View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Elsevier - Publisher Connector

Developmental Biology 340 (2010) 269–282

Contents lists available at ScienceDirect

Developmental Biology

journal homepage: www.elsevier.com/developmentalbiology

Genomes & Developmental Control

Cis-regulatory characterization of sequence conservation surrounding the Hox4 genes

Beena Punnamoottil [a,b,c], Carl Herrmann [d], Juan Pascual-Anaya [e], Salvatore D'Aniello [e], Jordi Garcia-Fernàndez [e], Altuna Akalin [b,f], Thomas S. Becker [b,c], Silke Rinkwitz [a,b,c,*]

[a] *Carl von Ossietzky University, Oldenburg, Germany*
[b] *Sars International Center, Bergen, Norway*
[c] *Brain and Mind Research Institute, University of Sydney, Sydney, Australia*
[d] *TAGC Inserm U928 and Université de la Méditerranée, Marseille, France*
[e] *Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Barcelona, Spain*
[f] *Computational Biology Unit, University of Bergen, 5008 Bergen, Norway*

ARTICLE INFO

Article history:
Received for publication 7 January 2009
Revised 17 January 2010
Accepted 30 January 2010
Available online 6 February 2010

Keywords:
Hox
Cis-regulation
CNE
Zebrafish
GFP
Transgenesis
Amphioxus
Sequence conservation
Cluster

ABSTRACT

Hox genes are key regulators of anterior–posterior axis patterning and have a major role in hindbrain development. The zebrafish Hox4 paralogs have strong overlapping activities in hindbrain rhombomeres 7 and 8, in the spinal cord and in the pharyngeal arches. With the aim to predict enhancers that act on the *hoxa4a*, *hoxb4a*, *hoxc4a* and *hoxd4a* genes, we used sequence conservation around the Hox4 genes to analyze all fish:human conserved non-coding sequences by reporter assays in stable zebrafish transgenesis. Thirty-four elements were functionally tested in GFP reporter gene constructs and more than 100 F1 lines were analyzed to establish a correlation between sequence conservation and *cis*-regulatory function, constituting a catalog of Hox4 CNEs. Sixteen tissue-specific enhancers could be identified. Multiple alignments of the CNEs revealed paralogous *cis*-regulatory sequences, however, the CNE sequence similarities were found not to correlate with tissue specificity. To identify ancestral enhancers that direct Hox4 gene activity, genome sequence alignments of mammals, teleosts, horn shark and the cephalochordate amphioxus, which is the most basal extant chordate possessing a single prototypical Hox cluster, were performed. Three elements were identified and two of them exhibited regulatory activity in transgenic zebrafish, however revealing no specificity. Our data show that the approach to identify *cis*-regulatory sequences by genome sequence alignments and subsequent testing in zebrafish transgenesis can be used to define enhancers within the Hox clusters and that these have significantly diverged in their function during evolution.

© 2010 Elsevier Inc. All rights reserved.

Introduction

Hox genes encode a family of transcription factors that are organized in clusters in most metazoan lineages. In humans and mice, four paralogous Hox clusters, comprising 39 genes, control cellular identity in the central nervous system (CNS), the skeleton, and inner organs along the embryonic antero-posterior axis (Deschamps and van Nes, 2005; Kessel and Gruss, 1991; Krumlauf et al., 1993). The Hox genes of group 1 to 4 are essential for patterning the hindbrain and the adjacent neural crest (reviewed by Kiecker and Lumsden, 2005), but also for the specification of neuronal phenotypes (Arenkiel et al., 2004; Gaufo et al., 2003, 2004). Hox genes have played major roles in driving key evolutionary innovations; as an example, the vertebrate limbs adopted the Hox gene system for regulating outgrowth and polarity. Enhancers regulating the development of extremities have been identified close to the genes they regulate (Morrison et al., 1997; Sharpe et al., 1998), but also outside of the gene cluster, namely the global control region (GCR) and the early limb control region (ELCR) upstream and downstream of the HoxD cluster (Spitz et al., 2003, 2005; Tarchini et al., 2006).

During bilaterian evolution, various rearrangements occurred in the primordial Hox cluster in distinct lineages, ranging from enlargement to split into single genes (for a review of nomenclature see Duboule, 2007; Monteiro and Ferrier, 2006). For example *Oikopleura dioica*, a species of urochordates, the sister group of the vertebrates (Bourlat et al., 2006; Delsuc et al., 2006) has nine Hox genes dispersed throughout the genome, but has retained, to some extent, antero-posterior expression domains (Seo et al., 2004). This suggests that in *Oikopleura* Hox genes are regulated through close-by flanking sequences, which may control their activity in a way comparable to that of developmental regulatory genes not organized in clusters. The cephalochordate amphioxus, *Branchiostoma floridae*,

* Corresponding author. Brain and Mind Research Institute, University of Sydney, Sydney, Australia. Fax: +61 293510731.
E-mail address: srinkwitz@med.usyd.edu.au (S. Rinkwitz).

0012-1606/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ydbio.2010.01.035

represents the most basal chordate lineage (Bourlat et al., 2006; Holland et al., 2008) and reveals many aspects of the ancestral chordate genomic organization (Brooke et al., 1998; Garcia-Fernandez and Holland, 1994; Putnam et al., 2008), which has subsequently been lost in the urochordates, while having been retained and consolidated in vertebrates (Duboule, 2007). Amphioxus has a single, prototypical Hox cluster containing 15 genes displaying temporal and spatial colinearity (Ferrier et al., 2000; Holland et al., 2008; Schubert et al., 2006; Wada et al., 1999). As discussed in Duboule (2007), the driving force behind retaining and consolidating a clustered structure of Hox genes in vertebrates may be the communal ability of the genes to fulfil a functional task that cannot be fulfilled by any of the genes in isolation. For example, nested transcriptional units and shared promoters (Hadrys et al., 2006) could result in novel proteins, whereas increased gene proximity may facilitate the coordinated transcription of neighbouring genes through global control regions located outside the cluster (Kmita et al., 2002a; Spitz et al., 2003). It is likely that the function of single genes became integrated into a broader control landscape and that the clusters, duplicated in vertebrates, facilitated dosage-specific and overlapping gene activities, ultimately resulting in compensatory, synergistic-, and overlapping functions (Condie and Capecchi, 1994; Dasen et al., 2003; Di-Poi et al., 2007; Gaufo et al., 2003; Kmita et al., 2002a, 2005; Manley and Capecchi, 1997). Most recent findings revealed that clustering is required to maintain full repression of the genes through chromatin modifications during very early stages of development (Soshnikova and Duboule, 2009).

With the increased availability of genome sequences, functional elements can be defined as conserved non-coding sequences (CNEs), highly conserved non-coding sequences (HCNEs) or ultraconserved regions (UCRs) depending on the degree of conservation (Bejerano et al., 2004; Sandelin et al., 2004). When tested in reporter gene assays in mice and zebrafish, the majority of HCNEs and UCRs show enhancer activity (de la Calle-Mustienes et al., 2005; McGaughey et al., 2008; Pennacchio et al., 2006). Enhancers of a given gene often display overlapping tissue-specific activity (Komisarczuk et al., 2009) and the deletion in mice of some UCRs with regulatory activity did not result in an obvious morphological or behavioural phenotype (Ahituv et al., 2007) pointing at a certain degree of redundancy. While functions of these elements other than *cis*-regulatory activity is difficult to address, there are hypotheses that some might participate in mediating DNA looping in long range gene regulatory processes in chromatin structure (Woltering and Duboule, 2009).

To identify *cis*-regulatory sequences that control the zebrafish Hox4 genes we used sequence conservation as a parameter to predict genomic sequences for testing in reporter gene constructs. Despite the genome duplication that occurred in the teleost lineage (3R) and led to seven Hox clusters in zebrafish (hoxaa, hoxab, hoxba, hoxbb, hoxca, hoxcb and hoxda), the Hox3 and Hox4 paralogs were retained together as single copies only on the "a"-duplicates (see Fig. 1), which might have occurred due to regulatory sequences that are shared by both genes. This represents most likely the original structure that existed in vertebrates before the split with teleosts. Vertebrate Hox3–Hox5 genomic sequence comparisons were used to identify conservation in the non-coding sequence surrounding the zebrafish *hoxa4a*, *hoxb4a*, *hoxc4a* and *hoxd4a* genes. Using human:zebrafish conservation as a threshold for functional testing, CNEs were defined for their ability to drive reporter gene expression in stable zebrafish transgenes. Because the four gene loci arose from a single ancestral cluster, we used multiple alignment tools to analyze the paralogous relationships of the CNEs and to establish a common function. We further searched for ancestral Hox4 regulatory information by aligning the Hox5–Hox3 genome sequence of the urochordate amphioxus against hornshark, tetrapod, mammal and teleost sequences. Two CNEs were identified upstream of *AmphiHox4*, a further one, located in the intron of the gene, was found to be conserved in all analyzed species, but had unspecific *cis*-regulatory activity in zebrafish.

Materials and methods

Maintenance of zebrafish and embryos

Zebrafish (*Danio rerio*) were maintained and obtained from our breeding colony under standard conditions according to a protocol developed in our laboratory (http://www.sars.no/facilities/fishRaisingprotocol.doc). Embryos for Tol2 transgenesis were obtained from crosses of wildtype TAB zebrafish. Potential transgenic founders were outcrossed to a Singapore spotty strain. Fertilized eggs

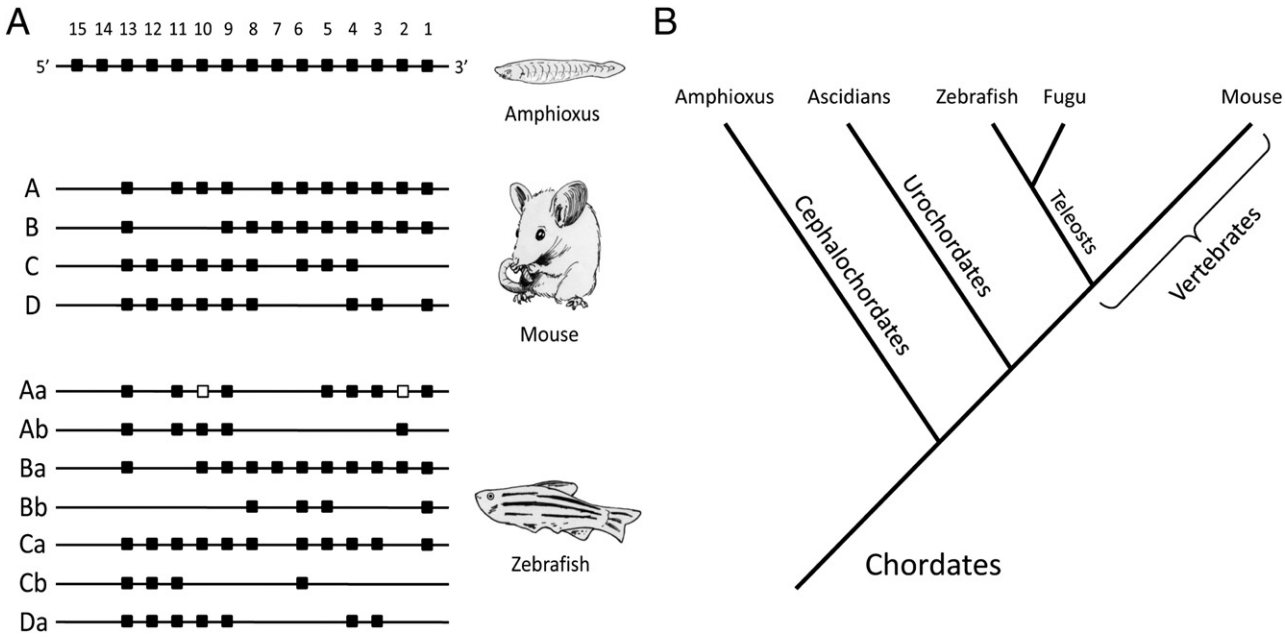

Fig. 1. A. Hox gene complement of amphioxus, mouse and zebrafish. Amphioxus is the only metazoan with a complete array of 15 genes. After two rounds of genome duplications that took place during vertebrate evolution, vertebrates present four paralogous clusters in mammals, and due to a third whole genome duplication up to 8 in teleosts. Hox gene coding sequences are represented as black boxes, and pseudogenes as white boxes. B. Phylogenetic relationship of the chordates and vertebrates.

were kept at 28.5 °C in E3 embryo medium with 0.003% 1-phenyl-2-thiourea to prevent pigmentation and were staged according to Kimmel et al. (Kimmel et al., 1995).

Identification of orthologous CNEs

Starting from the zebrafish zv6 CNEs, we used the liftOver tool from the UCSC browser to identify orthologous CNEs in human (hg18), mouse (mm8), chicken (galGal3) and xenopus (xenTro2). The corresponding locations can be viewed at http://tagc.univ-mrs.fr/zebrahox, which contains links to the UCSC locations.

Identification of paralogous relationships

Paralogous relationships between CNEs in the four zebrafish regions were determined by aligning all CNEs against each other using the YASS algorithm (Noe and Kucherov, 2005), which is adapted to local alignments of non-coding DNA. We run YASS with standard options, and keep alignments with *E*-values <0.1. Obtained alignments are displayed graphically in Fig. 8, where the width of the lines represents the length of the local alignment, and the color indicates the percentage identity.

Multiple alignments

Multiple alignments were performed using MUSCLE 3.6 and are displayed on the supplementary website. The full alignment of the intronic element was performed using MUSCLE and manually adjusted using the SeaView software. Motifs that appear in the multiple alignments are confirmed by using the GLAM2 software (Frith et al., 2008) which identifies gapped conserved motifs, and builds motif logos. The motifs obtained from GLAM2 are compared to motif databases using the Tomtom software (Gupta et al., 2007). As motif databases, we used Transfac and the database of homeodomain binding sites of (Berger et al., 2008). We used default parameters.

Phylogenetic analysis

The phylogenetic analysis was done using the phylogenetic pipeline available on www.phylogeny.fr (Dereeper et al., 2008). Multiple alignments were done using MUSCLE, and curated with Gblocks. The trees were built using the PhyML algorithm, and the bootstrap values were approximated using the aLRT method implemented in PhyML 3.0.

Cloning of conserved non-coding sequences and plasmid DNA injection

Sequences conserved between human and zebrafish were chosen for enhancer testing and 50 bp–100 bp were added 5′ and 3′ for PCR amplification. Fragments were amplified by PCR using Platinum *Taq* polymerase (Invitrogen) from zebrafish genomic DNA extracted from 24 hpf embryos using DNeasy Tissue Kit (Qiagen) and from amphioxus genomic DNA. The primer pairs for the amplification are listed together with the coordinates of the elements in Table Supplement 1. The amplified PCR products were subcloned into a pCR8/GW/Topo cloning Vector and were cloned afterwards into the destination vector (Tol2 GFP-Expression Vector) from which the element was sequenced for confirmation as described (Navratilova et al., 2009). The plasmid DNA was purified by phenol–chlorophorm extraction and diluted to 50 ng/µl with DNase/RNase free water. Transposase mRNA was synthesized and purified with the mMessage SP6 Kit (Ambion) from the pCS2_TP vector after digestion with NotI and purified according to instructions by the manufacturer. The injection mixture contained 25 ng/µl DNA and 25 ng/µl Transposase mRNA. An amount less than 2 nl was injected into the cytoplasm of each egg shortly after fertilization. The embryos were incubated at 28.5 °C and screened at 24 hpf for GFP fluorescence. Embryos that showed GFP expression were selected and raised.

Screening and expression pattern documentation

Injected adult zebrafish were outcrossed with wildtype fish and progeny were screened for GFP fluorescence. Positive embryos were collected for expression pattern documentation and raising. The corresponding parental founder fish were kept separately in the tank system. Three to five F1 lines were generated originating from two to five positive founders. In general, a CNE driving GFP in at least three zebrafish lines from three different founders in the same structures was considered to act as tissue-specific enhancer. A few elements had such clear Hox4 specific activity that two lines were deemed sufficient for evaluation. For tested CNEs that resulted in inconsistent expression patterns in the transgenic lines, no conclusion in regards to their function could be drawn.

Transgenic GFP expressing zebrafish embryos and larvae were screened and photographed using a GFP filter on a Nikon ECLIPSE TE2000-S inverse microscope, equipped with a 470/40 nm excitation filter and a 500 nm LP emission filter (Chroma) with Diagnostic Instruments Inc. SpotRT™ KE monochrome digital camera and Spot™ software. For orientation of the larvae during the imaging procedure, the embryos and larvae where anesthetized with 0.2% tricaine (3-amino benzoic acid ethylester, Sigma) and embedded in 3% methylcellulose.

Immunohistochemistry

Immunostaining and vibratome sectioning of F1 larvae were performed as described (Punnamoottil et al., 2008).

In situ hybridization

In situ hybridization was performed according to (Oxtoby and Jowett, 1993). The *hoxc4a* template was amplified by PCR from genomic DNA with primer sequences published at Zfin.

PCR-primers for *hoxc4a* sequence:

forward primer: GCTAGTAGGAGGGCTTTATGG, reverse primer: GGATCCATTAACCCTCACTAAAGGGAAAAAGGAGTGAGGCAGACG.

The amplified *hoxc4a* sequence (549 bp) was cloned into the Gateway pCR2.1 Vector and digested with the restriction enzyme HindIII. The antisense probe was synthesized with T7 RNA Polymerase using Digoxigenin labeled ribonucleotides.

Results

Conserved non-coding sequences around the Hox4 paralogous genes

The seven zebrafish Hox clusters originated from the four vertebrate Hox clusters by whole genome duplication and subsequent loss of single genes (all of them in the eighth cluster, hoxdb) (Amores et al., 1998; Hoegg and Meyer, 2005; Kuraku and Meyer, 2009). Whereas many Hox paralogs are dispersed on the duplicated “a”- and “b”-versions of the clusters, the linked Hox4 and Hox3 genes are always located on the a-clusters, and the duplicated gene pairs have been lost in the b-clusters. This suggests that the Hox4/Hox3 gene arrangement in zebrafish can be compared directly to the four Hox clusters in tetrapods and mammals.

The conservation within the non-coding sequences upstream and downstream of *hoxa4a*, *hoxb4a*, *hoxc4a* and *hoxd4a* was identified in the UCSC genome browser using the zebrafish March 2006 assembly (Fig. 2). We searched the intergenic regions starting downstream of Hox5 up to 9 kb downstream of Hox4 for elements of high fish–human conservation (CNEs). Whereas in the Hoxd cluster the *hoxd3a*–*hoxd4a* region is preserved, the *hoxd5a* to *hoxd8a* genes are lost,

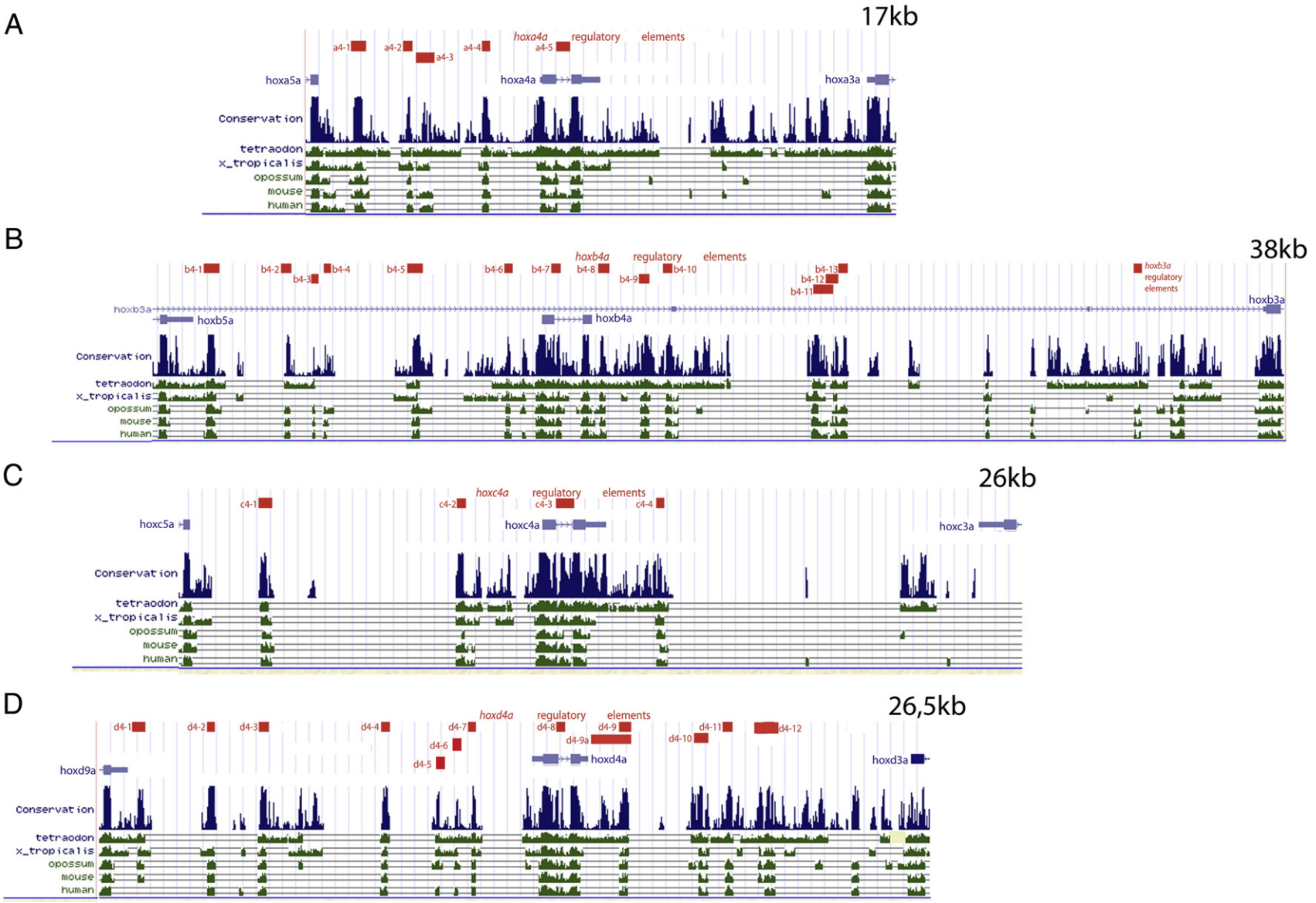

Fig. 2. Windows of the UCSC genome browser (ZV6) showing the Hox5 to Hox3 genomic region in the Hoxaa (A), the Hoxba (B), the Hoxca (C) and the Hoxda (D) clusters including the conservation tracks. The loci are of different sizes with the proportions kept in this figure. For better overview the *hoxa4a*, *hoxb4a*, *hoxc4a* and *hoxd4a* genes are centered. The red horizontal bars indicate the highly conserved non-coding elements (CNEs) conserved from teleosts to human. The elements in rectangles have been additionally tested. A. Genomic region from *hoxa5a* to *hoxa3a* located on chromosome 19, spanning 17 kb. B. A 38 kb genomic region from *hoxb5a* to *hoxb3a* on chromosome 3. C. The genomic region from *hoxc5a* to *hoxc3a* on chromosome 23, containing 26 kb. D. The region from *hoxd9a* to *hoxd3a* on chromosome 9 with around 27 kb.

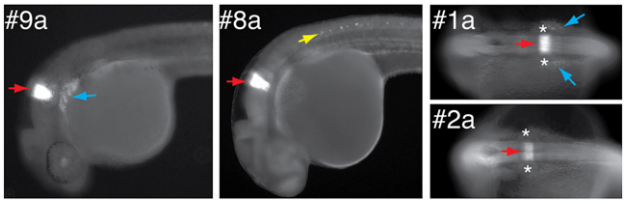

Fig. 3. The tested *hoxb3a* regulatory element is a tissue-specific enhancer. Lines from four different founders had strong GFP expression in hindbrain r5. Furthermore almost each founder showed another additional expression feature, which are also within the *hoxb3a* expression domain — founder #9a: pharyngeal endoderm (blue arrow), founder #8a: spinal cord (yellow arrow), founder #1a: migrating neural crest (blue arrow).

prompting us to take *hoxd9a* as a boundary for sequence conservation analysis. Thirty-two elements were identified that met our threshold (red bars in Fig. 2). Two further elements were included into the analyses, one fragment in the intron of *hoxa4a* and one element downstream adjacent to *hoxd9a*. An element upstream of *hoxb3a* is a known enhancer and was used as control in the functional experiments. The zebrafish elements as well as their orthologous sequences in xenopus, mouse, human and chicken are listed at http://tagc.univ-mrs.fr/zebrahox, a site linked to the UCSC genome browser. Multiple alignments of the orthologous sequences and the fish–human conservation level are included.

Conserved non-coding sequences with strong evolutionary constraint act as enhancers and have overlapping activities in Hox4 gene regulation

We next asked whether the high evolutionary constraint of the sequences is due to their regulatory functions. The sequences that were chosen for enhancer testing are illustrated in Fig. 2 by the red bars and are listed with size and corresponding primer pairs in supplementary Table 1. For PCR-based amplification of each element, up to 150 bp of flanking sequence were amplified at the 5′ and 3′ ends of the CNEs of interest. The fish–human conservation levels varied between 57% and 84% as listed for each element on the website. Altogether, 36 elements were amplified from zebrafish genomic DNA and subsequently cloned into a Tol2–Gata2–GFP reporter vector for injection into zebrafish embryos as described (Navratilova et al., 2009). The reporter gene activity in several independent transgenic insertions (see Materials and methods) was analyzed by fluorescent microscopy during 1 to 5 days post fertilization (dpf), to identify functional elements. To evaluate our experimental approach we tested a CNE that was chosen by the same computational approach as the other CNEs, but that includes a well analyzed *hoxb3a* regulatory element (Manzanares et al., 1997, 2002). This element drove expression in hindbrain rhombomere r5 in four different lines (Fig. 3). Surprisingly, almost every line showed an additional *hoxb3a* specific expression feature such as spinal cord (#8a), migrating neural crest (#1a) and pharyngeal endoderm (#9a), a regulatory function that might be dependent on the site of insertion. The reporter gene expression patterns were compared to the endogenous expression patterns of the Hox4 genes. *Hoxa4a*, *hoxb4a* and *hoxd4a* expression features were described in detail in a previous study (Punnamoottil et al., 2008), the *hoxc4a* expression pattern is provided in supplemental Fig. 1. The genes have overlapping activities within the hindbrain r7 and r8, the spinal cord and the pharyngeal arches 3–7, structures on which we have focussed in this study.

The *hoxa4a* locus contains five elements that were tested for enhancer activity (Fig. 4). The expression domain in the hindbrain was represented by elements a4-2 and a4-3, which drove tissue-specific expression in different subdomains in rhombomeres 7 and 8 (Fig. 4, red arrows). Element a4-2 mediated strong expression predominantly in the dorsal hindbrain with GFP expressing cells along the ventricular surface. Transverse sections through anti-GFP stained larvae revealed the localisation of the cell bodies along the dorsal and medial ventricular surface with projections towards the lateral and ventral hindbrain (a4-2″, light blue arrows). Element a4-3 mediated broad GFP expression in the medioventral hindbrain, in which the cytoplasmatic GFP allowed for the resolution of axonal projections

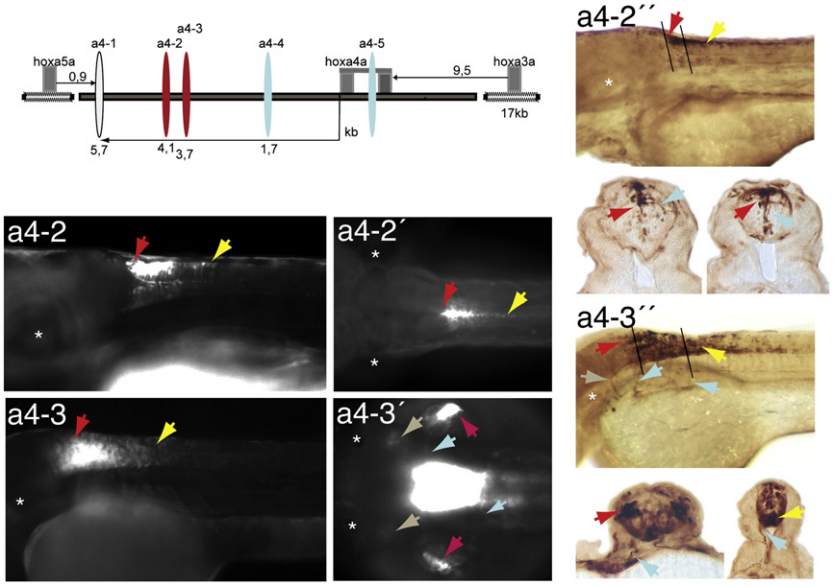

Fig. 4. Schematic of the genomic region from *hoxa5a* to *hoxa3a* with the tested elements in a color code (red: tissue-specific enhancer activity, blue: unspecific, white: no activity) and the distance of the element from the exons in kb. The expression of GFP under the regulatory control of the tested CNEs is shown in representative zebrafish larvae between stage 2 dpf to 4 dpf. The asterisk marks the level of the inner ear and the arrows follow a color code: *red*, hindbrain; *yellow*, spinal cord; *brown*, cranial ganglia; *light blue*, neuronal projection; *pink*, pectoral fin. a4-2. Expression of GFP in a dorsal and ventral subpopulation of cells in the most posterior part of the hindbrain and in the anterior spinal cord. a4-2′. Dorsal view of the posterior hindbrain and spinal cord with GFP expression in cells along the ventricular surface. a4-2″. Anti-GFP stained transgenic larva with sections trough the posterior hindbrain and anterior spinal cord (indicated by black lines). a4-3. Wide GFP expression in the medial posterior hindbrain and anterior spinal cord. a4-3′. Dorsal view that shows GFP expression in the vagal ganglion, in the pectoral fins and in fibers innervating the pectoral fins and the somites. a4-3″. Anti-GFP stained transgenic larva with sections showing signal in the hindbrain (red arrow points to the vagal nucleus) and in the spinal cord. The level of sections are indicated by the black lines. Neuronal projections are visible (1) from the vagal nuclei towards the vagal ganglion (brown arrow) and (2) into the pectoral fin and (3) along the somites.

into the pectoral fins (a4-3′ brown arrows and a4-3″ light blue arrows) and within the 10th nerve towards the vagal ganglia (a4-3″, brown arrow). Sections through the hindbrain of anti-GFP stained larvae revealed that GFP was localized in neurons of the vagal nuclei (a4-3″, red arrows). Both elements also mediated expression in the spinal cord (yellow arrow) — while element a4-2 driven GFP was localized along the ventricular surface with a decreasing intensity towards posterior, and a4-3 mediated expression in the lateral and ventral spinal cord (a4-3″), where the GFP labeled neurons sent out neuronal projections to the adjacent myotomes (a4-3′, a4-3″, light blue arrows). The other elements that were tested had either unspecific activity (a4-4, a4-5) or no activity (a4-1).

In summary, the main expression features of the *hoxa4a* gene are mediated by elements a4-2 and a4-3, which act as enhancers and regulated almost the same expression features seen in the *hoxa4a* enhancer trap line CLGY409 (Punnamoottil et al., 2008).

Within the *hoxb4a* locus 13 elements were tested (Fig. 5). Among these, five were found to function as tissue-specific enhancers: b4-2, b4-4 (data not shown), b4-5, b4-7 and b4-10. The other eight elements had unspecific activity and one element was considered as negative.

The expression patterns driven by the enhancers are shown in Fig. 5. Four of the enhancers, b4-2, b4-4, b4-5 and b4-10, drove GFP in the hindbrain r7 and r8, with two of them driving in addition expression in the pectoral fins and in the pharyngeal arches (b4-4 and b4-5). All five enhancers revealed activity in the spinal cord, with one of them, element b4-7, regulating this expression domain exclusively. Element b4-4 further mediated somitic expression and activity in the notochord. One element (b4-5) had activity in the cranial ganglia. In the hindbrain expression domain, element b4-10 mediated a sharp anterior boundary at r6, whereas b4-5 had a diffuse anterior expression within r7 (Fig. 5, red arrows). A region homologous to element b4-10 has been described in mice as Region A enhancer, and was found to regulate posterior hindbrain expression with a sharp anterior r6/r7 expression boundary (Whiting et al., 1991). It is noteworthy that the intronic element b4-7 was the only element that drove expression exclusively in the spinal cord. Our observations suggest that the enhancers in the *hoxb4a* locus have overlapping activities in the main expression domains and might regulate the main features of the gene in synergy.

Anti-GFP staining of b4-5 larvae (Fig. 5) and subsequent sectioning revealed that the GFP was distributed in a dorsal and lateral

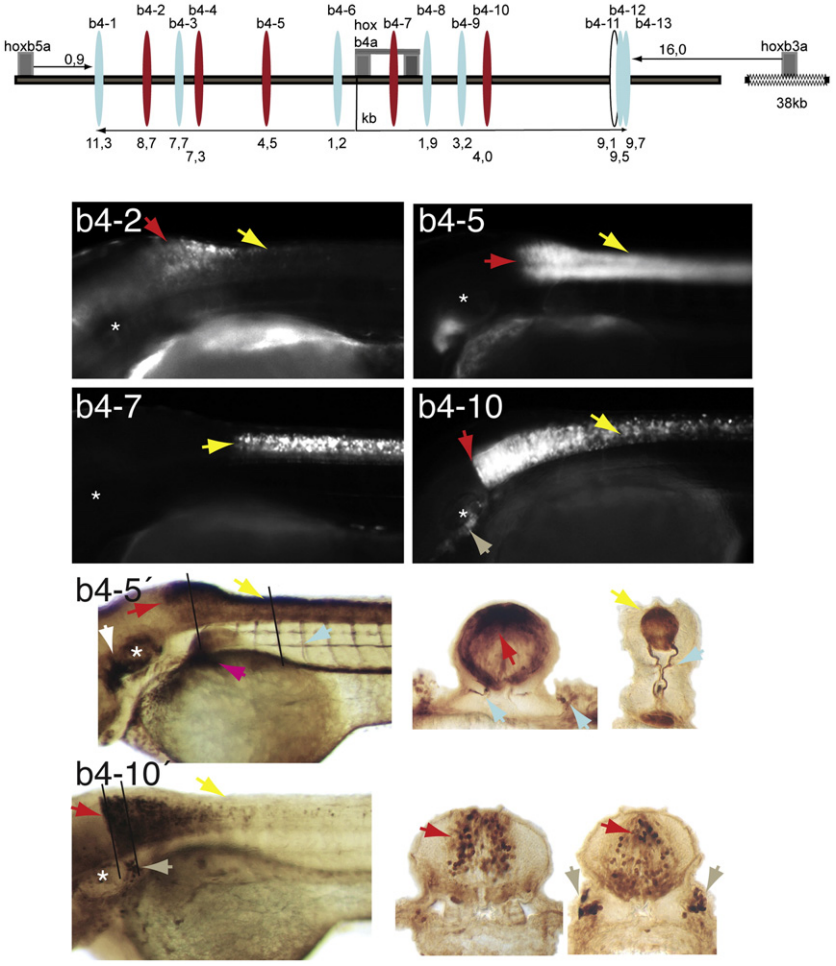

Fig. 5. Schematic of the *hoxb5a* to *hoxb3a* genomic region. Tested elements are indicated as ovals (red: tissue-specific enhancer activity, blue: unspecific, white: no activity) with its distance from the exons in kb. The regulatory activity of the elements is illustrated by representative larvae at 2 to 4 dpf. The asterisk marks the level of the inner ear. The arrows follow a color code: red, hindbrain; yellow, spinal cord; brown, cranial ganglia; light blue, neuronal projection; white, migrating neural crest. b4-2. GFP expression in the posterior hindbrain, extending with decreasing intensity into the anterior spinal cord. b4-5. Larva with strong and broad GFP expression in the hindbrain r7 and r8 extending posterior into the spinal cord. GFP is also visible in a pectoral fin neuron, in the developing pharyngeal arches and a cranial ganglion. b4-5′. Anti-GFP stained larva with sections at the level of the 2nd somite and the spinal cord. The level of sections are indicated by the black lines. Strong GFP expression is indicated in the dorsal and lateral domains of hindbrain r7 and r8 with projections into the pectoral fins. In the spinal cord the signal is also stronger dorsally with two clusters of motoneurons projecting laterally along the adjacent somites. b4-7. Larva with strong GFP expression in the entire spinal cord. b4-10. Larva with GFP expression in the hindbrain r7 and r8 with a sharp anterior expression boundary at r6. Expression extends posterior into the spinal cord. GFP is also visible in cranial ganglia around the inner ear. b4-10′. Anti-GFP stained larva of the same transgenic line with sections through the hindbrain r6/7 and r7, which shows the distribution of the GFP positive cells in the medial hindbrain and in cells of the posterior lateral line ganglion.

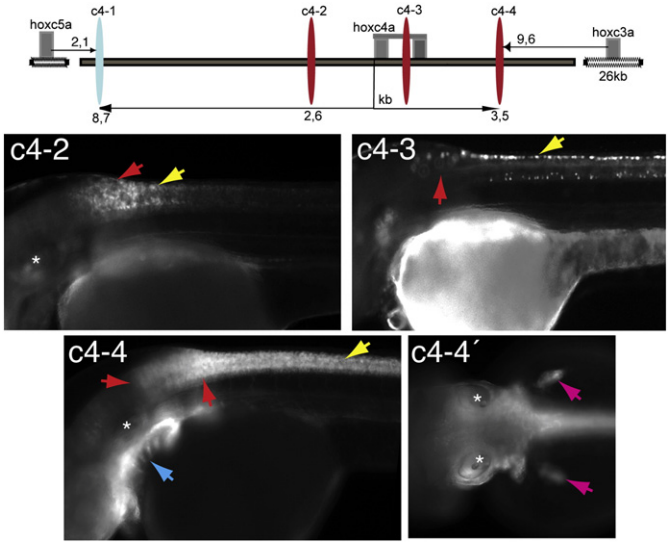

Fig. 6. Schematic of the *hoxc5a* to *hoxc3a* genomic region. Tested CNEs are illustrated as explained in Figs. 4 and 5. Representative GFP expressing larvae are shown at stages 2 to 4 dpf. The asterisk marks the level of the inner ear. The arrows follow a color code: red, hindbrain; yellow, spinal cord; dark blue, pharyngeal arches; pink, pectoral fins. c4-2. Larva with GFP expression in the hindbrain r7 and r8 extending weaker into the anterior spinal cord. c4-2′. Dorsal view of the same larva as in A with posterior hindbrain and pectoral fin expression. c4-3. Larva with GFP expression in the posterior hindbrain and in the spinal cord. c4-3′. Another line with higher magnification shows single GFP labeled cells in hindbrain r7 and further expression in the pharyngeal arches. c4-4. Larva with expression in the hindbrain r7 and r8, in the spinal cord and also in all pharyngeal arches. c4-4′. A dorsal view of a larva with focus on GFP expression in the pectoral fins.

subdomain of the hindbrain with some positive neuronal clusters located ventrally and GFP positive neurons sent out projections to innervate the pectoral fins (light blue arrows in b4-5'). Further, GFP labeled motor neurons from the spinal cord sent out projections towards the adjacent myotomes (light blue arrows in b4-5'). Sections through immunostained b4-10 larvae showed GFP positive cells in a medioventral subdomain in the hindbrain (b4-10'). Thus, despite indicated overlapping functions of enhancers in the *hoxb4a* locus, further single elements seem to have unique functions at cellular levels.

Within the *hoxc4a* locus four elements were tested for enhancer function (Fig. 6), where three of them (c4-2, c4-3 and c4-4) drove reproducible tissue-specific expression and the fourth had unspecific activity. Expression in the hindbrain r7 and r8 (red arrows) was mediated through all three tissue-specific elements. Element c4-2 regulated GFP expression in the medial r7 and r8, which became weaker towards anterior. Element c4-3 mediated weak expression in single cells of the lateral and ventral hindbrain and element c4-4 drove expression in r8 with GFP fading into r7 to form a weak boundary at r6. The *hoxc4a* spinal cord expression domain was also represented by all three tissue-specific elements (yellow arrows). In addition, c4-2 and c4-4 regulated expression in the pectoral fins (pink arrows) and c4-4 further mediated expression in the pharyngeal arches (dark blue arrow).

In summary, although the *hoxc4a* locus has fewer conserved sequences, most of them act as tissue-specific enhancers mediating the main expression features of the gene.

Within the *hoxd4a* locus twelve CNEs were tested for *cis*-regulatory activity (Fig. 7) with one of them (d4-9) also tested as a

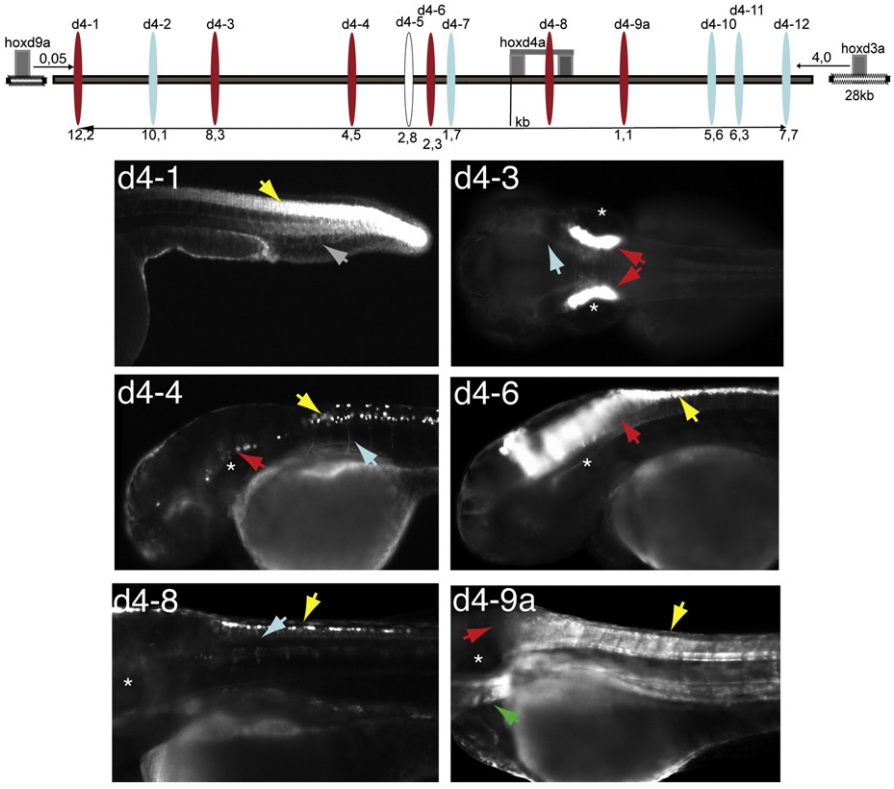

Fig. 7. Schematic of the *hoxd9a* to *hoxd3a* genomic region with the tested elements illustrated as described in Figs. 4 and 5. The arrow color code: red, hindbrain; yellow, spinal cord; grey, somites; light blue, neuronal projection; green, pharyngeal arches. Representative larvae with GFP expression under regulatory control of CNEs that surround *hoxd4a* are shown at 2 to 4 dpf. The asterisk marks the level of the inner ear. d4-1. Larva showing GFP expression in the posterior spinal cord and in the somites. d4-3. GFP expression in two bilateral columns in hindbrain r2 to r6, adjacent to the inner ear. d4-4. Larva with GFP expression in spinal cord motor neurons and their projections towards the somites. Some neurons in the hindbrain r4–6 are also GFP positive. d4-6. Larva with a strong expression boundary at cerebellum/midhindbrain boundary. Strong GFP expression in the hindbrain and in the dorsal spinal cord. d4-8. Element d4-8 is located in the intron of the gene. A larva with expression in the dorsal and ventral spinal cord with GFP expressing neurons projecting towards the somites. d4-9a. Larva shows GFP expression in the hindbrain with a week anterior boundary at r6/r7. The GFP distribution continues into the spinal cord and GFP expression is also visible in the pharyngeal arches.

longer fragment (compare Fig. 2, d4-9a). The *hoxd5a* to *hoxd8a* genes were lost in the zebrafish hoxda cluster, which might have resulted in a situation where retained regulatory sequences drive either the upstream located *hoxd9a* or the downstream located *hoxd4a* gene, maybe conferring unique expression features to both genes. Six elements regulated tissue-specific expression. The other elements (d4-2, d4-7, d4-10, d4-11) had unspecific activity and one element (d4-5) was negative. Element d4-1 is located only 50 bp downstream of the *hoxd9a* gene and drove expression in the posterior spinal cord (yellow arrow), most likely representing a *hoxd9a* expression domain, and further in the somites and pectoral fins. Element d4-3 had specific activity throughout r4 to r6 in two lateral columns adjacent to the inner ears (red arrows). From retrograde labeling studies in chick, frog and mice it is known that these regions include the vestibular nuclei that are innervated by the 8th nerve (Diaz and Glover, 2002; Pasqualetti et al., 2007; Straka et al., 2001). This is not an obvious feature of the *hoxd4a* gene, but it could be part of the broad expression pattern of *hoxd9a*. The d4-4 element drove expression in motor neurons within the spinal cord (yellow arrows) that sent projections towards the skeletal muscle (light blue arrow). An autoregulatory element (ARE) was described in the sequence that we have included in element d4-6 (Nolte et al., 2003). When tested, this element resulted in GFP expression in the hindbrain, cerebellum and spinal cord. Element d4-8 is located in the *hoxd4a* intron. This element drove consistent expression in r8 and in the spinal cord with labeled fibers that projected towards the adjacent myotome. In addition, activity in the *hoxd4a* expression domains pharyngeal arches and pectoral fins was observed. CNE d4-9 was tested as a 395 bp element. This fragment did not show regulatory activity. Because this element is located in the critical region that was identified previously as a hindbrain enhancer in mouse, as well as in zebrafish (Nolte et al., 2003, 2006), we aligned the d4-9 sequence with the 570 bp zebrafish enhancer sequence that was published. We realized that important 5′ flanking sequences including a DR5 RARE site were missing from the element that we had tested. The critical sequences are conserved only in fishes. When these were included into the fragment for regulatory testing, resulting in a 989 bp fragment (d4-9a — for primer and coordinates see Table suppl. 1), GFP was driven in the hindbrain, spinal cord and pharyngeal arches. The pattern did not show a strict anterior expression boundary; instead the GFP faded from r8 towards anterior through r7.

An overview of the main expression features that are mediated by the Hox4 CNEs is given in Fig. 8. The figure illustrates that the activity of the genes is mediated through several enhancers regulating broad expression in overlapping domains. To compare the regulatory information with enhancers that have been previously identified in the mouse, the comparable genomic fragments are indicated with a link to the literature.

Paralogous and orthologous relationships of Hox4 CNEs

Because the four Hox4 loci originated from a single Hox4 locus in ancestral chordates, functionally important sequences that surround an ancestral Hox4 gene might be evolutionary constrained in duplicated Hox4 loci. We therefore asked whether there are CNE similarities and if paralogous *cis*-regulatory sequences with similar functions could be identified. Multiple alignments of the paralogous zebrafish CNEs revealed that many of the conserved sequences are indeed very similar to each other. Three groups of CNEs could be established showing the highest level of similarity in between the paralogous loci. Their relationships are illustrated in Fig. 9, indicating that paralogous CNEs located 1–2 kb downstream of *hoxa5a*, *hoxb5a* and *hoxc5a* (a4-1, b4-1, c4-1) correspond to a CNE located approximately 8 kb downstream of *hoxd9a* (d4-4) (group “I” in Fig. 9). Whereas in the hoxda cluster this sequence regulates the 4.5 kb distant *hoxd4a* gene, in the other three clusters the paralogous elements are supposed to regulate the Hox5 and/or the Hox4 genes (Sharpe et al., 1998). Element d4-4, active in the spinal cord, therefore represents the beginning of the *hoxd4a* regulatory domain whereas elements d4-1 to d4-3 are likely to regulate the *hoxd9a* gene as this gene is widely expressed in the brain (zfin ID: ZDB-GENE-990415-121). We concluded that CNEs a4-1, b4-1, c4-1 and d4-4 are paralogous sequences, however, a *cis*-regulatory function could not be directly compared, because with exception of the d4-4 these elements did not function as tissue-specific enhancers. From previous studies (Hadrys et al., 2006) we know that the sequence b4-1 includes a promoter that initiates a long transcription unit containing exons of both genes, *hoxb4a* and *hoxb3a*. Taking this into account we suggest that the first Hox4 regulatory sequence (#1 in aa, ba, ca and #4 in da) could have been retained in all four loci, because this might include a promoter sequence for the initiation of a very long transcript starting downstream of Hox5 and including Hox4 and Hox3 exon sequences. The *hoxb4a*/*hoxb3a* transcript and promoter has been described and recently similar transcripts for *hoxa4a*/*hoxa3a* and *hoxc4a*/*hoxc3a* have been annotated in the Ensembl genome browser (www.ensembl.org). In the zebrafish hoxda cluster such a transcript is not described, but a comparable transcript is suggested in zebrafish since a long *Hoxd3* transcript starting downstream of *Hoxd8* (where *Hoxd5* to *Hoxd7* do not exist) is annotated in the mouse genome (ENSMUST00000111983). The four clusters share another conserved element upstream of the Hox4 genes, a4-4 in the *hoxa4a*, b4-5 in the *hoxb4a*-, c4-2 in the *hoxc4a* locus and d4-5 in the *hoxd4a* locus (group “II” in Fig. 9). A third group is formed by paralogous CNEs located in the intron sequences of the genes (“III” in Fig. 9), however, whereas there was strong conservation in the *hoxba*, *hoxca* and *hoxda* genes, we found this sequence only marginally conserved in *hoxa4a*. Further elements are either most conserved between the *hoxb4a* and *hoxd4a*, the *hoxb4a* and *hoxc4a* or the *hoxc4a* and *hoxd4a* loci. From this analysis we could group the clusters in two similarity groups: aa, ba, ca and ba, ca, da, revealing no striking CNE similarities between the *hoxa4a* locus and the *hoxd4a* locus.

We conducted a phylogenetic analysis with group I, II and III CNEs using the phylogeny.fr web-service (Dereeper et al., 2008).The orthologous sequences from human, mouse, tetraodon and xenopus were included. The phylogenetic trees for group II and III CNEs were not robust, due to short sequences and a lack of conservation in the *hoxa4a* intron. The tree resulting from alignments of group I CNEs revealed a very clear and robust picture, in which the a4-1 and b4-1 elements are grouped, while the c4-1 and d4-4 elements form a distinct sub-tree (Fig. Suppl. 3).

Finally we asked whether the most ancestral CNE might represent the ancient *cis*-regulatory sequence acting on Hox4. The *AmphiHox3* – *AmphiHox5* intergenic region from the cephalochordate amphioxus (*Branchiostoma floridae*) genome, which harbors the most closely extant representative of the ancestral, and extinct, chordate Hox cluster (Holland et al., 2008), was used for sequence comparisons with several vertebrate species, namely xenopus, zebrafish, fugu, horn shark, mouse, and human, using VISTA. An illustration of these comparisons is shown in supplementary Fig. 2. One amphioxus sequence block showed high sequence conservation within the intron of all vertebrate Hox4 genes (highlighted in the supplementary Fig. 2 and 10A by a red line and designated in the amphioxus sequence as bf4). Two further sequences in the amphioxus genome, upstream of the two *AmphiHox4* exons, were found to be conserved in zebrafish and fugu and were designated bf1 and bf2 (Fig. 10A, blue and green box). An additional conserved sequence in between bf2 and bf4 represents the previously described microRNA *mir-10* (Amemiya et al., 2008). A 674 bp Amphioxus genome fragment was amplified to test the element bf1 through zebrafish transgenesis. This element, only conserved with zebrafish *hoxa4a* and fugu *hoxa4a* and *hoxd4a*, had unspecific activity. A 548 bp fragment was amplified to test CNE bf2, but did not drive GFP expression. Element bf4, located in the

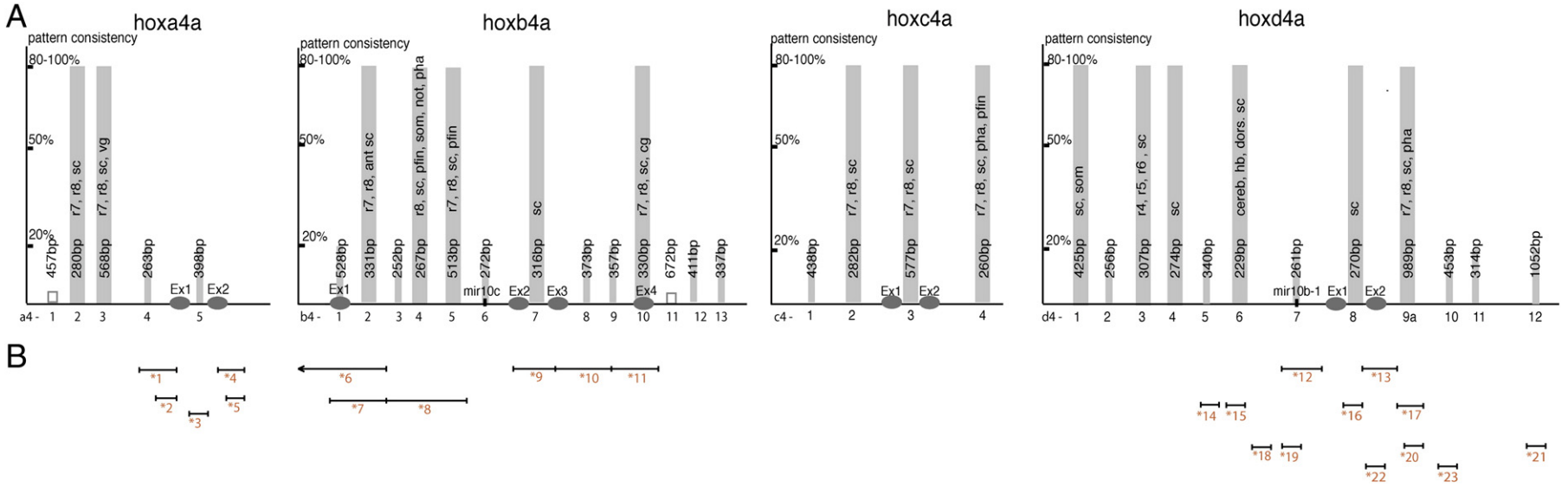

Fig. 8. A. Overview about the CNEs that were tested from the four Hox clusters. The diagram illustrates how the main expression features are regulated by CNEs functioning as tissue-specific enhancers. The *hoxa4a*, *hoxb4a*, *hoxc4a* and *hoxd4a* genes are shown on the *X*-axis with the grey ovals indicating the exon sequences. The grey bars illustrate the tested elements and their genomic localization relative to the gene. Slim, short bars indicate CNEs that resulted in unspecific or inconsistent GFP expression patterns. Long, broad bars indicate CNEs that had specific cis-regulatory functions — these include abbreviations of the expression features. Short, empty squares indicate negative elements. B. Enhancers previously identified in the mouse are indicated in the genomic positions that are comparable to the zebrafish loci. The number below points to the article in which the enhancers were described. *1, 2, 3, 4, 5: (Behringer et al., 1993): 5 kb in mouse, neuronal expression; 3 kb in mouse, mesodermal expression; 500 bp in mouse, CR1 element, no enhancer function; 3.3 kb in mouse, neuronal and mesodermal expression; 2.3 kb in mouse mesodermal expression. *6, 7, 8: (Sharpe et al., 1998): 6.1 kb in mouse, *Hoxb5* specific expression; 3.9 kb in mouse, Region E, neuronal expression; 4.5 kb in mouse, Region D, mesodermal expression. *9, 10, 11: (Brend et al., 2003; Gould et al., 1998, 1997; Whiting et al., 1991): 1 kb in mouse, Region C, CR1 element, posterior CNS and mesoderm; 1.45 kb in mouse, Region B, lung expression, somitic expression; 3 kb in mouse, Region A, 3′ RARE, hindbrain with anterior r6/7 boundary and spinal cord expression. *12: (Morrison et al., 1996): 2.9 kb human sequence tested in mouse, neuronal expression with anterior boundary in the rostral spinal cord. *13: (Morrison et al., 1997): 1.2 kb human sequence tested in mouse, *Hoxd4* specific mesodermal expression. *14: (Nolte et al., 2003; Popperl and Featherstone, 1992; Zhang et al., 1997): human *HOXD4* 5′ flanking sequence, murine 5′ RARE element, tested in mouse, mesodermal expression. *15: (Nolte et al., 2003): ARE auto regulatory element. *16: (Morrison et al., 1997; Nolte et al., 2003): human *HOXD4* intron element, CR1 element, no independent expression, consensus binding site for many homeodomain containing proteins, consensus binding site for CDX homeoprotein. *17: (Morrison et al., 1997; Zhang et al., 1997): 700 bp human 3′ *HOXD4* sequence, tested in mouse, *Hoxd4* specific neuronal expression. *18, 19: (Folberg et al., 1997; Nolte et al., 2003): 5.2 kb and 1.1 kb upstream of *Hoxd4* coding region located promoters, P1 and P2 respectively, P1 promotor conserved in zebrafish sequence. *20: (Nolte et al., 2003, 2006; Rastegar et al., 2004; Zhang et al., 2000): 579 bp zebrafish sequence tested in mouse, Region A, hindbrain and spinal cord specific *Hoxd4* expression, 3′ neuronal enhancer is required for retinoic acid response, contains 3′ RARE, required for initiation and maintenance of *Hoxd4* expression, significant higher and complete expression pattern only with enhancer and *Hoxd4* promoter, *Pax6* binding region and positive regulation of *Hoxd4a* expression in mice. *21: (Visel et al., 2009): sequence in mouse contains p300 element. *22, 23: (Brend et al., 2003): 1.4 kb in mouse likely to be somite 5/6 enhancer; 0.9 kb in mouse likely to be somite 5/6 enhancer. Abbreviations: cg, cranial ganglia; not, notochord; pha, pharyngeal arches; pfin, pectoral fin; r, rhombomere; sc, spinal cord; som, somites.

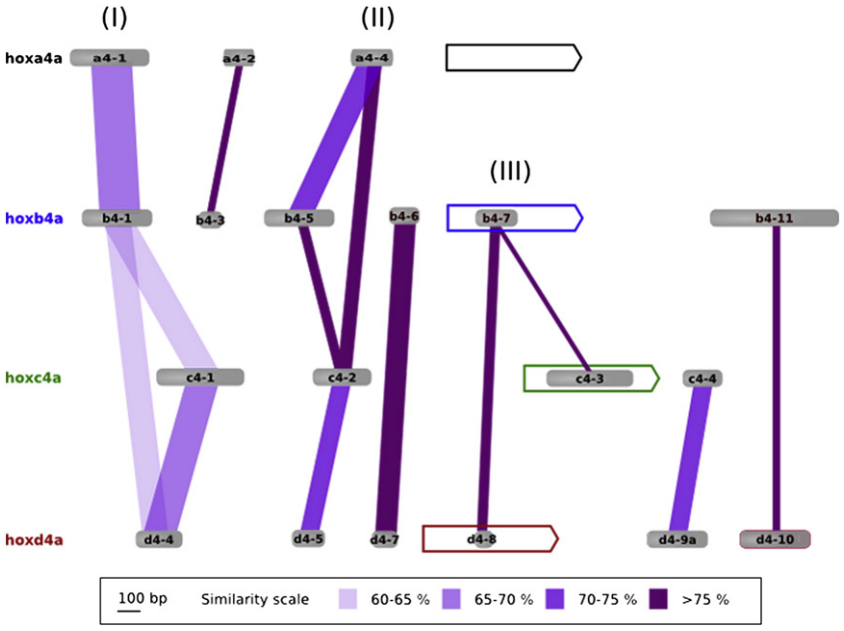

Fig. 9. Paralogous relationships between zebrafish CNEs. Lines represent local alignments between CNEs represented as grey boxes with a *p*-value <0.1 as given by the YASS alignment software. The width of the line is proportional to the length of the alignment, and the color represents the degree of similarity. The position of the genes is indicated by colored frames. Note that the relative spacing between elements does not reflect their genomic location.

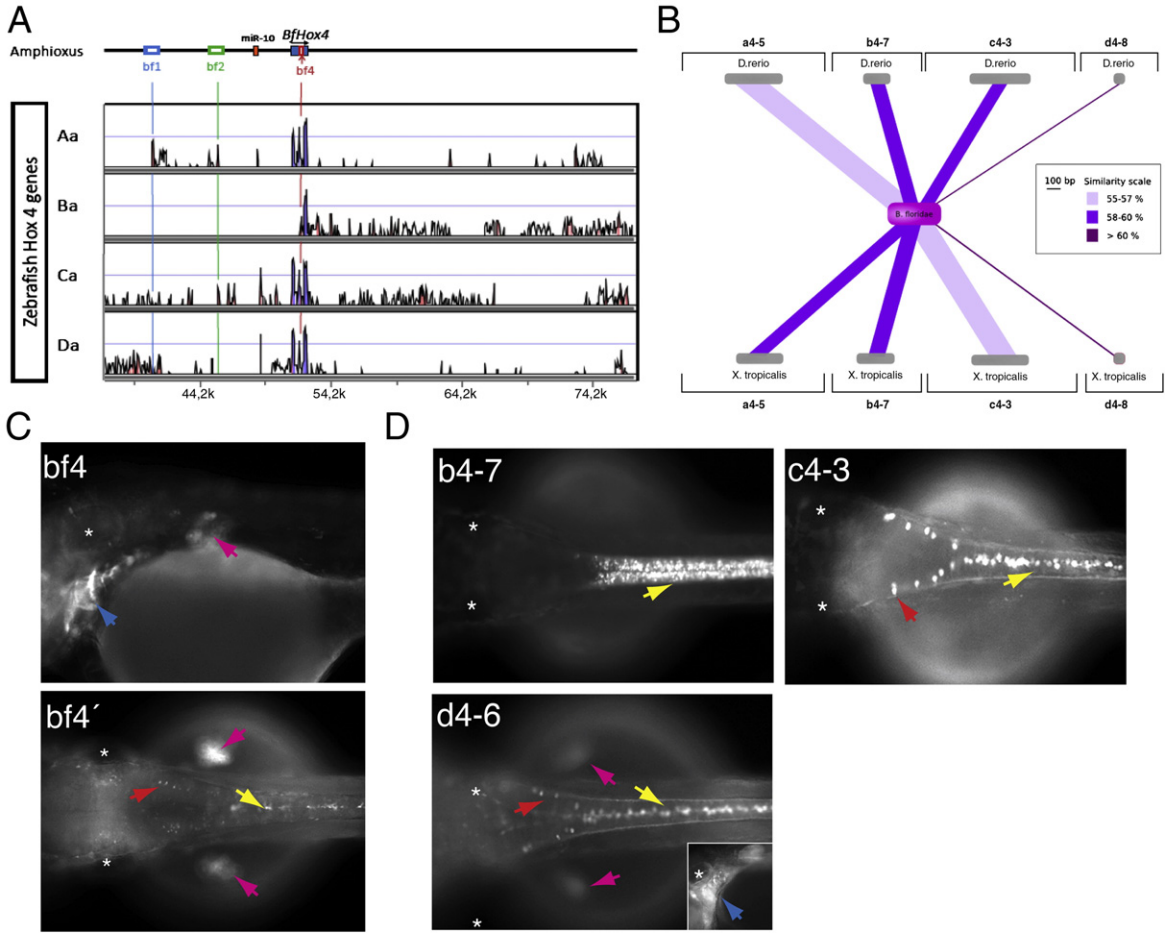

Fig. 10. A. Three conserved sequences, identified through VISTA alignments of the zebrafish Hox4 loci (bf1, bf2, bf4). B. Alignments of the amphioxus intronic element with the four paralogous zebrafish intronic elements (upper part) and the frog intronic elements (lower part). The width of the line is proportional to the length of the local alignment, the color indicates the level of similarity. C. The intronic element of the *AmphiHox4* gene, bf4, was tested in zebrafish transgenesis. Shown is are larvae with GFP expression in the pharyngeal arches (blue arrow) and the pectoral fins (pink arrows) as well as hindbrain (red arrow) and spinal cord (yellow arrow). D. For comparison the expression features that are regulated through the tested zebrafish intron elements are shown. The arrows use the following designated color code: red, hindbrain; yellow, spinal cord; blue, pharyngeal arches and pink, pectoral fins.

intron of *AmphiHox4*, was tested as a 374 bp fragment. We evaluated its activity as unspecific, because the GFP expression features were not consistent in the different lines, even though we observed some Hox-specific expression features such as hindbrain (Fig. 10C, red arrow), spinal cord (yellow arrow), pharyngeal arches (blue arrow) and pectoral fins (pink arrow). For comparison Fig. 10C shows the regulatory activity of the zebrafish Hox4 intronic elements — b4-7 in the spinal cord, c4-3 in the lateral hindbrain r7, r8 and in the spinal cord, d4-8 in the spinal cord, while a4-5 was unspecific.

Because of the high conservation of CNE bf4 and its presence in all vertebrate Hox4 genes as well as in *AmphiHox4* and even in *Drosophila* (Haerry and Gehring, 1997; Lou et al., 1995), we analyzed the similarity scale of the amphioxus and vertebrate intron CNEs (compare group III CNEs in Fig. 9) and further the presence of transcription factor binding sites according to software and motif databases that are described in detail in the Materials and methods section. The performed multiple alignments included all paralogous sequences from zebrafish and orthologous sequences from human, mouse, chicken, tetraodon and xenopus. For illustration, the similarity of *AmphiHox*4 CNE to xenopus and zebrafish Hox4-intron CNEs is indicated in Fig. 10B. Through the analysis an AT-rich conserved motif became obvious (Suppl. Fig. 4A), which is located within a previously described 72 bp sequence block identified as CB1 in mouse, chicken and puffer fish *Hoxb4* genes (Morrison et al., 1995). Two clustered Cdx-like binding motifs were identified, with a more contrasted conservation pattern among paralogs. Upstream of this Cdx cluster is a generic homeodomain binding motif (AATTA). Both intronic motifs (the homeodomain binding site and the Cdx cluster) were found previously to be conserved in the mouse *Hoxa4*, *Hoxb4*, *Hoxa7* and *Fugu Hoxb4* genes (Nolte et al., 2003).

Further multiple alignments of the paralogous group I and II CNEs (see Fig. 9) revealed that a Cdx-like transcription factor binding site can be also identified in group I CNEs. This highly conserved region includes two Cdx-like binding motifs (CCATTAAA and TTTAATGG) and further an E-box (CACGTG), representing a potential bHLH-binding motif, which is conserved in all aligned sequences (Suppl. Fig. 4B). While the second Cdx motif is strictly conserved, the first instance appears to have been lost in the *hoxa4a*, *hoxc4a* and *hoxd4a* CNEs due to an insertion in the third position of the motif, while it is present in the *hoxb4a* CNEs. It has been shown that *cdx1a*/*cdx4* indeed negatively regulate *hoxd4a*, when *cdx1a*/*cdx4* function was abolished *hoxd4a* expressed strongly and broadly in the spinal cord (Skromne et al., 2007). From the amphioxus comparative analyses it can be concluded that an ancestral Cdx-like binding motif located in the intron of Hox4 genes could have been retained or adopted by other conserved regulatory sequence(s) during evolution. Within group II CNEs a retinoic acid response element (RARE) consensus sequence was identified that has been described previously in mouse and zebrafish Hoxd4 genes (Nolte et al., 2003). These are indicated in figure supplement 4C. Further AT-rich motifs (Fig. Suppl. 4C) that are located downstream adjacent likely represent additional functional elements. Using the TRANSFAC based PROMO software many transcription factor binding sites were predicted with those enriched that bind molecules of the Pou and Hox families.

Discussion

Hox genes encode transcription factors with multiple functions during embryonic development and act to define morphology, cell identities and neuronal subtypes throughout the animal kingdom. Even though Hox genes have species-specific mutational phenotypes, the coding sequence of one gene can often substitute for the coding sequence of orthologous or even paralogous genes (Calonge et al., 2007; Horan et al., 1995a,b; Malicki et al., 1990; Manley and Capecchi, 1997, 1998; McGinnis et al., 1990), suggesting that, to a large degree, differences between Hox genes are found within their spatiotemporal (*cis*-) regulation.

Hox clusters are tightly arranged, and regulatory control regions located outside the clusters (Spitz et al., 2005; Tumpel et al., 2002, 2006), shared transcription units (Hadrys et al., 2006) as well as chromatin modifications (Soshnikova and Duboule, 2009) might suffice to explain the colinearity of the vertebrate Hox clusters. Even though single enhancers have been tested previously in reporter gene constructs (Hadrys et al., 2006; Kwan et al., 2001; Manzanares et al., 1999; Morrison et al., 1997; Zhang et al., 1997), it was thought that regulatory elements within the Hox clusters need their genomic context to function properly.

Our study contributes to this topic by defining the conserved regulatory content of the genomic region surrounding the Hox4 genes in zebrafish — first predicting putative enhancers through highest sequence conservation and then, using the main Hox4 gene expression features (hindbrain r7 and r8, spinal cord and pharyngeal arches), systematically testing the CNEs in reporter gene constructs in stable transgenesis. This study thereby established a list of the most conserved CNEs in the Hox4 loci and assessed their *cis*-regulatory function (if any). Establishing their paralogous relationships through multiple sequence alignments revealed that even though sequence similarities were retained during evolution, the CNEs have evolved in their function, underscoring the complexity of *cis*-regulatory mechanisms in the Hox clusters.

Hox4 gene activity is regulated through multiple regulatory sequences that act in an overlapping manner

About 50% of the tested sequences conserved from fish to human had enhancer activity. Especially from the *hoxb4a* and *hoxd4a* gene loci we could extrapolate that a broad range of control sequences contribute to the regulation of the Hox4 genes. As observed earlier to some extent in mouse (Morrison et al., 1997; Zhang et al., 1997) one surprising outcome of our study was that certain expression domains are mostly regulated through several enhancers. Such redundancies and overlapping activities of regulatory sequences have been hypothesized to ensure specific transcripts dosages and the coordination of highly specific expression patterns (Gavalas et al., 2003; Gerard et al., 1996; Komisarczuk et al., 2009; Sharpe et al., 1998).

The *hoxb4a* and *hoxd4a* genes appeared to be regulated in a more complex way when compared to *hoxa4a* and *hoxc4a*, and a higher number of regulatory sequences seem to control these genes. In the *hoxa4a* and the *hoxc4a* loci, only a few CNEs were found and their consistent activities seem to mediate a simpler and clearer regulation of the expression pattern of the endogenous gene. For example *hoxa4a* element a4-3 directed strong and robust expression to the posterior hindbrain, including motoneurons innervating the fin (compare Fig. 5), comparable to an enhancer trap line carrying an insertion 0.8 kb upstream of *hoxa4a* (Punnamoottil et al., 2008). The *hoxc4a* pattern appeared to be mainly mediated through c4-2 in the posterior hindbrain and pectoral fins and through c4-4 in pharyngeal arches, pectoral fins, posterior hindbrain and spinal cord (compare Fig. 6). Despite the overlapping activity of the majority of the tested enhancers in the *hoxb4a* locus, some elements clearly showed unique functions resulting in different expression at cellular levels. For example elements b4-5 and b4-10 were both active in the developing posterior hindbrain, where one element drove GFP expression in neurons located dorsally and laterally (Fig. 5, b4-5′), and the other drove expression in neurons located medially (Fig. 5, b4-10′).

Elements from the hoxda cluster revealed unexpected regulatory activities as for instance elements d4-1 and d4-3 likely regulated expression features of the widely expressed *hoxd9a* gene (published at ZFIN database), but remarkably both enhancers were very specific (Fig. 7). Testing of element d4-4 revealed *hoxd4a* specific expression features and, as revealed by establishing paralogous relationships, this

sequence was defined as the start of the *hoxd4a* regulatory domain. This element had activity in the spinal cord. The hindbrain expression domain was mediated by three tested sequences d4-6, d4-8 and d4-9a. Even though a 5′ RARE sequence is included in element d4-5, this element did not have *cis*-regulatory activity by itself. With element d4-9a we have tested an element that includes a previously identified neuronal enhancer of *hoxd4a*. Both, mouse and zebrafish sequences, were described to include a 3′RARE and regulated hindbrain r7 and r8 and spinal cord expression when tested in mice (Nolte et al., 2003). Substitutions within the RARE sequence did affect the neuronal expression, but did not exclusively shift the anterior expression boundary. Testing of element d4-9a that contained, in addition to the core sequence conservation (human:zebrafish), further sequences conserved only in fishes (pufferfish:zebrafish) including the 3′RARE resulted in GFP expression in r7, r8, spinal cord and branchial arches. A shorter fragment, CNE d4-9, missing the 3′RARE did not have regulatory activity and revealed the importance of species-specific sequence conservation in enhancer prediction.

The non-reproducible activity of many elements suggests that the high conservation threshold that we chose as a basis for testing might have excluded additional regulatory sequences flanking the CNEs. Some elements might be more dependent than others on the genomic context in which they land resulting in position-specific functions. Further, *cis*-regulatory sequences often interact with each other to generate complete *cis*-regulatory modules as recently demonstrated for notochord-specific enhancers of the *ctgf* and *sox9a* genes (Rastegar et al., 2008). To protect 5′ HoxD genes from the regulatory influence of the adjacent *Evx2* gene, two insulator sequences are required to act in combination (Kmita et al., 2002b).

Evolutionary relationships of the Hox4 genomic sequences

The four vertebrate Hox clusters originated through 2 rounds of genome duplications from a single ancestral Hox cluster (Fig. 1A). Some studies group the A- and D-cluster and the B- and C-cluster together (Bailey et al., 1997), whereas others favor the split of an ancestral Hox cluster into A and B followed by a further split of the A-cluster into A and C and the B-cluster into B and D (for overview see (Kuraku and Meyer, 2009). The total number of CNEs is different among the Hox4 paralogs: while we found up to 12 CNEs in the *hoxb4a* and *hoxd4a* loci, only four could be discerned in the *hoxa4a* and *hoxc4a* loci, which, at first approximation, might imply a higher similarity of the ba- and da-clusters. Based on the highest level of conservation, we established paralogous relationships for CNEs in the Hox4 loci as illustrated in the diagram in Fig. 9. Accordingly, we conclude that the first Hox4 regulatory sequence (#1 in aa, ba, ca and #4 in da) could have been retained in all four loci, because this might be a promoter/regulatory sequence for the initiation of a very long transcript starting downstream of Hox5 and including Hox4 and Hox3 exon sequences. Four elements, a4-4, b4-5, c4-2 and d4-5, are closely related and were defined as group II CNEs. Whereas the *hoxb4a* and *hoxc4a* elements mediated similar tissue-specific expression in the hindbrain r7 and r8, the *hoxa4a* element had unspecific *cis*-regulatory activity and d4-5 did not direct reporter expression at all. Further elements are most conserved between two loci, *hoxb4a* and *hoxd4a*, *hoxb4a* and *hoxc4a* and *hoxc4a* and *hoxd4a*. Many of these elements had unspecific or no activity revealing that retaining sequence conservation cannot be directly correlated with tissue-specific *cis*-regulatory activity as was also suggested previously (Kmita et al., 2002b). From the phylogenetic analysis of group I CNEs two similarity groups of the Hox4 loci have emerged: aa, ba, ca and ba, ca, da, revealing no high CNE similarities between the *hoxa4a* locus and the *hoxd4a* locus. This could be explained by an evolutionary scenario where the ancestral cluster first split into B and C followed by a split of the clusters into B and D and C and A having either B/A and D/C or D/B and AC as sister clusters. However the phylogenetic tree that we have established for group I CNEs favours D/C and B/A sister clusters.

The intronic CNEs (group III) seem to harbor an ancestral Hox4 paralog enhancer present even in Drosophila (Haerry and Gehring, 1997) and, as shown here, in the cephalochordate amphioxus. Since the tested *AmphiHox4* element that included just the strongest sequence conservation drove only inconsistent GFP expression in zebrafish transgenes, the original regulatory relevance of this element in the chordate ancestor cannot be deduced. In vertebrates *Hoxb4* and *Hoxd4* intron sequences were identified as enhancers regulating mesoderm, spinal cord and neural crest expression (Aparicio et al., 1995; Brend et al., 2003; Gould et al., 1997; Morrison et al., 1995) including a highly conserved sequence block with two Hox binding sites (Gehring et al., 1994; Morrison et al., 1995). The presence of Cdx binding sites in both, group III and I CNEs (Fig. Suppl. 4), might imply that particular CNEs may have been retained or recruited into new evolutionary contexts and as such may be responsible for the recruitment of developmental gene networks to new structures and morphological innovations.

A major conclusion drawn from our analysis is that even though *cis*-regulatory sequences in the Hox clusters have similarities and are conserved during evolution, their functions have significantly evolved, suggesting that novel expression features might arise through modification of conserved enhancers, in line with recent findings in insect dorsoventral patterning (Cande et al., 2009).

Acknowledgments

This work was supported by a PhD fellowship and a grant from the Deutsche Forschungsgemeinschaft to SR and Ministerior de Educación y Ciencia BMC2008-03776 to JGF. SR and TSB were supported by a core grant from the Sars Centre, and by a grant from the Institut du Cerveau et de la Moelle Epinière, Paris, France to TSB. CH thanks Céline Brochier for useful discussions.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ydbio.2010.01.035.

References

Ahituv, N., Zhu, Y., Visel, A., Holt, A., Afzal, V., Pennacchio, L.A., Rubin, E.M., 2007. Deletion of ultraconserved elements yields viable mice. PLoS Biol. 5, e234.

Amemiya, C.T., Prohaska, S.J., Hill-Force, A., Cook, A., Wasserscheid, J., Ferrier, D.E., Pascual-Anaya, J., Garcia-Fernandez, J., Dewar, K., Stadler, P.F., 2008. The amphioxus Hox cluster: characterization, comparative genomics, and evolution. J. Exp. Zoolog. B Mol. Dev. Evol. 310, 465–477.

Amores, A., Force, A., Yan, Y.L., Joly, L., Amemiya, C., Fritz, A., Ho, R.K., Langeland, J., Prince, V., Wang, Y.L., Westerfield, M., Ekker, M., Postlethwait, J.H., 1998. Zebrafish hox clusters and vertebrate genome evolution. Science 282, 1711–1714.

Aparicio, S., Morrison, A., Gould, A., Gilthorpe, J., Chaudhuri, C., Rigby, P., Krumlauf, R., Brenner, S., 1995. Detecting conserved regulatory elements with the model genome of the Japanese puffer fish, Fugu rubripes. Proc. Natl. Acad. Sci. U. S. A. 92, 1684–1688.

Arenkiel, B.R., Tvrdik, P., Gaufo, G.O., Capecchi, M.R., 2004. Hoxb1 functions in both motoneurons and in tissues of the periphery to establish and maintain the proper neuronal circuitry. Genes Dev. 18, 1539–1552.

Bailey, W.J., Kim, J., Wagner, G.P., Ruddle, F.H., 1997. Phylogenetic reconstruction of vertebrate Hox cluster duplications. Mol. Biol. Evol. 14, 843–853.

Behringer, R.R., Crotty, D.A., Tennyson, V.M., Brinster, R.L., Palmiter, R.D., Wolgemuth, D.J., 1993. Sequences 5′ of the homeobox of the Hox-1.4 gene direct tissue-specific expression of lacZ during mouse development. Development 117, 823–833.

Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S., Haussler, D., 2004. Ultraconserved elements in the human genome. Science 304, 1321–1325.

Berger, M.F., Badis, G., Gehrke, A.R., Talukder, S., Philippakis, A.A., Pena-Castillo, L., Alleyne, T.M., Mnaimneh, S., Botvinnik, O.B., Chan, E.T., Khalid, F., Zhang, W., Newburger, D., Jaeger, S.A., Morris, Q.D., Bulyk, M.L., Hughes, T.R., 2008. Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. Cell 133, 1266–1276.

Bourlat, S.J., Juliusdottir, T., Lowe, C.J., Freeman, R., Aronowicz, J., Kirschner, M., Lander, E.S., Thorndyke, M., Nakano, H., Kohn, A.B., Heyland, A., Moroz, L.L., Copley, R.R.,

Telford, M.J., 2006. Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. Nature 444, 85–88.
Brend, T., Gilthorpe, J., Summerbell, D., Rigby, P.W., 2003. Multiple levels of transcriptional and post-transcriptional regulation are required to define the domain of Hoxb4 expression. Development 130, 2717–2728.
Brooke, N.M., Garcia-Fernandez, J., Holland, P.W., 1998. The ParaHox gene cluster is an evolutionary sister of the Hox gene cluster. Nature 392, 920–922.
Calonge, W.M., Martinez, L., Lacadena, J., Fernandez-Dumont, V., Matesanz, R., Tovar, J.A., 2007. Expression of homeotic genes Hoxa3, Hoxb3, Hoxd3 and Hoxc4 is decreased in the lungs but not in the hearts of adriamycin-exposed mice. Pediatr. Surg. Int. 23, 419–424.
Cande, J., Goltsev, Y., Levine, M.S., 2009. Conservation of enhancer location in divergent insects. Proc. Natl. Acad. Sci. U. S. A. 106, 14414–14419.
Condie, B.G., Capecchi, M.R., 1994. Mice with targeted disruptions in the paralogous genes hoxa-3 and hoxd-3 reveal synergistic interactions. Nature 370, 304–307.
Dasen, J.S., Liu, J.P., Jessell, T.M., 2003. Motor neuron columnar fate imposed by sequential phases of Hox-c activity. Nature 425, 926–933.
de la Calle-Mustienes, E., Feijoo, C.G., Manzanares, M., Tena, J.J., Rodriguez-Seguel, E., Letizia, A., Allende, M.L., Gomez-Skarmeta, J.L., 2005. A functional survey of the enhancer activity of conserved non-coding sequences from vertebrate Iroquois cluster gene deserts. Genome Res. 15, 1061–1072.
Delsuc, F., Baurain, D., Philippe, H., 2006. Vertebrate origins: does the tunic make the man? Med. Sci. (Paris) 22, 688–690.
Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevente, F., Dufayard, J.F., Guindon, S., Lefort, V., Lescot, M., Claverie, J.M., Gascuel, O., 2008. Phylogeny.fr: robust phylogenic analysis for non-specialist. Nucleic Acids Res. 1, W465–W469.
Deschamps, J., van Nes, J., 2005. Developmental regulation of the Hox genes during axial morphogenesis in the mouse. Development 132, 2931–2942.
Di-Poi, N., Zakany, J., Duboule, D., 2007. Distinct roles and regulations for hoxd genes in metanephric kidney development. PLoS Genet. 3, e232.
Diaz, C., Glover, J.C., 2002. Comparative aspects of the hodological organization of the vestibular nuclear complex and related neuron populations. Brain Res. Bull. 57, 307–312.
Duboule, D., 2007. The rise and fall of Hox gene clusters. Development 134, 2549–2560.
Ferrier, D.E., Minguillon, C., Holland, P.W., Garcia-Fernandez, J., 2000. The amphioxus Hox cluster: deuterostome posterior flexibility and Hox14. Evol. Dev. 2, 284–293.
Folberg, A., Kovacs, E.N., Featherstone, M.S., 1997. Characterization and retinoic acid responsiveness of the murine Hoxd4 transcription unit. J. Biol. Chem. 272, 29151–29157.
Frith, M.C., Saunders, N.F.W., Kobe, B., Bailey, T.L., 2008. Discovering sequence motifs with arbitrary insertions and deletions. PLoS Comput. Biol. 4, e1000071.
Garcia-Fernandez, J., Holland, P.W., 1994. Archetypal organization of the amphioxus Hox gene cluster. Nature 370, 563–566.
Gaufo, G.O., Thomas, K.R., Capecchi, M.R., 2003. Hox3 genes coordinate mechanisms of genetic suppression and activation in the generation of branchial and somatic motoneurons. Development 130, 5191–5201.
Gaufo, G.O., Wu, S., Capecchi, M.R., 2004. Contribution of Hox genes to the diversity of the hindbrain sensory system. Development 131, 1259–1266.
Gavalas, A., Ruhrberg, C., Livet, J., Henderson, C.E., Krumlauf, R., 2003. Neuronal defects in the hindbrain of Hoxa1, Hoxb1 and Hoxb2 mutants reflect regulatory interactions among these Hox genes. Development 130, 5663–5679.
Gehring, W.J., Qian, Y.Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A.F., Resendez-Perez, D., Affolter, M., Otting, G., Wuthrich, K., 1994. Homeodomain–DNA recognition. Cell 78, 211–223.
Gerard, M., Chen, J.Y., Gronemeyer, H., Chambon, P., Duboule, D., Zakany, J., 1996. In vivo targeted mutagenesis of a regulatory element required for positioning the Hoxd-11 and Hoxd-10 expression boundaries. Genes Dev. 10, 2326–2334.
Gould, A., Itasaki, N., Krumlauf, R., 1998. Initiation of rhombomeric Hoxb4 expression requires induction by somites and a retinoid pathway. Neuron 21, 39–51.
Gould, A., Morrison, A., Sproat, G., White, R.A., Krumlauf, R., 1997. Positive cross-regulation and enhancer sharing: two mechanisms for specifying overlapping Hox expression patterns. Genes Dev. 11, 900–913.
Gupta, S., Stamatoyannopolous, J.A., Bailey, T., Noble, W.S., 2007. Quantifying similarity between motifs. Genome Biol. 8, R24.
Hadrys, T., Punnamoottil, B., Pieper, M., Kikuta, H., Pezeron, G., Becker, T.S., Prince, V., Baker, R., Rinkwitz, S., 2006. Conserved co-regulation and promoter sharing of hoxb3a and hoxb4a in zebrafish. Dev. Biol. 297, 26–43.
Haerry, T.E., Gehring, W.J., 1997. A conserved cluster of homeodomain binding sites in the mouse Hoxa-4 intron functions in Drosophila embryos as an enhancer that is directly regulated by Ultrabithorax. Dev. Biol. 186, 1–15.
Hoegg, S., Meyer, A., 2005. Hox clusters as models for vertebrate genome evolution. Trends Genet. 21, 421–424.
Holland, L.Z., Albalat, R., Azumi, K., Benito-Gutierrez, E., Blow, M.J., Bronner-Fraser, M., Brunet, F., Butts, T., Candiani, S., Dishaw, L.J., Ferrier, D.E., Garcia-Fernandez, J., Gibson-Brown, J.J., Gissi, C., Godzik, A., Hallbook, F., Hirose, D., Hosomichi, K., Ikuta, T., Inoko, H., Kasahara, M., Kasamatsu, J., Kawashima, T., Kimura, A., Kobayashi, M., Kozmik, Z., Kubokawa, K., Laudet, V., Litman, G.W., McHardy, A.C., Meulemans, D., Nonaka, M., Olinski, R.P., Pancer, Z., Pennacchio, L.A., Pestarino, M., Rast, J.P., Rigoutsos, I., Robinson-Rechavi, M., Roch, G., Saiga, H., Sasakura, Y., Satake, M., Satou, Y., Schubert, M., Sherwood, N., Shiina, T., Takatori, N., Tello, J., Vopalensky, P., Wada, S., Xu, A., Ye, Y., Yoshida, K., Yoshizaki, F., Yu, J.K., Zhang, Q., Zmasek, C.M., de Jong, P.J., Osoegawa, K., Putnam, N.H., Rokhsar, D.S., Satoh, N., Holland, P.W., 2008. The amphioxus genome illuminates vertebrate origins and cephalochordate biology. Genome Res. 18, 1100–1111.
Horan, G.S., Kovacs, E.N., Behringer, R.R., Featherstone, M.S., 1995a. Mutations in paralogous Hox genes result in overlapping homeotic transformations of the axial skeleton: evidence for unique and redundant function. Dev. Biol. 169, 359–372.
Horan, G.S., Ramirez-Solis, R., Featherstone, M.S., Wolgemuth, D.J., Bradley, A., Behringer, R.R., 1995b. Compound mutants for the paralogous hoxa-4, hoxb-4, and hoxd-4 genes show more complete homeotic transformations and a dose-dependent increase in the number of vertebrae transformed. Genes Dev. 9, 1667–1677.
Kessel, M., Gruss, P., 1991. Homeotic transformations of murine vertebrae and concomitant alteration of Hox codes induced by retinoic acid. Cell 67, 89–104.
Kiecker, C., Lumsden, A., 2005. Compartments and their boundaries in vertebrate brain development. Nat. Rev. Neurosci. 6, 553–564.
Kimmel, C.B., Ballard, W.W., Kimmel, S.R., Ullmann, B., Schilling, T.F., 1995. Stages of embryonic development of the zebrafish. Dev. Dyn. 203, 253–310.
Kmita, M., Fraudeau, N., Herault, Y., Duboule, D., 2002a. Serial deletions and duplications suggest a mechanism for the collinearity of Hoxd genes in limbs. Nature 420, 145–150.
Kmita, M., Tarchini, B., Duboule, D., Herault, Y., 2002b. Evolutionary conserved sequences are required for the insulation of the vertebrate Hoxd complex in neural cells. Development 129, 5521–5528.
Kmita, M., Tarchini, B., Zakany, J., Logan, M., Tabin, C.J., Duboule, D., 2005. Early developmental arrest of mammalian limbs lacking HoxA/HoxD gene function. Nature 435, 1113–1116.
Komisarczuk, A., Kawakami, K., Becker, T., 2009. *Cis*-regulation and chromosomal rearrangement of the fgf8 locus after the teleost/tetrapod split. Dev. Biol.
Krumlauf, R., Marshall, H., Studer, M., Nonchev, S., Sham, M.H., Lumsden, A., 1993. Hox homeobox genes and regionalisation of the nervous system. J. Neurobiol. 24, 1328–1340.
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, 765–773.
Kwan, C.T., Tsang, S.L., Krumlauf, R., Sham, M.H., 2001. Regulatory analysis of the mouse Hoxb3 gene: multiple elements work in concert to direct temporal and spatial patterns of expression. Dev. Biol. 232, 176–190.
Lou, L., Bergson, C., McGinnis, W., 1995. Deformed expression in the Drosophila central nervous system is controlled by an autoactivated intronic enhancer. Nucleic Acids Res. 23, 3481–3487.
Malicki, J., Schughart, K., McGinnis, W., 1990. Mouse Hox-2.2 specifies thoracic segmental identity in Drosophila embryos and larvae. Cell 63, 961–967.
Manley, N.R., Capecchi, M.R., 1997. Hox group 3 paralogous genes act synergistically in the formation of somitic and neural crest-derived structures. Dev. Biol. 192, 274–288.
Manley, N.R., Capecchi, M.R., 1998. Hox group 3 paralogs regulate the development and migration of the thymus, thyroid, and parathyroid glands. Dev. Biol. 195, 1–15.
Manzanares, M., Cordes, S., Ariza-McNaughton, L., Sadl, V., Maruthainar, K., Barsh, G., Krumlauf, R., 1999. Conserved and distinct roles of kreisler in regulation of the paralogous Hoxa3 and Hoxb3 genes. Development 126, 759–769.
Manzanares, M., Cordes, S., Kwan, C.T., Sham, M.H., Barsh, G.S., Krumlauf, R., 1997. Segmental regulation of Hoxb-3 by kreisler. Nature 387, 191–195.
Manzanares, M., Nardelli, J., Gilardi-Hebenstreit, P., Marshall, H., Giudicelli, F., Martinez-Pastor, M.T., Krumlauf, R., Charnay, P., 2002. Krox20 and kreisler co-operate in the transcriptional control of segmental expression of Hoxb3 in the developing hindbrain. Embo J. 21, 365–376.
McGaughey, D.M., Vinton, R.M., Huynh, J., Al-Saif, A., Beer, M.A., McCallion, A.S., 2008. Metrics of sequence constraint overlook regulatory sequences in an exhaustive analysis at phox2b. Genome Res. 18, 252–260.
McGinnis, N., Kuziora, M.A., McGinnis, W., 1990. Human Hox-4.2 and Drosophila deformed encode similar regulatory specificities in Drosophila embryos and larvae. Cell 63, 969–976.
Monteiro, A.S., Ferrier, D.E., 2006. Hox genes are not always colinear. Int. J. Biol. Sci. 2, 95–103.
Morrison, A., Ariza-McNaughton, L., Gould, A., Featherstone, M., Krumlauf, R., 1997. HOXD4 and regulation of the group 4 paralog genes. Development 124, 3135–3146.
Morrison, A., Chaudhuri, C., Ariza-McNaughton, L., Muchamore, I., Kuroiwa, A., Krumlauf, R., 1995. Comparative analysis of chicken Hoxb-4 regulation in transgenic mice. Mech. Dev. 53, 47–59.
Morrison, A., Moroni, M.C., Ariza-McNaughton, L., Krumlauf, R., Mavilio, F., 1996. In vitro and transgenic analysis of a human HOXD4 retinoid-responsive enhancer. Development 122, 1895–1907.
Navratilova, P., Fredman, D., Hawkins, T., Edwards, K., Lenhard, B., Becker, T., 2009. Systematic human/zebrafish comparative identification of *cis*-regulatory elements around vertebrate developmental transcription factor genes. Dev. Biol. 327, 526–540.
Noe, L., Kucherov, G., 2005. YASS: enhancing the sensitivity of DNA similarity search. Nucleic Acids Res. 33, W540–W543.
Nolte, C., Amores, A., Nagy Kovacs, E., Postlethwait, J., Featherstone, M., 2003. The role of a retinoic acid response element in establishing the anterior neural expression border of Hoxd4 transgenes. Mech. Dev. 120, 325–335.
Nolte, C., Rastegar, M., Amores, A., Bouchard, M., Grote, D., Maas, R., Kovacs, E.N., Postlethwait, J., Rambaldi, I., Rowan, S., Yan, Y.L., Zhang, F., Featherstone, M., 2006. Stereospecificity and PAX6 function direct Hoxd4 neural enhancer activity along the antero-posterior axis. Dev. Biol. 299, 582–593.
Oxtoby, E., Jowett, T., 1993. Cloning zebrafish krox-20 gene (krx-20) and its expression during hindbrain development. Nucleic Acids Res. 21, 1087–1095.
Pasqualetti, M., Diaz, C., Renaud, J.S., Rijli, F.M., Glover, J.C., 2007. Fate-mapping the mammalian hindbrain: segmental origins of vestibular projection neurons assessed using rhombomere-specific Hoxa2 enhancer elements in the mouse embryo. J. Neurosci. 27, 9670–9681.
Pennacchio, L.A., Ahituv, N., Moses, A.M., Prabhakar, S., Nobrega, M.A., Shoukry, M., Minovitsky, S., Dubchak, I., Holt, A., Lewis, K.D., Plajzer-Frick, I., Akiyama, J., De Val, S.,

Afzal, V., Black, B.L., Couronne, O., Eisen, M.B., Visel, A., Rubin, E.M., 2006. In vivo enhancer analysis of human conserved non-coding sequences. Nature 444, 499–502.
Popperl, H., Featherstone, M.S., 1992. An autoregulatory element of the murine Hox-4.2 gene. Embo J. 11, 3673–3680.
Punnamoottil, B., Kikuta, H., Pezeron, G., Erceg, J., Becker, T.S., Rinkwitz, S., 2008. Enhancer detection in zebrafish permits the identification of neuronal subtypes that express Hox4 paralogs. Dev. Dyn. 237, 2195–2208.
Putnam, N.H., Butts, T., Ferrier, D.E., Furlong, R.F., Hellsten, U., Kawashima, T., Robinson-Rechavi, M., Shoguchi, E., Terry, A., Yu, J.K., Benito-Gutierrez, E.L., Dubchak, I., Garcia-Fernandez, J., Gibson-Brown, J.J., Grigoriev, I.V., Horton, A.C., de Jong, P.J., Jurka, J., Kapitonov, V.V., Kohara, Y., Kuroki, Y., Lindquist, E., Lucas, S., Osoegawa, K., Pennacchio, L.A., Salamov, A.A., Satou, Y., Sauka-Spengler, T., Schmutz, J., Shin, I.T., Toyoda, A., Bronner-Fraser, M., Fujiyama, A., Holland, L.Z., Holland, P.W., Satoh, N., Rokhsar, D.S., 2008. The amphioxus genome and the evolution of the chordate karyotype. Nature 453, 1064–1071.
Rastegar, M., Kobrossy, L., Kovacs, E.N., Rambaldi, I., Featherstone, M., 2004. Sequential histone modifications at Hoxd4 regulatory regions distinguish anterior from posterior embryonic compartments. Mol. Cell. Biol. 24, 8090–8103.
Rastegar, S., Hess, I., Dickmeis, T., Nicod, J.C., Ertzer, R., Hadzhiev, Y., Thies, W.G., Scherer, G., Strahle, U., 2008. The words of the regulatory code are arranged in a variable manner in highly conserved enhancers. Dev. Biol. 318, 366–377.
Sandelin, A., Bailey, P., Bruce, S., Engstrom, P.G., Klos, J.M., Wasserman, W.W., Ericson, J., Lenhard, B., 2004. Arrays of ultraconserved non-coding regions span the loci of key developmental genes in vertebrate genomes. BMC Genomics 5, 99.
Schubert, M., Escriva, H., Xavier-Neto, J., Laudet, V., 2006. Amphioxus and tunicates as evolutionary model systems. Trends Ecol. Evol. 21, 269–277.
Seo, H.C., Edvardsen, R.B., Maeland, A.D., Bjordal, M., Jensen, M.F., Hansen, A., Flaat, M., Weissenbach, J., Lehrach, H., Wincker, P., Reinhardt, R., Chourrout, D., 2004. Hox cluster disintegration with persistent anteroposterior order of expression in Oikopleura dioica. Nature 431, 67–71.
Sharpe, J., Nonchev, S., Gould, A., Whiting, J., Krumlauf, R., 1998. Selectivity, sharing and competitive interactions in the regulation of Hoxb genes. Embo J. 17, 1788–1798.
Skromne, I., Thorsen, D., Hale, M., Prince, V.E., Ho, R.K., 2007. Repression of the hindbrain developmental program by Cdx factors is required for the specification of the vertebrate spinal cord. Development 134, 2147–2158.
Soshnikova, N., Duboule, D., 2009. Epigenetic temporal control of mouse Hox genes in vivo. Science 324, 1320–1323.
Spitz, F., Gonzalez, F., Duboule, D., 2003. A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. Cell 113, 405–417.
Spitz, F., Herkenne, C., Morris, M.A., Duboule, D., 2005. Inversion-induced disruption of the Hoxd cluster leads to the partition of regulatory landscapes. Nat. Genet. 37, 889–893.
Straka, H., Baker, R., Gilland, E., 2001. Rhombomeric organization of vestibular pathways in larval frogs. J. Comp. Neurol. 437, 42–55.
Tarchini, B., Duboule, D., Kmita, M., 2006. Regulatory constraints in the evolution of the tetrapod limb anterior–posterior polarity. Nature 443, 985–988.
Tumpel, S., Cambronero, F., Wiedemann, L.M., Krumlauf, R., 2006. Evolution of *cis* elements in the differential expression of two Hoxa2 coparalogous genes in pufferfish (Takifugu rubripes). Proc. Natl. Acad. Sci. U. S. A. 103, 5419–5424.
Tumpel, S., Maconochie, M., Wiedemann, L.M., Krumlauf, R., 2002. Conservation and diversity in the *cis*-regulatory networks that integrate information controlling expression of Hoxa2 in hindbrain and cranial neural crest cells in vertebrates. Dev. Biol. 246, 45–56.
Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., Afzal, V., Ren, B., Rubin, E.M., Pennacchio, L.A., 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature 457, 854–858.
Wada, H., Garcia-Fernandez, J., Holland, P.W., 1999. Colinear and segmental expression of amphioxus Hox genes. Dev. Biol. 213, 131–141.
Whiting, J., Marshall, H., Cook, M., Krumlauf, R., Rigby, P.W., Stott, D., Allemann, R.K., 1991. Multiple spatially specific enhancers are required to reconstruct the pattern of Hox-2.6 gene expression. Genes Dev. 5, 2048–2059.
Woltering, J.M., Duboule, D., 2009. Conserved elements within open reading frames of mammalian Hox genes. J. Biol. 8, 17.
Zhang, F., Nagy Kovacs, E., Featherstone, M.S., 2000. Murine hoxd4 expression in the CNS requires multiple elements including a retinoic acid response element. Mech. Dev. 96, 79–89.
Zhang, F., Popperl, H., Morrison, A., Kovacs, E.N., Prideaux, V., Schwarz, L., Krumlauf, R., Rossant, J., Featherstone, M.S., 1997. Elements both 5′ and 3′ to the murine Hoxd4 gene establish anterior borders of expression in mesoderm and neurectoderm. Mech. Dev. 67, 49–58.