

Revisiting the Origin of the Vertebrate Hox14 by Including Its Relict Sarcopterygian Members

NATHALIE FEINER^{1,2}, ROLF ERICSSON³, AXEL MEYER^{1,2,4},
AND SHIGEHIRO KURAKU^{1,2,4*}

¹Department of Biology, University of Konstanz, Konstanz, Germany

²International Max-Planck Research School (IMPRS) for Organismal Biology,
University of Konstanz, Konstanz, Germany

³Biological Sciences, Macquarie University, Sydney, New South Wales, Australia

⁴Konstanz Research School Chemical Biology (KoRS-CB), University of Konstanz, Konstanz, Germany

Bilaterian *Hox* genes play pivotal roles in the specification of positional identities along the anteroposterior axis. Particularly in vertebrates, their regulation is tightly coordinated by tandem arrays of genes [paralogy groups (PGs)] in four gene clusters (HoxA-D). Traditionally, the uninterrupted *Hox* cluster (*Hox1-14*) of the invertebrate chordate amphioxus was regarded as an archetype of the vertebrate *Hox* clusters. In contrast to Hox1-13 that are globally regulated by the "Hox code" and are often phylogenetically conserved, vertebrate Hox14 members were only recently revealed to be present in an African lungfish, a coelacanth, chondrichthyans and a lamprey, and decoupled from the Hox code. In this study we performed a PCR-based search of Hox14 members from diverse vertebrates, and identified one in the Australian lungfish, *Neoceratodus forsteri*. Based on a molecular phylogenetic analysis, this gene was designated *NfHoxA14*. Our real-time RT-PCR suggested its hindgut-associated expression, previously observed also in cloudy catshark *HoxD14* and lamprey *Hox14 α* . It is likely that this altered expression scheme was established before the *Hox* cluster quadruplication, probably at the base of extant vertebrates. To investigate the origin of vertebrate Hox14, by including this sarcopterygian Hox14 member, we performed focused phylogenetic analyses on its relationship with other vertebrate posterior Hox PGs (Hox9-13) as well as amphioxus posterior *Hox* genes. Our results confirmed the hypotheses previously proposed by other studies that vertebrate Hox14 does not have any amphioxus ortholog, and that none of 1-to-1 pairs of vertebrate and amphioxus posterior *Hox* genes, based on their relative location in the clusters, is orthologous.

Bilaterian *Hox* genes instruct the specification of regional identities along the anteroposterior axis. They are arranged in tandem arrays of genes, and their regulation is tightly coordinated in a colinear fashion: the closer a gene is to the 3' end of the *Hox* cluster, the earlier and more anteriorly it is expressed during embryogenesis (Lewis, '78; McGinnis and Krumlauf, '92; Duboule, '94; Kmita and Duboule, 2003). Although all invertebrate bilaterians basically have one *Hox* gene cluster, vertebrates typically possess four clusters (Hox A D) that are derived from two rounds of whole genome duplication (2R WGD; Graham et al., '89; reviewed in Kuraku and Meyer, 2009).

Present address of Rolf Ericsson is The Natural History Museum, Palaeontology Department, Cromwell Road, London SW7 5BD, United Kingdom.

Grant Sponsors: University of Konstanz, International Max Planck Research School (IMPRS) for Organismal Biology; German Research Foundation; Grant number: KU2669/1 1.

*Correspondence to: Shigehiro Kuraku, Department of Biology, University of Konstanz, Universitätsstrasse 10, 78457 Konstanz, Germany.
E mail: shigehiro.kuraku@uni-konstanz.de

The 13 paralogy groups (PGs) (Hox1-13) were recognized by the end of the 20th century in all vertebrate species whose *Hox* clusters were fully sequenced (Zeltser et al., '96; reviewed in Garcia Fernandez, 2005). The only exception identified as late as 2004 was the Hox14 group reported for the coelacanth *Latimeria menadoensis* (*HoxA14*) and the horn shark *Heterodontus francisci* (*HoxD14*) (Powers and Amemiya, 2004; see Fig. 1 for phylogenetic positions of these species and others mentioned below). *HoxD14* was later identified in the cloudy catshark *Scyliorhinus torazame* (Kuraku et al., 2008) and the lesser spotted dogfish *S. canicula* (Oulion et al., 2010) as well as in the elephant shark *Callorhynchus milii* (also called ghost shark or elephant fish) (Venkatesh et al., 2007; Ravi et al., 2009). Only very recently, *HoxA14* was identified in the African lungfish *Protopterus annectens* (Liang et al., 2011). In the Japanese lamprey *Lethenteron japonicum*, a Hox14 member, designated *Hox14 α* , was identified by RT-PCR (Kuraku et al., 2008). As is the case for many non *Hox* genes, orthology of *Hox* gene clusters of this animal to the four jawed vertebrate *Hox* clusters is ambiguous (Kuraku and Meyer, 2009). Therefore, it is not clear which cluster *Hox14 α* belongs to. Additionally, a *HoxA14* pseudogene was identified in both elephant shark and horn shark, and also a pseudogenized *HoxB14* was found in the elephant shark (Powers and Amemiya, 2004; Ravi et al., 2009).

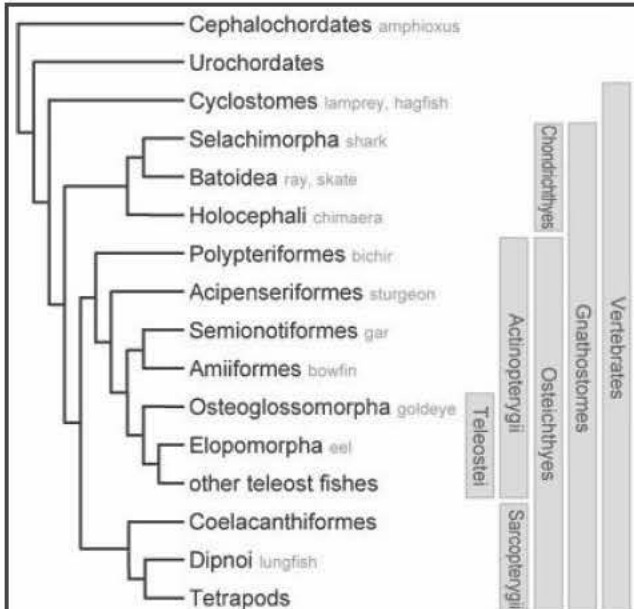


Figure 1. Phylogenetic relationships among the major chordate lineages. Relationships are based on previous molecular phylogenetic analyses (Inoue et al., 2003; Kikugawa et al., 2004; Delsuc et al., 2006; reviewed in Meyer and Zardoya, 2003). English common names of species included in this study are shown in gray beside their taxon names.

The presence of pseudogenized, but still recognizable *HoxA14* orthologs in lineages that diverged more than 400 million years ago (Heinicke et al., 2009; Inoue et al., 2010) indicates independent pseudogenization processes in elasmobranchs and chimaeras (Ravi et al., 2009). Expression patterns of the Hox14 members were investigated only in the lamprey, the cloudy catshark and the lesser spotted dogfish, species whose embryonic resources are accessible in the laboratory, and they were shown to exhibit hindgut associated expressions (Kuraku et al., 2008; Oulion et al., 2011). Notably, they are neither expressed in derivatives of the neural crest nor in the neural tube, somites, or fins, in which at least a subset of Hox1-13 genes is known to be expressed (Kuraku et al., 2008).

The cephalochordate amphioxus possesses a single *Hox* cluster, which is often regarded as “archetypal” and considered to display the prequadruplicated ground state of vertebrate *Hox* clusters (Amemiya et al., 2008). Only recently, Holland et al. (2008) reported that the cluster possesses an additional *Hox* gene designated *AmphiHox15* as well as previously known *Hox14*. However, unambiguous assignment of 1 to 1 orthologies between amphioxus and vertebrate posterior *Hox* genes cannot be established without further data (Ferrier, 2004; Amemiya et al., 2008; Hueber et al., 2010). This observation can be explained by an elevated evolutionary rate of the posterior *Hox* genes which has been termed the “deuterostome posterior flexibility” (Ferrier et al., 2000). For instance, the non orthology between the amphioxus *Hox14* gene and the vertebrate *Hox14* genes has been supported by phylogenetic analysis (Kuraku et al., 2008) as well as a non tree based study (Thomas Chollier et al., 2010). The identical name of the amphioxus and vertebrate genes is simply derived from the same relative location in the cluster, but does not reflect true orthology. Interestingly, orthology between *AmphiHox15* and vertebrate PG13 was previously suggested (Holland et al., 2008; Thomas Chollier et al., 2010), despite their nonsyntenic location in the cluster. However, the support for this grouping is poor, possibly because of the large data sets used in these studies.

In this study, we performed a PCR scan of *Hox14* members in the Australian lungfish, a non tetrapod sarcopterygian, and in silico searches of *Hox14* members in diverse vertebrates. We report the identification of a *Hox14* member in the lungfish, designated *NfHoxA14*, and suggest its embryonic expression in the hindgut. The hindgut associated expressions, observed also in the cloudy catshark *HoxD14* and the lamprey *Hox14 α* , should have been retained since the prevertebrate era when *Hox* genes existed in an ancestral single cluster. Importantly, our phylogenetic analysis indicated that the amphioxus *Hox* cluster contains no ortholog of the vertebrate *Hox14* genes. Our analysis suggested that the amphioxus *Hox* cluster is not an archetype representing a condition before the 2R WGD in the vertebrate lineage. Thus, the vertebrate *Hox* cluster has a unique composition of PGs, compared with invertebrate counterparts, one of which is *Hox14*.

MATERIALS AND METHODS

Animal

Embryos of *N. forsteri* were obtained from the breeding colony established by Jean Joss at Macquarie University in Sydney, Australia (Macquarie University Animal Ethics Committee approval number: 2003/001). The embryos were kept in sterile pond water until they reached required stages, which were determined according to Kemp's normal table (Kemp, '82) and other supporting materials (<http://www.bio.mq.edu.au/dept/centres/lungfish/development/lungfishSQL.php>). Specimens used for RNA extraction were shipped in RNAlater (Qiagen, Hilden, Germany). Animals that were subjected to in situ hybridization were stored in methanol after fixation in 4% paraformaldehyde.

PCR

Total RNA was extracted using TRIzol (Invitrogen, Karlsruhe, Germany) from a whole embryo at stage 35. This RNA was reverse transcribed into cDNA using SuperScript III (Invitrogen), following the instructions of 3' RACE System (Invitrogen). This cDNA was used as template for a degenerate PCR using forward primers, which were designed based on amino acid stretches shared among Hox14 sequences of the Japanese lamprey, coelacanth, horn shark and elephant shark. Primer sequences were 5' CC GAR MGN CAR GTN AAR ATH TGG TT 3' (TERQVKIWF) for the first reaction and 5' G GTC AAR ATH TGG TTY CAR AAY CA 3' (QVKIWFQNO) for the nested reaction. The 5' end of the cDNA was obtained using the GeneRacer Kit (Invitrogen). These cDNA fragments were used as templates for the riboprobes for in situ hybridization. The assembled full length *N. forsteri* HoxA14 cDNA sequence is deposited in EMBL under the accession number FR751091.

cDNAs of the *eukaryotic translation elongation factor 1 α 1* (*EF 1 α 1*; often imprecisely designated *EF 1 α*) and *glyceraldehyde 3 phosphate dehydrogenase (GAPDH)* genes were isolated by degenerate PCR. The initial protein coding *EF 1 α 1* cDNA fragment was isolated using the forward primer 5' TC TAY AAR TGY GGN GGN ATH GAY AA 3' (YKCGGIDK) and the reverse primer 5' C ATA TCT CTT ACN GCR AAN CKN CCN A 3' (LGRFAVRDM). The 3' end of this cDNA was amplified with 3' RACE using a gene specific forward primer, 5' CACTGCTCA CATTGCCTGC 3'. The *GAPDH* sequence was amplified with 3' RACE using the forward primer 5' ATA WSW GCA CCW WSW GCN GAY GC 3' (ISAPSADA) in the first reaction and the forward primer 5' A CCT WSW GCW GAY GCN CCN ATG 3' (APSADAPM) in the nested reaction. These partial cDNA sequences are deposited in EMBL under the accession numbers FR751092 (*EF 1 α 1*) and FR751093 (*GAPDH*).

Retrieval of Non-lungfish Hox Sequences

Sequences of posterior *Hox* genes were retrieved from the Ensembl genome database (version 60; <http://www.ensembl.org>;

Hubbard et al., 2009) and NCBI Protein database, by running Blastp (Altschul et al., '97) using the newly identified lungfish HoxA14 peptide sequence as a query. An optimal multiple alignment of the retrieved amino acid sequences including the query was constructed using the alignment editor XCED, in which the MAFFT program is implemented (Katoh et al., 2005). For a list of sequences used in this study, see S Table 1.

Molecular Phylogenetic Analysis

Molecular phylogenetic trees were inferred using alignment of the 60 amino acids of the homeodomain, unless otherwise stated. To investigate phylogenetic relationships within the Hox14 PG (shown in Fig. 3A), we used PhyML (Guindon and Gascuel, 2003) for both neighbor joining (NJ) (Saitou and Nei, '87) and maximum likelihood (ML) tree inference, and MrBayes 3.1 (Huelsenbeck and Ronquist, 2001). Because the LG substitution matrix (Le and Gascuel, 2008) is not implemented in MrBayes 3.1, a transformed matrix, compatible with MrBayes 3.1, was obtained (<http://code.google.com/p/garli/source/browse/garli/trunk/example/LGmodel.mod?r=742>). The data set for this analysis contained all six vertebrate *Hox14* genes available (see *Introduction*) and the four human *Hox13* genes as outgroup, and this resulted in 112 amino acid residues that could be unambiguously aligned. The *P. annectens* HoxA14 gene was excluded from the phylogenetic analyses because of its incomplete homeodomain.

Similarly, we conducted a molecular phylogenetic analysis to compare the likelihood of two previously reported scenarios (S Fig. 1C and D, respectively; Holland et al., 2008; Thomas Chollier et al., 2010) and the two simple hypotheses (Fig. 5A and B; also see S Fig. 1A and B; Ferrier et al., 2000) for the evolution of the posterior *Hox* genes. The per site log likelihoods of the ML trees under these four scenarios (S Fig. 1) as well as the ML tree in a heuristic search (Fig. 5C) were calculated in RAxML v7.2.8 (Stamatakis, 2006). For this purpose, an enriched data set (the data set used below in the analysis on the possible orthology between vertebrate PG13 and *AmphiHox15*, plus all other amphioxus posterior *Hox* genes) was divided into eight operational taxonomic units (OTUs) as described in Results, and the ML trees under the two simple scenarios, out of 10,395 tree topologies, were exhaustively searched. The topologies of the ML trees under each scenario are depicted in S Figure 1.

To assess the statistical support for the orthology between *AmphiHox15* and the vertebrate PG13, all 10,395 possible tree topologies resulting from eight OTUs were assessed. ML trees were inferred using RAxML, assuming LG+F+ Γ_4 model (shape parameter of the gamma distribution $\alpha = 0.36$; Yang, '94). The data set used in this analysis consisted of *AmphiHox15*, all human posterior *Hox* genes (Hox9-13), and all six vertebrate *Hox14* genes. *Abd B* genes of two ecdysozoans (*Drosophila melanogaster* and *Priapulus caudatus*), and *Post2* genes of two lophotrochozoans (*Euprymna scolopes* and *Neanthes virens*) served as outgroup.

In both analyses, phylogenetic relationships within individual OTUs were constrained according to generally accepted phylogenetic relationships of relevant species. Relationships within the human posterior PGs were constrained based on the 1 2 4 pattern of the 2R WGD assuming that the A and B, and the C and D clusters are “sister clusters” [namely, ((A,B),(C,D))]; Amores et al., '98; see also Ravi et al., 2009].

Alternative tree topologies were statistically tested using CONSEL (Shimodaira and Hasegawa, 2001). *P* values of the approximately unbiased (AU) and the Shimodaira Hasegawa (SH) tests were calculated for selected tree topologies that supported particular scenarios, and compared with the ML trees.

In Situ Hybridization

The aforementioned 5' and 3' cDNA fragments were used as templates for the riboprobes used in situ hybridization. Whole mount and paraffin embedded section in situ hybridizations using *N. forsteri* embryos were performed as previously described (Murakami et al., 2001; Kuraku et al., 2005).

Real-Time RT-PCR

Three *N. forsteri* embryos (one embryo at stage 35 and two at stage 40) were dissected as shown in Figure 4A, resulting in eleven tissue fractions, designated a to k. Total RNA was extracted from each of these tissues using TRIzol (Invitrogen). The RNA was treated with DNase I (10 units for 1 µg of total RNA) for 15 min at room temperature. The DNA digestion was terminated by adding 1 µl EDTA (25 mM) and incubating at 65°C for 15 min.

In order to compare the expression level of *NfHoxA14* between the eleven tissue samples, the genes *GAPDH* and *EF 1α1* were used as internal controls (Van Hiel et al., 2009). Gene specific primers to amplify approximately 200 bp long cDNA fragments of *N. forsteri HoxA14*, *EF 1α1*, and *GAPDH*, respectively, were designed utilizing OligoPerfect™ (<http://tools.invitrogen.com/content.cfm?pageid=9716>). Sequences for the primers were: 5' GAGGAACAATGGTCTCTGAA 3' (forward) and 5' TGACATGTTTTGGTCATTGT 3' (reverse) for *EF 1α1*; 5' CTGTTTCATCAATGCTCCAT 3' (forward) and 5' TCACACAG CAGGTTTTGTT 3' (reverse) for *GAPDH*; and 5' GCTGCCTCAA TTTAAGAAAGT 3' (forward) and 5' AAAAGGCCAACACAGTAG 3' (reverse) for *HoxA14*. After confirming specificity of the primers in a test run, the analysis run was conducted using the Bio Rad CFX96 real time PCR system. A pre denaturation of 3 min at 95°C was followed by 50 cycles of three steps at different temperatures (95°C for 10 sec, 56°C for 10 sec, 72°C for 30 sec). A melting curve from 95 to 56°C was recorded for each reaction to monitor homogeneity of the amplified products.

The parameter used in the statistical evaluation was the threshold cycle [$C_{(t)}$], which was set by the Bio Rad CFX Manager software. The $C_{(t)}$ value gives the number of cycles in which the amplification curve of a given reaction reaches a fixed threshold

level in its exponential phase. Thus, the smaller the $C_{(t)}$ value of a reaction is, the higher the amount of initial cDNA template was.

Statistical evaluation was conducted with two data sets comprising two different internal control genes, *GAPDH* and *EF 1α1*. First, the average $C_{(t)}$ of the three replicates for each reaction was calculated, and then this value of the control gene [$C_{(t)ctrl}$] was subtracted from that of *NfHoxA14* [$C_{(t)HoxA14}$]. Thus, one value $\Delta C_{(t)}$ for each of the eleven samples and for each control gene was obtained. As these values are on an exponential scale, they had to be processed to make them linearly comparable. Additionally, the reciprocal value was calculated in order to produce the smallest final value for the reaction with the least initial *NfHoxA14* copy number. This processing after (Keegan et al., 2002) describes the formula:

$$2^{-[C_{(t)HoxA14} - C_{(t)ctrl}]}$$

The resulting values and their standard error of the mean were then plotted for each control gene (Fig. 4B).

RESULTS

Identification of a Hox14 cDNA in *N. forsteri*

By means of RT PCR, the full length cDNA of *N. forsteri HoxA14*, including 5' and 3' UTRs was sequenced. The affiliation of this gene to the vertebrate PG14 was suggested in a Blastx search against NCBI nonredundant protein sequences (nr) and confirmed by the program HoxPred (URL: <http://cege.vub.ac.be/hoxpred/>; Thomas Chollier et al., 2007) with the posterior probability of 1.0. A sequence alignment containing the six vertebrate *Hox14* genes available and the four human *Hox13* genes was constructed (Fig. 2). A high level of sequence conservation in the home odomain was revealed, and we identified in the *N. forsteri* sequence four amino acids that are exclusively shared by the Hox14 members, indicating their close relationship (Fig. 2).

There were three amino acid mismatches between the newly identified *N. forsteri* sequence and the previously reported *HoxA14* of the African lungfish *P. annectens* (accession number in NCBI Nucleotide, HQ441267) (Fig. 2). Between these two sequences we observed the number of synonymous substitution per site (K_s) of 0.79 ± 0.27 based on the method by Yang and Nielsen (2000) implemented in PAML (Yang, '97). In comparison to other pairs of species (Kuraku and Kuratani, 2006), the nonsaturated synonymous substitution between the two lungfish sequences indicates that they split much more recently than the early vertebrate era when the multiple *Hox* clusters were generated. For this reason, the two lungfish *HoxA14* genes should be orthologous.

Survey of Hox14 Members in Other Vertebrate Species

To search for members of the Hox14 PG within the mammalian and teleost lineages, tBlastn searches were performed online using the *N. forsteri HoxA14* peptide sequence as a query. First,

human <i>HoxA13</i>	GRKKRVPYTKVQLKELEREYATNKFITKDKRRRISATTNLSEKQVTIWFQNRVKEKKVI
human <i>HoxB13</i>	GRKKRIPYSGQLRELEREYAANKFITKDKRRKISAATSLSERQITWFQNRVKEKKVL
human <i>HoxC13</i>	GRKKRVPYTKVQLKELEKEYAASKFITKEKRRRISATTNLSEKQVTIWFQNRVKEKKVV
human <i>HoxD13</i>	GRKKRVPYTKLQLKELENEYAINKFINKDKRRRISAATNLSEKQVTIWFQNRVVKDKKIV
Australian lungfish <i>HoxA14</i>	QRKKRVPYTKYQITTELEKAFEVNRFLTPESRQHIKAVKLGKGLTERQVKIWFQNRQKEKKLL
African lungfish <i>HoxA14</i>	QRKKRVPYTKYQIAELEKAFEENRFLTPESRQVIKAVKLGKGLTERQVK
coelacanth <i>HoxA14</i>	QRKKRVPYSKHQITELERAFEENRFLTPETIRQNISVKLGKGLTERQVKIWFQNRQKEKKLL
elephant shark <i>HoxD14</i>	QRKKRIPYSKLQITTELEKAFENNRFLTPETIRLNISLKLGLTERQVKIWFQNRQKEKKLL
horn shark <i>HoxD14</i>	QRKKRVPYKQIQIAELEMAYENNRFLTPEVRLNISFKLGLTERQVKIWFQNRQKEKKLL
lesser spotted dogfish <i>HoxD14</i>	QRKKRIPYKQIQITTELEMAFENNRFLTPEVRLNISFKLGLTERQVKIWFQNRQKEKKLL
Japanese lamprey <i>Hox14α</i>	PRKKRVPYKQIQISELERAFDENRFLTPELRLSISHRLSLTERQVKIWFQNRQKEKKLM

Figure 2. Alignment of the 60 amino acid residues of the homeodomains of human *Hox13* and vertebrate *Hox14* genes. Amino acid residues specific to the PG14, based on comparison with human PG1-13 (Kuraku et al., 2008), are shown in bold. Note that the sequence of *HoxA14* of the African lungfish (accession number in NCBI Nucleotide, HQ441267; Liang et al., 2011) is incomplete.

we performed a search in NCBI dbEST as well as in nr/nt databases of all mammals (taxon ID: 40674) and teleost fishes (taxon ID: 32443). Second, we performed tBlastn searches against nucleotide genomic sequences of species included in the Ensembl Genome Browser. These searches resulted in no *Hox14* sequences in all available tetrapods and teleost fishes.

We also attempted to identify *Hox14* with RT PCR in chondrichthyans (*Raja clavata* and *Scyliorhinus canicula*), sturgeons (*Huso dauricus* and a hybrid between *Huso huso* and *Acipenser ruthenus*), a gar (*Lepisosteus platyrhinchus*), a bichir (*Polypterus senegalus*), and a hagfish (*Eptatretus burgeri*), but this survey resulted in no additional *Hox14* members (see Fig. 1 for phylogenetic positions of these species). This should be confirmed by the anticipated whole genome sequences of these species.

Phylogenetic Relationship Within Vertebrate *Hox14*

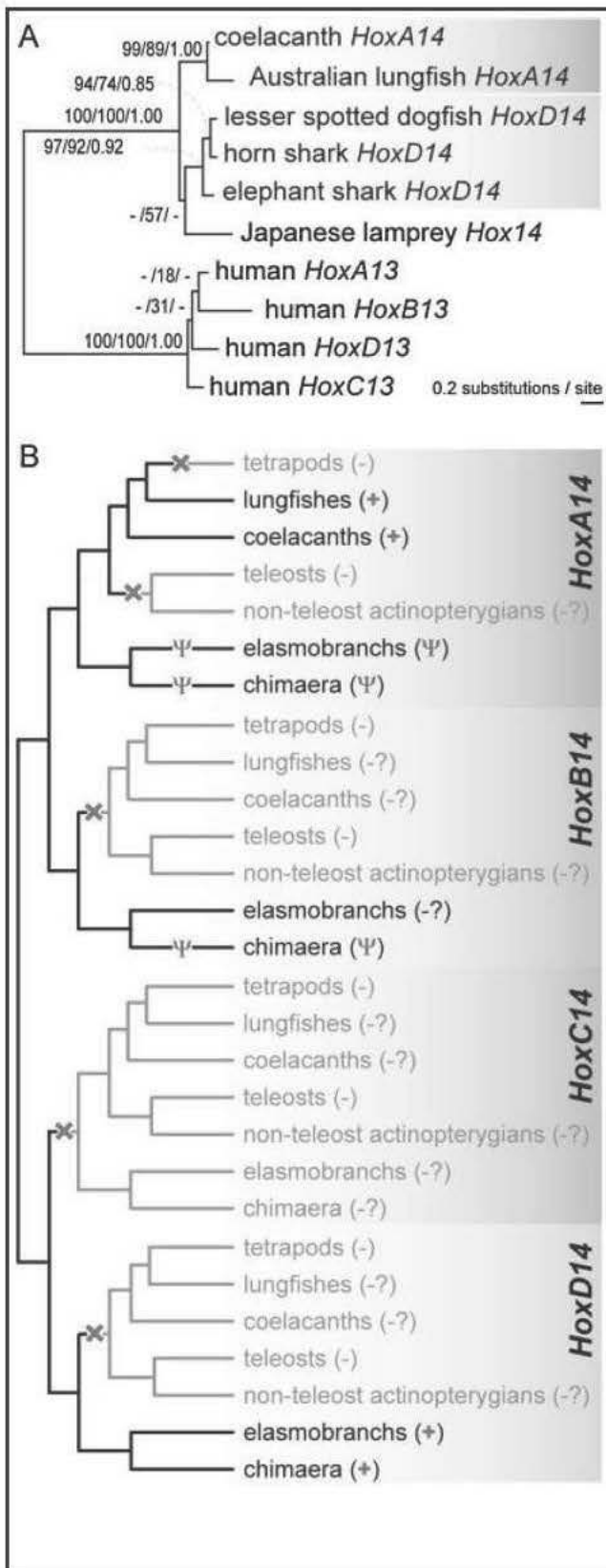
A sequence data set containing all six vertebrate *Hox14* sequences available and human *Hox13* genes as outgroup was used to reconstruct the phylogenetic relationships within the *Hox14* PG. The ML tree heuristically inferred (Fig. 3A) shows the high affinity of the newly identified Australian lungfish *HoxA14* to the coelacanth *HoxA14* gene (bootstrap probabilities of 99 in NJ, 89 in ML, and Bayesian posterior probability 1.00).

To further assess the statistical support for the close relationship of the newly identified Australian lungfish *Hox14* gene with the coelacanth *HoxA14*, an exhaustive analysis of all possible tree topologies resulting from seven OTUs (horn shark *HoxD14*, lesser spotted dogfish *HoxD14*, elephant shark *HoxD14*, coelacanth *HoxA14*, Australian lungfish *HoxA14*, Japanese lamprey *Hox14α* and four human *Hox13* genes) was conducted. The ML tree and alternative tree topologies with similar likelihood values

placed the newly identified *N. forsteri* sequence closest to coelacanth *HoxA14*. The tree topology with the largest likelihood which violates this lungfish coelacanth clustering was identified in this exhaustive search, and compared with the ML tree. This comparison provided *P* values of 0.18 in the AU test and 0.19 in the SH test for the tree violating the closest relationship between *N. forsteri HoxA14* and coelacanth *HoxA14* (ML tree: *P* = 0.82 in AU test, *P* = 0.81 in SH test). Although the non orthology of the Australian lungfish *Hox14* gene to the coelacanth *HoxA14* is not rejected at the 5% significance level, our analysis supported their orthology. Even though the most straightforward interpretation of the resultant tree topology (Fig. 3A) is that the lungfish *Hox14* gene belongs to the Hox A cluster, it is also possible that the lungfish gene belongs to the Hox B cluster as this would result in an identical tree topology. To confirm the putative genomic linkage of *NfHoxA14* with other HoxA members, a screening of a BAC library targeting the genomic region containing *NfHoxA14* was carried out but was unsuccessful (C. Amemiya, personal communication).

Embryonic Expression of *HoxA14* in the Australian Lungfish

Embryonic expression patterns of *N. forsteri HoxA14* were first analyzed with whole mount and section in situ hybridizations in stages 35 and 44. We observed strong ubiquitous expression signals of the *EF 1α1* gene, included as a positive control, but no signals were observed for the *HoxA14* gene, probably because of its possibly low expression level. Thus, differences in expression levels of *HoxA14* between various tissues were quantified by real time RT PCR (Fig. 4A). The result clearly showed that the expression level of *HoxA14* is highest in the sample including the hindgut at stage 40 (Fig. 4B). This observation is consistent between the two internal control genes. In the experiment with the

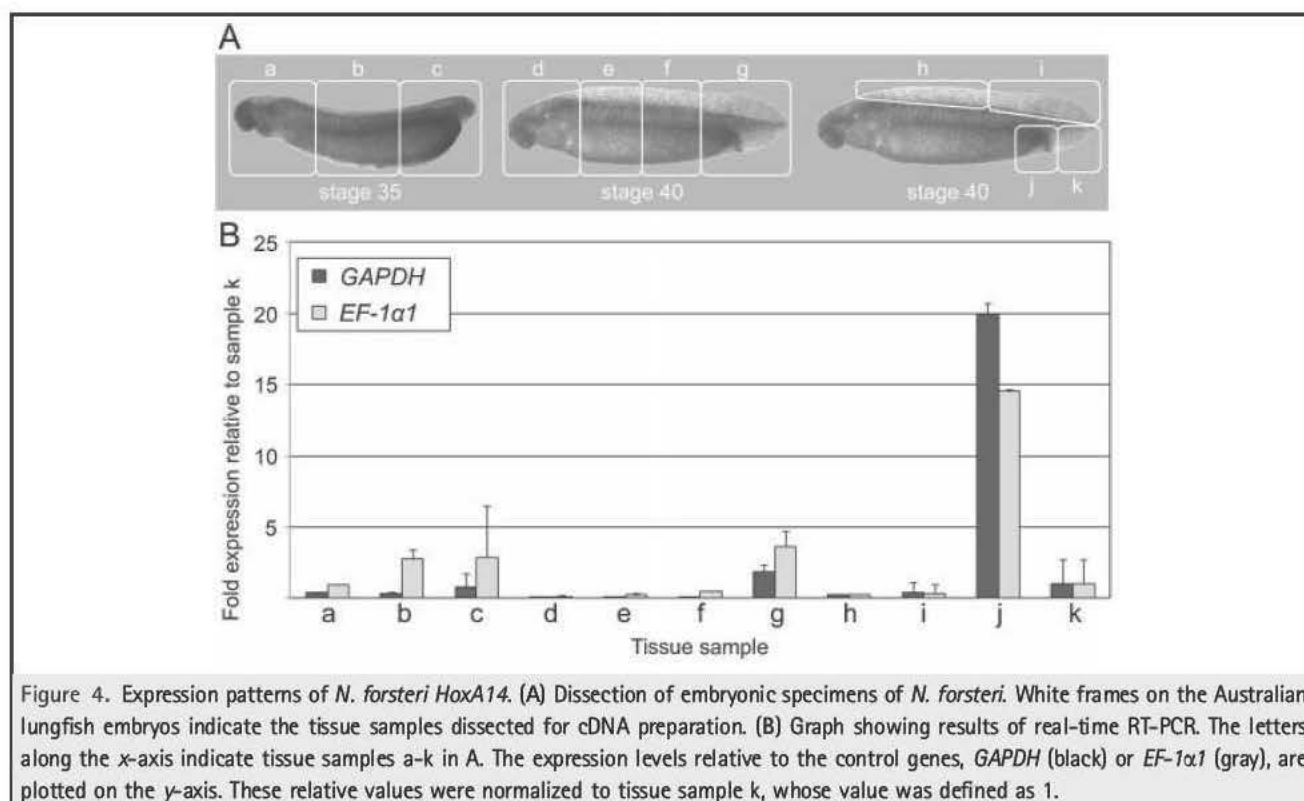


GAPDH gene included as an internal control, the up regulation of *HoxA14* in the hindgut region (sample j) compared with the tail bud region of the same stage (sample k) was 20 fold (Fig. 4B). The same comparison with *EF 1 α 1* as control showed a 14 fold up regulation. The sample g also exhibited a slightly higher level of *HoxA14* amplification, probably because of *HoxA14* expression in the hindgut region included in this tissue sample.

Orthology/Paralogy of Posterior *Hox* Genes Between Amphioxus and Vertebrates

To address the evolutionary origin of the vertebrate posterior PGs, the phylogenetic relationships between amphioxus and human posterior *Hox* genes were investigated. With an enriched sequence data set including amphioxus *Hox14* and *Hox15*, and the vertebrate PG14, we revisited the two simple hypotheses analyzed originally by Ferrier et al. (2000). Hypothesis A assumes independent (tandem) duplications in the amphioxus and the vertebrate lineage (Fig. 5A), whereas hypothesis B is based on a hypothetical last common ancestor which already possessed a tandemly duplicated set of posterior *Hox* genes, and thus each amphioxus posterior *Hox* gene is orthologous to one particular vertebrate PG (Fig. 5B). The ML tree under each hypothesis was inferred by constraining the following relationships: in hypothesis A, first the human genes were constrained arbitrarily, and the topology of the amphioxus posterior *Hox* genes was optimized in an exhaustive ML analysis. Second, the resulting ML tree topology was used to constrain the amphioxus posterior *Hox* genes in another exhaustive search for the best topology within the human posterior *Hox* genes. This process was repeated until

Figure 3. Phylogenetic relationships within the vertebrate PG14 and the inferred scenario of vertebrate *Hox14* evolution. (A) Molecular phylogeny of the six vertebrate *Hox14* genes for which the complete homeodomain sequence was available and human *Hox13* genes based on 112 amino acids. *Protopterus annectens HoxA14* (Liang et al., 2011) was excluded because of its insufficient length. Support values are shown for each node in order, bootstrap probabilities in the NJ and in the ML analysis, and Bayesian posterior probabilities. The LG +F+ Γ_4 model (shape parameter of gamma distribution $\alpha = 0.62$) was assumed. The human *Hox13* genes were chosen as outgroup because the PG13 is the one which is phylogenetically closest to the PG14 (see Fig. 5C). (B) Phylogenetic distribution of vertebrate *Hox14* genes and their taxon-specific absence. The timings of secondary gene losses (marked by "X") were inferred based on most parsimonious interpretation (also see Results). Pseudogenization events are denoted by the symbol "Ψ". "-" marks lineages without fully sequenced genomes in which no *Hox14* member was identified to date. Note that the elephant shark *Hox* clusters contain one intact *HoxD14* gene, and pseudogenized *HoxA14* and *HoxB14* (Ravi et al., 2009), and the horn shark has an intact *HoxD14* gene as well as a pseudogenized *HoxA14* (Powers and Amemiya, 2004).



no more changes in topologies were observed. In hypothesis B, eight OTUs were defined, namely one for each *Hox9* to *Hox15* (each amphioxus gene was grouped together with its assumed human orthologs) and an outgroup. The assumptions about orthologous relationships among the *Hox* genes were based on their relative locations in the *Hox* cluster assuming conserved synteny between amphioxus and human. For optimized ML tree topologies of the two scenarios, see S Figure 1A and B. When we compared these ML trees under these two simple hypothesis with the heuristic ML tree (Fig. 5C), statistical tests significantly rejected the tree topologies based on the two simple hypotheses at the 5% level (Hypothesis A: $P < 0.01$ in AU test, $P = 0.02$ in SH test; Hypothesis B: $P = 0.02$ in AU test, $P = 0.03$ in SH test; ML tree: $P = 0.88$ in AU test, $P = 0.98$ in SH test; see S Table 2). The two previously proposed hypotheses (S Fig. 1C and D) were not clearly rejected at the 5% level (S Table 2).

Possible Orthology Between *AmphiHox15* and Vertebrate PG13

We also assessed the possible 1 to 1 orthology between *AmphiHox15* and vertebrate PG13, suggested previously (Holland et al., 2008; Thomas Chollier et al., 2010). An exhaustive ML analysis was performed with eight OTUs, namely the human posterior PG9 13, vertebrate *Hox14* genes, *AmphiHox15* and outgroup (S Table 3). The ML tree supported the orthology between *AmphiHox15* and vertebrate *Hox13* ($P = 0.88$ in AU test

and $P = 1.00$ in SH test; S Table 3). The best tree violating this relationship (rank 18 in S Table 3) favored the orthology between *AmphiHox15* and vertebrate *Hox14*. The 1 to 1 comparison between the ML tree and the best tree topology violating the orthology between *AmphiHox15* and vertebrate *Hox13* revealed that the latter was not significantly rejected by the AU test ($P = 0.14$), and the SH test ($P = 0.17$) at the 5% level.

DISCUSSION

Phylogenetic Distribution of Vertebrate *Hox14*

The vertebrate PG14 failed to be identified until 2004, because it is not present in the tetrapod and teleost lineages, which contain virtually all of the fully sequenced vertebrate genomes to date. The *Hox14* members identified to date are restricted to more basal vertebrates, such as lamprey (Kuraku et al., 2008), chondrichthyans (Powers and Amemiya, 2004; Ravi et al., 2009), lungfish (Liang et al., 2011 and this study) and coelacanth (Powers and Amemiya, 2004). Interestingly, no single vertebrate species has been found to possess more than one functional *Hox14* gene (Fig. 3B). The restricted phylogenetic distribution implies that the evolutionary history of the vertebrate PG14 is characterized by frequent secondary gene losses (Fig. 3B). For example, no *HoxC14* gene has been identified to date, and was most likely lost immediately after the 2R WGD (Fig. 3B). In contrast to *HoxC14*, the timings of gene loss events of other

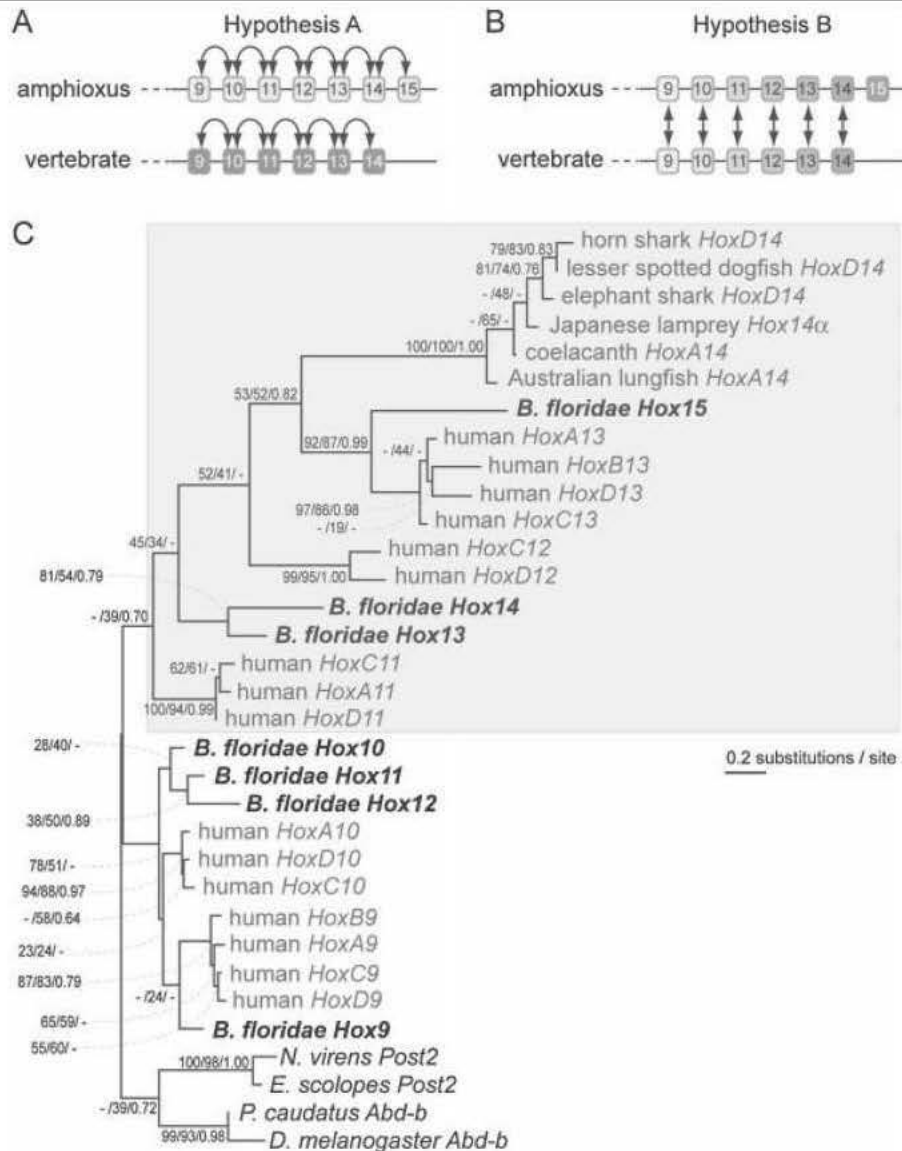


Figure 5. Phylogenetic relationships between the chordate posterior *Hox* genes. (A) Hypothesis A. This scenario is based on independent tandem duplications (arrows) in the amphioxus and the vertebrate *Hox* cluster which gave rise to the posterior *Hox* genes. (B) Hypothesis B. This scenario implies a fully duplicated set of posterior *Hox* genes which existed already before the split between cephalochordates and vertebrates. Based on syntenic relationships, the orthology (arrows) between each amphioxus posterior *Hox* gene and its putative vertebrate counterpart is assigned. Likelihood values of the best tree topology of each proposed hypothesis was calculated assuming LG+F+ Γ_4 model (shape parameter $\alpha = 0.45$). (C) The ML tree obtained in a topology analysis. Note that the tree topology is significantly different from those in A and B. The gray background indicates the part of the tree whose topology is identical to previous reports (Holland et al., 2008; Thomas-Chollier et al., 2010). Support values at nodes are shown in order, bootstrap probabilities in the NJ and the ML analysis, and Bayesian posterior probabilities. See S-Table 1 for accession IDs of the included sequences.

Hox14 genes cannot be precisely mapped onto the vertebrate species tree. More sequence data of non teleost actinopterygians (bichir, sturgeon, paddlefish, gar and bowfin) or cyclostomes (hagfish and lamprey) could potentially reveal more cryptic

Hox14 genes, which would lead to a more detailed picture of vertebrate PG14 evolution.

Although the *Hox* clusters of the crown teleosts (Clupeocephala) were investigated genome wide in great detail, our current

knowledge about the *Hox* clusters of nonteleost actinopterygians and basal teleost fish species (Osteoglossomorpha and Elopomorpha; see Fig. 1) is sparse. The only studies performed to date are PCR surveys of *Hox* gene repertoires in the basal teleosts, Japanese eel (*Anguilla japonica*; Elopomorpha) and the goldeye (*Hiodon alosoides*; Osteoglossomorpha), and a basal actinopterygian, a bichir (*Polypterus palmas*) (Ledje et al., 2002; Chambers et al., 2009; Guo et al., 2010). To gain a full picture of the phylogenetic distribution of vertebrate Hox14, genome wide resources for these animals are still awaited. In contrast, abundant sequence data is available for laboratory teleost fish models, and the absence of any Hox14 sequence from whole genome data and EST databases is convincing evidence for the loss of the PG14, likely early in teleost or actinopterygian fish evolution (Fig. 3B).

Functional Evolution of Hox14

The lungfish Hox14 member we identified in this study belongs most likely to the Hox A cluster (Fig. 3A). It should be noted that this interpretation could be misled by so called “hidden paralogy” if more gene losses than estimated by the most parsimonious scenario had occurred (Kuraku, 2010). Unfortunately, an attempt to screen a *N. forsteri* BAC library failed to isolate clones containing *N. forsteri HoxA14*, and thus the physical linkage of *N. forsteri HoxA14* to the Hox A cluster still needs to be proven (C. Amemiya, personal communication).

Our phylogenetic analysis revealed a high affinity of lamprey *Hox14 α* to jawed vertebrate *HoxD14* genes (Fig. 3A). This possible orthology suggests that the lamprey once experienced or still maintains a condition with four *Hox* clusters, and that cyclostomes diverged after the quadruplication of the ancestral vertebrate genome (Kuraku et al., 2009; reviewed in Kuraku, 2008).

Expression data of *Hox14* genes has been revealed to date only in the cloudy catshark, lesser spotted dogfish and the lamprey (Kuraku et al., 2008; Oulion et al., 2011). Lamprey *Hox14 α* and *HoxD14* of the two sharks share the hindgut associated expression (Kuraku et al., 2008; Oulion et al., 2011). Our real time PCR analysis in lungfish embryos indicated significant expression of *NfHoxA14* in the hindgut containing tissue sample, but not in other tissue samples (Fig. 4). Thus, all vertebrate *Hox14* genes analyzed to date show no significant expression in the CNS, somites or fin buds, in which at least some of PG1-13 genes are expressed in a colinear fashion (Dolle et al., '89; Hunt et al., '91). *HoxA14* of lungfish (Fig. 4) and *HoxD14* of the cloudy catshark (Kuraku et al., 2008) share the hindgut associated expression despite their assignment to different *Hox* clusters (A and D, respectively). This suggests an early establishment of this shared expression pattern and its decoupling from the Hox code already before the 2R WGD.

Independent Origins of Vertebrate and Amphioxus Hox14

The ancient decoupling of vertebrate Hox14 from the Hox code (Kuraku et al., 2008) raises the question about the phylogenetic

origin of vertebrate Hox14. The single *Hox* cluster of the cephalochordate amphioxus also possesses a gene called *Hox14*. This gene was previously shown to be not orthologous to the vertebrate PG14, but rather to be derived from a tandem duplication in the amphioxus lineage (Powers and Amemiya, 2004; Kuraku et al., 2008). Our analysis also suggests that there is no ortholog of the vertebrate PG14 in amphioxus (Fig. 5C). This result can be explained by two alternative scenarios. The first scenario is that the origin of vertebrate PG14 dates back to a vertebrate specific tandem duplication before the 2R WGD, but after the cephalochordate vertebrate split. The second scenario is that an ortholog of vertebrate Hox14 already existed in the last common ancestor of chordates, but was secondarily erased from the amphioxus *Hox* cluster. Previous studies supported the second scenario (Holland et al., 2008; Thomas Chollier et al., 2010). In fact, our phylogenetic analysis, based on the more up to date and focused dataset, also favors the second scenario (Fig. 5C).

Previous studies supported the orthology between vertebrate PG13 and *AmphiHox15* (Holland et al., 2008; Thomas Chollier et al., 2010). Our analysis also strongly supports this orthology (S Table 3; Fig. 5C), although this result is not significantly supported. Overall, the present study does not support any 1 to 1 orthology of posterior *Hox* genes between amphioxus and vertebrates based on their relative location in the cluster (Fig. 5C; S Table 2). Our phylogenetic analysis, based on the enriched data set, statistically rejected the two simple scenarios which assume either independent tandem duplications after the split between amphioxus and vertebrate lineages (Fig. 5A) or full retention of genes derived from tandem duplications before the split between amphioxus and vertebrate lineages (Fig. 5B; S Table 2). Hence, as in previous studies (Holland et al., 2008; Hueber et al., 2010; Thomas Chollier et al., 2010), our analysis contradicts the paradigm of the “deuterostome posterior flexibility” that postulates obscured 1 to 1 orthologies (Hypothesis B in Fig. 5B; see *Introduction*; Ferrier et al., 2000). If the “posterior flexibility” is true, it would violate the modern methodological framework of molecular phylogenetics on which the convincing results of unlikelihood of the scenario (S Table 2) is based. The enriched data set does not contain sufficient phylogenetic signals to confidently support a particular scenario, but at least contains sufficient information to rule out the possibility of ancient tandem duplications before the separation of amphioxus and vertebrate (Fig. 5B) as well as the hypothesis in Figure 5A. The most likely scenario could be that the posterior *Hox* genes of amphioxus and vertebrates are derived from an ancestor which possessed a subset of posterior *Hox* genes, and that lineage specific tandem duplications, and secondary gene losses shaped the *Hox* clusters differentially between these two lineages. Thus, the amphioxus *Hox* cluster should not be regarded as “archetypal” (Amemiya et al., 2008). From the viewpoint of vertebrate evolution, the hindgut associated expression of Hox14 and its decoupling from the Hox code are special features as a PG unique

to vertebrates. It is also remarkable that Hox14 is, to our knowledge, always represented by at most one gene per species, because one of the specialties of vertebrate Hox functions is the redundancy achieved by multiplied clusters, seen in almost all PGs in species examined to date (except for PG7 in teleost fishes and PG12 in *Xenopus tropicalis*; see Kuraku and Meyer, 2009).

ACKNOWLEDGMENTS

This study was supported by the Young Scholar Fund of the University of Konstanz and the research grant (KU2669/1 1) from German Research Foundation (DFG) to S.K., and also by the International Max Planck Research School (IMPRS) for Organismal Biology to N.F. We thank Jean M.P. Joss for generous supply of *N. forsteri* embryos, Ursula Topel, Adina J. Renz, and Elke Hespeler for technical support in cDNA cloning and in situ hybridization, Tereza Manousaki for assistance in molecular phylogenetic analyses, and Pincelli Hull for critically reading the manuscript.

LITERATURE CITED

- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402.
- Amemiya CT, Prohaska SJ, Hill-Force A, Cook A, Wasserscheid J, Ferrier DE, Pascual-Anaya J, Garcia-Fernandez J, Dewar K, Stadler PF. 2008. The amphioxus Hox cluster: characterization, comparative genomics, and evolution. *J Exp Zool B Mol Dev Evol* 310:465–477.
- Amores A, Force A, Yan YL, Joly L, Amemiya C, Fritz A, Ho RK, Langeland J, Prince V, Wang YL, Westerfield M, Ekker M, Postlethwait JH. 1998. Zebrafish *hox* clusters and vertebrate genome evolution. *Science* 282:1711–1714.
- Chambers KE, McDaniell R, Raincrow JD, Deshmukh M, Stadler PF, Chiu CH. 2009. Hox cluster duplication in the basal teleost *Hiodon alosoides* (Osteoglossomorpha). *Theory Biosci* 128:109–120.
- Delsuc F, Brinkmann H, Chourrout D, Philippe H. 2006. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* 439:965–968.
- Dolle P, Izpisua-Belmonte JC, Falkenstein H, Renucci A, Duboule D. 1989. Coordinate expression of the murine Hox-5 complex homeobox-containing genes during limb pattern formation. *Nature* 342:767–772.
- Duboule D. 1994. Temporal colinearity and the phylotypic progression: a basis for the stability of a vertebrate bauplan and the evolution of morphologies through heterochrony. *Dev Suppl* 135–142.
- Ferrier DE. 2004. Hox genes: did the vertebrate ancestor have a Hox14? *Curr Biol* 14:R210–R211.
- Ferrier DE, Minguillon C, Holland PW, Garcia-Fernandez J. 2000. The amphioxus hox cluster: deuterostome posterior flexibility and Hox14. *Evol Dev* 2:284–293.
- Garcia-Fernandez J. 2005. The genesis and evolution of homeobox gene clusters. *Nat Rev Genet* 6:881–892.
- Graham A, Papalopulu N, Krumlauf R. 1989. The murine and drosophila homeobox gene complexes have common features of organization and expression. *Cell* 57:367–378.
- Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52: 696–704.
- Guo B, Gan X, He S. 2010. Hox genes of the Japanese eel *Anguilla japonica* and Hox cluster evolution in teleosts. *J Exp Zool B Mol Dev Evol* 314:135–147.
- Heinicke MP, Naylor GP, Hedges SB. 2009. Cartilaginous fishes (Chondrichthyes). In: Hedges SB, Kumar S, editors. *The timetree of life*. New York: Oxford University Press. p 320–327.
- Holland LZ, Albalat R, Azumi K, Benito-Gutierrez E, Blow MJ, Bronner-Fraser M, Brunet F, Butts T, Candiani S, Dishaw LJ, Ferrier DE, Garcia-Fernandez J, Gibson-Brown JJ, 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 GW, McHardy AC, Meulemans D, Nonaka M, Olinski RP, Pancer Z, Pennacchio LA, Pestarino M, Rast JP, 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 JK, Zhang Q, Zmasek CM, de Jong PJ, Osoegawa K, Putnam NH, Rokhsar DS, Satoh N, Holland PW. 2008. The amphioxus genome illuminates vertebrate origins and cephalochordate biology. *Genome Res* 18:1100–1111.
- Hubbard TJ, Aken BL, Ayling S, Ballester B, Beal K, Bragin E, Brent S, Chen Y, Clapham P, Clarke L, Coates G, Fairley S, Fitzgerald S, Fernandez-Banet J, Gordon L, Graf S, Haider S, Hammond M, Holland R, Howe K, Jenkinson A, Johnson N, Kahari A, Keefe D, Keenan S, Kinsella R, Kokocinski F, Kulesha E, Lawson D, Longden I, Megy K, Meidl P, Overduin B, Parker A, Pritchard B, Rios D, Schuster M, Slater G, Smedley D, Spooner W, Spudich G, Trevanion S, Vilella A, Vogel J, White S, Wilder S, Zadissa A, Birney E, Cunningham F, Curwen V, Durbin R, Fernandez-Suarez XM, Herrero J, Kasprzyk A, Proctor G, Smith J, Searle S, Flicek P. 2009. Ensembl 2009. *Nucleic Acids Res* 37:D690–697.
- Hueber SD, Weiller GF, Djordjevic MA, Frickey T. 2010. Improving Hox protein classification across the major model organisms. *PLoS One* 5:e10820.
- Huelsenbeck JP, Ronquist F. 2001. MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–755.
- Hunt P, Whiting J, Nonchev S, Sham MH, Marshall H, Graham A, Cook M, Allemann R, Rigby PW, Gulisano M, Faiella A, Boncinelli E, Krumlauf R. 1991. The branchial Hox code and its implications for gene regulation, patterning of the nervous system and head evolution. *Dev Suppl* 2:63–77.
- Inoue JG, Miya M, Tsukamoto K, Nishida M. 2003. Basal actinopterygian relationships: a mitogenomic perspective on the phylogeny of the “ancient fish”. *Mol Phylogenet Evol* 26:110–120.
- Inoue JG, Miya M, Lam K, Tay BH, Danks JA, Bell J, Walker TI, Venkatesh B. 2010. Evolutionary origin and phylogeny of the

- modern holocephalans (Chondrichthyes: Chimaeriformes): a mitochondrial perspective. *Mol Biol Evol* 27:2576–2586.
- Katoh K, Kuma K, Toh H, Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res* 33:511–518.
- Keegan BR, Feldman JL, Lee DH, Koos DS, Ho RK, Stainier DY, Yelon D. 2002. The elongation factors Pandora/Spt6 and Foggy/Spt5 promote transcription in the zebrafish embryo. *Development* 129:1623–1632.
- Kemp A. 1982. The embryological development of the Queensland lungfish, *Neoceratodus forsteri* (Kreffl). *Mem Qld Mus* 20:553–597.
- Kikugawa K, Katoh K, Kuraku S, Sakurai H, Ishida O, Iwabe N, Miyata T. 2004. Basal jawed vertebrate phylogeny inferred from multiple nuclear DNA-coded genes. *BMC Biol* 2:3.
- Kmita M, Duboule D. 2003. Organizing axes in time and space; 25 years of colinear tinkering. *Science* 301:331–333.
- Kuraku S. 2008. Insights into cyclostome phylogenomics: pre-2R or post-2R? *Zool Sci* 25:960–968.
- Kuraku S. 2010. Palaeogenomics of the vertebrate ancestor—impact of hidden paralogy in hagfish and lamprey gene phylogeny. *Integr Comp Biol* 50:124–129.
- Kuraku S., Kuratani S. 2006. Time scale for cyclostome evolution inferred with a phylogenetic diagnosis of hagfish and lamprey cDNA sequences. *Zool Sci* 23:1053–1064.
- 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.
- Kuraku S, Usuda R, Kuratani S. 2005. Comprehensive survey of carapacial ridge-specific genes in turtle implies co-option of some regulatory genes in carapace evolution. *Evol Dev* 7:3–17.
- Kuraku S, Takio Y, Tamura K, Aono H, Meyer A, Kuratani S. 2008. Noncanonical role of Hox14 revealed by its expression patterns in lamprey and shark. *Proc Natl Acad Sci USA* 105:6679–6683.
- Kuraku S, Meyer A, Kuratani S. 2009. Timing of genome duplications relative to the origin of the vertebrates: did cyclostomes diverge before or after? *Mol Biol Evol* 26:47–59.
- Le SQ, Gascuel O. 2008. An improved general amino acid replacement matrix. *Mol Biol Evol* 25:1307–1320.
- Ledje C, Kim CB, Ruddle FH. 2002. Characterization of Hox genes in the bichir, *Polypterus palmis*. *J Exp Zool (Mol Dev Evol)* 294:107–111.
- Lewis EB. 1978. A gene complex controlling segmentation in *Drosophila*. *Nature* 276:565–570.
- Liang D, Wu R, Geng J, Wang C, Zhang P. 2011. A general scenario of Hox gene inventory variation among major sarcopterygian lineages. *BMC Evol Biol* 11:25.
- McGinnis W, Krumlauf R. 1992. Homeobox genes and axial patterning. *Cell* 68:283–302.
- Meyer A, Zardoya R. 2003. Recent advances in the (molecular) phylogeny of vertebrates. *Annu Rev Ecol Evol Syst* 34:311–338.
- Murakami Y, Ogasawara M, Sugahara F, Hirano S, Satoh N, Kuratani S. 2001. Identification and expression of the lamprey Pax6 gene: evolutionary origin of the segmented brain of vertebrates. *Development* 128:3521–3531.
- Oulion S, Debiais-Thibaud M, d'Aubenton-Carafa Y, Thermes C, Da Silva C, Bernard-Samain S, Gavory F, Wincker P, Mazan S, Casane D. 2010. Evolution of Hox gene clusters in gnathostomes: insights from a survey of a shark (*Scyliorhinus canicula*) transcriptome. *Mol Biol Evol* 27:2829–2838.
- Oulion S, Borday-Birraux V, Debiais-Thibaud M, Mazan S, Laurenti P, Casane D. 2011. Evolution of repeated structures along the body axis of jawed vertebrates, insights from the *Scyliorhinus canicula* hox code. *Evol Dev* 13:247–259.
- Powers TP, Amemiya CT. 2004. Evidence for a Hox14 paralog group in vertebrates. *Curr Biol* 14:R183–R184.
- Ravi V, Lam K, Tay BH, Tay A, Brenner S, Venkatesh B. 2009. Elephant shark (*Callorhynchus milii*) provides insights into the evolution of Hox gene clusters in gnathostomes. *Proc Natl Acad Sci USA* 106:16327–16332.
- Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425.
- Shimodaira H, Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* 17:1246–1247.
- Stamatakis A. 2006. RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.
- Thomas-Chollier M, Leyns L, Ledent V. 2007. HoxPred: automated classification of hox proteins using combinations of generalised profiles. *BMC Bioinformatics* 8:247.
- Thomas-Chollier M, Ledent V, Leyns L, Vervoort M. 2010. A non-tree-based comprehensive study of metazoan Hox and ParaHox genes prompts new insights into their origin and evolution. *BMC Evol Biol* 10:73.
- Van Hiel MB, Van Wielendaele P, Temmerman L, Van Soest S, Vuerinckx K, Huybrechts R, Broeck JV, Simonet G. 2009. Identification and validation of housekeeping genes in brains of the desert locust *Schistocerca gregaria* under different developmental conditions. *BMC Mol Biol* 10:56.
- Venkatesh B, Kirkness EF, Loh YH, Halpern AL, Lee AP, Johnson J, Dandona N, Viswanathan LD, Tay A, Venter JC, Strausberg RL, Brenner S. 2007. Survey sequencing and comparative analysis of the elephant shark (*Callorhynchus milii*) genome. *PLoS Biol* 5:e101.
- Yang Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J Mol Evol* 39:306–314.
- Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 13:555–556.
- Yang Z, Nielsen R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol Biol Evol* 17:32–43.
- Zeltser L, Desplan C, Heintz N. 1996. Hoxb-13: a new Hox gene in a distant region of the HOXB cluster maintains colinearity. *Development* 122:2475–2484.