Hox clusters are unusually large. Several conserved noncoding elements (CNEs) were predicted in the Hox clusters of lamprey, elephant shark, and human. Transgenic zebrafish assay indicated the potential of CNEs to function as enhancers. Interestingly, CNEs in individual lamprey Hox clusters are frequently conserved in multiple Hox clusters in elephant shark and human, implying a many-to-many orthology relationship between lamprey and gnathostome Hox clusters. Such a relationship suggests that the first two rounds of genome duplication may have occurred independently in the lamprey and gnathostome lineages.

Hox genes encode transcription factors that specify the identities of body segments along the anteroposterior axis of metazoan embryos. Because of the crucial role of Hox proteins in defining the identities of body segments, Hox genes are attractive candidates for understanding the morphological diversity of animals (1–3). In most metazoan genomes, Hox genes are organized into clusters. Invertebrates typically possess a single cluster that is either intact (e.g., amphioxus), split into fragments (e.g., fruit fly), or atomized (e.g., *Oikopleura*) (4). By contrast, vertebrates contain multiple Hox clusters because of whole-genome duplication events that occurred at different stages of their evolutionary history. For example, tetrapods, elephant shark, and coelacanth contain four Hox clusters (5) because of the two rounds of whole-genome duplication events (denoted as "1R" and "2R") that occurred early during the evolution of vertebrates (6). Most teleost fishes contain seven or eight Hox clusters as a result of an additional "teleost-specific" genome duplication (TSGD) event in the ray-finned fish lineage (7). The Atlantic salmon, whose lineage has experienced a more recent tetraploidization event on top of the TSGD, contains 13 Hox clusters (8). A feature of Hox cluster genes is the collinearity between their positions in the cluster and their expression pattern along the anteroposterior axis of developing embryos. This phenomenon, known as "spatial collinearity," is conserved in invertebrates and gnathostomes (9–11). In addition, gnathostome Hox genes also exhibit "temporal collinearity" whereby anterior genes are expressed earlier than posterior genes (12).

The cyclostomes, comprising the jawless vertebrates lampreys and hagfishes, are the sister group of extant gnathostomes. However, the two groups differ significantly in their morphological traits and physiological systems. Cyclostomes contain a single, medially located dorsal nostril as opposed to the two ventrally located nostrils in gnathostomes. In addition, cyclostomes lack mineralized tissues, hinged jaws, paired appendages, pancreas, and spleen that are characteristic of gnathostomes. Cyclostomes also possess a physiologically distinct adaptive immune system that lacks antibodies. Instead, they make use of variable lymphocyte receptors for antigen recognition (13). These contrasting features combined with the unique phylogenetic position of cyclostomes make them a critical group for understanding the evolution and diversity of vertebrates.

In contrast to the detailed information available for Hox gene clusters in various gnathostome taxa, the number of Hox clusters in cyclostomes is unclear. A PCR-based survey of the Pacific hagfish (*Eptatretus stoutii*) provided evidence for seven *Hox9* genes, suggesting the presence of at least seven Hox clusters in hagfish (14). Similar surveys of the sea lamprey (*Petromyzon marinus*) and Japanese lamprey (*Lethenteron japonicum)* have identified up to four fragments of the Hox paralogous group (PG) 5/6, implying the presence of at least four Hox clusters in lampreys (15–18). The recent assembly and analysis of the somatic

Significance

Lampreys and hagfishes (cyclostomes) are the only living group of jawless vertebrates and therefore are important for the study of vertebrate evolution. We have characterized Hox clusters in the Japanese lamprey (*Lethenteron japonicum*), and shown that it contains at least six Hox clusters as compared with four Hox clusters in tetrapods. This suggests that the lamprey lineage has undergone an additional round of genome duplication compared with tetrapods. Several conserved noncoding elements (CNEs) were predicted in the Hox clusters of lamprey, elephant shark, and human. Transgenic assay of CNEs demonstrated their potential to function as *cis*-regulatory elements. Thus, these CNEs may represent part of the core set of *cis*-regulatory elements that were present in the common ancestor of vertebrates.

Author contributions: S.B. and B.V. designed research; T.K.M., V.R., S. Yamasaki, A.P.L., B.-H.T., S.T., and A.T. performed research; S. Yanai contributed new reagents/analytic tools; T.K.M., V.R., A.P.L., M.M.L., S.B., and B.V. analyzed data; and B.V. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. KF318001–KF318029 and APJL00000000).

[1]T.K.M. and V.R. contributed equally to this work.

[2]To whom correspondence may be addressed. E-mail: mcbbv@imcb.a-star.edu.sg or sydney.brenner@hotmail.com.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1315760110/-/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1315760110

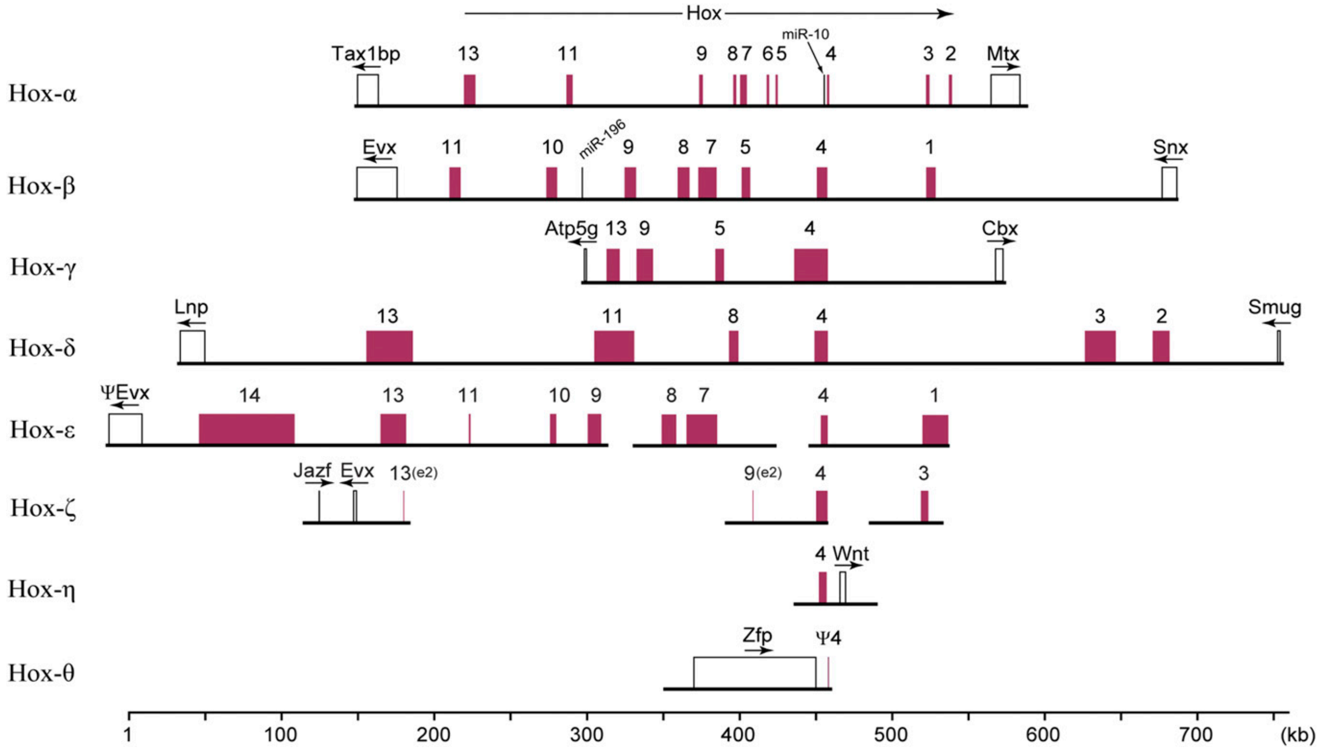

Fig. 1. Hox gene loci in the Japanese lamprey. Genes are represented as boxes and arrows denote the direction of transcription. Pseudogenes are denoted by the ψ symbol. Hox gene pairs *Hox-ε8*/*ε7* and *Hox-ε4*/*ε1* are putatively assigned to be part of Hox-ε locus comprising *Hox-ε14* to *Hox-ε9* genes. Likewise, Hox genes *Hox-ζ13*, *Hox-ζ9*/*ζ4*, and *Hox-ζ3* are putatively assigned to be part of a single locus. More data are required to confirm whether they are really part of such loci/clusters. e2, second exon (only the second exon could be identified for these genes).

genome of the sea lamprey has revealed the presence of two Hox clusters and eight other Hox genes that could not be assigned to any cluster (19). Interestingly, previous phylogenetic analysis of sea lamprey Hox genes had suggested that the Hox clusters of sea lamprey and gnathostomes arose from independent duplications and that the last common ancestor of cyclostomes and gnathostomes had a single Hox cluster (20). However, recent analyses of several gene families and the whole genome of sea lamprey have concluded that the exclusive clustering of lamprey genes in phylogenetic trees, suggestive of independent duplications in the lamprey lineage, is likely to be an artifact owing to a guanine and cytosine (GC) bias in the lamprey genome that affects codon use and amino acid composition of lamprey proteins (21, 22). In this study, we have carried out an exhaustive search for Hox genes in the Japanese lamprey genome by probing three BAC libraries and by generating a 20.5× coverage 454-based genome assembly using DNA from the testis. Japanese lamprey and sea lamprey are Northern hemisphere lampreys (subfamily Petromyzontidae) that diverged about 30–10 Mya (23). Our analyses provide evidence for the presence of at least six Hox clusters in the Japanese lamprey genome, suggesting that the lamprey lineage has experienced an additional round of whole-genome duplication compared with tetrapods.

Results and Discussion

Hox Gene Clusters in the Japanese Lamprey. We probed three different BAC libraries (IMCB_Testis1, IMCB_Testis2, and IMCB_Blood1) extensively for Hox genes and completely sequenced selected positive BACs (32 in all) (*Methods*). In addition, a 20.5× coverage genome assembly was generated on Roche 454 systems using the same stock of DNA as that of the IMCB_Testis2 BAC library (*Methods*). The genome assembly has an N50 scaffold size of 1.05 Mb and spans 1.03 Gb of the estimated 1.6-Gb germline genome of the Japanese lamprey. We could identify 43 Hox genes in the combined dataset of BAC sequences and the genome assembly (*SI Appendix*, Fig. S1). This set includes all 18 previously known Japanese lamprey Hox genes (*SI Appendix*, Table S1) and orthologs of all 31 sea lamprey Hox genes available in GenBank (*SI Appendix*, Table S2), indicating that we have identified most of the Hox genes in the Japanese lamprey genome. Notably, there are eight *Hox4* genes in the Japanese lamprey, indicating that its genome potentially contains eight Hox clusters. We could identify four complete Hox clusters (i.e., clusters of Hox genes flanked by genes known to be linked to Hox clusters in other vertebrates) in addition to clusters of two or more Hox genes or singleton Hox genes (Fig. 1). Because we could not assign orthology between the lamprey Hox clusters and the four gnathostome Hox clusters (see the following section), we designated the four complete lamprey Hox clusters as Hox-α, -β, γ, and -δ. The remaining Hox genes/clusters were tentatively organized into four loci and designated as Hox-ε, -ζ, -η, and -θ loci, as shown in Fig. 1. Despite the incompleteness of Hox-ε, -ζ, -η, and -θ loci, there is sufficient evidence that Japanese lamprey contains at least six Hox clusters. The singleton *Hox4* gene in the Hox-η locus (*Hox-η4*) is unique in that it comprises four coding exons compared with two exons in most Hox genes and three exons in *Hox13* and *Hox14* genes (5, 24–26). The coding sequence of the singleton *Hox4* gene in Hox-θ locus (*Hox-θ4*) is highly similar to *Hox-η4*, but its first two exons are not identifiable and the last exon contains a premature stop codon. Additionally, sequences outside the coding regions of these two *Hox4* genes are quite divergent. Thus, these two *Hox4* genes represent distinct genes and not different alleles of a single gene.

Interestingly, Japanese lamprey orthologs of the 31 Hox genes known in the sea lamprey are distributed across six Hox clusters/loci (*SI Appendix*, Fig. S2*A*). This indicates that the sea lamprey Hox genes are also likely to be organized into at least six Hox clusters. A notable finding is that the *Hox12* gene is totally absent in both the Japanese lamprey and sea lamprey genomes. This gene is present in the single Hox cluster of amphioxus and in two different Hox clusters of most gnathostomes (*HoxC12* and *HoxD12*). Thus, *Hox12* seems to have been lost very early during the duplication history of lamprey Hox clusters. In zebrafish,

Table 1. CNEs between the Japanese lamprey Hox-α, Hox-β, Hox-γ, and Hox-δ clusters and the four Hox clusters (HoxA, B, C, D) of elephant shark and human

	Gnathostome CNEs (length; % identity)								
	HoxA cluster		HoxB cluster		HoxC cluster		HoxD cluster		
Lamprey CNE ID	*C. milii*	Human	*C. milii*	Human	*C. milii*	Human	*C. milii*	Human	Location
Hox-αCNE1	112 bp (71%)	105 bp (67%)	108 bp (69%)	92 bp (74%)	—	—	—	—	Hox8-7 intergenic
Hox-αCNE2	184 bp (72%)	159 bp (74%)	160 bp (79%)	160 bp (78%)	145 bp (68%)	50 bp (78%)	52 bp (81%)	96 bp (74%)	Hox7-6 intergenic
Hox-αCNE3	109 bp (69%)	81 bp (73%)	101 bp (72%)	61 bp (77%)	—	—	—	—	Hox6-5 intergenic
Hox-αCNE4	—	—	—	—	75 bp (59%)	73 bp (59%)	71 bp (69%)	55 bp (71%)	Hox5-4 intergenic
Hox-αCNE5	91 bp (66%)	91 bp (64%)	83 bp (64%)	62 bp (71%)	—	—	—	—	Hox4 intron
Hox-αCNE6	52 bp (69%)	58 bp (72%)	52 bp (70%)	53 bp (68%)	—	—	—	—	Hox4-3 intergenic
Hox-αCNE7	71 bp (83%)	71 bp (85%)	—	—	—	—	—	—	Hox3-2 intergenic
Hox-βCNE1	158 bp (67%)	129 bp (71%)	150 bp (71%)	144 bp (65%)	—	—	65 bp (69%)	53 bp (68%)	Hox7-5 intergenic
Hox-γCNE1	73 bp (77%)	93 bp (73%)	93 bp (69%)	98 bp (66%)	56 bp (75%)	51 bp (67%)	58 bp (71%)	64 bp (64%)	Hox9-5 intergenic
Hox-δCNE1	—	—	52 bp (69%)	52 bp (69%)	—	—	—	—	Hox8-4 intergenic

Each lamprey Hox cluster sequence was aligned with all of the four Hox clusters of elephant shark and human using MLAGAN and CNEs were predicted using VISTA. *C. milii*, *Callorhinchus milii.*

Hoxd12a is expressed in the pectoral fin (27), whereas in mouse, *HoxD12* is implicated in the development of forelimbs (28, 29). Thus, the *HoxD12* gene retained in gnathostome Hox clusters may have been coopted for the development of gnathostome-specific paired appendages.

A striking feature of the Japanese lamprey Hox clusters/loci is their large size and high repeat content compared with gnathostome Hox clusters. The intact, single Hox clusters in invertebrates are generally large (e.g., amphioxus ~400 kb) (25), whereas gnathostome Hox clusters are more compact (~100–210 kb) and contain very little interspersed repetitive elements (~1–8%) (5). The four complete Japanese lamprey Hox clusters range from 145 to 526 kb and contain unusually high levels of interspersed repetitive elements (23–33%) that are higher than the overall repeat sequence content of the whole genome (21%). Some lamprey intergenic and intronic sequences are extraordinarily large. For example, the intergenic region between *Hox-δ4* and *Hox-δ3* is as large (168 kb) as a single gnathostome Hox cluster. The first intron of the lamprey *Hox-ε14* is 56 kb, whereas its homologous introns in elephant shark *HoxD14* and coelacanth *HoxA14* genes are only 0.7 kb and 3.3 kb, respectively (5, 26). Thus, the organization of lamprey Hox clusters is more invertebrate-like than gnathostome-like.

To determine orthology relationships between the four gnathostome Hox clusters (A–D) and the supernumerary lamprey Hox clusters/loci, we carried out phylogenetic analysis using two different methods [maximum likelihood (ML) and Bayesian inference (BI)]. Because the lamprey *Hox4* PG contains eight members (seven full-length and one partial), phylogenetic analysis of PG4 should be most informative. However, phylogenetic analyses of *Hox4* genes (full-length or second exon only, coding sequence, protein sequence, or first and second codon positions) indicated that the lamprey genes cluster together away from the gnathostome genes (*SI Appendix*, Fig. S3–S5). A similar clustering of lamprey genes was observed when phylogenetic analysis was carried out for Hox PGs 8, 9, 11, and 13 that contain four or five members in the Japanese lamprey (*SI Appendix*, Fig. S6–S11). This pattern of clustering of lamprey Hox genes suggests that lamprey genes were duplicated independently in the lamprey lineage after it diverged from the gnathostome lineage. However, a similar clustering has been reported previously for sea lamprey Hox and non-Hox genes (20, 21) and was interpreted as an artifact because of the GC bias of the sea lamprey genome that affected the codon use pattern and amino acid composition of protein sequences (21, 22). We analyzed the GC content of the Japanese lamprey genome and also found it to be unusually high (48% GC content), and in fact higher than that of the sea lamprey genome (46%). It is therefore likely that the Japanese lamprey genome also has a GC bias that might be causing the exclusive clustering of the Japanese lamprey Hox genes. Thus, the phylogenetic analysis was inconclusive in assigning orthology to the Japanese lamprey Hox clusters.

We next attempted to use the synteny of genes flanking the Hox clusters to infer the orthology of lamprey and gnathostome Hox clusters. Interestingly, although the synteny of some genes flanking the lamprey Hox clusters is conserved in the gnathostome Hox cluster loci, the complement of genes linked to each cluster is different between lamprey and gnathostomes (*SI Appendix*, Fig. S2). For example, the cluster of *Mtx-Cbx-Hnrp* genes located downstream of the lamprey Hox-α cluster is not linked in its entirety to any of the human Hox clusters. This indicates that the genes linked to the lamprey Hox clusters have experienced a different history of secondary loss compared with that in humans. In addition, some genes flanking the lamprey Hox clusters seem to have been flipped around. For instance, in the human HoxD locus, the *AGPS* gene is found downstream and *ATF* is located upstream of the Hox cluster, respectively, but their lamprey homologs are located at opposite ends of the lamprey Hox-γ cluster. Thus, genes flanking the lamprey Hox clusters were also not informative for inferring orthology between the lamprey and gnathostome Hox clusters.

CNEs. To identify potential ancient vertebrate *cis*-regulatory elements in Hox clusters, we predicted conserved noncoding elements (CNEs) in the lamprey and gnathostome Hox clusters. Transgenic reporter assays of CNEs have shown that many of them have the potential to function as tissue-specific enhancers (30, 31). CNEs are typically predicted based on alignment of orthologous or paralogous gene loci. However, because we could not assign exact orthology between the lamprey and gnathostome

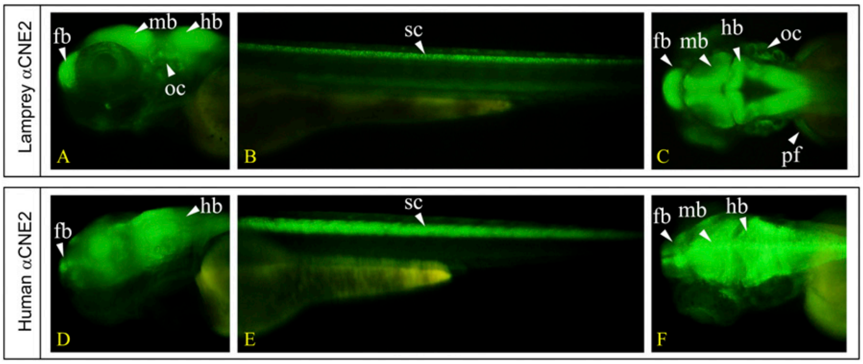

Fig. 2. Expression patterns driven by lamprey Hox-αCNE2 and its human homolog in 3 dpf F1 generation zebrafish embryos. (*A*, *B*, *D*, and *E*) Lateral views. (*C* and *F*) Dorsal views. fb, forebrain; hb, hindbrain; mb, midbrain; oc, otic capsule; pf, pectoral fin; sc, spinal cord.

Hox clusters, we aligned each of the four complete lamprey Hox clusters (Hox-α, -β, -γ, and -δ) with all of the four Hox clusters of selected gnathostomes (elephant shark and human) using the global alignment program MLAGAN (32) and predicted CNEs using VISTA (33). Overall, we identified very few CNEs between lamprey and the two gnathostome Hox clusters: seven in the Hox-α cluster and one each in the Hox-β, -γ, and -δ clusters (Table 1 and *SI Appendix*, Figs. S12 and S13). The paucity of CNEs in the lamprey Hox clusters is consistent with the low number of CNEs identified between the whole genomes of the sea lamprey and gnathostomes (22). Interestingly, despite the few CNEs, each lamprey Hox cluster (except Hox-δ cluster) shares CNEs with two or more elephant shark and human paralogous Hox clusters. Notably, the lamprey Hox-α cluster, which has lost very few Hox genes (and hence seems to be evolving slowly), shares CNEs across all four paralogous Hox clusters of elephant shark and human. In addition, the single CNE in the lamprey Hox-γ cluster is also conserved in all four paralogous Hox clusters of the two gnathostomes. This pattern of CNEs between individual lamprey Hox clusters and multiple paralogous gnathostome Hox clusters suggests that each lamprey Hox cluster is a common ortholog of all of the four gnathostome Hox clusters rather than an ortholog of a single gnathostome Hox cluster. If it were the latter case, each lamprey Hox cluster would have shared CNEs predominantly with a single gnathostome Hox cluster in a manner similar to the Hox clusters of elephant shark and human (*SI Appendix*, Fig. S14*A*) and fugu and human (*SI Appendix*, Fig. S14*B*).

To determine the biological significance of the lamprey CNEs, we tested Hox-αCNE2 (located between *Hox7* and *6*) that is conserved in all of the four paralogous Hox clusters of elephant shark and human, and its human HoxB cluster homolog in transgenic zebrafish. Both lamprey and human CNEs drove reproducible expression in the spinal cord (Fig. 2), where some *Hox6* genes are known to express. For example, the lamprey *Hox-α6* (known in literature as *LjHox6W*) is expressed in the neural tube (17), whereas zebrafish *Hox6* genes are expressed in the spinal cord (27, 34). These findings suggest that the CNEs in the lamprey and gnathostome Hox clusters have the potential to function as enhancers mediating the expression patterns of Hox genes. Interestingly, the lamprey CNE drove expression in the pectoral fin (Fig. 2), an organ that is absent in lamprey. Zebrafish *Hoxc6a* is expressed in the pectoral fin (34).

Genome Duplications in the Lamprey Lineage. The presence of at least six Hox clusters in the Japanese lamprey and the sea lamprey suggests that the lamprey lineage has experienced an additional round of genome duplication after 1R and 2R. Although it is possible that the additional lamprey Hox loci are the result of gene or segmental duplications, this possibility is less likely given that some of the lamprey Hox loci (Hox-ε and -ζ loci) comprise multiple Hox genes and/or known Hox cluster-linked genes. To date there is no evidence for such large-scale segmental duplication of Hox loci in any other vertebrate. Additional Hox clusters have been found so far only in ray-finned fishes and they are all associated with additional genome duplication event(s) (7, 8, 35). To verify our inference of an additional genome duplication in the lamprey lineage, we compared the ages of lamprey and human paralogs by calculating the rate of transversion at fourfold degenerate sites (4DTv rates), which is insensitive to variation in local GC content. For comparison, we also determined the relative age of paralogs in coelacanth and a teleost fish, the stickleback, which has undergone the additional TSGD. These comparisons showed that the ancient duplicate genes (4DTv value ≥ 0.2) of lamprey, like those of stickleback, are younger than those of human and coelacanth (Fig. 3). This pattern indicates that the lamprey lineage, like the stickleback lineage, has experienced large-scale gene duplication more recently than the human and coelacanth lineages. This is consistent with the view that, like the stickleback lineage, the lamprey lineage has experienced an additional whole-genome duplication after 1R and 2R.

A major unresolved issue in the study of early evolution of vertebrates is the timing of 1R and 2R in relation to the divergence of cyclostome and gnathostome lineages. The timings of 1R and 2R have been previously assessed by molecular phylogenetic analysis of lamprey and gnathostome genes (36). More recently, with the availability of the whole-genome sequence of the sea lamprey, this issue has been addressed by the frequency of retained duplicate genes and their synteny pattern in the genomes of sea lamprey and gnathostomes (22). The consensus of these analyses was that both 1R and 2R probably occurred before the cyclostome and gnathostome split. Given our present finding that the ancestor of the Japanese lamprey and sea lamprey lineages experienced an additional genome duplication, the timing of 1R and 2R needs to be reexamined taking into account that lamprey genomes contain a large number of paralogs resulting from the third round of a genome duplication event. We sought clues about the timings of 1R and 2R in the Hox clusters of lamprey and gnathostomes. The complement of syntenic genes linked to each lamprey Hox cluster is different from those linked to gnathostome Hox clusters (*SI Appendix*, Fig. S2). Although this pattern of synteny of duplicated genes can be explained by shared duplication history followed by independent gene losses, this scenario can also be due to independent histories of duplication and secondary loss of genes in the lamprey and gnathostome lineages. Another interesting feature of the lamprey Hox clusters is that they share CNEs across the four paralogous Hox clusters of gnathostomes (Table 1), rather than predominantly with one gnathostome cluster as discussed previously. This pattern of CNE distribution suggests a many-to-many orthology relationship between lamprey and gnathostome Hox clusters, which would be the case if the lamprey and gnathostome Hox clusters duplicated after the divergence of the two lineages. Thus, it is possible that the first two rounds of whole-genome duplication may have occurred independently in the lamprey and gnathostome lineages. This unconventional possibility, however, needs to be verified through detailed comparisons of the germ-line genomes of the sea lamprey and Japanese lamprey with representative gnathostome genomes.

Methods

Ethical Statement. All animal experiments were approved by the Institutional Animal Care and Use Committee of the Biological Resource Centre, Agency for Science, Technology and Research (A*STAR), Singapore (protocol #100520).

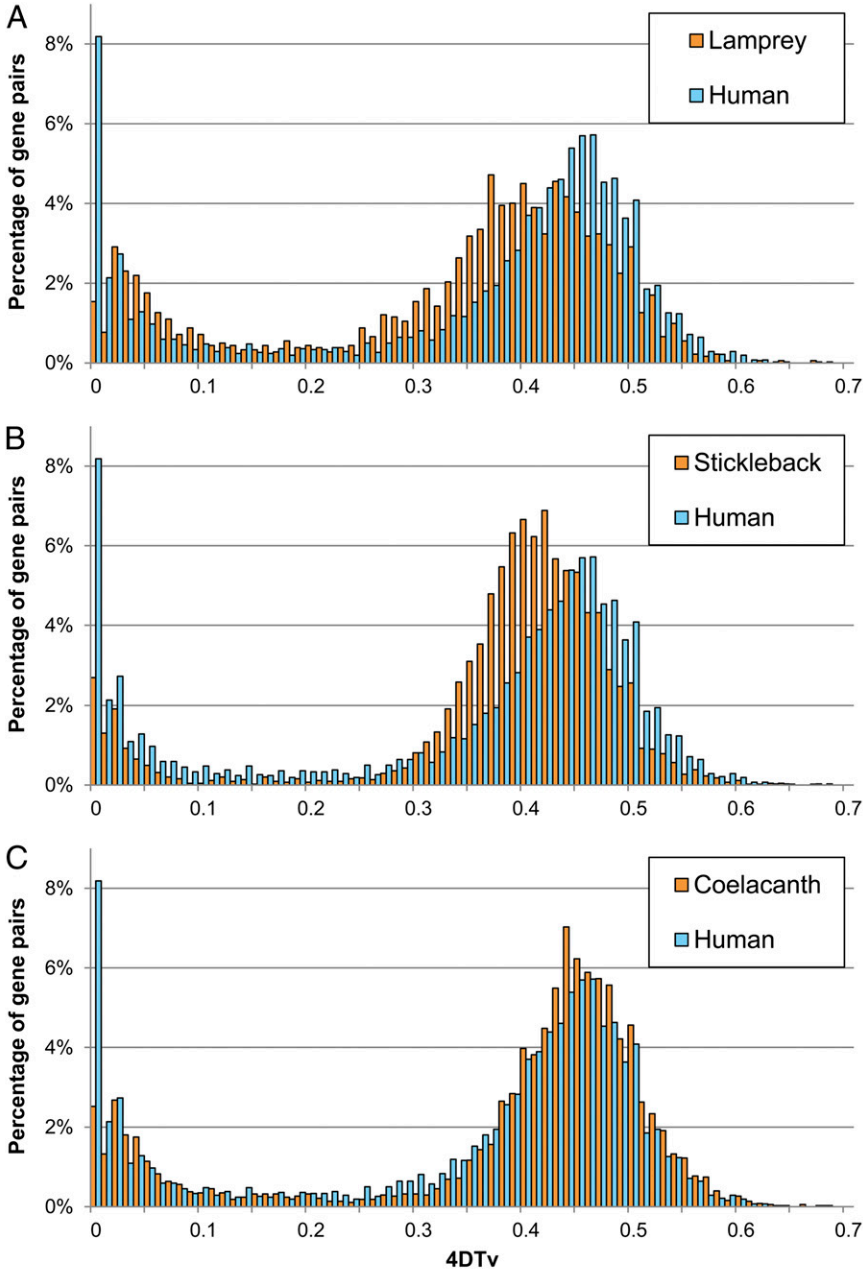

Fig. 3. The 4DTv of paralogs from lamprey and other vertebrates compared with human. (*A*) Japanese lamprey and human; (*B*) stickleback and human; (*C*) coelacanth and human. The 4DTv rates of paralogs in each genome showed a bimodal distribution, with 77–92% of the gene pairs having a 4DTv value of 0.2 or more. These gene pairs are composed mainly of ancient duplicate genes. The median 4DTv values for these gene pairs in human, lamprey, stickleback, and coelacanth are 0.437, 0.402, 0.409, and 0.444, respectively. The 4DTv rates of coelacanth are similar to that of human, consistent with the notion that human and coelacanth shared the last round of whole-genome duplication (i.e., 2R). On the other hand, the 4DTv rates of lamprey and stickleback are lower than that of human. For stickleback, this result is consistent with the teleost-specific additional genome duplication. For lamprey, this implies that there exists one round of whole-genome duplication that is more recent than 2R.

Identification and Sequencing of BACs. Three different Japanese lamprey BAC libraries (IMCB_Testis1: *Eco*RI, 92,160 clones, average insert size 100 kb; IMCB_Testis2: *Hin*dIII, 165,888 clones, average insert size 115 kb; and IMCB_Blood1: *Hin*dIII, 119,808 clones, average insert size 115 kb) were used to identify Hox-containing BAC clones. At an estimated genome size of 1.6 Gb, the three libraries provide a fold coverage of 5.7×, 11.9×, and 8.6×, respectively. The BAC libraries were screened using standard radioactive probing methods and probes specific for 18 Japanese lamprey Hox genes in GenBank. Selected positive BACs ("seed" BACs) were sequenced completely using a standard shotgun sequencing method and gap filling by PCR or primer walking. Sequencing was done using the BigDye Terminator Cycle Sequencing Kit (Applied Biosystems) on ABI 3730xl capillary sequencers (Applied Biosystems). Chromatograms were processed and assembled using Phred-Phrap and Consed (www.phrap.org/phredphrapconsed.html). Sequences of seed BACs were extended by identifying and sequencing overlapping BACs. In total, 638 potentially positive BACs were identified, of which 32 were sequenced completely (*SI Appendix*, Fig. S1) (GenBank accession nos. KF318001–KF318013). Because it is known that ~20% of germ-line DNA is lost in the somatic tissues of lamprey resulting from a developmentally programmed genome rearrangement (19), the blood library was used only to ensure complete coverage of the lamprey Hox gene repertoire. A combination of *ab initio* and homology-based methods was used to predict genes. The exon–intron structure of two lamprey *Hox* genes, *Hox-ζ4* and *Hox-η4*, were confirmed by 5′RACE using mRNA from gill and kidney tissues, respectively.

Generation of Lamprey Genome Sequence. Genomic DNA was extracted from the mature testis of a single individual. This stock of DNA was used for making the IMCB_Testis2 BAC library as well as for generating whole-genome shotgun sequences using the Roche 454 GS FLX Titanium and GS FLX+ systems. A combination of the following libraries was used to produce a 20.5× coverage genome assembly: six shotgun libraries (15.6×), two 3-kb paired-end libraries (1.8×), one 8-kb paired-end library (1.0×), one 12-kb paired-end library (1.2×), one 16-kb paired-end library (0.9×), and one BAC library (IMCB_Testis2; 38,210 BAC ends). The combined dataset was assembled using the Newbler assembler (ver. 2.7, Roche/454 Life Sciences). The assembly has been submitted to GenBank (accession no. APJL00000000) and is accessible at http://jlampreygenome.imcb.a-star.edu.sg/. The types and extent of repetitive sequences in the genome were predicted by RepeatMasker using Repbase library supplemented with lamprey-specific repeats. The latter were identified as reads that aligned to >500 other reads in a random pool of 100,000 reads, and assembling them using Phrap. Sequences of the Hox-BACs were mapped to the assembly, and additional Hox genes on the scaffolds that lie outside the BACs were identified using homology-based methods.

Phylogenetic Analysis. Phylogenetic analysis was carried out for paralogous Hox genes that were present in four or more copies. All lamprey sequences used for phylogenetic analysis have been submitted to GenBank (accession nos. KF318014–KF318029). Multiple alignments were generated using MUSCLE (www.ebi.ac.uk/Tools/msa/muscle/). Codon-based alignments of corresponding nucleotide coding sequences were generated based on the amino acid alignment using PAL2NAL (http://coot.embl.de/pal2nal/). Alignments were refined by manual inspection and trimming using BioEdit sequence alignment editor (www.mbio.ncsu.edu/BioEdit/BioEdit.html). The best-fit substitution models for the alignments were deduced using MEGA-CC (www.megasoftware.net/). ML and BI methods were used for phylogenetic analysis using the best-fit substitution model. MEGA-CC was used for batch ML analyses and 100 bootstrap replicates were used for node support. For BI analyses, we used MrBayes 3.2.1 (http://mrbayes.csit.fsu.edu/). Two independent runs starting from different random trees were run for 1 million generations with sampling every 100 generations. A consensus tree was built from all sampled trees excluding the first 2,500 (burn-in). Samples not reaching "stationarity" after 1 million generations were run for 5 million generations and the "burn-in" was adjusted accordingly.

Identification and Analysis of CNEs. Elephant shark and human Hox cluster sequences were extracted from GenBank and Ensembl, respectively. Repetitive sequences were identified and masked using RepeatMasker (www.repeatmasker.org/) based on Repbase (www.girinst.org/repbase/). Multiple alignments of Japanese lamprey, elephant shark, and human Hox cluster loci sequences were generated using the global alignment program MLAGAN (http://genome.lbl.gov/vista/index.shtml) with lamprey as the reference sequence. CNEs were predicted using a cutoff of ≥65% identity across 50-bp windows and visualized using VISTA (http://genome.lbl.gov/vista/index.shtml).

Functional Assay of CNEs. Selected CNEs were amplified by PCR using genomic DNA as a template. The products were cloned into a miniTol2 transposon donor plasmid linked to the mouse *cFos* basal promoter (McFos) and coding sequence of GFP. The CNE-containing McFos-miniTol2 construct and transposase mRNA were coinjected into the yolk of zebrafish embryos at the late one-cell or early two-cell stage. Embryos were screened for transient GFP expression and reared until maturity. Mature F0 (founder generation) adults were out-crossed with wild-type zebrafish to produce F1 progeny.

Relative Age of Paralogs. 4DTv is an indicator of the relative age of duplicate genes (37). The lamprey proteome was obtained by whole-genome annotation using Maker version 2.27-beta and by BLAST search of the lamprey genome against human, elephant shark, and sea lamprey proteomes. Lamprey paralogs were identified as reciprocal best hits in a BLAST search (E < 1e-3) of lamprey proteome against itself, whereas in-species paralogs for the other genomes were identified as genes with reciprocally highest identities in Ensembl Biomart release 71 (www.ensembl.org). 4DTv rates were identified from paralog pairs containing at least 50 fourfold degenerate sites (lamprey, 1,823 pairs; human, 4,214 pairs; stickleback, 4,445 pairs; and coelacanth, 3,771 pairs). A protein alignment of each gene pair was carried out using Clustal-Omega version 1.2.0 (38) and the alignment was converted to a coding sequence alignment using PAL2NAL version 14. Fourfold degenerate sites were extracted using Rphast (39) and the 4DTv rates were calculated using an in-house Perl script. A distribution of the 4DTv rates for each genome's set of paralogs was plotted and compared against that of human paralogs.

ACKNOWLEDGMENTS. We thank the Agency for Science, Technology and Research (A*STAR) Computational Resource Centre for the use of its high-performance computing facilities. This work was funded by the Okinawa Institute of Science and Technology Graduate University, Japan and the Biomedical Research Council of A*STAR, Singapore.

1. Carroll SB, Weatherbee SD, Langeland JA (1995) Homeotic genes and the regulation and evolution of insect wing number. *Nature* 375(6526):58–61.
2. Holland PW, Garcia-Fernàndez J (1996) Hox genes and chordate evolution. *Dev Biol* 173(2):382–395.
3. Wagner GP, Amemiya C, Ruddle F (2003) Hox cluster duplications and the opportunity for evolutionary novelties. *Proc Natl Acad Sci USA* 100(25):14603–14606.
4. Duboule D (2007) The rise and fall of Hox gene clusters. *Development* 134(14): 2549–2560.
5. Amemiya CT, et al. (2010) Complete HOX cluster characterization of the coelacanth provides further evidence for slow evolution of its genome. *Proc Natl Acad Sci USA* 107(8):3622–3627.
6. Putnam NH, et al. (2008) The amphioxus genome and the evolution of the chordate karyotype. *Nature* 453(7198):1064–1071.
7. Kuraku S, Meyer A (2009) The evolution and maintenance of Hox gene clusters in vertebrates and the teleost-specific genome duplication. *Int J Dev Biol* 53(5-6):765–773.
8. Mungpakdee S, et al. (2008) Differential evolution of the 13 Atlantic salmon Hox clusters. *Mol Biol Evol* 25(7):1333–1343.
9. Duboule D, Dollé P (1989) The structural and functional organization of the murine HOX gene family resembles that of Drosophila homeotic genes. *EMBO J* 8(5):1497–1505.
10. Gaunt SJ, Sharpe PT, Duboule D (1988) Spatially restricted domains of homeo-gene transcripts in mouse embryos: Relation to a segmented body plan. *Development* 104(Supplement):169–179.
11. Graham A, Papalopulu N, Krumlauf R (1989) The murine and Drosophila homeobox gene complexes have common features of organization and expression. *Cell* 57(3):367–378.
12. Izpisúa-Belmonte JC, Falkenstein H, Dollé P, Renucci A, Duboule D (1991) Murine genes related to the Drosophila AbdB homeotic genes are sequentially expressed during development of the posterior part of the body. *EMBO J* 10(8):2279–2289.
13. Pancer Z, et al. (2004) Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. *Nature* 430(6996):174–180.
14. Stadler PF, et al. (2004) Evidence for independent Hox gene duplications in the hagfish lineage: A PCR-based gene inventory of Eptatretus stoutii. *Mol Phylogenet Evol* 32(3):686–694.
15. Irvine SQ, et al. (2002) Genomic analysis of Hox clusters in the sea lamprey Petromyzon marinus. *J Exp Zool* 294(1):47–62.
16. Force A, Amores A, Postlethwait JH (2002) Hox cluster organization in the jawless vertebrate Petromyzon marinus. *J Exp Zool* 294(1):30–46.
17. Takio Y, et al. (2004) Evolutionary biology: Lamprey Hox genes and the evolution of jaws. *Nature* 429(6989):1 p following 262.
18. Takio Y, et al. (2007) Hox gene expression patterns in Lethenteron japonicum embryos—insights into the evolution of the vertebrate Hox code. *Dev Biol* 308(2):606–620.
19. Smith JJ, Antonacci F, Eichler EE, Amemiya CT (2009) Programmed loss of millions of base pairs from a vertebrate genome. *Proc Natl Acad Sci USA* 106(27):11212–11217.
20. Fried C, Prohaska SJ, Stadler PF (2003) Independent Hox-cluster duplications in lampreys. *J Exp Zoolog B Mol Dev Evol* 299(1):18–25.
21. Qiu H, Hildebrand F, Kuraku S, Meyer A (2011) Unresolved orthology and peculiar coding sequence properties of lamprey genes: The KCNA gene family as test case. *BMC Genomics* 12:325.
22. Smith JJ, et al. (2013) Sequencing of the sea lamprey (Petromyzon marinus) genome provides insights into vertebrate evolution. *Nat Genet* 45(4):415–421, e1–e2.
23. Kuraku S, Kuratani S (2006) Time scale for cyclostome evolution inferred with a phylogenetic diagnosis of hagfish and lamprey cDNA sequences. *Zoolog Sci* 23(12):1053–1064.
24. Kuraku S, et al. (2008) Noncanonical role of Hox14 revealed by its expression patterns in lamprey and shark. *Proc Natl Acad Sci USA* 105(18):6679–6683.
25. Amemiya CT, et al. (2008) The amphioxus Hox cluster: Characterization, comparative genomics, and evolution. *J Exp Zoolog B Mol Dev Evol* 310(5):465–477.
26. Ravi V, et al. (2009) Elephant shark (Callorhinchus milii) provides insights into the evolution of Hox gene clusters in gnathostomes. *Proc Natl Acad Sci USA* 106(38):16327–16332.
27. Thisse B, Thisse C (2005) High throughput expression analysis of ZF-models consortium clones. ZFIN Direct Data Submission (http://zfin.org).
28. Davis AP, Capecchi MR (1996) A mutational analysis of the 5′ HoxD genes: Dissection of genetic interactions during limb development in the mouse. *Development* 122(4): 1175–1185.
29. Kondo T, Dollé P, Zákány J, Duboule D (1996) Function of posterior HoxD genes in the morphogenesis of the anal sphincter. *Development* 122(9):2651–2659.
30. Pennacchio LA, et al. (2006) In vivo enhancer analysis of human conserved non-coding sequences. *Nature* 444(7118):499–502.
31. Woolfe A, et al. (2005) Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol* 3(1):e7.
32. Brudno M, et al.; NISC Comparative Sequencing Program (2003) LAGAN and Multi-LAGAN: Efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res* 13(4):721–731.
33. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I (2004) VISTA: Computational tools for comparative genomics. *Nucleic Acids Res* 32(Web Server issue):W273-279.
34. Thisse B, Thisse C (2004) Fast release clones: A high throughput expression analysis. ZFIN Direct Data Submission (http://zfin.org*)*.
35. Crow KD, Smith CD, Cheng JF, Wagner GP, Amemiya CT (2012) An independent genome duplication inferred from Hox paralogs in the American paddlefish—a representative basal ray-finned fish and important comparative reference. *Genome Biol Evol* 4(9):937–953.
36. Kuraku S, Meyer A, Kuratani S (2009) Timing of genome duplications relative to the origin of the vertebrates: Did cyclostomes diverge before or after? *Mol Biol Evol* 26(1): 47–59.
37. Verde I, et al.; International Peach Genome Initiative (2013) The high-quality draft genome of peach (Prunus persica) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nat Genet* 45(5):487–494.
38. Sievers F, et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7:539.
39. Hubisz MJ, Pollard KS, Siepel A (2011) PHAST and RPHAST: Phylogenetic analysis with space/time models. *Brief Bioinform* 12(1):41–51.