BMC Genetics

RESEARCH ARTICLE Open Access

Hox genes reveal genomic DNA variation in tetraploid hybrids derived from *Carassius auratus* red var. (female) × *Megalobrama amblycephala* (male)

Y. D. Wang[1,2†], Q. B. Qin[1,2†], R. Yang[1,2†], W. Z. Sun[1,2†], Q. W. Liu[1,2], Y. Y. Huo[1,2], X. Huang[1,2], M. Tao[1,2], C. Zhang[1,2], T. Li[1,2] and S. J. Liu[1,2*]

Abstract

Background: Allotetraploid F_1 hybrids ($4nF_1$) (AABB, 4n = 148) were generated from the distant hybridization of *Carassius auratus* red var. (RCC) (AA, 2n = 100) (♀) × *Megalobrama amblycephala* (BSB) (BB, 2n = 48) (♂). It has been reported that *Hox* gene clusters are highly conserved among plants and vertebrates. In this study, we investigated the genomic organization of *Hox* gene clusters in the allotetraploid F_1 hybrids and their parents to investigate the polyploidization process.

Results: There were three copies of *Hox* genes in the $4nF_1$ hybrids, two copies in RCC and one copy in BSB. In addition, obvious variation and pseudogenization were observed in some *Hox* genes from $4nF_1$.

Conclusion: Our results reveal the influence of polyploidization on the organization and evolution of *Hox* gene clusters in fish and also clarify some aspects of vertebrate genome evolution.

Keywords: Allotetraploid, *Hox gene*, polyploidization, pseudogenization

Background

Polyploidization is a widespread mechanism for speciation in eukaryotes, especially plants and vertebrates [1–5]. Polyploids with duplicated genomes may originate from a single species (autopolyploidy) or from different species through interspecific hybridization (allopolyploidy) [6]. Allopolyploids are prevalent in nature, suggesting there is an evolutionary advantage to obtaining multiple sets of genetic material for adaptation and development [7]. However, the molecular mechanisms underlying the processes and consequences of allopolyploidy remain unclear [8]. Polyploidy is relatively rare in animals compared with plants, and the influence of polyploidization on intragenomic variation in polyploid animals is poorly understood. In our earlier study, we successfully obtained fertile tetraploid hybrids from *Carassius auratus* red var. (RCC)(♀) × *Megalobrama amblycephala* (BSB)(♂) [9, 10]. RCC has 100 chromosomes and belongs to the Cyprinidae subfamily, while BSB has 48 chromosomes and belongs to the Cultrinae subfamily [11]. These new polyploid hybrids represent unique specimens for studying genomic changes in F_1 hybrids and could significantly contribute to our understanding of evolution.

Hox genes, a set of important developmental regulatory genes, are highly conserved and typically organized cluster [12]. In vertebrates, *Hox* genes encode two exons, and the highly conserved homeodomain (60 aa) is encoded by the second exon [13]. Recent research has shown that gene duplication, sequence variation, and selective pressure played crucial roles in the origin and evolution of *Hox* genes [14]. The earliest indications of genome duplication came from the comparative analysis of *Hox* genes and clusters from different chordate lineages [15–18].

* Correspondence: lsj@hunnu.edu.cn
†Equal contributors
[1]State Key Laboratory of Developmental Biology of Freshwater Fish, Hunan Normal University, Changsha 410081, Hunan, People's Republic of China
[2]College of Life Sciences, Hunan Normal University, Changsha 410081, Hunan, People's Republic of China

© The Author(s). 2017 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

Table 1 PCR amplification bands in RCC, BSB and $4nF_1$

Species	Locus	Size (bp)	Exon1 (bp)	Intron (bp)	Exon 2 (bp)	GenBank accession no.
RCC	*HoxA4ai*	1181	89-500	501-974	975-1181	JX282274
	HoxA4aii	1184	89-500	501-977	978-1184	JX282275
	HoxA9ai	867	1-381	382-670	671-867	JX282276
	HoxA9aii	819	1-381	382-622	623-819	JX282277
	HoxA2bi	1486	1-314	315-901	902-1486	JX282278
	HoxA2bii	1448	1-314	315-863	864-1448	JX282279
	HoxD4ai	960	1-315	316-735	736-960	JX282280
	HoxD4aii	952	1-315	316-719	728-952	JX282281
BSB	*HoxA4a*	1188	89-500	501-981	982-1188	JX282282
	HoxA9a	879	1-381	382-682	683-879	JX282283
	HoxA2b	1479	1-311	312-894	895-1479	JX282284
	HoxD4a	911	1-306	307-686	687-911	JX282285
$4nF_1$	*HoxA4ai*	1183	89-500	501-976	977-1183	JQ901468
	HoxA4aii	1169	89-500	501-962	963-1169	JX282286
	HoxA4aiii	1177	89-500	501-970	971-1177	JX282287
	HoxA9ai	867	1-381	382-670	671-867	JX282288
	HoxA9aii	817	1-381	382-620	621-817	JX282289
	HoxA9aiii	863	1-381	382-666	667-863	JX282290
	HoxA2bi	1486	1-314	315-901	902-1486	JX282291
	HoxA2bii	1448	1-314	315-863	864-1448	JX282292
	HoxA2biii	1475	1-314	315-890	891-1475	JX282293
	HoxA11bi	1251	3-590	591-1153	1154-1251	JX282294
	HoxA11bii	1411	3-590	591-1313	1314-1411	JX282295
	HoxA11biii	1437	3-590	591-1339	1340-1437	JX282296
	HoxB1bi	733	1-477	478-567	568-733	JX282297
	HoxB1bii	734	1-477	478-568	569-734	JX282298
	HoxB1biii	731	1-477	478-565	566-731	JX282299
	HoxB5bi	1196	1-561	562-990	991-1196	JX282300
	HoxB5bii	1195	1-561	562-989	990-1196	JX282301
	HoxB5biii	1190	1-561	562-984	985-1190	JX282302
	HoxB6bi	807	1-169	170-667	668-807	JX282303
	HoxB6bii	819	1-169	170-679	680-819	JX282304
	HoxB6biii	812	1-169	170-672	673-819	JX282304
	HoxC4ai	1176	1-410	411-935	936-1176	JX282306
	HoxC4aii	1173	1-410	411-932	933-1173	JX282307
	HoxC4aiii	1169	1-410	411-928	929-1169	JX282308
	HoxC4a-1	1179	1-410	411-938	939-1179	JX282309
	HoxD4ai	960	1-315	316-735	736-960	JX282310
	HoxD4aii	952	1-315	316-719	728-952	JX282311
	HoxD4aiii$^{\Psi}$	942	-	-	-	JX282312
	HoxD9a$^{\Psi}$	897	-	-	-	JX282313
	HoxD10a$^{\Psi}$	1481	-	-	-	JX282314
	HoxD10aii	1554	1-589	590-1324	1325-1554	JX282315
	HoxD10aiii	1495	1-592	593-1265	1266-1495	JX282316

Ψ denotes a pseudogene

In general, polyploidization plays an important role in fish evolution [19]. The purpose of this research was to study the effects of allopolyploidization on *Hox* gene organization and evolution. In this article, three distinct *Hox* duplicates were observed in the $4nF_1$ genome, compared with two copies in RCC and one copy in BSB. Our data reveal the genetic variation and evolutionary characteristics of the *Hox* gene family in $4nF_1$ and provide new insights into their evolutionary patterns.

Results

Sequence information for RCC, BSB and $4nF_1$ clones

Using 11 pairs of degenerate primers (Additional file 1: Table S1), we obtained partial sequence information for eight putative *Hox* genes from RCC, four putative *Hox* genes from BSB, and 32 putative *Hox* genes from the $4nF_1$. All these fragments were between 700 and 1500 bp long and included the exon1-intron-exon2 region (Table 1). To avoid biased amplification of only one *Hox* gene copy, we selected 20 clones of each gene from $4nF_1$, 20 clones of each gene from RCC and 80 clones from BSB (20 clones for each *Hox* gene PCR fragment). All fragments from RCC, BSB and the $4nF_1$ were confirmed to be *Hox* gene sequences, and each included the homeobox. All *Hox* sequences have been submitted to GenBank; their accession numbers are listed in Table 1.

Molecular organization of the *Hox* gene sequence

We comparatively analysed the inferred amino acid sequences of the *Hox* genes in $4nF_1$ with those in zebrafish, fugu, medaka, and BSB (Additional file 1: Table S2), which indicated that the $4nF_1$ sequences were similar to those of the other species. The organization of the *Hox* clusters in $4nF_1$ is shown in Fig. 1. The clusters can be summarized as *HoxAai*, *HoxAaii*, *HoxAaiii*, *HoxAbi*, *HoxAbii*, *HoxAbiii*, *HoxBai*, *HoxBaii*, *HoxBaiii*, *HoxBbi*, *HoxBbii*, *HoxBbiii*, *HoxCai*, *HoxCaii*, *HoxCaiii*, *HoxCbi*, *HoxCbii*, *HoxCbiii*, *HoxDai*, *HoxDaii*, and *HoxDaiii* (Table 1). Among these copies, we found that *HoxD4aiiiΨ*, *HoxD9aΨ*, and *HoxD10aΨ* in $4nF_1$ were pseudogenes (Fig. 2). Two deletions at codons 316 and 317 in the coding region of *HoxD4aiiiΨ* suggested that it was a pseudogene. The alignment of the putative *HoxD4a* sequences is shown in Fig. 2a. *HoxD9aΨ* has become a pseudogene because a stop codon prematurely terminates expression of the full-length functional product (Fig. 2b). An insertion was observed at codon 593 in the *HoxD10aΨ* coding region; alignment of the putative *HoxD10a* duplicated sequences is shown in Fig. 2c. *HoxD10aΨ* had an inserted G nucleotide compared with *HoxD10aiii*, whereas a T in *HoxD4aΨ* was replaced by a G compared with *HoxD4ai*. Thus, non-functionalization is a possible fate for some duplicated *Hox* genes. The GC levels of the pseudogenes tended to be lower than that of their counterpart genes (Additional file 1: Table S3). For instance, in $4nF_1$, the exons of the pseudogene *HoxD4aiiiΨ* exhibited a GC content of 50.1%, which was lower than that of its functional counterparts *HoxD4ai* and *HoxD4aii* (51.3%, 52.1%). As shown in Additional file 1: Table S3, the exon GC content of the pseudogene *HoxD10aiΨ* was 49.4%, which was lower than those of its putative functional counterparts *HoxD10aii* and *HoxD10aiii* (49.6% and 49.9%, respectively) in $4nF_1$. Similarly, the exon GC content of the pseudogene *HoxD9aΨ* (43.3%) was slightly lower than that of its putative functional *HoxB1b* paralogues (50.1%, 50.2%, and 50.2%). During duplication, one copy typically remains functional, whereas the other copy may lose its function, which generally leads to a decreased GC level for the non-functional gene.

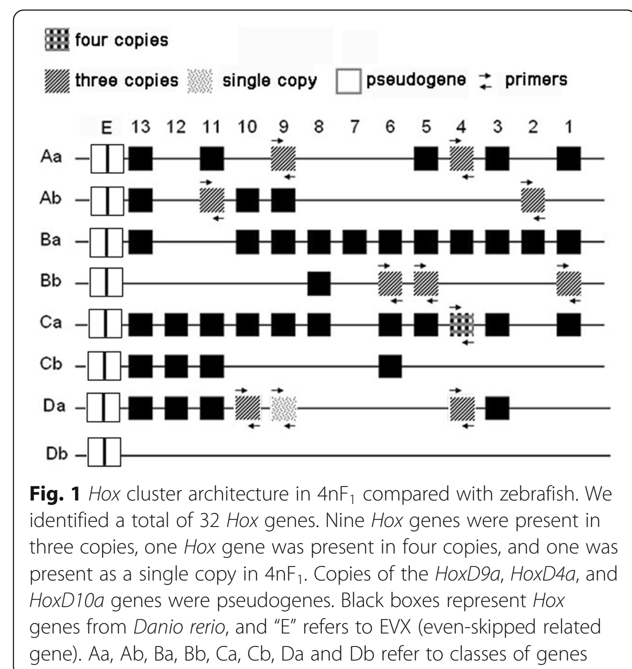

Fig. 1 *Hox* cluster architecture in $4nF_1$ compared with zebrafish. We identified a total of 32 *Hox* genes. Nine *Hox* genes were present in three copies, one *Hox* gene was present in four copies, and one was present as a single copy in $4nF_1$. Copies of the *HoxD9a*, *HoxD4a*, and *HoxD10a* genes were pseudogenes. Black boxes represent *Hox* genes from *Danio rerio*, and "E" refers to EVX (even-skipped related gene). Aa, Ab, Ba, Bb, Ca, Cb, Da and Db refer to classes of genes

Phylogenetic relationships

For most genes, such as *HoxA4a*, *HoxB1b*, and *HoxD10a*, three distinct orthologues of the zebrafish genes were identified in $4nF_1$. These duplicated genes shared a high identity percentage for the deduced amino acid sequences (Additional file 1: Tables S2 and S3). An identity analysis of the putative amino acid sequences suggested that the duplicated sequences were more closely related to each other than to the reported zebrafish orthologues except for the *HoxC4aiii* sequences. For instance, the percentage nucleotide identity between the *HoxA11bi*, *HoxA11bii*, and *HoxA11biii* orthologues from $4nF_1$ and *HoxA11b* from zebrafish was only 89.9%, 89.9%, and 92.4%, respectively. Conversely, the identity between the paralogues *HoxA11bi* and *HoxA11bii*,

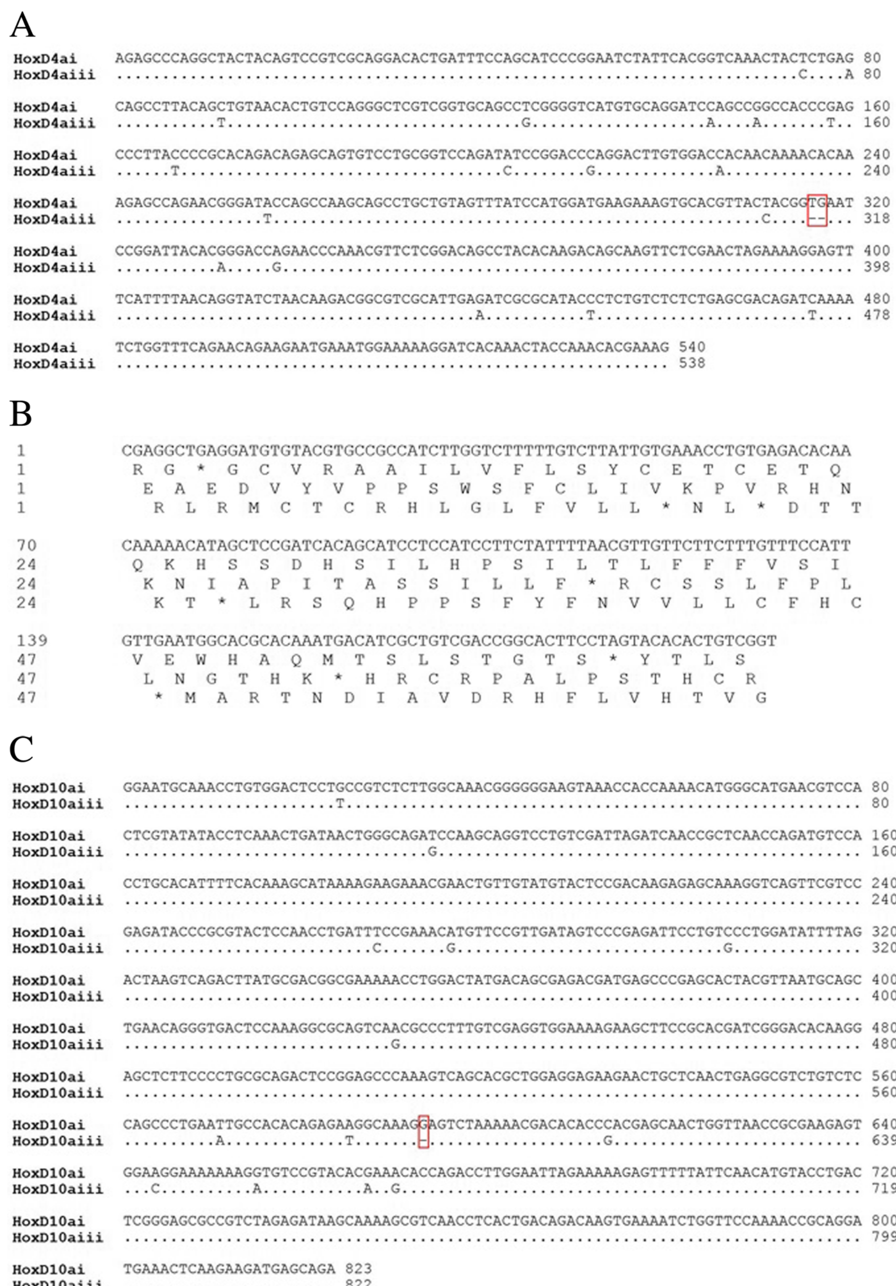

Fig. 2 The pseudogenes *HoxD9a*, *HoxD4aiii*, and *HoxD10ai* in $4nF_1$. **a** Two deletions occurred in codons 316 and 317 in the coding region of *HoxD4aiiiΨ*. **b** *HoxD9aΨ* became a pseudogene because a stop codon prematurely terminated expression of the full-length functional product. **c** One insertion occurred in codon 593 in the coding region of *HoxD10aiΨ*

HoxA11bi and *HoxA11biii*, and *HoxA11bii* and *HoxA11-biii* in $4nF_1$ was 98.6%, 96.4%, and 96.0%, respectively (Additional file 1: Table S2 and Fig. 2a). The identity between *HoxB1bi* and *HoxB1bii*, *HoxB1bi* and *HoxB1biii*, and *HoxB1bii* and *HoxB1biii* was 99.5%, 95.7% and 96.2%, whereas the similarity to their zebrafish orthologues was 91.0%, 90.6% and 91.5% (Additional file 1: Table S2 and Fig. 3b). These results showed that *Hox-A11bi*, *HoxA11bii*, and *HoxA11biii* as well as *HoxB1bi*, *HoxB1bii* and *HoxB1biii* all share a mostly closed ancestral cluster and are true orthologues of the zebrafish genes *HoxA11b* and *HoxB1b*. Analysis of the sequences obtained for *HoxC4a* suggested that four distinct copies of this gene exist in $4nF_1$, which were named *HoxC4ai*, *HoxC4aii*, *HoxC4aiii* and *HoxC4a-1*. The putative amino acid sequence of *HoxC4a-1* shares approximately 100%, 100% and 99% similarity to those of *HoxC4ai*, *HoxC4aii*, and *HoxC4aiii*, respectively. However, the nucleotide

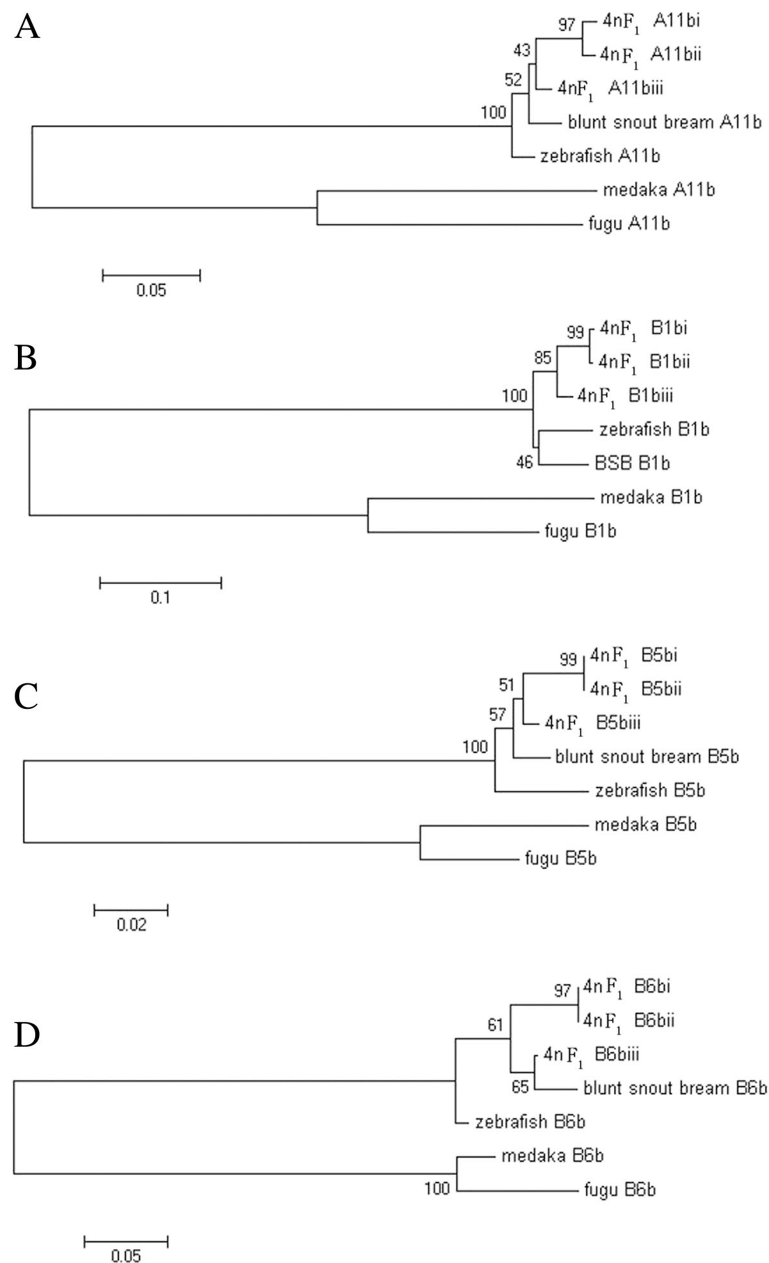

Fig. 3 Maximum likelihood tree inferred from the alignment of amino acid sequences. This tree illustrates the phylogenetic relationships among putative *Hox* sequences in $4nF_1$ and reported orthologues from blunt snout bream (*Megalobrama amblycephala*), zebrafish (*Danio rerio*), fugu (*Fugu rubripes*), and medaka (*Oryzias latipes*). The numbers at the branch nodes indicate percentage bootstrap support for each node based on 1000 replicates. **a**–**d** *HoxA11b*, *HoxB1b*, *HoxB5b*, and *HoxB6b*

similarity to all three sequences is 100%, which suggests the mutation was synonymous.

To evaluate the speciation of $4nF_1$, the nucleotide identity percentages among all known representatives of the *HoxA4a*, *HoxA9a*, *HoxA2b*, and *HoxD4a* gene groups in RCC, BSB, and the $4nF_1$ were examined (Table 2, Fig. 4). The identities of orthologous 'i' or 'ii' genes between $4nF_1$ and RCC were much higher than those between $4nF_1$ and BSB. For example, the nucleotide identity percentages of the orthologous *HoxA4ai*, *HoxA9ai*, *HoxA2bi*, and *HoxD4ai* genes between $4nF_1$ and RCC were 99.5%, 99.4%, 99.6% and 99.6%, respectively. Conversely, the similarity of these genes between $4nF_1$ and BSB was 97.0%, 92.3.0%, 97.2%, and 93.7%, respectively. Although similarly high identity was observed, the 'iii' gene in $4nF_1$ did not exhibit higher similarity to the gene in RCC or BSB for all four *Hox* sequence groups, suggesting no

Table 2 Percentage nucleotide identity (on the left) and percentage amino acid identity (on the right) between duplicated *Hox* coding regions in $4nF_1$, RCC, and BSB

	HoxA4a (%)	*HoxA9a* (%)	*HoxA2b* (%)	*HoxD4a* (%)
$4nF_1$ i:$4nF_1$ ii	97.4/97.5	91.8/89.5	96.9/96.9	98.3/97.7
:$4nF_1$ iii	98.5/98.0	98.9/98.9	97.5/97.6	96.2/61.6
:RCC i	99.5/99.5	99.4/99.4	99.6/100.0	99.6/100.0
:RCC ii	97.2/96.6	92.3/90.1	96.5/96.6	98.3/97.7
:BSB	97.0/96.6	92.3/91.6	97.2/97.9	93.7/94.9
$4nF_1$ ii:$4nF_1$ iii	98.3/98.5	91.6/88.5	96.9/95.9	95.3/60.5
:RCC i	97.5/98.0	92.0/89.0	96.8/96.9	98.3/97.7
:RCC ii	98.8/98.5	94.9/92.1	98.3/97.6	100.0/100.0
:BSB	96.7/97.0	91.0/86.4	96.4/96.3	93.3/92.7
$4nF_1$ iii:RCC i	98.7/98.5	99.1/98.4	97.4/97.6	95.9/61.6
:RCC ii	98.2/98.5	92.2/89.0	96.5/95.9	95.3/60.5
:BSB	98.0/98.5	92.2/90.6	97.7/97.3	94.0/59.4
RCC i:RCCii	97.4/97.0	92.5/89.5	96.6/96.6	98.3/97.7
:BSB	97.2/97.0	92.5/91.1	97.1/97.9	94.0/94.9
RCC ii:BSB	96.9/97.0	91.5/88.0	96.2/96.6	93.3/92.7

Values before slashes (/) denote nucleotide identity, values after slashes denote amino acid identity

obvious orthologous relationship between the two species. Thus, we speculated that the 'iii' genes were variants of RCC or BSB genes. For example, the *HoxA4aiii*, *HoxA9aiii*, *HoxA2biii*, and *HoxD4aiii* genes from $4nF_1$ and the *HoxA4a*, *HoxA9a*, *HoxA2b*, and *HoxD4a* genes from BSB shared 98.0%, 92.2%, 97.7%, and 94.0% identity (Table 2).

Discussion

The structure of cloned *Hox* gene sequences

Prior PCR surveys and genomic library screening have identified interesting variability in *Hox* gene content among teleosts [12, 15, 16, 20, 21]. Luo et al. [22] estimated there were 14–16 *Hox* gene clusters in goldfish. Our data suggested 18–21 *Hox* gene clusters were present in $4nF_1$, with each was located on a different acrocentric chromosome. The *Hox* gene clusters in $4nF_1$ were approximately the sum of the clusters in RCC and BSB, except that some clusters were lost. The topology of the *Hox* gene maximum likelihood tree (Fig. 3) further suggested that some of the *Hox* genes orthologous to zebrafish genes were present in two copies in RCC, one copy in BSB, and three copies in $4nF_1$. However, the third copy did not exhibit notably higher similarity to the gene in RCC or BSB. We speculated that variation and reorganization of the genome likely occurred during polyploidization, resulting in new copies in $4nF_1$. This might be evidence that allopolyploidization induces a variety of rapid genomic changes in a $4nF_1$ population [23, 24]. Using sequence alignment in $4nF_1$, we isolated 32 fragments that can be characterized as *HoxA*, *HoxB*, *HoxC*, and *HoxD* family genes. However, amplified RCC and BSB DNA were only characterized as the *HoxA* and *HoxD* genes. We speculated that the increase in the number of $4nF_1$ genes might be related to polyploidization. This situation was also observed in our previous study [25, 26]; the number of $4nF_1$ fragments increased, and some genes from RCC and BSB were lost. At present, although we have no precise data explaining this outcome, we speculate that allotetraploidization might lead to rapid changes in $4nF_1$ genome diversity. Our study is the first to evaluate the organization of *Hox* clusters in a $4nF_1$ population. This theory is also strongly supported by other studies examining *Hox* genes [22], other gene families [27], and DNA content [28].

The significance of polyploidization

Polyploidization likely increases genomic variation rates, which can result in the formation of new polyploid species [29]. First, the process of polyploidization can itself generate species that are reproductively isolated from their diploid progenitors, increasing the number of species as a by product. For example, several studies have indicated that a polyploidization event occurred in an ancestor of teleost fish shortly after this lineage diverged from the lineage leading to tetrapods [30–32]. Second, an entirely different trait can result in increased rates of polyploidization [6]. Synonymous mutations increase genomic variation. For example, the putative amino acid sequence of *HoxC4a-1* shares approximately 100%, 100%, and 99% similarity with those of *HoxC4ai*, *HoxC4aii*, and *HoxC4aiii*, respectively. The identity of their nucleotide sequences is 100%. In the polyploidization process, genome duplication produces abundant genomic DNA, so the organism maintains the dosage balance or rapidly stabilizes the duplicated genomes via retention/exclusion of redundancy. Lynch et al. [33] suggested there are three outcomes in the evolution of duplicate genes: non-functionalization, neo-functionalization and subfunctionalization. Interestingly, we found some pseudogenes in $4nF_1$, such as *HoxD4aiiiΨ*, *HoxD9aΨ* and *HoxD10aΨ*. Pseudogenes are formed either by random mutations that create stop codons and prematurely terminate full-length functional product expression or by insertions/deletions that shift the reading frame, rendering the translated protein non-functional. We speculate that dosage effects generated selection pressure from the loss of *Hox* genes or the formation of pseudogenes after whole genome duplication. This pressure is consistent with the expectation that there are *Hox* clusters in the $4nF_1$ genome that have lost functional *Hox* genes due to the reduction of redundancy following the polyploidization event. However, $4nF_1$ required genetic recombination, mutation, and pseudogenization to reduce the amount of incompatible genetic material and improve fertility [34]. Thus, we unexpectedly

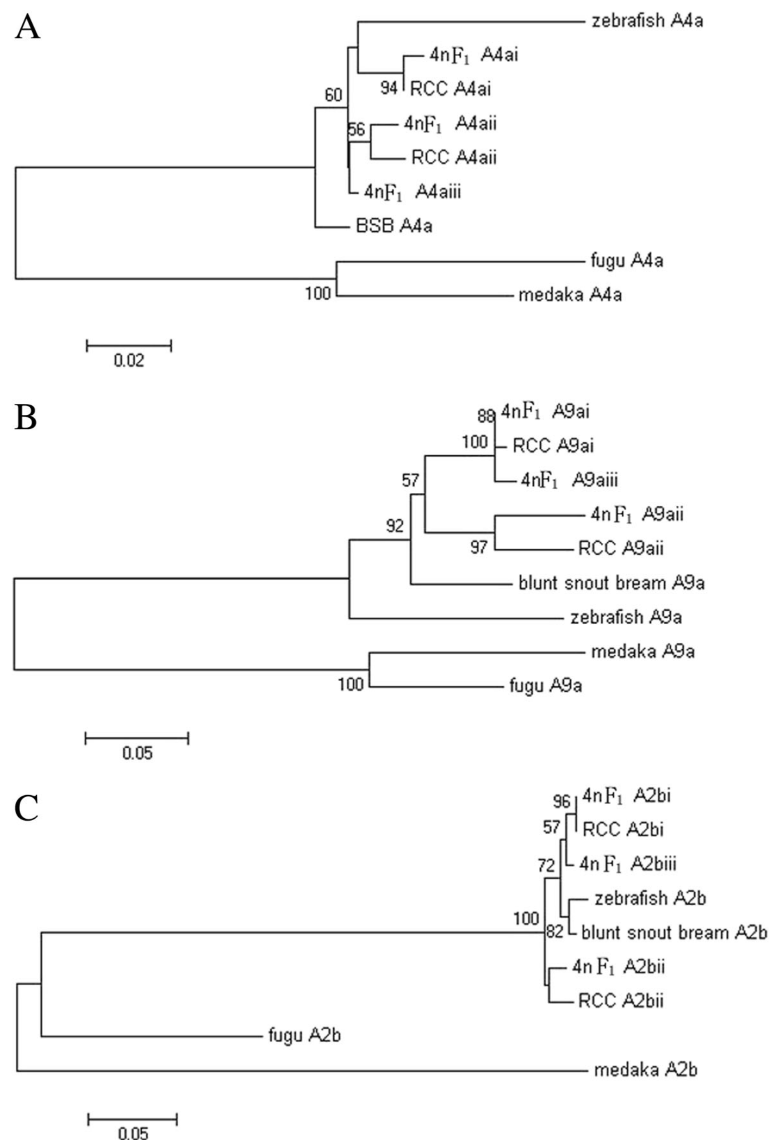

Fig. 4 Maximum likelihood tree inferred from the alignment of amino acid sequences. This tree illustrates the phylogenetic relationships among putative *Hox* sequences in 4nF$_1$, BSB, and RCC and the reported orthologues from blunt snout bream (*Megalobrama amblycephala*), zebrafish (*Danio rerio*), and medaka (*Oryzias latipes*). The number at each branch node indicates the percentage bootstrap support for that node based on 1000 replicates. **a**–**c** *HoxA4a*, *HoxA9a* and *HoxA2b*.

obtained autotetraploids with greater fertility among the 4nF$_1$ progeny, and we successfully established an autotetraploid fish line [35]. Our characterization of the *Hox* gene clusters in tetraploid hybrids improves our understanding of the evolutionary processes occurring after *Hox* gene duplication in vertebrates.

Conclusions

We identified three copies of *Hox* genes in 4nF$_1$, two copies in RCC and one copy in BSB. In addition, obvious variation and pseudogene generation were observed in some 4nF$_1$ *Hox* genes. These results reveal the effects of polyploidization on the organization and evolution of *Hox* gene clusters in fish and also help to clarify aspects of vertebrate genome evolution.

Methods

DNA extraction

Specimens of 4nF$_1$ (4n = 148), RCC (2n = 100), and BSB (2n = 48) were obtained from the Engineering Research Center of Polyploid Fish Breeding and Reproduction of the State Education Ministry at Hunan Normal University.

Fish treatments were carried out according to the regulations for protected wildlife and the Administration of Affairs Concerning Animal Experimentation, and approved by the Science and Technology Bureau of China. Approval from the Department of Wildlife Administration was not required for the experiments conducted in this paper. The fish were deeply anesthetized with 100 mg/L MS-222 (Sigma-Aldrich, St Louis, MO, USA) before dissection. Narcotic drugs was fed before blood sampling. Total genomic DNA was isolated from peripheral blood cells using the standard phenol chloroform extraction procedures described by Sambrook et al. [36].

Cloning and sequencing of *Hox* genes

We amplified fragments of *Hox* genes from genomic DNA by PCR amplification using several combinations of degenerate primers (Table 1). PCR was performed in 50-μL reaction volumes using Taq DNA polymerase (TaKaRa, Dalian, China). A typical PCR programme consisted of a denaturation step at 94°C for 5 min; 35 cycles of 98°C for 15 s, 55°C for 45 s and 72°C for 1 min; and a final elongation step at 72°C for 10 min. PCR products were cloned into a T vector and sequenced with an automated ABI 3700 DNA sequencer. The sequences were BLAST searched against the non-redundant protein database maintained at the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov) to determine their identity.

Sequence comparison and analysis

Sequence homology and variation among the fragments amplified from RCC, BSB and the $4nF_1$ were analysed in BioEdit [37, 38]. Partial DNA sequences for each gene were verified using a BLASTx search. To increase the probability of detecting duplicated paralogues and circumventing PCR errors, we sequenced 20 clones for each gene from $4nF_1$, RCC and BSB. The obtained sequences were screened for *Hox* gene fragments using different BLAST searches (BLASTn, BLASTp, and BLASTx) against GenBank (http://www.ncbi.nlm.gov/Blast.cgi). Then, we evaluated the organization of the $4nF_1$ *Hox* clusters compared to RCC and BSB to characterize the *Hox* genes.

Phylogenetic analysis

Using Clustal X 1.81, the derived amino acid sequences of these fragments were aligned with the *Hox* genes from BSB, zebrafish, fugu, medaka and other teleosts retrieved from GenBank [38]. Regions of sequences that were difficult to align were removed from the alignment. Gaps were also removed from the alignment. The maximum likelihood method implemented in the online software RAxML was used to construct a phylogenetic tree [39].

Additional file

Additional file 1: Table S1. The PCR Gene-specific degenerate primers. Gene-specific degenerate primers designed based on the alignment and identification of consensus orthologous *Hox* gene sequences from zebrafish (*Danio rerio*), medaka (*Oryzias latipes*), pufferfish (*Fugu rubripes*), mouse (*Mus musculus*), cichlids, and humans (*Homo sapiens*). **Table S2.** The Percentage of the amino acid. Percentage amino acid identity between paralogous *Hox* sequences obtained from $4nF_1$ and reported orthologues from zebrafish, fugu, and medaka. **Table S3.** Comparison of GC levels among duplicated genes. (DOCX 20 kb)

Abbreviations

$4nF_1$: Allotetraploid F_1 hybrids; BSB: Blunt snout bream; RCC: Red crucian carp

Acknowledgements

We would like to sincerely thank many researchers who help to complete this manuscript, including Drs. Yao Zhanzhou and Zhao Rurong.

Funding

This research was supported by National Natural Science Foundation of China Grants 31210103918, 31430088 and 31730098, the earmarked fund for China Agriculture Research System (CARS-45), the Natural Science Foundation of Hunan Province for Distinguished Young Scholars (Grants No. 2017JJ1022), the Major Program of Educational Commission of Hunan Province (Grant No. 17A133), the Cooperative Innovation Center of Engineering and New Products for Developmental Biology of Hunan Province (20134486), the Construction Project of Key Discipline of Hunan Province and China, Hunan Normal University Graduate Research Innovation Project (CX2017B184).

Availability of data and materials

The dataset(s) supporting the conclusions of this article is (are) available in the GenBank repository with access No. JX282274- JX282316 and hyperlink to dataset(s) in https://www.ncbi.nlm.nih.gov/nuccore/JX282274-JX282316. The dataset(s) supporting the conclusions of this article is (are) included within the article (and its additional file(s)).

Authors' contributions

YW, QQ and SL designed the study, performed the analyses, performed the technical discussions, and prepared and drafted the manuscript. WS, QL, XH and YH participated in data simulation and discussions. RY, TL, CZ and MT were involved in the statistical analysis. All authors have read and approved the final manuscript.

Ethics approval

All experiments were approved by the Animal Care Committee of Hunan Normal University and followed the guidelines statement of the Administration of Affairs Concerning Animal Experimentation of China. All samples were raised in natural ponds, all dissections were performed under sodium pentobarbital anaesthesia, and all efforts were made to minimize suffering.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 20 May 2017 Accepted: 13 September 2017
Published online: 11 October 2017

References

1. Ohno S, Wolf U, Atkin NB. Evolution from fish to mammals by gene duplication. *Hereditas*. 1968;59(1):169–87.

2. Masterson J. Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. *Science*. 1994;264(5157):421–4.
3. Wolfe KH, Shields DC. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature*. 1997;387(6634):708–13.
4. Comai L. Genetic and epigenetic interactions in allopolyploid plants. *Plant molecular biology*. 2000;43(2-3):387–99.
5. Liu S, Liu Y, Zhou G, Zhang X, Luo C, Feng H, He X, Zhu G, Yang H. The formation of tetraploid stocks of red crucian carp× common carp hybrids as an effect of interspecific hybridization. *Aquaculture*. 2001;192(2):171–86.
6. Otto SP. The evolutionary consequences of polyploidy. *Cell*. 2007;131(3):452–62.
7. Mallet J. Hybrid speciation. *Nature*. 2007;446(7133):279–83.
8. Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJ, Bierne N, Boughman J, Brelsford A, Buerkle CA, Buggs R, et al. Hybridization and speciation. *Journal of evolutionary biology*. 2013;26(2):229–46.
9. Liu S, Qin Q, Xiao J, Lu W, Shen J, Li W, Liu J, Duan W, Zhang C, Tao M, et al. The formation of the polyploid hybrids from different subfamily fish crossing and its evolutionary significance. *Genetics*. 2007;176
10. Qin Q, Wang Y, Wang J, Dai J, Liu Y, Liu S. Abnormal chromosome behavior during meiosis in the allotetraploid of Carassius auratus red var. (female symbol)xMegalobrama amblycephala (male symbol). *BMC genetics*. 2014;15:95.
11. Yu X-J, Zhou T, Li Y-C, Li K, Zhou M. Chromosomes of Chinese fresh-water fishes. In: *Beijing (in Chinese): Science Publishing House*; 1989.
12. Santini S, Bernardi G. Organization and base composition of tilapia Hox genes: implications for the evolution of Hox clusters in fish. *Gene*. 2005;346: 51–61.
13. Koh EG, Lam K, Christoffels A, Erdmann MV, Brenner S, Venkatesh B. Hox gene clusters in the Indonesian coelacanth, Latimeria menadoensis. *Proceedings of the National Academy of Sciences of the United States of America*. 2003;100(3):1084–8.
14. Mallo M, Alonso CR. The regulation of Hox gene expression during animal development. *Development*. 2013;140(19):3951–63.
15. Amores A, Force A, Yan YL, Joly L, Amemiya C, Fritz A, Ho RK, Langeland J, Prince V, Wang YL, et al. Zebrafish hox clusters and vertebrate genome evolution. *Science*. 1998;282(5394):1711–4.
16. Amores A, Suzuki T, Yan YL, Pomeroy J, Singer A, Amemiya C, Postlethwait JH. Developmental roles of pufferfish Hox clusters and genome evolution in ray-fin fish. *Genome Res*. 2004;14(1):1–10.
17. Chiu CH, Amemiya C, Dewar K, Kim CB, Ruddle FH, Wagner GP. Molecular evolution of the HoxA cluster in the three major gnathostome lineages. *Proceedings of the National Academy of Sciences of the United States of America*. 2002;99(8):5492–7.
18. Chiu CH, Dewar K, Wagner GP, Takahashi K, Ruddle F, Ledje C, Bartsch P, Scemama JL, Stellwag E, Fried C, et al. Bichir HoxA cluster sequence reveals surprising trends in ray-finned fish genomic evolution. *Genome Res*. 2004; 14(1):11–7.
19. Lyu J. Crop evolution: After allopolyploidization. *Nature plants*. 2016;2(10): 16156.
20. Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehal P, Christoffels A, Rash S, Hoon S, Smit A, et al. Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. *Science*. 2002;297(5585):1301–10.
21. Naruse K, Tanaka M, Mita K, Shima A, Postlethwait J, Mitani H. A medaka gene map: the trace of ancestral vertebrate proto-chromosomes revealed by comparative gene mapping. *Genome Res*. 2004;14(5):820–8.
22. Luo J, Stadler PF, He S, Meyer A. PCR survey of hox genes in the goldfish Carassius auratus auratus. *Journal of experimental zoology Part B, Molecular and developmental evolution*. 2007;308(3):250–8.
23. Qin Q, Lai Z, Cao L, Xiao Q, Wang Y, Liu S. Rapid genomic changes in allopolyploids of Carassius auratus red var. (female symbol) x Megalobrama amblycephala (male symbol). *Scientific reports*. 2016;6:34417.
24. Stephens SG. Possible Significance of Duplication in Evolution. *Advances in Genetics*. 1951;4:247–65.
25. Qin Q, He W, Liu S, Wang J, Xiao J, Liu Y. Analysis of 5S rDNA organization and variation in polyploid hybrids from crosses of different fish subfamilies. *J Exp Zool B Mol Dev Evol*. 2010;314
26. Chen L, Li W, Liu S, Tao M, Long Y, Duan W, Zhang C, Xiao J, Qin Q, Luo K, et al. Novel genetic markers derived from the DNA fragments of Sox genes. *Molecular and cellular probes*. 2009;23(3-4):157–65.
27. Malaga-Trillo E, Laessing U, Lang DM, Meyer A, Stuermer CA. Evolution of duplicated reggie genes in zebrafish and goldfish. *Journal of molecular evolution*. 2002;54(2):235–45.
28. Ohno S, Muramoto J, Christian L, Atkin NB. Diploid-tetraploid relationship among old-world members of the fish family Cyprinidae. *Chromosoma*. 1967;23(1):1–9.
29. Comai L. The advantages and disadvantages of being polyploid. *Nat Rev Genet*. 2005;6(11):836–46.
30. Meyer A, Schartl M. Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Current opinion in cell biology*. 1999;11(6):699–704.
31. Panopoulou G, Hennig S, Groth D, Krause A, Poustka AJ, Herwig R, Vingron M, Lehrach H. New evidence for genome-wide duplications at the origin of vertebrates using an amphioxus gene set and completed animal genomes. *Genome Res*. 2003;13(6A):1056–66.
32. Seoighe C, Wolfe KH. Extent of genomic rearrangement after genome duplication in yeast. *Proceedings of the National Academy of Sciences of the United States of America*. 1998;95(8):4447–52.
33. Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. *Science*. 2000;290(5494):1151–5.
34. Qin Q, He W, Liu S, Wang J, Xiao J, Liu Y. Analysis of 5S rDNA organization and variation in polyploid hybrids from crosses of different fish subfamilies. *Journal of experimental zoology Part B, Molecular and developmental evolution*. 2010;314(5):403–11.
35. Qin Q, Wang Y, Wang J, Dai J, Xiao J, Hu F, Luo K, Tao M, Zhang C, Liu Y, et al. The autotetraploid fish derived from hybridization of Carassius auratus red var. (female) x Megalobrama amblycephala (male). *Biology of reproduction*. 2014;91(4):93.
36. J S, EF F, T M. Molecular cloning: a laboratory manual. 2nd ed; 1989.
37. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In: *Nucl Acids Symp Ser. 1999*:95–8.
38. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. *Nucleic acids research*. 1997;25(24):4876–82.
39. Stamatakis A, Hoover P Fau – Rougemont J, Rougemont J: A rapid bootstrap algorithm for the RAxML Web servers. (1076-836X (Electronic)).

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit