Development 135, 3543-3553 (2008) doi:10.1242/dev.025817

Functional resolution of duplicated *hoxb5* genes in teleosts

Olga Jarinova[1,2], Gary Hatch[1], Luc Poitras[1], Christelle Prudhomme[3], Magdalena Grzyb[1], Josée Aubin[3], Félix-Antoine Bérubé-Simard[3], Lucie Jeannotte[3] and Marc Ekker[1,2,*]

The duplication-degeneration-complementation (DDC) model predicts that subfunctionalization of duplicated genes is a common mechanism for their preservation. The additional Hox complexes of teleost fish constitute a good system in which to test this hypothesis. Zebrafish have two *hoxb* complexes, with two *hoxb5* genes, *hoxb5a* and *hoxb5b*, the expression patterns of which suggest subfunctionalization of an ancestral *hoxb5* gene. We characterized conserved non-coding elements (CNEs) near the zebrafish *hoxb5* genes. One CNE, J3, is only retained in the *hoxb5a* locus, whereas the others, J1 and J2, are present in both *hoxb5* loci. When tested individually, the enhancer activity of individual CNEs, including J3, extensively overlapped and did not support a role in subfunctionalization. By contrast, reporter transgene constructs encompassing multiple CNEs were able to target reporter gene expression to unique domains of *hoxb5a* and *hoxb5b* expression. The deletion of J3 from the *hoxb5a* locus resulted in expression that approached that of *hoxb5b*, whereas its insertion in the *hoxb5b* locus increased reporter expression and rendered it more similar to that of *hoxb5a*. Our results highlight the importance of interactions between CNEs in the execution of complementary subfunctions of duplicated genes.

KEY WORDS: CNS development, Gene duplication, Mouse embryogenesis, Transcriptional regulation, Zebrafish embryogenesis, cis-acting regulatory elements

INTRODUCTION

Duplication events, which can affect anything from single genes to the whole genome, are thought to contribute to adaptive evolution and promote increased animal complexity (Ohno, 1970). Duplicated genes can be lost by accumulating deleterious mutations (nonfunctionalization), or they may acquire a new adaptive function (neofunctionalization) or divide an ancestral function (subfunctionalization). The duplication-degeneration-complementation (DDC) model predicts that subfunctionalization is the most common mechanism for the preservation of duplicated genes (Force et al., 1999). According to the DDC model, functional modules of duplicated genes are likely to undergo complementary degeneration. The division of a single functional unit into separate subunits makes the complete loss of the original function through nucleotide substitutions, deletions or other types of DNA alterations, an unlikely event. Moreover, the subfunctionalized duplicates are under relaxed evolutionary constraints compared with the ancestral gene. Therefore, subfunctionalization of the duplicates increases an organism's chances of survival and, in rare cases, may be a transitional state leading to neofunctionalization (Lynch and Force, 2000).

The Hox gene family underwent significant expansion during evolution and is ideally suited to address the question of how genes acquire new functions. Hox genes are involved in the control of regional identity of the embryonic axes of metazoans. They are organized in clusters with the most anteriorly expressed genes located at the 3′ end of the cluster and the posteriorly expressed genes at the 5′ end. All vertebrates contain several Hox clusters that are thought to have resulted from the sequential duplications of a proHox cluster early in metazoan evolution. Thus, mammals have four Hox gene clusters (*HoxA*, *HoxB*, *HoxC* and *HoxD*), whereas zebrafish possess seven (*hoxaa*, *hoxab*, *hoxba*, *hoxbb*, *hoxca*, *hoxcb* and *hoxd*). The additional clusters in zebrafish are likely to have resulted from a duplication event that happened after the divergence of the fish and tetrapod lineages, somewhere between 300 and 450 million years ago (Amores et al., 1998; Taylor et al., 2001). The presence of additional Hox clusters is common, but not ubiquitous, in teleosts, including those species with a small genome such as *Takifugu rubripes* and *Spheroides nephelus* (Amores et al., 2004; Chiu et al., 2004). Comparison of the structure of the teleost Hox clusters with those of mammals reveals the loss of individual genes within some of the duplicated clusters. For instance, the genome of the pufferfish, *Takifugu rubripes*, contains two *hoxb3* genes, *hoxb3a* and *hoxb3b*, whereas only *hoxb3a* is found in the zebrafish genome, suggesting the loss of one of the *hoxb3* duplicates following the divergence of zebrafish and *Takifugu* lineages (Amores et al., 2004).

There are also examples in which both paralogous genes were retained after the duplication event, as is the case for the *hoxb5* genes in both zebrafish and *Takifugu* (Amores et al., 1998; Amores et al., 2004). Bruce and colleagues (Bruce et al., 2001) determined that two zebrafish *hoxb5* genes, *hoxb5a* and *hoxb5b*, are expressed in overlapping yet distinct domains in the notochord, neural tube and somites. Thus, their combined expression domains are strikingly similar to that of the single *Hoxb5* gene in the mouse. Furthermore, the same study reported the biochemical equivalence of the Hoxb5a and Hoxb5b proteins (Bruce et al., 2001). Combined, these results suggest that the *hoxb5a* and *hoxb5b* genes underwent subfunctionalization through loss of cis-acting regulatory elements. To determine whether the subfunctionalization of *Hoxb5* genes is reflected in changes in non-coding regions, we have compared the *Hoxb5* loci of human, mouse, zebrafish and *Takifugu*. This allowed us to identify blocks of highly conserved non-coding elements (CNEs). We compared the regulatory activity of the CNEs in transgenic assays in which they were tested collectively and individually.

[1]Center for Advanced Research in Environmental Genomics, Department of Biology, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada. [2]Department of Cellular and Molecular Medicine, University of Ottawa, Ottawa, Ontario K1H 8M5, Canada. [3]Centre de recherche en cancérologie de l'Université Laval, Centre Hospitalier Universitaire de Québec, L'Hôtel-Dieu de Québec, Québec G1R 2J6, Canada.

*Author for correspondence (e-mail: mekker@uottawa.ca)

Accepted 5 September 2008

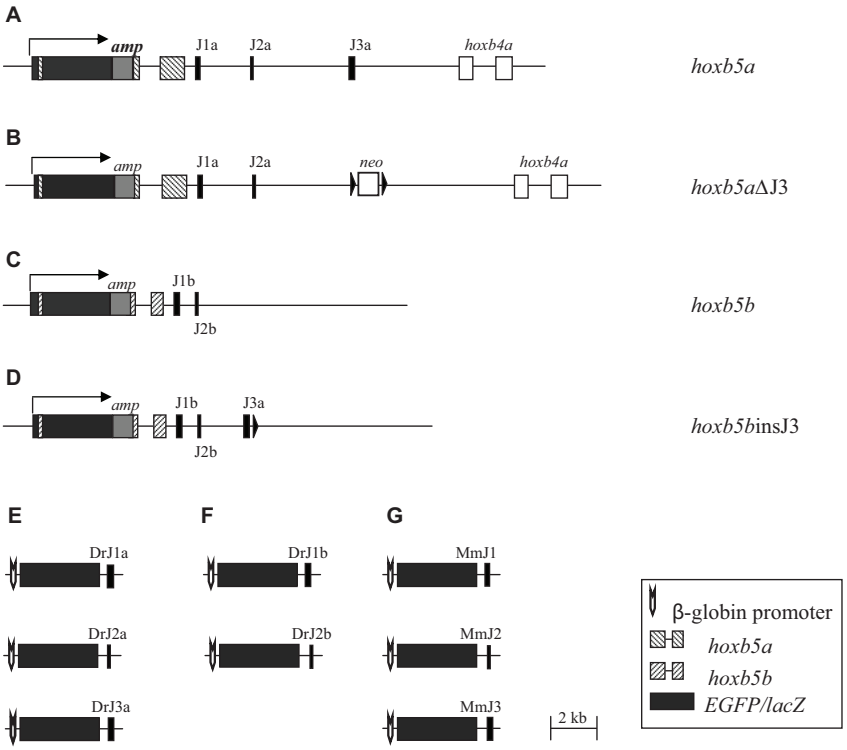

Fig. 1. Constructs used to generate transgenic animals. (**A-D**) The reporter transgene was inserted into the first exon of the zebrafish *hoxb5a* (A) or *hoxb5b* (C) gene. (B) The J3 element was deleted from the *hoxb5a* construct. (D) J3 was inserted into the *hoxb5b* construct. (**E-G**) Individual conserved non-coding elements (CNEs) from mouse and zebrafish *Hoxb5* loci were coupled to a β-globin minimal promoter and the *EGFP* or *lacZ* reporter genes (black rectangles). a and b represent the *hoxba* and *hoxbb* zebrafish complexes, respectively. In A-D, hatched rectangles indicate exon sequences; black triangles, loxP sites; *amp*, ampicillin resistance gene; *neo*, neomycin resistance gene. Mm, *Mus musculus*; Dr, *Danio rerio*.

This led to the finding that one CNE specific to the *hoxb5a* locus, named J3, accounts for the differences in expression between the paralogs, but that interactions between J3 and other CNEs are essential to achieve correct expression. This has important implications for the experimental approaches chosen to determine the patterns of regulatory evolution of duplicated genes.

MATERIALS AND METHODS

Sequence alignments and phylogenetic analysis

Sequences surrounding the *Hoxb5* genes of *Danio rerio*, *Takifugu rubripes*, *Homo sapiens* and *Mus musculus* were obtained from Ensembl Genome Data Resources (http://www.sanger.ac.uk). *Danio rerio hoxba*, contig 11157.4 (zv.3), position 40334-67614; *Danio rerio hoxbb*, contig BX001014.6 (zv.4), position 25612-57312; *Takifugu rubripes hoxba*, contig Scaffold_706, position 25001-50001; *Takifugu rubripes hoxbb*, contig Scaffold_1245, position 50965-70027; *Mus musculus Hoxb*, contig SuperContig NT_165773, position 96162969-96184091; *Homo sapiens HOXB* GenBank accession AC103702.3.1.187386, position 84742-109241.

Multiple sequence alignments were computed with MultiPipMaker (http://bio.cse.psu.edu/pipmaker) or with the Pretty Results program of the GCG Wisconsin package. The boundary of each CNE was determined as the longest region of high sequence similarity between mammalian and teleost sequences. Phylogenetic and molecular evolutionary analyses were conducted with the Kimura two-parameter model using the neighbor-joining method and bootstraping (Kumar et al., 2001). Searches for potential transcription factor binding sites in the CNEs were performed with Patch_Search (http://www.gene-regulation.com/pub/programs.html).

Constructs

Reporter constructs encompassing large regions of the zebrafish *hoxb5a* and *hoxb5b* loci were generated by homologous recombination in bacteria (Copeland et al., 2001). Recombination was carried out in *Escherichia coli* DY380 and EL350 strains provided by Dr Neal Copeland (US National Cancer Institute, Frederick, MD).

PACs containing the *hoxb5a* and *hoxb5b* zebrafish loci were obtained from RZPD (plate positions 254O17 and 227H9, respectively). For *hoxb5a*, a PAC fragment spanning from 1353 bp upstream of the *hoxb5a* initiation codon to 1316 bp downstream of the *hoxb4a* termination codon (total size 18,495 bp), was first inserted into a plasmid vector using homology arms of ~200 bp. Similarly, we created a plasmid containing a 12,169 bp fragment from the *hoxb5b* locus that spans from 1154 bp upstream of the *hoxb5b* initiation codon to 9763 bp downstream of the *hoxb5b* termination codon. We inserted the *lacZ* and *EGFP* reporter genes, encoding β-galactosidase and enhanced green fluorescent proteins, respectively, 33 bp downstream of either the *hoxb5a* or the *hoxb5b* initiation codon, in frame, to generate the *hoxb5alacZ*, *hoxb5aEGFP*, *hoxb5blacZ* and *hoxb5bEGFP* constructs (Fig. 1A,C). Constructs containing *lacZ* and *EGFP* were used for the production of transgenic mice and zebrafish, respectively.

Additional reporter constructs derived from *hoxb5alacZ* and *hoxb5aEGFP*, from which the J3 enhancer sequence was deleted, were produced through an additional round of homologous recombination. The homology arms in the targeting cassette corresponded to sequences located directly upstream and downstream of J3. The deletion constructs are referred to as *hoxb5aΔJ3lacZ* and *hoxb5aΔJ3EGFP* (Fig. 1B).

Finally, the J3 sequence from *hoxb5a* was introduced into the *hoxb5blacZ* and *hoxb5bEGFP* plasmids to generate *hoxb5b*insJ3*lacZ* and *hoxb5b*insJ3*EGFP*. The J3 sequence was inserted at a position that would correspond to that of J3 if this enhancer existed in the *hoxb5b* locus (Fig. 1D). Although the *hoxb5b* locus appears to have diminished in size after duplication, elements existing in both zebrafish loci (J1, J2 and *mir10a*) are present in the same order and at the same relative positions. By extrapolation, if present, J3 would be located ~3.1 kb downstream of the *hoxb5b* termination codon. To excise the *neo* gene adjacent to the introduced J3 element, EL350 cells hosting *hoxb5b*insJ3*lacZ* or *hoxb5b*insJ3*EGFP* were grown in arabinose-containing medium to induce recombination between loxP sites flanking the *neo* gene.

To produce transgene constructs containing individual CNEs, these were PCR-amplified and inserted downstream of the reporter gene in a way that mimics their position and orientation in their original genomic context (Fig. 1E-G). The *p1229* and *βpEGFP* vectors, which contain a human β-globin minimal promoter coupled to either *lacZ* or *EGFP*, respectively, were used for the production of transgenic mice and zebrafish (Cormack et al., 1996; Yee and Rigby, 1993; Zerucha et al., 2000).

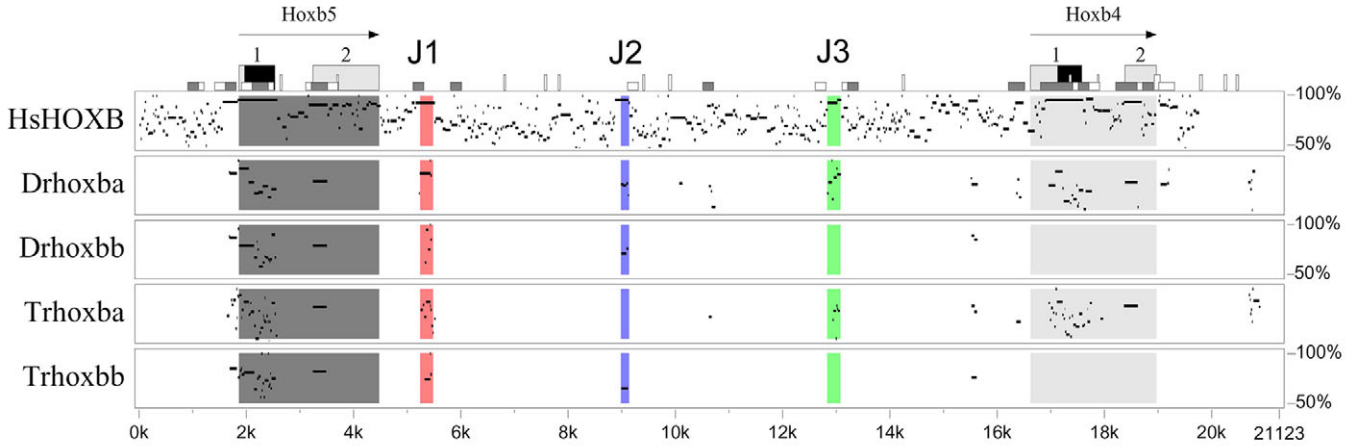

Fig. 2. Conserved non-coding elements in the *Hoxb5* genes of vertebrates. Sequences from mouse, human, zebrafish and *Takifugu* were compared using Pipmaker. The mouse *Hoxb5* locus is used as a reference and is shown on the horizontal axis. Sequences from human (Hs), zebrafish (Dr) and *Takifugu rubripes* (Tr) are compared and the similarity is shown on the vertical axis. CNE positions are indicated by shading: red, J1; blue, J2; green, J3. A sequence similarity lower than 50% was observed between the J2 elements from *Takifugu hoxb5a* and mouse *Hoxb5* using a different algorithm. Identical results to those shown were obtained when teleost sequences were used as a reference. The annotations at the top refer to the mouse reference sequence: exons (black boxes), simple repeats (vertical rectangles) and CpG islands (white and gray rectangles).

Production and genotyping of transgenic animals

All experiments were performed according to the guidelines of the Canadian Council on Animal Care and were approved by the institutional animal care committees. Transgene preparation and microinjection into fertilized eggs were according to standard procedures (Hogan, 1986). Genotyping of animals and analysis of *lacZ* expression were as previously described (Fraidenraich et al., 1998; Larochelle et al., 1999). Production of transgenic zebrafish and monitoring of EGFP expression were as described (Amsterdam et al., 1995). *EGFP* reporter constructs were microinjected at 100 ng/μl into more than 100 zebrafish embryos and those showing at least two *EGFP*-expressing cells were referred to as primary transgenic embryos.

In situ hybridization

In situ hybridization was performed as described on whole-mount preparations (Thisse et al., 1993) or on paraffin sections (Jaffe et al., 1990).

Quantitative RT-PCR

One hundred transient-transgenic zebrafish embryos were generated for each of the constructs and were randomly divided in three pools of 33 embryos. Total RNA was extracted from each pool using the RNAeasy Kit (Qiagen) and cDNA synthesis was carried out using oligo(dT) primers. Relative amounts of *EGFP*, *hoxb5a* and *hoxb5b* transcripts were estimated by PCR amplification of ~150 bp regions of each gene using the QuantiTect SYBR Green PCR Kit (Qiagen) with standard curves inferred from five different concentrations of a *hoxb5aEGFP* cDNA standard. Relative *EGFP* transcript numbers were quantified for each group and normalized to endogenous *hoxb5a* expression.

RESULTS

Identification of putative regulatory elements in the non-coding regions of paralogous group 5 Hox genes

In order to determine how the regulatory mechanisms that control the expression of the two teleost *hoxb5* genes can be compared with that of the single *Hoxb5* gene of mammals, we carried out a phylogenetic footprinting analysis of the sequences flanking the *hoxb5* transcription units from two mammalian species, mouse and human, and from the two teleosts, zebrafish and *Takifugu*. Using this approach, we identified three conserved sequences, named J1, J2 and J3, located in the 3′ flanking region of *Hoxb5* loci (Fig. 2 and see Fig. S1 in the supplementary material; GenBank accession numbers BK006730-BK006737). The length of these sequences varied from ~120 to ~260 bp. J1 and J2 sequences are both present in all the *Hoxb5* loci we examined. In mouse and human, the J1 and J2 sequences are 1.7 kb and 5.5 kb downstream of the *Hoxb5* stop codon, respectively (Fig. 2).

The sequence identity varied from 56 to 99% for the J1 element, depending on the species compared (see Table S1 in the supplementary material). Identity for J2 sequences varied between 46 and 90% in pairwise comparisons. The J2 sequences from *Takifugu* were more divergent and not initially detected by the Pipmaker algorithm (Fig. 2 and see Table S1 in the supplementary material). The third conserved sequence, J3, was found further downstream of the *Hoxb5* genes of mammals and teleosts, but was absent from the teleost *hoxb5b* loci (Fig. 2). This element showed identity that varied from 47 to 91% in pairwise comparisons.

We identified two additional conserved non-coding sequences in the *Hoxb5* loci. The first, located between J2 and J3, is found in all the mammalian and teleost species examined. It is found in *hoxb5a*, but not in *hoxb5b* loci (Fig. 2). This conserved sequence is smaller than the others (<50 bp). Another conserved region is located downstream of J3 (Fig. 2) and represents a previously identified microRNA gene (*mir10a*) that codes for a 22 nt non-coding RNA molecule that is suggested to be involved in the regulation of Hox gene expression (Mansfield et al., 2004).

Finally, the immediate 5′ flanking region of *Hoxb5* genes was well conserved in all species examined (69 to 94%, Fig. 2; see Table S1 in the supplementary material).

The conserved non-coding elements found in the 3′ flanking region of *Hoxb5* fall within three large fragments (3-4.5 kb) of the mouse *Hoxb5/Hoxb4* intergenic region previously shown to have regulatory activity that recapitulates aspects of endogenous *Hoxb5* and *Hoxb4* expression (Sharpe et al., 1998). The zebrafish J1, J2 and J3 CNEs also correspond to sequences reported by Hadrys and colleagues in their comparative analysis of the *Hoxb* clusters in mammals and teleosts (Hadrys et al., 2004).

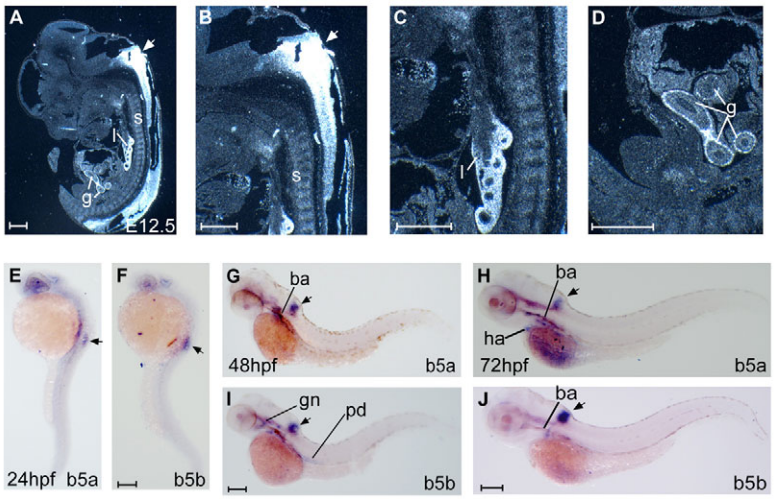

Fig. 3. Embryonic expression of mouse and zebrafish *Hoxb5* genes. (**A-D**) Sagittal sections of E12.5 mouse embryos. The antisense *Hoxb5* probe consisted of a 430 bp fragment from the *Hoxb5* cDNA that includes the N-terminal part of the Hoxb5 protein (Krumlauf et al., 1987). (**E-J**) Whole-mount in situ hybridization on (E,F) 24 hpf, (G,I) 48 hpf and (H,J) 72 hpf zebrafish embryos using antisense riboprobes for *hoxb5a* (b5a, 550 bp probe) and *hoxb5b* (b5b, 600 bp probe). Zebrafish embryos are shown as lateral views. Arrows point to the embryonic neural tube. ba, branchial arches; g, gut; gn, trigeminal ganglion; l, lung; pd, pronephric duct; s, somites. Scale bars: 500 μm in A-D; 250 μm in E-J.

Transgenic analysis of cis-regulatory sequences found near the zebrafish *hoxb5* loci

We analyzed the activity of non-coding elements found near the zebrafish *hoxb5* loci by inserting the *lacZ* or *EGFP* reporter genes in frame with the first exon of either *hoxb5a* or *hoxb5b* (see Materials and methods). The constructs encompassed 18.5 kb and 12.2 kb of the respective loci (Fig. 1A,C) and included ~1.2-1.3 kb of 5′ flanking sequences. In the 3′ direction, the *hoxb5a* construct spanned ~12 kb and encompassed the J1-J3 conserved non-coding sequences. It also contained the *hoxb4a* exons and intron, as well as 1.3 kb of *hoxb4a* 3′ flanking sequence. Similarly, the *hoxb5b* construct extended over a distance of ~8 kb downstream of the last exon and encompassed the J1 and J2 conserved non-coding sequences.

Expression of the *hoxb5alacZ* and *hoxb5blacZ* transgenes was examined in mouse embryos at E10.5, E12.5 and E13.5. Transgenic mouse embryos, either primary or from established lines, showed *lacZ* expression in regions that coincide with domains of mouse *Hoxb5* expression (Fig. 3A-D, Fig. 4I-L) (Hogan et al., 1988). At E10.5, *hoxb5alacZ* embryos exhibited β-galactosidase activity along the embryonic neural tube, with the anterior border of reporter gene expression reaching the caudal hindbrain (Fig. 4A,B, arrows). By E12.5, the anterior boundary of *lacZ* expression shifted rostrally, extending to the caudal limit of the otic vesicle (Fig. 4I,J, arrows). This rostral extension of the expression domain might be mediated by a retinoic acid response element (RARE), previously characterized by Sharpe and colleagues (Sharpe et al., 1998), that resides within the J2 CNE. Furthermore, the *hoxb5alacZ* construct consistently targeted reporter gene expression to the paraxial mesoderm/anterior somites at all developmental stages examined (Fig. 4A,B,I,J, asterisks). Analysis of the *hoxb5alacZ* expression patterns on sagittal sections of E13.5 embryos revealed β-galactosidase activity in the cartilage primordia of embryonic ribs (Fig. 5A,B, cp).

At E10.5, the anterior boundary of *hoxb5blacZ* reporter expression in the central nervous system (CNS) was located at the caudal limit of the embryonic forelimbs, that is, more posteriorly than the CNS expression boundary observed for *hoxb5alacZ* (Fig. 4A-D, arrows). By E12.5, the rostral boundary of *hoxb5blacZ* reporter gene expression in the CNS extended up to the caudal limit of the otic vesicle, similar to that detected in the *hoxb5alacZ* transgenic embryos (Fig. 4J,L, arrows). In addition to the CNS, *hoxb5blacZ* was expressed in neural crest derivatives such as cranial and dorsal root ganglia and associated nerve fibers. At E10.5 *hoxb5blacZ* expression was detected in the nodose, tenth cranial nerve, facio-acoustic and dorsal root ganglia (Fig. 4C, ng, fg). Expression persisted in these domains until at least E13.5 (Fig. 5D,E).

Dissection of internal organs from *hoxb5alacZ* mouse embryos at E13.5 revealed β-galactosidase activity in the midgut in a punctuate distribution (Fig. 5C,I). This might reflect a dot-like pattern of *Hoxb5* expression in the enteric nervous system similar to that reported for the *Hoxa5* gene (Larochelle et al., 1999). In addition, *hoxb5alacZ* was expressed in the stomach, meta- and mesonephros and adrenal gland (Fig. 5I, st, k, ag). A β-galactosidase signal of lower intensity was detected in the stomach, midgut and adrenal gland in *hoxb5blacZ* embryos (Fig. 5F,J).

Expression of *hoxb5alacZ* was stable and consistent between embryos from established lines at all embryonic stages examined (Fig. 4 and data not shown). By contrast, embryos obtained from *hoxb5blacZ* males showed patterns of transgene expression that differed between embryos collected from the same mating (see Fig. S2 in the supplementary material). Only embryos with the strongest X-Gal staining are shown in Figs 4, 5. This phenomenon was observed until at least the F2 generation obtained from an outcross of transgenic F1 males to wild-type females.

Similar constructs containing *EGFP* as the reporter were tested in primary transgenic zebrafish embryos. The *hoxb5aEGFP* and *hoxb5bEGFP* constructs recapitulated aspects of the endogenous *hoxb5a* and *hoxb5b* expression. In most cases, the anterior border of reporter gene expression corresponded to the posterior hindbrain, reflecting the rostral restriction of endogenous *hoxb5* gene expression detected by in situ hybridization (Fig. 3E-J, Fig. 6A-I).

At 24-36 hours post-fertilization (hpf), *hoxb5aEGFP* expression was observed in cells of the developing CNS (Fig. 6A, arrows) and somites (Fig. 6A, arrowheads), mirroring *hoxb5a* expression (Fig. 3 and data not shown). At later stages, strong reporter expression persisted in the developing CNS and was also observed in muscle cells along the anterior-posterior axis (Fig. 6B,C). At 48 hpf, 10% of primary *hoxb5aEGFP* embryos expressed the transgene in the heart and the cardial veins (data not shown).

In zebrafish, the *hoxb5b* sequences directed reporter gene expression preferentially to cells of neuronal origin at all developmental stages examined (Fig. 6D,E), similar to what

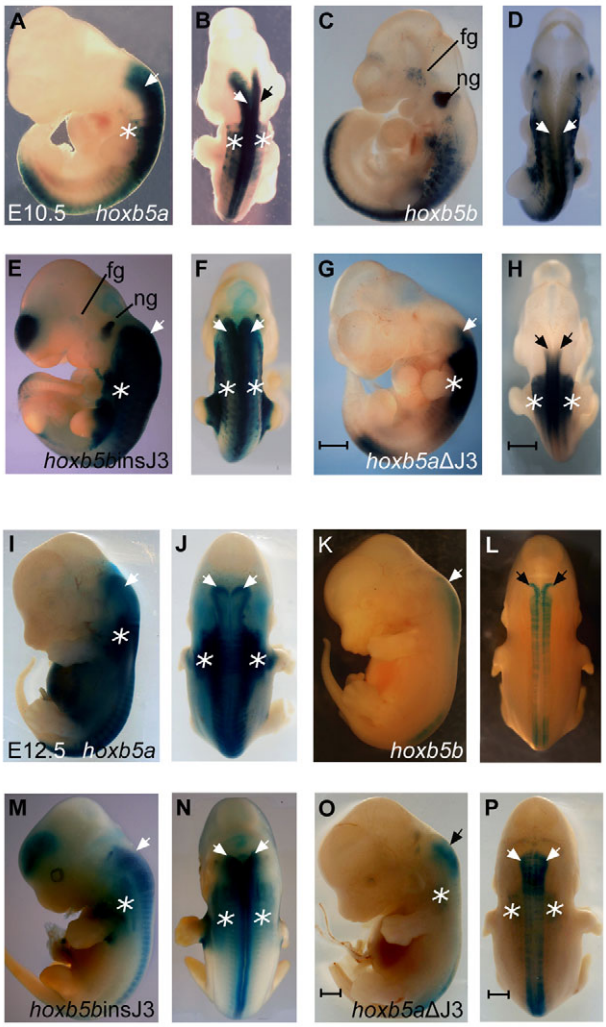

Fig. 4. Expression of *lacZ* reporter transgenes in mouse embryos. E10.5 (**A-H**) and E12.5 (**I-P**) transgenic mouse embryos, as indicated. A,C,E,G,I,K,M,O are lateral views; B,D,F,H,J,L,N,P are dorsal views. Arrows indicate the neural tube and asterisks indicate somites. fg, facioacoustic ganglion; ng, nodose ganglion. Scale bars: 500 μm.

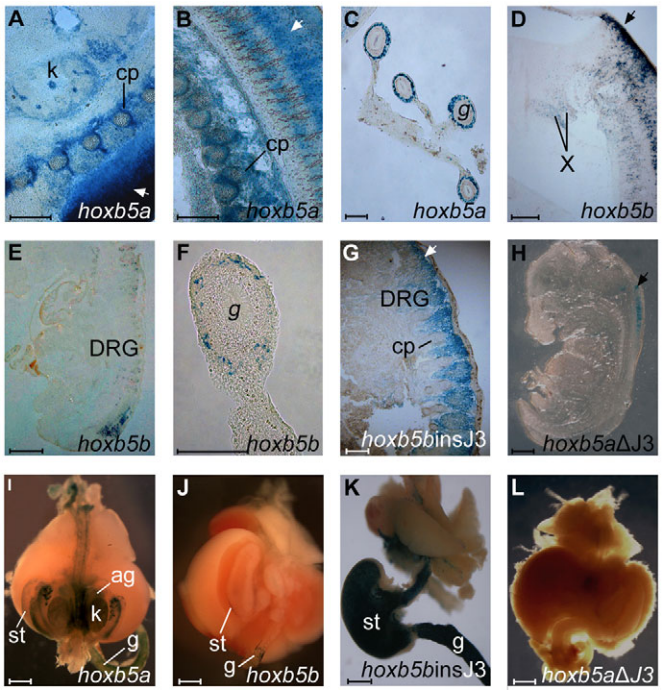

Fig. 5. Analysis of reporter transgene expression on sagittal sections and in the internal organs of mouse embryos. (**A-H**) Sagittal sections from E13.5 transgenic mouse embryos, as indicated. Arrows indicate the neural tube. (**I-L**) Internal organs from E13.5 transgenic embryos, as indicated. cp, cartilage primordium; DRG, dorsal root ganglia; g, gut; k, kidney; st, stomach; ag, adrenal glands; X, trigeminal ganglia. Scale bars: 330 μm in A-D,F,G; 1000 μm in E,H; 125 μm in I,J,L; 250 μm in K.

was found in transgenic mouse embryos. This contrasted with the mesodermal activity recorded for the paralogous *hoxb5a* locus (Fig. 6A-C, arrowheads). Occasionally, *hoxb5bEGFP* expression was observed in cells of hematopoietic origin (data not shown).

Consistent with data obtained in transgenic mice, the levels of *hoxb5aEGFP* expression generally appeared to be higher than those seen in *hoxb5bEGFP* embryos, based on visual inspection and on the total number of cells expressing the transgenes. Furthermore, the *hoxb5aEGFP* construct yielded a much larger proportion of *EGFP*-positive embryos (defined as embryos containing two or more *EGFP*-expressing cells) than the *hoxb5bEGFP* construct (60% as compared with 30%). To further address these differences, we performed real-time quantitative RT-PCR on RNA extracted from 100 primary transgenic embryos for each construct. The relative levels of *EGFP* transcripts in each group were estimated and normalized to those of endogenous *hoxb5a*. As shown in Fig. 6J, *hoxb5aEGFP* embryos had 104-fold higher *EGFP* transcript levels than *hoxb5bEGFP* embryos.

Combined, the results obtained from the mouse and zebrafish transgenesis experiments demonstrate that the DNA sequences included in the constructs are able to recapitulate most of the *hoxb5a* and *hoxb5b* expression domains except for the notochord and branchial arches. Furthermore, the differences in expression of the *hoxb5a* and *hoxb5b* constructs appear to be associated with those aspects of expression that differ between paralogs. For example, one of the most striking differences observed between *hoxb5a* and *hoxb5b* expression is the exclusive expression of *hoxb5b* in structures of the peripheral nervous system, such as the nerve ganglia, and this was mimicked by *hoxb5blacZ* but not by *hoxb5alacZ* (Fig. 4C, Fig. 5D,E). Similarly, the high expression levels of *hoxb5a* in somites and their derivatives (Bruce et al., 2001) were recapitulated by *hoxb5aEGFP* (Fig. 4B,J, asterisks; Fig. 5A,B, cp; Fig. 6A,C, arrowheads).

As the J3 element is only present in *hoxb5a*, and presumably degenerated in *hoxb5b*, we hypothesized that it might be involved in directing unique aspects of *hoxb5a* expression. To test this, we deleted J3 from the *hoxb5alacZ* and *hoxb5aEGFP* reporter constructs, yielding *hoxb5aΔJ3lacZ* and *hoxb5aΔJ3EGFP*. Conversely, we also inserted J3 into the *hoxb5blacZ* and *hoxb5bEGFP* constructs to examine whether this would render *hoxb5b* expression more similar to that of *hoxb5a*; the resulting constructs were named *hoxb5b*insJ3*lacZ* and *hoxb5b*insJ3*EGFP*.

In mouse, the *hoxb5aΔJ3lacZ* construct targeted *lacZ* expression to several domains of *Hoxb5* expression where its counterpart, *hoxb5alacZ*, was also expressed. However,

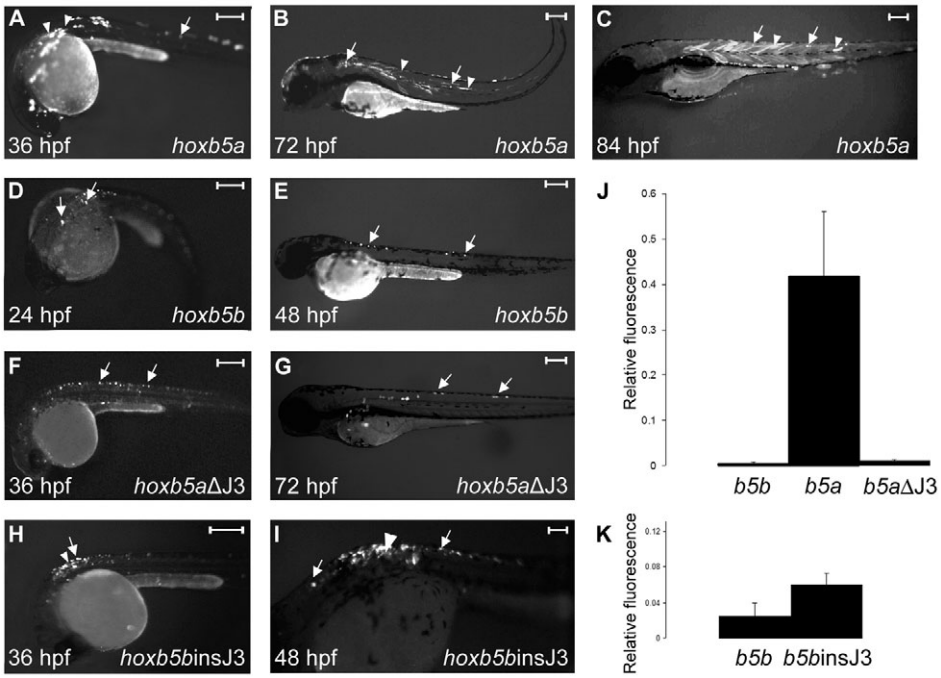

Fig. 6. Expression of *EGFP* reporter transgenes in zebrafish embryos. (**A-I**) Lateral views (anterior to the left) of primary transgenic zebrafish embryos injected with (A-C) *hoxb5aEGFP*, (D,E) *hoxb5bEGFP*, (F,G) *hoxb5a*Δ*J3EGFP* or (H,I) *hoxb5b*insJ*3EGFP*. Developmental stages are indicated. Arrows indicate cells of the developing nervous system; arrowheads mark somites and muscle cells. Approximately 60% of the embryos injected with *hoxb5bEGFP* and 50% of embryos injected with *hoxb5b*insJ*3EGFP* showed *EGFP* expression, as compared with only 30% of *hoxb5bEGFP* and 37.5% of *hoxb5a*Δ*J3EGFP* embryos. Scale bars: 250 μm. (**J**,**K**) Relative copy numbers of *EGFP* transcripts detected by RT-PCR in primary transgenic embryos injected with *hoxb5aEGFP* (*b5a*), *hoxb5bEGFP* (*b5b*), *hoxb5a*Δ*J3EGFP* (*b5a*Δ*J3*) or *hoxb5b*insJ*3EGFP* (*b5b*insJ*3*) constructs. Column heights represent relative copy numbers of *EGFP* transcripts after normalization to endogenous *hoxb5a* expression. Each bar represents the average of three groups of transgenic embryos and error bars indicate s.e.m.

expression levels were markedly diminished. Thus, at E10.5, *hoxb5a*Δ*J3lacZ* was expressed in the spinal cord, with anterior limits of expression reaching the base of the hindbrain (Fig. 4G,H, arrows), but the expression was reduced compared with that observed in *hoxb5alacZ* embryos (Fig. 4A,B; data not shown). Similar to *hoxb5alacZ*, the *hoxb5a*Δ*J3lacZ* construct was expressed in somites and their derivatives at E10.5 and E12.5 (Fig. 4G,H,O,P, asterisks).

Similar observations were made in transgenic zebrafish embryos. Thus, the *hoxb5a*Δ*J3EGFP* construct (Fig. 6F,G) exhibited both neuronal and mesodermal expression similar to that recorded for its intact counterpart, *hoxb5aEGFP* (Fig. 6B; data not shown). However, considerably fewer *EGFP*-expressing cells were observed. Furthermore, transgene expression appeared significantly diminished in cells of mesodermal origin as compared with that observed in *hoxb5aEGFP* zebrafish embryos (data not shown). When quantified by RT-PCR, transcript levels for *hoxb5a*Δ*J3EGFP* resembled those seen in *hoxb5bEGFP* transgenic embryos, rather than those characteristic of *hoxb5aEGFP* embryos (Fig. 6J).

The rostral boundary of *hoxb5b*insJ3*lacZ* expression in the E10.5 mouse neural tube was shifted anteriorly compared with that of *hoxb5blacZ* (Fig. 4E,F). Furthermore, high levels of *hoxb5b*insJ3*lacZ* expression were detected in the gut and stomach (Fig. 5K, st, g), similar to *hoxb5alacZ*. Finally, the *hoxb5b*insJ3*lacZ* construct drove reporter gene expression in somites and their derivatives (Fig. 4E,F,M,N, asterisks), whereas the *hoxb5blacZ* did not (Fig. 4C,D,K,L).

Levels of the *hoxb5b*insJ3*lacZ* reporter were apparently higher than those driven by the *hoxb5blacZ* construct (Fig. 4K-N). Moreover, the *hoxb5b*insJ3*lacZ* construct drove more consistent patterns of reporter gene expression than its parental counterpart, and the variegated patterns of expression seen with *hoxb5blacZ* were only observed in 10% of *hoxb5b*insJ3*lacZ* embryos (data not shown). When injected into zebrafish embryos, the *hoxb5*binsJ3*EGFP* construct drove reporter expression in both neural and mesodermal cells (Fig. 6I, arrows and arrowhead), whereas *hoxb5bEGFP* only exhibited neural expression (Fig. 6D,E, arrows). The *hoxb5b*insJ3*EGFP* construct also produced greater numbers of *EGFP*-positive embryos (Fig. 6H,I) and levels of *EGFP* expression were increased 2.4-fold compared with *hoxb5bEGFP* (Fig. 6K).

Taken together, the above results suggest that the J3 element is necessary to maintain higher levels of *hoxb5* gene expression and is at least partially responsible for the exclusive expression of *hoxb5a* in cells of mesodermal origin (Bruce et al., 2001). Deleting the J3 sequence from the *hoxb5a* locus reduced overall reporter transgene expression, whereas introduction of J3 into the *hoxb5b* locus resulted in higher expression levels and induced reporter expression in cells uniquely associated with *hoxb5a* expression.

Transgenic analysis of individual non-coding elements

To assess the individual contributions of the J1, J2 and J3 CNEs to overall *Hoxb5* expression, we generated reporter constructs carrying individual CNEs directing expression of *lacZ* or *EGFP* from a β-globin minimal promoter (Fig. 1E-G). Primary transgenic mouse embryos were examined at E12.5-14.5 and primary transgenic zebrafish embryos were examined at 24-96 hpf.

When tested in transgenic mouse embryos, each of the conserved sequences from the mouse *Hoxb5* locus or from the zebrafish *hoxb5a* and *hoxb5b* loci drove reporter transgene expression in domains representative of endogenous *Hoxb5* expression (Oosterveen et al., 2003; Sakach and Safaei, 1996). A transgene containing the mouse J1 element (MmJ1) directed *lacZ* expression in structures of the paraxial mesoderm (Fig. 7A) and neural tube (Table 1, data not shown). The mouse J2 element, MmJ2, induced reporter gene expression in the neural tube and in prevertebrae along the entire anterior-posterior axis (Fig. 7D, Table 1). Finally, the construct containing MmJ3 drove *lacZ* expression in the neural tube, developing vertebrae and posterior somites (Fig. 7G, Table 1).

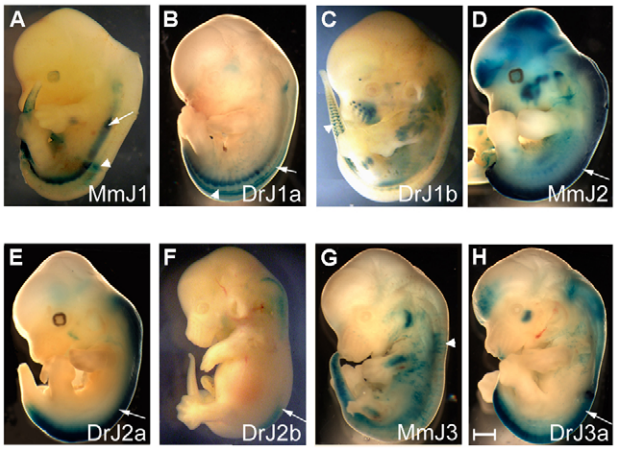

Fig. 7. Enhancer activity of individual *Hoxb5* non-coding elements in transgenic mouse embryos. (**A-H**) Reporter transgene expression was analyzed in primary transgenic embryos whose age varied from E12.5 to E14.5. The name and origin of the tested CNE are shown at the bottom right-hand corner of each panel. Arrowheads mark somite derivatives; arrows point to the neural tube. Mm, *Mus musculus*; Dr, *Danio rerio*; a and b refer to the *hoxba* and *hoxbb* zebrafish complexes, respectively. Scale bar: 1000 μm.

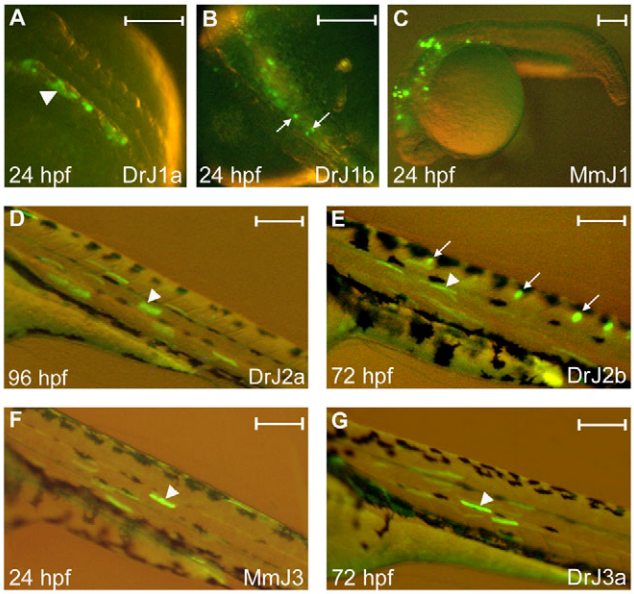

Fig. 8. Enhancer activity of individual *Hoxb5* non-coding elements in transgenic zebrafish. (**A-G**) Lateral views, anterior to the left. Arrows in B and E indicate cells of the developing CNS. Arrowheads indicate somites in A and muscle cells in D-G. The name and origin of each regulatory element are indicated as in Fig. 7. Age, in hpf, is indicated. Scale bars: 500 μm.

Individual zebrafish elements from either *hoxb5a* or from *hoxb5b* induced transgene expression in prevertebrae, in the neural tube and, occasionally, in the dorsal root ganglia and/or associated nerves (Fig. 7, Table 1). No major differences were observed between the activity of the zebrafish CNEs and their mouse counterparts, except for an inability of zebrafish J3 (DrJ3) to target expression to the somites/prevertebrae (Table 1).

When tested in transgenic zebrafish, the murine CNEs showed a spectrum of activities comparable to those observed in transgenic mice (Fig. 8, Table 1). MmJ1 targeted reporter expression to neural and mesodermal cells (Fig. 8C, Table 1; data not shown). MmJ2 drove expression almost exclusively in cells of mesodermal origin, in contrast to the strong neural activity detected in transgenic mice (Fig. 7D, Table 1). The mouse J3 sequence also targeted transgene expression almost exclusively to cells of mesodermal origin. The individual zebrafish CNEs targeted *EGFP* expression in transgenic zebrafish with patterns resembling the endogenous *hoxb5a* and *hoxb5b* genes. Activity was generally similar to that of the mouse CNEs. Exceptions included the observation that zebrafish J1A was less active in neural cells than mouse J1 or zebrafish J1B, and that zebrafish J2 sequences targeted expression to neural cells much better than did the orthologous mouse sequence. No other major differences were observed between the activity of an individual element from the *hoxb5a* locus and its paralogous sequence from the *hoxb5b* locus.

The above comparative analysis revealed that regulatory activity conferred by the cognate CNEs from duplicated zebrafish loci extensively overlapped and, in general, corresponded to expression patterns common to both *hoxb5* orthologs.

DISCUSSION

Gene duplication has long been considered a major source of evolutionary novelty. Paralogous genes resulting from duplication events have the opportunity to diverge and acquire new functions.

Table 1. Expression of reporter constructs in primary transgenic mouse and zebrafish embryos

	Mouse injections				Zebrafish injections	
		Expression sites				
Element tested	# Expressing/# Tg	Neural tube	DRG	Somites/prevertebrae	Neural	Mesodermal
MmJ1	5/5	4	2	4	20% (11/55)	100% (55/55)
DrJ1a	4/7	3	2	2	2% (1/48)	100% (48/48)
DrJ1b	3/10	2	1	2	30% (22/74)	80% (59/74)
MmJ2	3/6	3	2	3	0% (0/42)	100% (42/42)
DrJ2a	2/2	2	1	1	13% (7/54)	93% (50/54)
DrJ2b	4/8	3	1	1	71% (52/73)	67% (49/73)
MmJ3	4/8	2	1	2	2% (1/49)	100% (49/49)
DrJ3a	6/8	5	2	0	4% (2/54)	98% (53/54)

DRG, dorsal root ganglion.

Table 2. Potential transcription factor binding sites identified in the conserved *Hoxb5* non-coding sequences

Sequence (5′ to 3′)	Transcription factor	CNS	Position on Mm *Hoxb* contig (NT_165773)
TTACCT	Rarβ	5′ region of *hoxb5a*	96164708-96164713
AGAATTT	c/EBP	5′ region of *hoxb5a*	96164730-96164736
TTTACGA	Hoxa9	5′ region of *hoxb5a*	96164734-96164740
CACGTG	Smad3/c-Myc	5′ region of *hoxb5a*	96164755-96164760
CCATATTTGG	Mcm1/Srf	5′ region of *hoxb5a*	96164798-96164807
TGACATT	Hoxa9/Meis1a	J1	96168219-96168225
TCGTAAA	Hoxa9/Meis1a	J1	96168265-96168271
CACGTG	c-Myc	J1	96168318-96168323
TTTATGG	Hoxa9/Meis1a	J1	96168328-96168334
CATAAAGTG	Pax2.1	J1	96168329-96168337
TTTTATGGTTTA	Ubx	J1	96168339-96168351
TTTATGG	Hoxa9/Meis1a	J1	96168340-96168347
TAACTG	c-Myb	J1	96168408-96168413
ATGAGA	XPF (Ercc4)	J2	96172038-96172044
TGATCC	GR (Nr3c1)	J2	96172057-96172062
AGGTCA	Rarα1	J2	96172060-96172065
GGGTGA	Rxrα	J3	96175913-96175918
TGAACC	Rxrα	J3	96175916-96175921
AGGTCA	Rarα1	J3	96175924-96175929
TAAAAT	Pou1F1a	J3	96175947-96175952
AACAAAG	Lef1, Sox3, Sox5	J3	96175984-96175990
AGATTA	Gata1/4	J3	96176005-96176010

The 6-12 bp sequences listed are highly conserved in all the species examined (mouse, human, zebrafish and *Takifugu*) and coincide with known transcription factor binding sites.

Divergence in the expression patterns of duplicated genes has been proposed to be the first step in this process. The DDC model originally proposed by Force and colleagues states that subfunctionalization is a common mechanism of duplicated gene preservation (Force et al., 1999). The model makes the following predictions with regards to molecular mechanisms of subfunctionalization. (1) Subfunctionalization happens by means of changes in the non-coding regions of duplicated genes. (2) Resolution of functional redundancy occurs through complementary degeneration of individual regulatory elements. This requires the regulation of an ancestral gene to be modular in nature; the gene subfunctions have to be independent from each other and executed by discrete regulatory elements. (3) The process of subfunctionalization should be completed within 12.5 million years after a duplication event.

This model has received a lot of attention in the field of evolutionary developmental biology. The fates of several gene duplicates were investigated in an attempt to bridge the mathematical predictions of the DDC model with experimental observations. Studies of the paralogous *Nudt10* and *Nudt11* genes in mammals, the *sox9a* and *sox9b* and the *mitf-m* and *mitf-b* genes in teleost (Altschmied et al., 2002; Hua et al., 2003; Kluver et al., 2005) confirmed that complementary loss of independent subfunctions (subfunctionalization) may indeed constitute a common mechanism of resolution of functional redundancy between duplicated genes. However, these studies did not address the molecular changes that occurred at the level of the regulatory elements that instigated subfunctionalization.

To clarify how the dynamics of cis-regulatory sequence evolution support the DDC model, Santini et al. (Santini et al., 2003) and Wolfe and Elgar (Wolfe and Elgar, 2007), performed comparative sequence analyses of multiple gene loci that are duplicated in teleosts but are present at single copy in mammals. They found that paralogous genes in teleost often retain differential subsets of putative regulatory elements, consistent with the notion of regulatory subfunctionalization (Santini et al., 2003; Woolfe and Elgar, 2007).

We characterized the structure and function of regulatory elements from the subfunctionalized zebrafish *hoxb5a* and *hoxb5b* genes to determine whether the evolution of these paralogs occurred in accordance with the predictions of the DDC model. We identified conserved non-coding elements and tested their activity collectively in a context that is close to their natural environment. Changes occurring in CNEs from duplicated genes were previously shown to be associated with the differential domains of expression of co-paralogs (Kleinjan et al., 2008; Tumpel et al., 2006). In these two studies, CNEs were tested individually in reporter constructs that were used to produce transgenic animals in the endogenous species or in more-amenable model species. Although this approach can reveal changes in cis-regulatory elements that might account for some aspects of the differential expression patterns of the co-paralogs, it does have distinct limitations, such as a failure to reveal regulatory interactions as demonstrated by the results of the present study.

Execution of complementary subfunctions of the *hoxb5a* and *hoxb5b* genes may rely on interactions between multiple cis-regulatory elements

When tested in transgenic animals, large fragments from the zebrafish *hoxb5a* and *hoxb5b* loci targeted reporter gene expression to domains of endogenous *Hoxb5* expression that were consistent with the differences detected in the expression patterns of the two paralogs (Figs 3, 4, 5) (Bruce et al., 2001). For example, the rostral boundary of *hoxb5blacZ* expression in the neural tube was shifted posteriorly at E10.5 as compared with *hoxb5alacZ* embryos (Fig. 4). We then questioned if, consistent with the DDC model, complementary degenerative changes within individual elements are responsible for changes in expression. The J3 element is only retained in the *hoxb5a* locus and is lost from *hoxb5b*. Transgenic data obtained in both zebrafish and mice suggest that the loss of J3 might have contributed to divergence in expression between the zebrafish *hoxb5a* and *hoxb5b* genes (Figs 4, 6).

By contrast, functional tests of individual regulatory elements from zebrafish *hoxb5a* and *hoxb5b* did not reveal clear differences in activity (Figs 7, 8). MmJ2, DrJ2a and DrJ2b targeted transgene

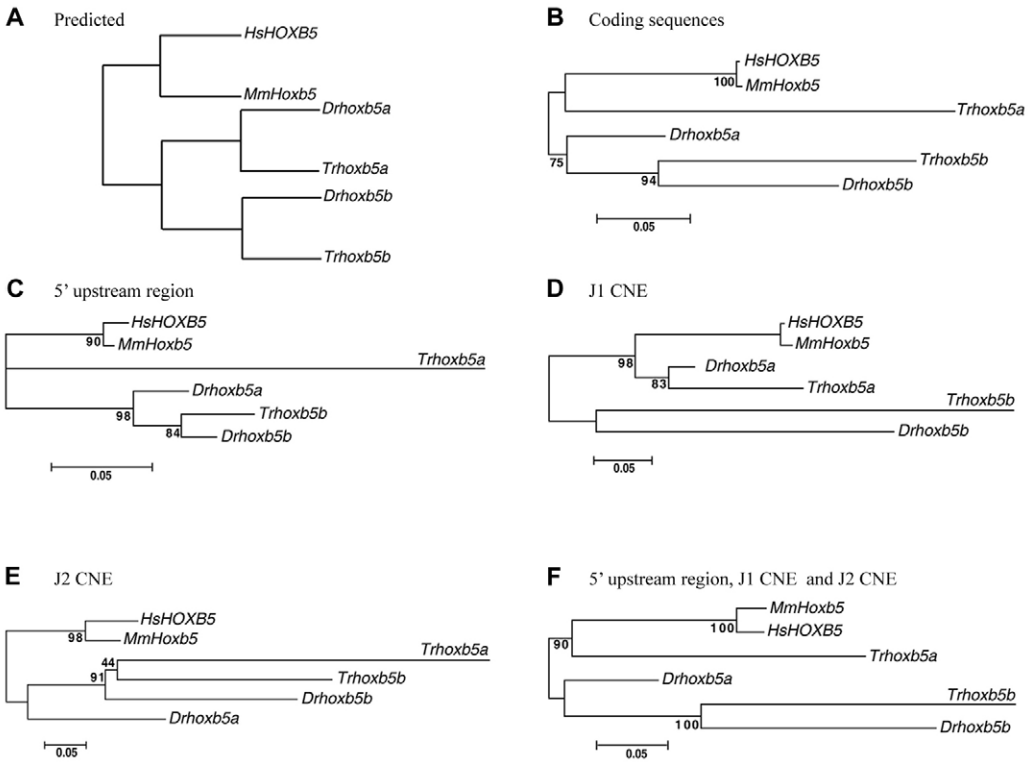

Fig. 9. Phylogenetic analysis of *Hoxb5* loci from human, mouse, zebrafish and *Takifugu*. (**A**) Phylogenetic tree predicted by the DDC model. Its topology is based on the assumption that *hoxba* and *hoxbb* teleost complexes resulted from a duplication event that happened after the separation of the fish and tetrapod lineages and before the divergence of *Takifugu* and zebrafish lineages. (**B-F**) Phylogenetic trees built with MEGA3 based on the following sequences: (B) *Hoxb5* coding sequences (850 bp); (C) 5′ upstream region of *Hoxb5* genes (130 bp); (D) J1 CNE (255 bp); (E) J2 CNE (150 bp); (F) CNEs present in all *Hoxb5* loci (535 bp), for which the 5′ upstream region, J1 and J2 were joined in a continuous single sequence for each *Hoxb5* locus. As the J3 CNE was only found in the *hoxb5a* complexes of zebrafish and *Takifugu*, it was not included in this analysis. Hs, human; Mm, mouse; Dr, zebrafish; Tr, *Takifugu*. The scale bar indicates an estimated evolutionary distance of 0.05 nucleotide substitutions per site. Numbers at nodes indicate the bootstrap values.

expression to similar domains in the mouse neural tube with a correct anterior border of expression. In addition, when tested individually, *hoxb5a* elements occasionally showed regulatory activity associated with domains exclusive to *hoxb5b*. For example, the DrJ3a and DrJ1a elements occasionally drove *lacZ* expression in the dorsal root ganglia and associated nerve fibers (Table 1, data not shown), an expression domain that correlates with zebrafish *hoxb5b* rather than *hoxb5a* expression.

Thus, contrary to the results of experiments involving large fragments from the zebrafish *hoxb5* loci, the functional analysis of individual enhancer elements did not reveal clear complementation in activities between CNEs from the duplicated zebrafish *hoxb5* genes. Therefore, we propose that some complementary *hoxb5* subfunctions rely on interactions between regulatory elements, rather than being executed by individual enhancers. Alternatively, the activity of J1, J2 and J3 might be fine-tuned by other regulatory elements that are included in the *hoxb5a* and *hoxb5b* reporter constructs but which have not been identified in this study.

The dynamics of change in the cis-acting regulatory elements of *hoxb5* duplicates are more complex than predicted by the DDC model

The DDC model predicts that regulatory elements of duplicated genes will undergo rapid complementary degeneration. Partition of an original CNE regulatory function, although not suggested by our transgenic analysis, could also be supported by phylogenetic analysis.

We aligned the J1, J2 and the immediate 5′ flanking sequences from all *Hoxb5* loci and examined the alignments for the presence of short (6-12 bp) highly conserved (>91%) sequences. Such sequences, which are evolving at an exceptionally slow rate, are likely to be functionally important. Although we identified short DNA sequences that were conserved in cognate elements of mammals and teleosts (Table 2), we were unable to find sequences that were specific for one of the two teleost paralogs (e.g. *hoxb5a* from zebrafish or *Takifugu*), but divergent or absent in the other teleost paralogs (e.g. *hoxb5b*) (data not shown). Thus, the data do not provide evidence for a simple complementary degeneration of regulatory subfunctions within the elements present in both teleost *hoxb5* loci.

We also scanned *Hoxb5* loci for putative transcription factor binding sites (Table 2). Among those that were found, some corresponded to transcription factors known to be involved in Hox gene regulation. For example, a RARE known to be essential for correct anterior expression of both *Hoxb8* and *Hoxb5* genes in the mouse neural tube (Oosterveen et al., 2003) was found within J2. Another putative RARE was found in one of the highly conserved regions of J3 (Table 2).

The DDC model also predicts that degeneration of redundant subfunctions should occur within 4.0-12.5 million years after the duplication event (Force et al., 1999). Thus, functional specialization of the *hoxb5a* and *hoxb5b* genes in teleosts would have preceded the divergence of the zebrafish and *Takifugu* lineages, and one might

expect higher sequence conservation between orthologous functional modules (i.e. *hoxb5a* from zebrafish and *hoxb5a* from *Takifugu*) than between paralogous modules (i.e. zebrafish *hoxb5a* and *hoxb5b*) (Fig. 9). We examined phylogenies based on the analysis of coding sequences of *Hoxb5* genes from human, mouse, zebrafish and *Takifugu* (Fig. 9B), as well as sequences from the 5′ flanking regions of *Hoxb5* genes (Fig. 9C) and of the J1 and J2 elements (Fig. 9D,E). The tree built for the J1 CNE matched the branching pattern of the hypothetical model (Fig. 9A,D), whereas phylogeny for other functional modules did not (Fig. 9A-C,E). We also joined the sequences of the CNEs in a continuous sequence for each locus. The topology of the resulting tree (Fig. 9F) was similar to that calculated for the coding sequences and the 5′ flanking regions of *Hoxb5* genes, suggesting that the coding region and most regulatory regions of *Hoxb5* loci are under common selective constraints or are drifting at the same rate. Overall, the dynamics of the molecular changes involved in the evolution of *hoxb5a* and *hoxb5b* teleost genes differ from those anticipated for duplicated genes that undergo subfunctionalization (Force et al., 1999). Alternatively, the DDC model might be imperfect in its timing aspects. In fact, the regulatory elements of *hoxb5a* and *hoxb5b* duplicates appear to diverge slower than anticipated.

Deviations from the predicted rate of divergence have also been reported for the duplicated teleost *sox9a* and *sox9b* genes (Cresko et al., 2003). Differences in the expression patterns of *sox9a* and *sox9b* in zebrafish and stickleback led to the conclusion that, even though partitioning of most *sox9* subfunctions occurred before the divergence of the teleost lineages, some gene subfunctions might have assorted differently in the two teleost species.

In addition to its enhancer activity, the J3 element may be required for proper maintenance of *hoxb5* expression

The absence of the J3 regulatory element in one of the two *hoxb5* loci is consistent with the DDC model. Sharpe and colleagues have previously referred to a large fragment of the mouse *Hoxb5/Hoxb4* intergenic region, encompassing J3, as the ‘mesodermal enhancer region’ (Sharpe et al., 1998). Our experiments revealed that the J3 element is not only responsible for the mesodermal *Hoxb5* expression but may also be required for maintaining high levels of *Hoxb5* expression. Indeed, reporter constructs containing J3 (*hoxb5a* and *hoxb5b*insJ3) appear to show higher expression levels than their J3-deleted counterparts (*hoxb5b* and *hoxb5aΔ*J3). This might also be associated with the intriguing observation that mouse embryos carrying *hoxb5blacZ* and *hoxb5aΔ*J3*lacZ*, the two transgenes lacking J3, showed variegated patterns of transgene expression among embryos from the same mating (see Fig. S2 in the supplementary material). Primary transgenic embryos also showed a similar variability in the transgene expression pattern (data not shown). As the number of independent integration events and the use of established lines argue against positional effects or mosaicism, we propose that one of the functions of the J3 element is to maintain gene expression. This additional function of J3 might involve interactions with proteins such as members of the Polycomb and Trithorax groups, which are known to regulate and maintain the transcriptional status of Hox genes through modification of chromatin structure.

Conclusions

Our data suggest that the patterns of regulatory evolution of teleost *hoxb5* duplicates involve mechanisms additional to those suggested by the DDC model (Force et al., 1999). Although phylogenetic filtering and functional tests of individual elements from the *hoxb5a* and *hoxb5b* loci did not reveal clear signs of complementation between the regulatory elements retained in both zebrafish loci, our results highlight the importance of interactions between CNEs in the execution of complementary subfunctions of duplicated genes.

We thank Adrianna Gambarotta, Marion Alfero, Marcelle Carter, Lucille Joly, Margot Lemieux and Vishal Saxena for technical help; Fabien Avaron for useful discussions; and Marie-Andrée Akimenko and David Lohnes for comments on the manuscript. This work was supported by a grant from NSERC, by OGS and NSERC awards to O.J., by a FRSQ Chercheur National Award to L.J. and by an Investigator award from CIHR to M.E.

Supplementary material

Supplementary material for this article is available at http://dev.biologists.org/cgi/content/full/135/21/3543/DC1

References

Altschmied, J., Delfgaauw, J., Wilde, B., Duschl, J., Bouneau, L., Volff, J. N. and Schartl, M. (2002). Subfunctionalization of duplicate mitf genes associated with differential degeneration of alternative exons in fish. *Genetics* **161**, 259-267.

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). Zebrafish hox clusters and vertebrate genome evolution. *Science* **282**, 1711-1714.

Amores, A., Suzuki, T., Yan, Y. L., Pomeroy, J., Singer, A., Amemiya, C. and Postlethwait, J. H. (2004). Developmental roles of pufferfish Hox clusters and genome evolution in ray-fin fish. *Genome Res.* **14**, 1-10.

Amsterdam, A., Lin, S. and Hopkins, N. (1995). The Aequorea victoria green fluorescent protein can be used as a reporter in live zebrafish embryos. *Dev. Biol.* **171**, 123-129.

Bruce, A. E., Oates, A. C., Prince, V. E. and Ho, R. K. (2001). Additional hox clusters in the zebrafish: divergent expression patterns belie equivalent activities of duplicate hoxB5 genes. *Evol. Dev.* **3**, 127-144.

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). Bichir HoxA cluster sequence reveals surprising trends in ray-finned fish genomic evolution. *Genome Res.* **14**, 11-17.

Copeland, N. G., Jenkins, N. A. and Court, D. L. (2001). Recombineering: a powerful new tool for mouse functional genomics. *Nat. Rev. Genet.* **2**, 769-779.

Cormack, B. P., Valdivia, R. H. and Falkow, S. (1996). FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **173**, 33-38.

Cresko, W. A., Yan, Y. L., Baltrus, D. A., Amores, A., Singer, A., Rodriguez-Mari, A. and Postlethwait, J. H. (2003). Genome duplication, subfunction partitioning, and lineage divergence: Sox9 in stickleback and zebrafish. *Dev. Dyn.* **228**, 480-489.

Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L. and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**, 1531-1545.

Fraidenraich, D., Lang, R. and Basilico, C. (1998). Distinct regulatory elements govern Fgf4 gene expression in the mouse blastocyst, myotomes, and developing limb. *Dev. Biol.* **204**, 197-209.

Hadrys, T., Prince, V., Hunter, M., Baker, R. and Rinkwitz, S. (2004). Comparative genomic analysis of vertebrate Hox3 and Hox4 genes. *J. Exp. Zoolog. B Mol. Dev. Evol.* **302**, 147-164.

Hogan, B., Constantini, F. and Lacy, E. (1986). *Manipulating the Mouse Embryo: a Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Hogan, B. L., Holland, P. W. and Lumsden, A. (1988). Expression of the homeobox gene, Hox 2.1, during mouse embryogenesis. *Cell Differ. Dev.* **25 Suppl**, 39-44.

Hua, L. V., Hidaka, K., Pesesse, X., Barnes, L. D. and Shears, S. B. (2003). Paralogous murine Nudt10 and Nudt11 genes have differential expression patterns but encode identical proteins that are physiologically competent diphosphoinositol polyphosphate phosphohydrolases. *Biochem. J.* **373**, 81-89.

Jaffe, L., Jeannotte, L., Bikoff, E. K. and Robertson, E. J. (1990). Analysis of beta 2-microglobulin gene expression in the developing mouse embryo and placenta. *J. Immunol.* **145**, 3474-3482.

Kleinjan, D. A., Bancewicz, R. M., Gautier, P., Dahm, R., Schonthaler, H. B., Damante, G., Seawright, A., Hever, A. M., Yeyati, P. L., van Heyningen, V. et al. (2008). Subfunctionalization of duplicated zebrafish pax6 genes by cis-regulatory divergence. *PLoS Genet.* **4**, e29.

Kluver, N., Kondo, M., Herpin, A., Mitani, H. and Schartl, M. (2005). Divergent expression patterns of Sox9 duplicates in teleosts indicate a lineage specific subfunctionalization. *Dev. Genes Evol.* **215**, 297-305.

Krumlauf, R., Holland, P. W., McVey, J. H. and Hogan, B. L. (1987). Developmental and spatial patterns of expression of the mouse homeobox gene, Hox 2.1. *Development* **99**, 603-617.

Kumar, S., Tamura, K., Jakobsen, I. B. and Nei, M. (2001). MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**, 1244-1245.
Larochelle, C., Tremblay, M., Bernier, D., Aubin, J. and Jeannotte, L. (1999). Multiple cis-acting regulatory regions are required for restricted spatio-temporal Hoxa5 gene expression. *Dev. Dyn.* **214**, 127-140.
Lynch, M. and Force, A. (2000). The probability of duplicate gene preservation by subfunctionalization. *Genetics* **154**, 459-473.
Mansfield, J. H., Harfe, B. D., Nissen, R., Obenauer, J., Srineel, J., Chaudhuri, A., Farzan-Kashani, R., Zuker, M., Pasquinelli, A. E., Ruvkun, G. et al. (2004). MicroRNA-responsive 'sensor' transgenes uncover Hox-like and other developmentally regulated patterns of vertebrate microRNA expression. *Nat. Genet.* **36**, 1079-1083.
Ohno, S. (1970). *Evolution by Gene Duplication*. Heidelberg, Germany: Springer-Verlag.
Oosterveen, T., Niederreither, K., Dolle, P., Chambon, P., Meijlink, F. and Deschamps, J. (2003). Retinoids regulate the anterior expression boundaries of 5' Hoxb genes in posterior hindbrain. *EMBO J.* **22**, 262-269.
Sakach, M. and Safaei, R. (1996). Localization of the HoxB5 protein in the developing CNS of late gestational mouse embryos. *Int. J. Dev. Neurosci.* **14**, 567-573.
Santini, S., Boore, J. L. and Meyer, A. (2003). Evolutionary conservation of regulatory elements in vertebrate Hox gene clusters. *Genome Res.* **13**, 1111-1122.
Sharpe, J., Nonchev, S., Gould, A., Whiting, J. and Krumlauf, R. (1998). Selectivity, sharing and competitive interactions in the regulation of Hoxb genes. *EMBO J.* **17**, 1788-1798.
Taylor, J. S., Van de Peer, Y., Braasch, I. and Meyer, A. (2001). Comparative genomics provides evidence for an ancient genome duplication event in fish. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **356**, 1661-1679.
Thisse, C., Thisse, B., Schilling, T. F. and Postlethwait, J. H. (1993). Structure of the zebrafish snail1 gene and its expression in wild-type, spadetail and no tail mutant embryos. *Development* **119**, 1203-1215.
Tumpel, S., Cambronero, F., Wiedemann, L. M. and Krumlauf, R. (2006). Evolution of cis elements in the differential expression of two Hoxa2 coparalogous genes in pufferfish (Takifugu rubripes). *Proc. Natl. Acad. Sci. USA* **103**, 5419-5424.
Woolfe, A. and Elgar, G. (2007). Comparative genomics using Fugu reveals insights into regulatory subfunctionalization. *Genome Biol.* **8**, R53.
Yee, S. P. and Rigby, P. W. (1993). The regulation of myogenin gene expression during the embryonic development of the mouse. *Genes Dev.* **7**, 1277-1289.
Zerucha, T., Stuhmer, T., Hatch, G., Park, B. K., Long, Q., Yu, G., Gambarotta, A., Schultz, J. R., Rubenstein, J. L. and Ekker, M. (2000). A highly conserved enhancer in the Dlx5/Dlx6 intergenic region is the site of cross-regulatory interactions between Dlx genes in the embryonic forebrain. *J. Neurosci.* **20**, 709-721.