

PRIMARY RESEARCH

Open Access

Salamander *Hox* clusters contain repetitive DNA and expanded non-coding regions: a typical *Hox* structure for non-mammalian tetrapod vertebrates?

Stephen Randal Voss^{1,2*}, Srikrishna Putta^{1,2}, John A Walker^{1,2}, Jeramiah J Smith¹, Nobuyasu Maki³ and Panagiotis A Tsonis⁴

Abstract

Hox genes encode transcription factors that regulate embryonic and post-embryonic developmental processes. The expression of *Hox* genes is regulated in part by the tight, spatial arrangement of conserved coding and non-coding sequences. The potential for evolutionary changes in *Hox* cluster structure is thought to be low among vertebrates; however, recent studies of a few non-mammalian taxa suggest greater variation than originally thought. Using next generation sequencing of large genomic fragments (>100 kb) from the red spotted newt (*Notophthalmus viridescens*), we found that the arrangement of *Hox* cluster genes was conserved relative to orthologous regions from other vertebrates, but the length of introns and intergenic regions varied. In particular, the distance between *hoxd13* and *hoxd11* is longer in newt than orthologous regions from vertebrate species with expanded *Hox* clusters and is predicted to exceed the length of the entire *HoxD* clusters (*hoxd13–hoxd4*) of humans, mice, and frogs. Many repetitive DNA sequences were identified for newt *Hox* clusters, including an enrichment of DNA transposon-like sequences relative to non-coding genomic fragments. Our results suggest that *Hox* cluster expansion and transposon accumulation are common features of non-mammalian tetrapod vertebrates.

Keywords: *Hox*, Salamander, Genome, Evolution

Background

Bilateria body plans are determined in part by DNA transcription factors called *Hox* genes [1-4]. Excepting fish, vertebrate *Hox* genes are ordered among four unlinked clusters that each span relatively short segments of genomic DNA (generally 100–200 Kb). The arrangement of *Hox* genes on chromosomes is co-linear with their pattern of transcription along the anterior-posterior and proximal-distal body axes during embryonic development [5,6]. The organization and structure of *Hox* gene clusters and associated non-coding regulatory elements are mostly conserved across vertebrates [7,8]. However, as genomic studies extend to non-

genetic model organisms, variations in *Hox* cluster structure are being discovered, including variations in gene number, repetitive sequence content, cluster length, and non-coding sequence conservation [9-15]. These variations suggest that the evolution of *Hox* cluster structure may correlate with phylogeny, unique modes of vertebrate development, and/or derived morphological characteristics.

In tetrapod vertebrates, stereotypic patterns of *Hox* expression are observed along the proximal-distal axes of developing limbs [16]. In most species, *Hox* developmental genetic programs are only expressed during limb development. However, salamanders reactivate *Hox* gene expression throughout life to correctly pattern tissues within regenerating limbs [17-21]. While some patterns of *Hox* expression in regenerating limbs recapitulate the expression pattern in developing limbs, spatial and temporal differences are observed [18-21]. This raises the

* Correspondence: srvooss@uky.edu

¹Department of Biology, University of Kentucky, Lexington, KY 40506, USA

²Spinal Cord and Brain Injury Research Center, University of Kentucky, Lexington, KY 40506, USA

Full list of author information is available at the end of the article

possibility that salamander *Hox* clusters may contain non-coding elements that uniquely regulate post-embryonic, tissue regeneration; such elements may not be expected within *Hox* clusters of vertebrates incapable of limb regeneration. There is another reason to suspect that salamander *Hox* clusters may differ from other vertebrate taxa—salamanders as a group have extremely large genomes. An average sized salamander genome is approximately 10× larger than the *Homo sapiens* genome; some salamanders have genomes that are 30× larger [22]. This larger genome size is reflected in the structure of genes, as salamander introns are longer on average than orthologous introns in other vertebrates [23,24].

Belleville et al. [25] reported that two pairs of adjacent *Hox* cluster genes from the red spotted newt (*Notophthalmus viridescens*) presented highly conserved coding and non-coding sequences relative to orthologous mammalian *Hox* sequences. These results suggested that *Hox* cluster evolution is constrained even within the context of a very large vertebrate genome (>20 pg/haploid nucleus) [22]. However, the results that we present below show that newt *Hox* clusters are more variable than originally thought. Sequencing of large genomic fragments (>100 Kb) reveals regional variation in length across newt *Hox* cluster regions and higher proportions of DNA transposon-like sequences within *Hox* introns and intergenic sequences than non-coding genomic regions. Our results show that expanded non-coding regions and relatively high repetitive DNA sequence content are typical of *Hox* clusters in amphibians and other non-mammalian tetrapod vertebrates.

Results and discussion

BAC library screening, sequencing, assembly, and annotation

A bacterial artificial clone (BAC) library of 41,472 clones was constructed for newt, and pools were screened by polymerase chain reaction (PCR) to identify clones that contained *Hox* genes. Two BACs containing *HoxC* (NV_H3_75P19; [GenBank:JF490017.1]) and *HoxD* (NV_H3_85F1; [GenBank:JF490018.1]) orthologs and two additional BACs containing only non-coding genomic DNA (NV_H3_28J3; [GenBank:JF490019.1] and NV_H3_32L5; [GenBank:JF490020.1]) were purified and sequenced to an average depth of 220 bp sequence reads per nucleotide position. The reads for NV_H3_75P19 were assembled into three large contigs with the breaks occurring between *hoxc5* and *hoxc4*, and a position 3' of *hoxc4*. The reads for NV_H3_28J3 were re-assembled into two large contigs with the break occurring 5' of *hoxd11*. The reads for the BACs that contained non-coding genomic DNA generated more than three contigs and could not be ordered; these were randomly concatenated for analyses described below. BLASTx searches revealed that the BACs containing *Hox* sequences contained some, but not all of *Hox* gene members from each cluster:

NV_H3_75P19 contained *hoxc11*, *hoxc10*, *hoxc9*, *hoxc8*, *hoxc6*, *hoxc5*, and *hoxc4*, and NV_H3_85F1 contained *hoxd11*, *hoxd10*, *hoxd9*, and *hoxd8*. The order of newt *Hox* genes was conserved relative to orthologs in other vertebrate genomes, as were coding and non-coding sequences, and exon/intron boundaries (Figures 1 and 2). High sequence identity was observed for *Hox* genes, which is typical of transcription factors that function in highly conserved developmental pathways. Conserved non-coding sequences (CNS) were identified from regions flanking *Hox* exons; these likely correspond to enhancer elements and non-coding RNAs that function in the regulation of *Hox* gene expression. For example, two CNSs that were identified downstream of newt *hoxd11* (40 kb) correspond to enhancer elements VIII and IX from Gerard et al. [26], and a CNS upstream (3 kb) of newt *hoxc8* corresponds to an enhancer from Shashikant et al. [27]. Also, a CNS identified downstream of *hoxc10* (28 kb) corresponds to human miRNA-196a, and a canonical miR-196a seed-pairing site is predicted 268 bp from the end of newt *hoxd8* [28]. Thus, elements that are known to regulate *Hox* gene functions in other vertebrate species show identity to sequences in newt *Hox* genomic regions.

While the general organization of newt *Hox* genes was conserved relative to other vertebrates, extensive variation was observed in the lengths of intergenic and intronic sequences (Table 1; Figures 3 and Figure 4). The length of the newt *hoxc11-c4* region was longer than orthologous mammalian (*H. sapiens*, *Mus musculus*) and zebrafish (*Danio rerio*) regions, but shorter than regions from *Anolis carolinensis* (lizard) and *Xenopus tropicalis* (frog), which are known to have expanded *HoxC* clusters [13]. While newt *HoxC* introns were also longer than mammalian and fish introns, frog and lizard introns also exceed the length of their mammalian counterparts. This supports the idea that salamander genes typically contain long introns [23,24], although we did not observe the same pattern for *HoxD* genes. While long lizard and frog *HoxD* introns were observed, newt *hoxd11-9* introns were typically shorter than orthologous mammalian introns. Thus, while non-mammalian tetrapod vertebrates, and especially the anolis lizard, have *Hox* genes with long introns, relative intron length varies among paralogous members of newt *HoxD* and *HoxC* clusters.

In annotating *HoxD* cluster genes, we discovered that *hoxd11* was located approximately 73 kb from the terminus of NV_H385F1. This distance, which provides a minimum estimate to the expected position of *hoxd13* (*hoxd12* is not known for amphibians [15,26]), predicts the newt *hoxd13-11* segment to be > 4.5× and 1.5× longer than orthologous *HoxD* regions from frog and lizard. It also exceeds the length of *hoxd11-13* segments in the coelacanth and a caecilian amphibian (*Typhlonectes natans*) [29], which until this study was thought to be longest among vertebrates (Figure 4). While it is possible

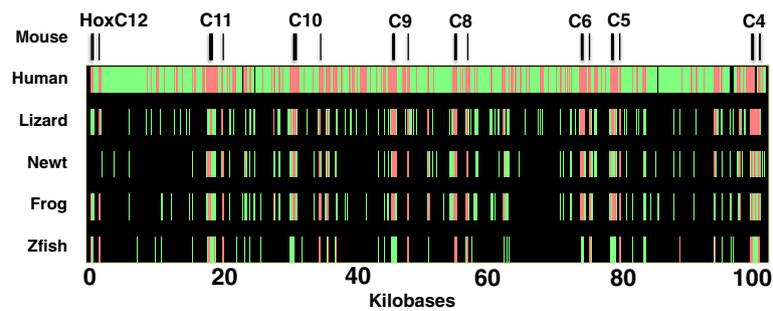


Figure 1 Multispecies alignment of orthologous *HoxC* genomic regions among selected vertebrate species. The black bars at the top of the figure show the positions of exons in the mouse sequence, which was used as the reference sequence for alignment. Red bars indicate strongly aligned regions—at least 100 bp in length without a gap and >70% nucleotide identity. Green bars indicate all other aligned regions. Zfish, zebrafish.

that the expanded region is explained by an evolutionary loss of the newt *hoxd13* gene, this seems unlikely because *hoxd13* orthologs are known for related salamanders [15], and we did not detect the presence of a pseudogene nucleotide signature. Because expanded *Hox* clusters have been shown for a representative caecilian [26] and anuran species [13], parsimony suggests the expansion of the *hoxd11–13* region to be a shared derived characteristic of amphibians, with convergent expansion of the same region in lizard.

Interspersed repeat sequences in BACs

Previous studies have shown that interspersed repetitive DNA sequences are rarely observed within *Hox* clusters of mammals and some reptiles, but are more abundant in species with expanded *Hox* clusters [15,26]. To test this idea, we searched *Hox* and non-*Hox* genomic clones for repeats that are catalogued in RepBase (Genetic Information Research Institute, Mountain View, USA) [30], and also aligned genomic sequences using MultiPipmaker [31] to identify direct and indirect repeats unique to the newt. In many cases, we found that both approaches identified repetitive sequences for the same segments of DNA;

however, more newt specific repeats were identified overall (Additional file 1: Table S1). The annotated (i.e., RepBase) interspersed repetitive sequence content of *HoxC* and *HoxD* genomic sequences was approximately two to three times lower than the content of the two, non-protein coding genomic clones (Table 2). Considering annotated and newt-specific repeats, 77% of the non-coding genomic sequence was identified as repetitive, compared to 24% and 32% for *HoxC* and *HoxD* sequences (Additional file 1: Table S1). These results suggest that the fixation probability for repetitive element accumulation is lower for *Hox* clusters, presumably because these regions are evolutionarily constrained by the functional sequences they encode. Repeats were more frequent in regions flanking genes, with the large intergenic regions flanking terminal *Hox* loci showing the greatest accumulation (Figures 3 and 4). Repeats were predicted for introns, and a higher density of DNA transposon-like sequences were predicted within *HoxC* and *HoxD* clusters than within non-coding genomic clones. Interestingly, the enrichment of DNA transposon-like sequences was about 20-fold for *HoxC* but only 2-fold for *HoxD* (Table 2). While this may reflect sampling bias, the more expanded of the two newt

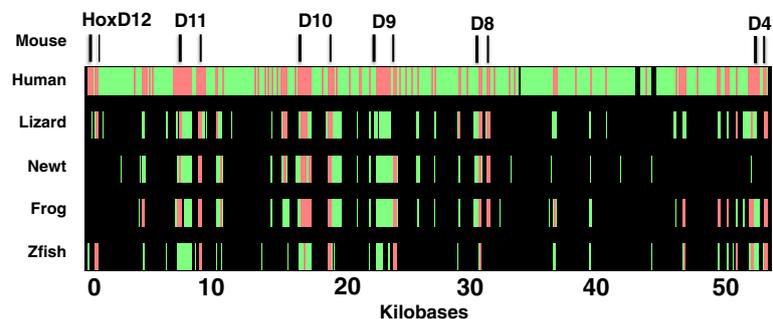


Figure 2 Multispecies alignment of orthologous *HoxD* genomic regions among selected vertebrate species. The black bars at the top of the figure show the positions of exons in the mouse sequence, which was used as the reference sequence for alignment. Red bars indicate strongly aligned regions—at least 100 bp in length without a gap and >70% nucleotide identity. Green bars indicate all other aligned regions. Zfish, zebrafish.

Table 1 Species comparison of *HoxC* and *HoxD* intron lengths

Gene ID	Human	Mouse	Lizard	Frog	Newt	Zebrafish
<i>hoxc11</i>	1257	1261	2025	1334	2907	1368
<i>hoxc10</i>	3158	3189	4591	2498	3211	1766
<i>hoxc9</i>	1703	1704	2139	1560	2225	1059
<i>hoxc8</i>	1368	1347	2075	1619	1393	1265
<i>hoxc6</i>	733	728	902	763	1283	610
<i>hoxc5</i>	701	692	927	713	753	818
<i>hoxc4</i>	488	475	1006	448	502	520
<i>hoxd11</i>	770	735	1667	748	675	863
<i>hoxd10</i>	1375	1366	1628	1980	1233	715
<i>hoxd9</i>	348	346	-	341	345	316
<i>hoxd8</i>	373	395	2474	1733	434	-

The longest intron is italicized for each Hox gene.

Hox clusters does not contain a higher proportion of DNA transposon-like sequences; instead, *HoxD* contains a moderately higher proportion of long interspersed retroelement-like sequences, simple repeats, and newt-specific repeats. Observation of a higher frequency of DNA transposon-like sequences, within arguably a more functionally constrained *HoxC* cluster, suggests an insertion bias for *Hox* genic regions. While this speculation awaits further study, our results support the idea that repetitive sequences, and in particular DNA transposon-like sequences, are more abundant within *Hox* clusters of non-mammalian tetrapod vertebrates [13] than is indicated by analysis of mammalian *Hox* clusters.

Conclusions

Salamander *Hox* genomic regions show elements of conservation and diversity in comparison to other vertebrate species. Whereas the structure and organization of *Hox* coding genes is conserved, newt *Hox* clusters show

variation in the lengths of introns and intergenic regions, and the *hoxd13-11* region exceeds the lengths of orthologous segments even among vertebrate species with expanded *Hox* clusters. We posit that the *hoxd13-11* expansion predated a basal salamander genome size increase that occurred approximately 180 million years ago [32] as it is preserved in all three extant amphibian groups. Over more recent timescales, additional evidence supports the idea that *Hox* clusters are amenable to structural evolution: there is variation in the lengths of introns and intergenic regions, relatively high numbers of repetitive sequences, and non-random accumulations of DNA transposons in newts and lizards. The non-random accumulation of DNA-like transposons could potentially alter developmental programming by creating sequence motifs for transcriptional regulation [33-35]. Overall, available data from several non-mammalian tetrapods suggest that *Hox* structural flexibility is the rule, not the exception. We speculate that such flexibility may contribute to developmental variation across non-mammalian taxa, both in embryogenesis and during the re-deployment of *Hox* genes during post-embryonic developmental processes, such as metamorphosis and regeneration.

Methods

BAC library construction, screening, and sequencing

The Clemson University Genomics Institute constructed a BAC library from partially restriction digested and size-selected genomic DNA that was isolated from the erythrocytes of a single *Notophthalmus viridescens* female (University of Dayton Institutional Animal Care and Use Committee Protocol # 011-12). A total of 41,472 clones were arrayed in 108 × 384 well plates. Superpools of clones were made by combining clones from twelve 384 well plates into a single pool. DNA was extracted from 400 ml of overnight cultures of

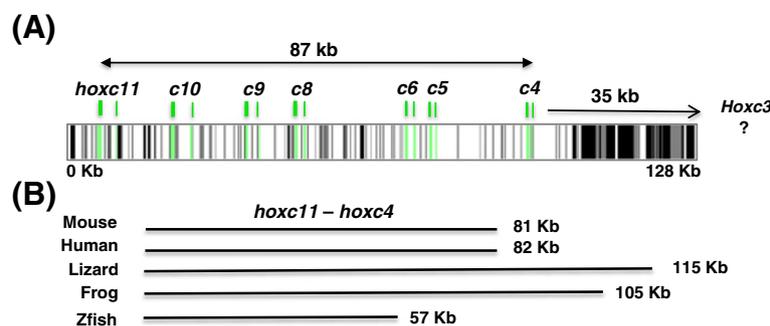


Figure 3 Structure and repetitive sequence content of newt *HoxC*, and total length of orthologous *hoxc11-c4*. (A) The structure and repetitive sequence content of newt *HoxC*. Green bars indicate the positions of exons, gray bars indicate the positions of RepBase repeats, and black bars indicate the positions of unique newt repetitive sequences. (B) The total length of orthologous *hoxc11-c4* non-coding segments from representative vertebrates.

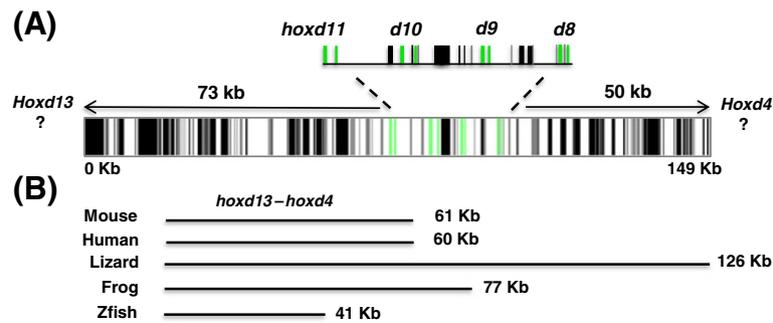


Figure 4 Structure and repetitive sequence content of newt *HoxD* and total length of orthologous *hoxd11-d8*. (A) The structure and repetitive sequence content of newt *HoxD*. Green bars indicate the positions of exons, gray bars indicate the positions of RepBase repeats, and black bars indicate the positions of unique newt repetitive sequences. (B) The total length of orthologous *hoxd11-d8* non-coding segments from representative vertebrates.

superpools using the Plasmid MaxiPrep kit (Qiagen, Valencia, CA, USA), and the DNA pellet was re-suspended in 250 μ l of water. PCR primers for newt *hoxc10* (forward: CAAAGAGAAAACGCGGAAAG; reverse: CGATACCGTCCCTCCATAA) and *hoxd10* (forward: TTTCCATTGTCGGTTTTTCC; reverse: TCCTACCACGGACATTACCC) were used to identify two *Hox* gene-containing BACs and two BACs that did not contain protein-coding sequence. The four clones were grown in 400 ml L-broth, and DNA was isolated using the Qiagen Large Construction Kit (Qiagen); genomic DNA contamination was reduced using Plasmid-Safe DNase treatment (Epicentre Biotechnologies, Madison, USA). The Roche GS FLX Titanium platform (Basel, Switzerland) was used to sequence BACs; the work was accomplished by the staff of the University of Iowa Sequencing Core Facility. The termini of BAC inserts were end-sequenced using Sanger technology and ABI Big-Dye 3.1 (Invitrogen, Grand Island, USA).

Table 2 Percent coverage of salamander genomic sequences by newt-specific and RepBase repetitive elements

	<i>HoxC</i>	<i>HoxD</i>	Non-coding
Newt-specific repeats	24	32	77
Total RepBase repeats	5.34	4.25	13.34
Long terminal repeat retrotransposons	2.75	2.70	9.50
Non-long terminal repeat retrotransposons	2.30	3.92	2.42
DNA transposons	2.54	0.32	0.12
Unclassified	0.04	0.00	1.30
Satellites	0.00	0.50	0.31
Simple repeats	0.88	1.88	0.87
Low complexity	0.62	0.38	0.31

DNA sequence assembly and annotation

Sequences were screened to trim vector, adapters, and contaminating *Escherichia coli* sequences. After an initial assembly using GS De Novo Assembler (454 Life Sciences, Branford, USA), contigs and singletons were assembled further using DNASTAR SeqMan (DNASTAR, Inc., Madison, USA). Contiguous sequences of assembled BACs were searched (*blastn*) against salamander expressed sequence tagged contigs at Sal-Site [36]; non-redundant nucleotide and protein databases at NCBI (*blastx* and *tblastp*) [37] were used to identify and annotate gene regions. For multispecies comparisons, genomic sequences for *H. sapiens* (GRCh37.10), and *M. musculus* (GRCh38.1), were obtained from NCBI. *Anolis carolinensis* (AnoCar 2.0) and *D. rerio* (Zv9) were obtained from Ensembl [38]. *X. tropicalis* (build 7.1) was obtained from Xenbase [39]. Sequences were aligned using MultiPipMaker [28]. Annotated repeats were identified by searching re-assembled BAC clones against all deposited repeats in RepBase [30]. Newt-specific repeats were identified using MultiPipmaker [28] by aligning re-assembled BAC clones against each other and by performing self-self BAC alignments. The “search both strands” and “high sensitivity” options were used in MultiPipmaker to identify significantly similar non-coding sequences that are located to different positions either within or between BACs. The terminal base pair positions for these alignments were recorded to denote the positions of repetitive sequences within BACs. If the two repeats occurred within 50 bp of each other, they were compiled as a single repetitive sequence with the most terminal base positions denoting the repeat span. The base pair coordinates for newt-specific repetitive sequences were combined with base pair coordinates for RepBase repetitive sequences to generate an underlay file (Additional file 1: Table S1), and this was used to create maps of repetitive elements for the *HoxD* and *HoxC* genomic regions.

Additional file

Additional file 1: Table S1. Base pair coordinates for RepBase and newt-specific repeats identified from BAC clones with *HoxC* and *HoxD* genomic regions.

Competing interest

The authors declare that they have no competing interests.

Authors' contributions

SRV, JJS, NM, and PAT conceived and designed the experiment, and analyzed and interpreted the results. SP assembled sequence reads corresponding to BACs and analyzed and annotated the sequence data. JAW performed PCR, isolated BACs for sequencing, and analyzed and annotated the sequence data. SRV drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgments

The research was supported by grant R24-OD010435 (SRV) and EY-10540 (PAT) from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH). The project also used resources developed under Multidisciplinary University Research Initiative grant (W911NF-09-1-0305) from the Army Research Office (SRV) and resources from the *Ambystoma* Genetic Stock Center, which is funded by grant DBI-0951484 from the National Science Foundation (SRV). The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official views of NCRR, NIH, ARO, or NSF.

Author details

¹Department of Biology, University of Kentucky, Lexington, KY 40506, USA. ²Spinal Cord and Brain Injury Research Center, University of Kentucky, Lexington, KY 40506, USA. ³Institute of Protein Research, Osaka University, 3-2 Yamadaoka, Suita-Shi, Osaka 565-0871, Japan. ⁴Department of Biology, University of Dayton, Dayton, OH, USA.

Received: 10 January 2013 Accepted: 25 January 2013

Published: 5 April 2013

References

- Lewis EB: A gene complex controlling segmentation in *Drosophila*. *Nature* 1978, **276**:565–570.
- Krumlauf R: Hox genes in vertebrate development. *Cell* 1994, **78**:191–201.
- Duboule D: The rise and fall of Hox gene clusters. *Development* 2007, **134**:2549–2560.
- Lemons D, McGinnis W: Genomic evolution of Hox gene clusters. *Science* 2006, **313**:1918–1922.
- Kmita M, Duboule D: Organizing axes in time and space; 25 years of colinear tinkering. *Science* 2003, **301**:331–333.
- Tschopp P, Duboule D: A genetic approach to the transcriptional regulation of Hox gene clusters. *Annu Rev Genet* 2011, **45**:145–166.
- Garcia-Fernandez J: The genesis and evolution of homeobox gene clusters. *Nat Rev Genet* 2005, **6**:881–892.
- Duboule D: Temporal colinearity and the phylotypic progression: a basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony. *Dev Suppl* 1994, **1994**:135–142.
- Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehal P, Christoffels A, Rash S, Hoon S, Smit A, Gelpke MD, Roach J, Oh T, Ho IY, Wong M, Detter C, Verhoeve F, Predki P, Tay A, Lucas S, Richardson P, Smith SF, Clark MS, Edwards YJ, Doggett N, Zharkikh A, Tavtigian SV, Pruss D, Barnstead M, Evans C, et al: Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* 2002, **297**:1301–1310.
- Hoegg S, Meyer A: Hox clusters as models for vertebrate genome evolution. *Trends Genet* 2005, **21**:421–424.
- Kurosawa G, Takamatsu N, Takahashi M, Sumitomo M, Sanaka E, Yamada K, Nishii K, Matsuda M, Asakawa S, Ishiguro H, Miura K, Kurosawa Y, Shimizu N, Kohara Y, Hori H: Organization and structure of Hox gene loci in Medaka genome and comparison with those of pufferfish and zebrafish genomes. *Gene* 2006, **370**:75–82.
- Woltering JM, Vonk FJ, Müller H, Bardine N, Tuduze IL, de Bakker MAG, Knöchel W, Sirbu IO, Durston AJ, Richardson MK: Axial patterning in snakes and caecilians: evidence for an alternative interpretation of the Hox code. *Dev Biol* 2009, **332**:82–89.
- Di-Poi N, Montoya-Burgos JI, Duboule D: A typical relaxation of structural constraints in Hox gene clusters of the green anole lizard. *Genome Res* 2009, **19**:602–610.
- Matsunami M, Sumiyama K, Saitou N: Evolution of conserved non-coding sequences within the vertebrate Hox clusters through the two-round whole genome duplications revealed by phylogenetic footprinting analysis. *J Mol Evol* 2010, **71**:427–436.
- Liang D, Wu R, Geng J, Chaolin Wang C, Zhang P: A general scenario of Hox gene inventory variation among major sarcopterygian lineages. *BMC Evol Biol* 2011, **11**:25.
- Zakany J, Duboule D: The role of Hox genes during vertebrate limb development. *Curr Opin Genet Dev* 2007, **17**:359–366.
- Savard P, Gates PB, Brocques JP: Position dependent expression of a homeobox gene transcript in relation to amphibian limb regeneration. *EMBO J* 1988, **7**:4275–4282.
- Gardiner DM, Blumberg B, Komine Y, Bryant SV: Regulation of HoxA expression in developing and regenerating axolotl limbs. *Development* 1995, **121**:1731–1741.
- Torok MA, Gardiner DM, Shubin NH, Bryant SV: Expression of HoxD genes in developing and regenerating axolotl limbs. *Dev Biol* 1998, **200**:225–233.
- Khan PA, Tsilifidis C, Liversage RA: Hox C6 expression during development and regeneration of forelimbs in larval *Notophthalmus viridescens*. *Dev Genes Evol* 1999, **209**:323–329.
- Carlson MR, Komine Y, Bryant SV, Gardiner DM: Expression of Hoxb13 and Hoxc10 in developing and regenerating axolotl limbs and tails. *Dev Biol* 2001, **229**:396–406.
- Gregory TR: *Animal Genome Size Database 2010*. <http://www.genomesize.com>.
- Casimir CM, Gates PB, Ross-Macdonald PB, Jackson JF, Patient RK, Brocques JP: Structure and expression of a newt cardio-skeletal myosin gene: implications for the C value paradox. *J Mol Biol* 1992, **202**:287–296.
- Smith JJ, Putta S, Zhu W, Pao GM, Verma IM, Hunter T, Bryant SV, Gardiner DM, Harkins TT, Voss SR: Genic regions of a large salamander genome contain long introns and novel genes. *BMC Genomics* 2009, **10**:19.
- Belleville S, Beauchemin M, Tremblay M, Noiseux N, Savard P: Homeobox-containing genes in the newt are organized in clusters similar to other vertebrates. *Gene* 1992, **114**:179–186.
- Gerard M, Duboule D, Zakany J: Structure and activity of regulatory elements involved in the activation of the Hoxd-11 gene during late gastrulation. *EMBO J* 1993, **12**:3539–3550.
- Shashikant CS, Bolanowsky SA, Anand S, Anderson SM: Comparison of diverged Hoxc8 early enhancer activities reveals modification of regulatory interactions at conserved cis-acting elements. *J Exp Zool B Mol Dev Evol* 2007, **308**:242–249.
- Hornstein E, Mansfield JH, Yekta S, Hu JK, Harfe BD, McManus MT, Baskerville S, Bartel DP, Tabin CJ: The microRNA miR-196 acts upstream of Hoxb8 and Shh in limb development. *Nature* 2005, **438**:671–674.
- Mannaert A, Amemiya CT, Bossuyt F: Comparative analyses of vertebrate posterior HoxD clusters reveal atypical cluster architecture in the caecilian *Typhlonectes natans*. *BMC Genomics* 2010, **11**:658.
- Smit AFA, Hubley R, Green P: *RepeatMasker Open-3.0 1996–2010*. <http://www.repeatmasker.org>.
- Schwartz S, Zhang Z, Frazer KA, Smit A, Riemer C, Bouck J, Gibbs R, Hardison R, Miller W: PipMaker—a resource for aligning two genomic sequences. *Gen Res* 2000, **10**:577–586.
- Zhang P, Wake DB: Higher-level salamander relationships and divergence dates inferred from complete mitochondrial genomes. *Mol Phylogenet Evol* 2009, **53**:492–508.
- Lowe CB, Bejerano G, Haussler D: Thousands of human mobile element fragments undergo strong purifying selection near developmental genes. *Proc Natl Acad Sci USA* 2007, **104**:8005–8010.
- Mikkelsen TS, Wakefield MJ, Aken B, Amemiya CT, Chang JL, Duke S, Garber M, Gentles AJ, Goodstadt L, Heger A, Jurka J, Kamal M, Mauceli E, Searle SM, Sharpe T, Baker ML, Batzer MA, Benos PV, Belov K, Clamp M, Cook A, Cuff J, Das R, Davidow L, Deakin JE, Fazzari MJ, Glass JL, Grabherr M, Greally JM, Gu W, et al: Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature* 2007, **447**:167–177.
- Borque G: Transposable elements in gene regulation and in the evolution of vertebrate genomes. *Curr Opin Genet Dev* 2009, **19**:607–612.

36. Smith JJ, Putta S, Walker JA, Kump DK, Samuels AK, Monaghan JR, Weisrock DW, Staben C, Voss SR: **Sal-Site: integrating new and existing ambystomatid salamander research and informational resources.** *BMC Genomics* 2005, **6**:181.
37. *National Center for Biotechnology Information.* <http://www.ncbi.nlm.nih.gov>.
38. *Ensembl.* <http://www.ensembl.org>.
39. *Xenbase: A Xenopus Laevis and Xenopus Tropicalis Resource.* <http://www.xenbase.org>.

doi:10.1186/1479-7364-7-9

Cite this article as: Voss et al.: Salamander *Hox* clusters contain repetitive DNA and expanded non-coding regions: a typical *Hox* structure for non-mammalian tetrapod vertebrates?. *Human Genomics* 2013 **7**:9.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

