


RESEARCH ARTICLE

Open Access



# *Hox* genes reveal variations in the genomic DNA of allotetraploid hybrids derived from *Carassius auratus* red var. (female) × *Cyprinus carpio* L. (male)

Rurong Zhao<sup>1,2†</sup>, Yude Wang<sup>1,2†</sup>, Li Zou<sup>3†</sup>, Yaxin Luo<sup>1,2</sup>, Huifang Tan<sup>1,2</sup>, Jiajun Yao<sup>1,2</sup>, Minghe Zhang<sup>1,2</sup> and Shaojun Liu<sup>1,2\*</sup> 

## Abstract

**Background:** *Hox* transcription factors are master regulators of animal development. Although highly conserved, they can contribute to the formation of novel biological characteristics when modified, such as during the generation of hybrid species, thus potentially serving as species-specific molecular markers. Here, we systematically studied the evolution of genomic sequences of *Hox* loci in an artificial allotetraploid lineage (4nAT, 4n = 200) derived from a red crucian carp (♀, RCC, 2n = 100) × common carp (♂, CC, 2n = 100) cross and its parents (RCC and CC).

**Results:** PCR amplification yielded 23 distinct *Hox* gene fragments from 160 clones in 4nAT, 22 fragments from 90 clones in RCC, and 19 fragments from 90 clones in CC. Sequence alignment of the *HoxA3a* and *HoxC10a* genes indicated both the inheritance and loss of paternal genomic DNA in 4nAT. The *HoxA5a* gene from 4nAT consisted of two subtypes from RCC and two subtypes from CC, indicating that homologous recombination occurred in the 4nAT hybrid genome. Moreover, 4nAT carried genomic pseudogenization in the *HoxA10b* and *HoxC13a* loci. Interestingly, a new type of *HoxC9a* gene was found in 4nAT as a hybrid sequence of CC and RCC by recombination in the intronic region.

**Conclusion:** The results revealed the influence of *Hox* genes during polyploidization in hybrid fish. The data provided insight into the evolution of vertebrate genomes and might be benefit for artificial breeding programs.

**Keywords:** PCR survey, *Hox* gene, Allotetraploid hybrid, Evolution

## Background

Genome duplication, and therefore polyploidy, can occur during animal evolution [1, 2]. Two types of polyploidy have been identified based on genetic origin: autopolyploidy, which originates from the genome duplication of a single species, and allopolyploidy, which occurs via inter-specific hybridization and results in haploid chromosomes inherited from different species. Allopolyploidy is prevalent in nature, suggesting the evolutionary advantage of

maintaining genetic heterogeneity [3]. Allopolyploidy plays an important role in facilitating the activation of cryptic mobile elements and rapid genomic change [4]. This 'genomic shock' has been reported in many allopolyploid plants and occurs through gene loss, chromosome mispairing, retrotransposon activation, altered methylation, or rearrangements between parental genomes. These can lead to novel gene sequences or differential homologous gene expression in hybrids [5]. After a short period of laboratory experiments, chromosomal rearrangements, duplications and deletions of chromosome segments, and shifts in ploidy have been observed and these appear to be adaptive [6]. Changes in genome structure usually have a direct impact on the phenotype and can contribute to individual adaptation [7]. Rapid genomic DNA changes have been

\* Correspondence: [lsj@hunnu.edu.cn](mailto:lsj@hunnu.edu.cn)

<sup>†</sup>Rurong Zhao, Yude Wang and Li Zou contributed equally to this work.  
<sup>1</sup>State Key Laboratory of Developmental Biology of Freshwater Fish, Hunan Normal University, Changsha 410081, Hunan, People's Republic of China  
<sup>2</sup>College of Life Sciences, Hunan Normal University, Changsha 410081, Hunan, People's Republic of China

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

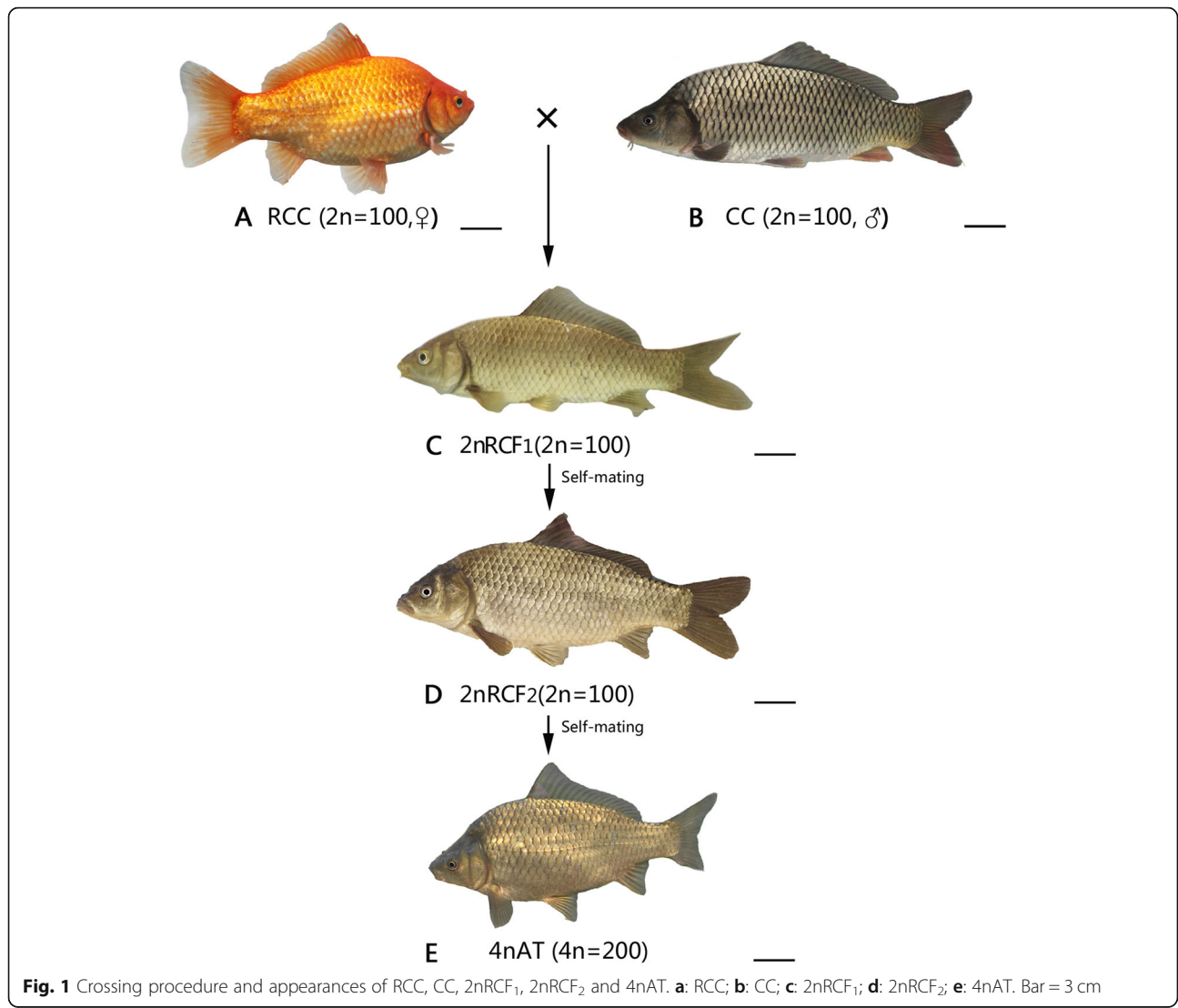


described in several allopolyploid plants [8]. Compared to plants, animals rarely undergo allopolyploidy. Therefore, the genome variations during the animal allopolyploidy process are unclear [9].

In a previous study, we obtained fertile allotetraploid hybrids from the interspecific hybridization of red crucian carp (RCC; *Carassius auratus* red var., ♀, 2n = 100) with common carp (CC; *Cyprinus carpio* L., ♂, 2n = 100) [10]. Although the F<sub>1</sub> and F<sub>2</sub> progenies were diploid hybrids (2n = 100), fertile allotetraploid offspring of both sexes were produced from the F<sub>3</sub> and later generations (Fig. 1) [11]. These allotetraploid offspring were used to generate rapidly growing and strongly resistant infertile triploids by hybridization with diploid fishes [12]. Hybrid polyploid fish have wide applications in the Chinese fish industry and they also provide an opportunity to study the molecular and genetic mechanisms that underlie the origination of evolutionary novelties including genome evolution and

adaptation. Previous studies confirmed the rapid genomic DNA changes in 4nAT [13], and in this study we investigated the genetic elements susceptible to rapid genomic changes.

*Hox* genes belong to a large family of transcription factors. They are organized in clusters within the genome and are characterized by a homeobox, which encodes a DNA-binding motif known as the homeodomain [14, 15]. *Hox* genes and *Hox* clusters provide evidence that rounds of genome duplications (the one-to-four (-to-eight-to-more in fish) rule) occur in vertebrates [16]. Recent studies have demonstrated that *Hox* gene clusters are fragmented, reduced, or expanded in many animals. These findings correlate with morphological changes occurring during evolution [17]. The *HoxC9* is a regulator of body patterning and development. Its expression has been described in the hind limb blood vessels of mice [18] and the axial vasculature of zebrafish [19]. *HoxC9* may also be involved in the



**Fig. 1** Crossing procedure and appearances of RCC, CC, 2nRCF<sub>1</sub>, 2nRCF<sub>2</sub> and 4nAT. a: RCC; b: CC; c: 2nRCF<sub>1</sub>; d: 2nRCF<sub>2</sub>; e: 4nAT. Bar = 3 cm

differentiation of bone marrow derived stem cells to endothelial cells because it was upregulated during stem cell differentiation [20]. *Hox* genes are clustered in the common ancestors of chordates, arthropods and nematodes and play an important role in embryogenesis by encoding transcription factors [21]. Therefore, *Hox* genes and *Hox* gene clusters may provide information about the process of evolution at the molecular level [22].

In this study, we compared the *Hox* genes of artificially derived 4nAT lineages with those of the original RCC and CC parents to obtain information on the genomic evolution of allopolyploid animals. Exploration of 4nAT and its related molecular genetic relationship with the parental lines can provide insight into the evolution of

selected species. Specific genome replication provides an example for the evolution of fish.

**Results**

**Sequence information for RCC, CC, and 4nAT**

Using two pairs of degenerate primers, we obtained 90 sequences of PCR clones from RCC, 90 sequences from CC, and 160 sequences from 4nAT. To assign names to the *Hox* genes, all acquired sequences were screened for *Hox* gene fragments using BLAST searches against NCBI and were aligned using Clustal X (2.0) software for verification. According to BLAST searches and sequence alignment analysis, all but 17 sequences could be

**Table 1** Summary of the PCR fragments from the 4nAT and its original parents

PG	Gene name	The total number of <i>Hox</i> gene fragments		
		RCC	CC	4nAT
1	<i>HoxB1b</i>	2	1	2
2	<i>HoxA2b</i>		1	
3	<i>HoxA3a</i>	13	10	16
	<i>HoxB3a</i>	7	9	8
4	<i>HoxA4a</i>	1		
5	<i>HoxA5a</i>	11	15	35
	<i>HoxC5a</i>	3	1	4
6	<i>HoxC6a</i>	1		
7	<i>HoxB7a</i>	4	2	5
8	<i>HoxB8a</i>		3	3
9	<i>HoxA9a</i>	1		2
	<i>HoxA9b</i>	3		
	<i>HoxB9a</i>	3	2	5
	<i>HoxC9a</i>	5	5	13
10	<i>HoxA10b</i>	7	2	5
	<i>HoxB10a</i>	1		3
	<i>HoxC10a</i>	13	14	22
	<i>HoxD10a</i>			1
11	<i>HoxA11a</i>	3	2	3
	<i>HoxA11b</i>			1
	<i>HoxC11a</i>			1
	<i>HoxD11a</i>			1
12	<i>HoxC12a</i>	2		
	<i>HoxD12a</i>		1	1
13	<i>HoxA13a</i>	1	3	1
	<i>HoxA13b</i>	1	1	3
	<i>HoxB13a</i>	1	1	
	<i>HoxC13a</i>	4	5	12
	<i>HoxD13a</i>	1	5	5
Total		88	83	152

assigned unambiguously to 29 distinct *Hox* genes. A summary of the analyses and proposed gene identification was presented in Table 1. The combined data of the PCR products yielded 22 distinct *Hox* gene fragments in RCC, 19 in CC, and 23 in 4nAT. Fifteen *Hox* gene fragments were found both in 4nAT and its original parents. Due to amplification bias during PCR amplification, we acquired more clones of *HoxA3a*, *HoxA5a*, and *HoxC10a* gene fragments than from other *Hox* genes. These *Hox* sequences were analyzed further. There were four isoforms in 4nAT (4nAT-I, 4nAT-II, 4nAT-III, and 4nAT-IV), two isoforms in RCC (RCC-I and RCC-II) and two isoforms in CC (CC-I and CC-II). Among the four isoforms of the *HoxA5a* gene, two were consistent with the *HoxA5a* gene in CC and the other two were consistent with the *HoxA5a* genes in RCC (Fig. 2a). The similarities between 4nAT-I and CC-I, 4nAT-II and RCC-I, 4nAT-III and CC-II, and 4nAT-IV and RCC-II were all 100%. We detected two isoforms of *HoxA3a* and *HoxC10a* in the genome of CC and RCC. However, we observed only three isoforms of *HoxA3a* (Fig. 2b) and *HoxC10a* (Fig. 2c) in 4nAT. For *HoxA3a*, the identities between 4nAT-I and CC-I, 4nAT-I and RCC-I, 4nAT-II and CC-II, and 4nAT-II and RCC-II were all 100%. For *HoxC10a*, the identities between 4nAT-I and CC-I, 4nAT-I and RCC-I, 4nAT-II and CC-II and 4nAT-III and RCC-II were all 100%. Almost all orthologous sequences in both RCC and CC showed a higher percentage of similarity to each other than to their duplicated homeologous genes within species.

DNA fragments (~ 1800 bp) were amplified from RCC, CC, and 4nAT using the primer pair *HC9aF-HC9aR*. The full-length *HoxC9a* gene in RCC, CC, and 4nAT was 1837 bp (GenBank no: MN584925), 1834 bp (GenBank no: MN584926) and 1835 bp (GenBank no: MN584927), respectively. The homeodomain sequence of the *HoxC9a* gene was identical to the homeobox fragment *HoxC9a* isolated in the PCR survey. Interestingly, in 4nAT, a new fragment of the *HoxC9a* gene was found, whose former part of nucleotide sequence was identical to that of the CC and the latter part was identical to that of RCC; the position of the exchange recombination occurred in the intron region (Fig. 3).

#### Hybrid identification

To determine if 4nAT was hybridized from RCC and CC, we examined the genomic composition of RCC, CC, and 4nAT using STRUCTURE. The genetic composition estimation was consistent with our sample collection. The two parental lines (RCC and CC) corresponded to the two clusters, and they were assigned exclusively to one of the two clusters (Fig. 4). All 4nAT individuals showed mixed ancestries (Fig. 4). This suggested that the 4nAT individuals genetically originated from RCC and CC.

#### Phylogenetic relationships

The identified *Hox* genes belonged to the paralogous groups 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 and 13; however, we did not detect paralogous groups 4 and 6 in CC, paralogous groups 2 and 8 in RCC, and paralogous groups 2, 4 and 6 in 4nAT. The phylogenetic analysis of the protein-coding nucleotide sequences assigned the obtained fragments to the expected orthologous genes among the three species (Fig. 4). Forty-two 4nAT, 29 RCC and 24 CC different subtypes of *Hox* gene fragments were analyzed using the HKY + G + I model in MEGA 5.0 software to estimate the genetic divergence between the species. The phylogenetic tree of the genes was constructed by the ML method, according to genetic distance, and the confidence level (1000 bootstrap replicates) was tested with 1000 replicate samplings (Fig. 5). In this phylogenetic tree, essentially all of the 4nAT *Hox* genes were first aggregated with their related parents, RCC or CC, and then combined with other subtypes of this gene in 4nAT (Fig. 5). The phylogenetic tree showed that the genetic material of 4nAT originated from RCC and CC.

#### Analysis of pseudogenes in 4nAT and the related parents

A pseudogene was a sequence that existed in the genome of a specific population, and was typically characterized by being very similar to one or more paralogous genes, yet was nonfunctional. This loss of function was due to a failure in transcription or translation, or production of a protein that did not have the same function as the protein encoded by a normal paralog gene. Among the 152 *Hox* gene sequences in 4nAT, we found that the *HoxA10b* and *HoxC13a2* were pseudogenes. Two bases were missing in *HoxA10b* and the length of the *HoxC13a* gene subtype (*HoxC13a2*) with two inserted bases was 161 bp. Among the 88 *Hox* gene sequences of RCC, the *HoxA11a2* was pseudogene. A base-deletion was found in the subtype (*HoxA11a2*) of *HoxA11a* gene and the length of the *HoxA11a2* with base deletion was 158 bp. Using Jellyfish software (Version: V1.4\_4383) to translate all the conserved sequence of these *HoxA10b* genes into amino acids, the translation termination appeared in the coded region, in every case. The GC percentage (45.22%) of the pseudogene *HoxA10b* was significantly lower than AT percentage (54.78%). Compared with the amino acid sequence of the *HoxC13a* gene (*HoxC13a1*) and *HoxA11a* gene (*HoxA11a1*), the amino acid sequence of *HoxC13a2* with base insertion and *HoxA11a2* with base deletion had changed greatly after these frameshift mutations, respectively (Fig. 6). The amino acid length of the pseudogenes (*HoxA11a2*: 51; *HoxC13a2*: 52) tended to be lower than that of their counterpart genes (*HoxA11a1*: 53; *HoxC13a1*: 53).



(See figure on previous page.)

**Fig. 2** Comparison of the *HoxA5a*, *HoxA3a*, *HoxC10a* sequences from 4nAT, RCC and CC. CC-I and CC-II were two isoforms of *HoxA5a* from CC, RCC-I and RCC-II were two isoforms of *HoxA5a* from RCC, and 4nAT-I, 4nAT-II, 4nAT-III and 4nAT-IV were four isoforms of *HoxA5a* found in 4nAT. Dots indicated identical nucleotides at a given site (a). Comparisons of the *HoxA3a* (b) and *HoxC10a* (c) sequences from 4nAT, RCC and CC were found in B and C, respectively. CC-I and CC-II were two isoforms from CC, RCC-I and RCC-II were from RCC, and 4nAT-I, 4nAT-II and 4nAT-III were found in 4nAT. Dots indicated identical nucleotides at a given site

### Molecular organization of the *HoxC9a* gene sequence

We used the sequence from 4nAT to compare the obtained DNA sequences with the *HoxC9a* cDNA sequence of zebrafish (*Danio rerio*, BC165307) in GenBank (Fig. 7). Both ends of the sequence from 4nAT, RCC and CC resembled the *HoxC9a* cDNA sequence of zebrafish (NM\_131528.3), while the intermediate sequence in 4nAT was an extra sequence, compared to the cDNA sequence in zebrafish (Fig. 7). The ends of the extra sequence were 'GT' and 'AG' using Jellyfish software alignment (539th position and 540th position: GT; 1616th position and 1617th position: AG). The intron junctions were consistent with the "GT-AG" rule in most species [23, 24]. The similarity of the *HoxC9a* gene exon sequence between 4nAT and zebrafish, CC and zebrafish and RCC and zebrafish was 92.70, 92.80 and 91.90%, respectively. The rule also indicated that the intermediate sequence in 4nAT, CC and RCC were an intron, and the ends of the sequence were exon1 and exon2. We obtained the cDNA of the *HoxC9a* gene by excising the intron and ligating exon1 and exon2. Then, the sequence was translated into putative amino acid sequences using Bioedit software (Version 7.0). The putative amino acid sequences of *HoxC9a* in 4nAT, RCC and CC shared approximately 94.6, 93.8, and 94.6% similarity with that of zebrafish, respectively. All of our findings determined the identity of the PCR clones and the exon1-intron-exon2 structure of the *HoxC9a* gene. The full-length *HoxC9a* genes in 4nAT, RCC and CC were obtained, and these sequences were compared and analyzed (Fig. 3). Sequence analysis indicated that the *HoxC9a* gene included exon1, intron and exon2. The intron junctions were consistent with the "GT-AG" rule. The two isoforms of the *HoxC9a* gene in 4nAT were from its original parents. Moreover, a recombinant isoform of the *HoxC9a* gene was found in 4nAT (Fig. 3).

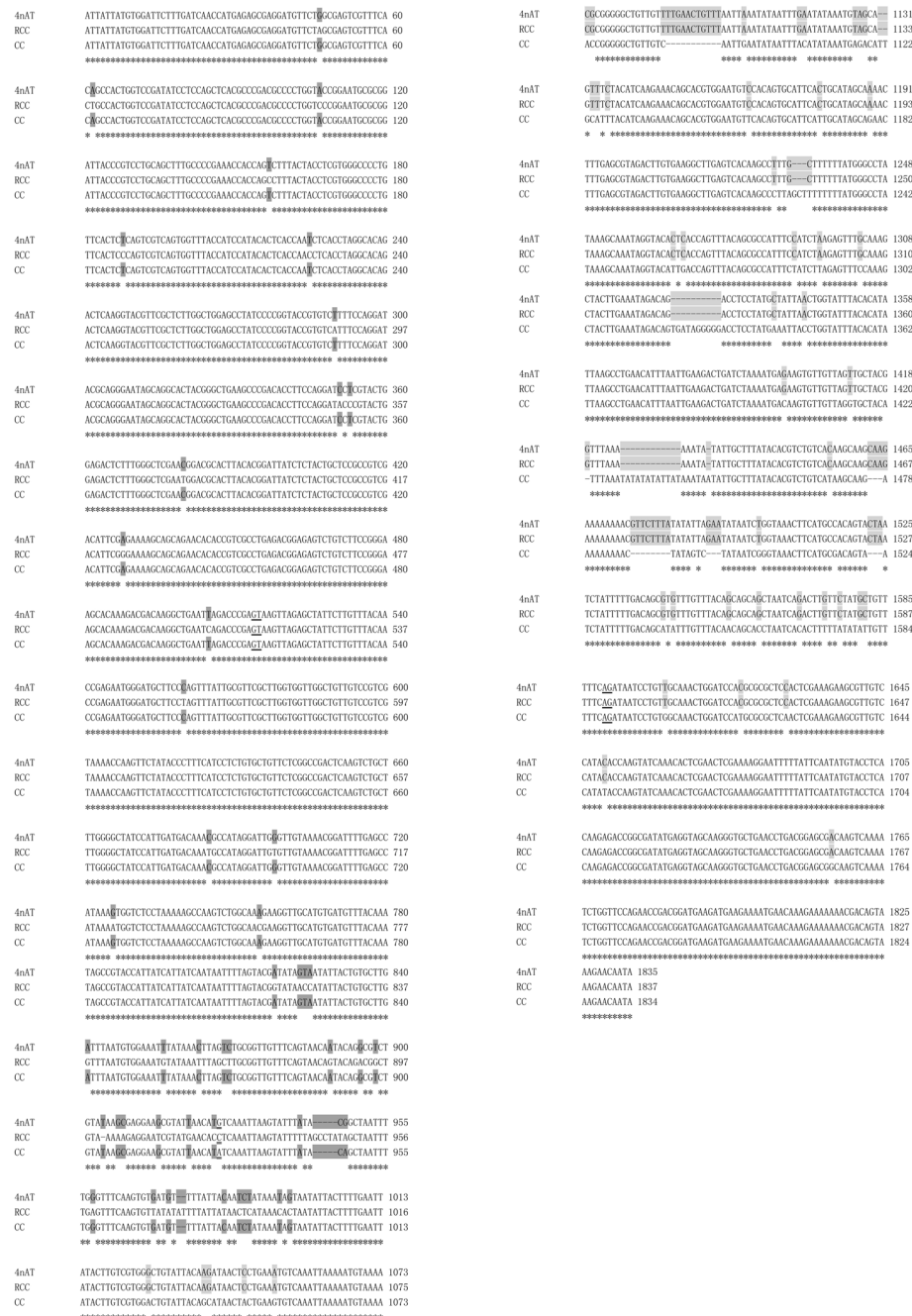
### Discussion

In this study, 29 *Hox* genes were screened in 4nAT and its parents (RCC and CC) using PCR. Among these, 23, 22 and 19 *Hox* genes were detected in 4nAT, RCC and CC, respectively. Compared with the *Hox* genes in zebrafish [25], all of the *Hox* genes detected here had homologous genes in zebrafish. The *Hox* genes in 4nAT, CC and RCC were also found in goldfish with twin tails [26]. In tetraodon nigroviridis

[27], fugu rubripes [28], and medaka [29], the identified *Hox* genes did not have complete identity with the *Hox* genes in RCC, CC, and 4nAT. The *HoxB10a* and *HoxD13a* genes were absent from the tetraodon nigroviridis, fugu rubripes, and medaka, but these genes were present in 4nAT and its related parents (RCC and CC). The retention and loss of genes may be associated with the evolutionary process [30]. These data allowed us to study the evolution of 4nAT after the doubling of the *Hox* gene cluster.

Many cloned fragments were obtained from *Hox* genes, such as *HoxA3a*, *HoxA5a* and *HoxC10a*, by PCR amplification. The length of these genes was 149 bp in 4nAT and its parents. In 4nAT, there were four subtypes of *HoxA5a*. Two had the same subtype as the *HoxA5a* gene of RCC, and two had the same subtype as the *HoxA5a* gene in CC. Three subtypes of the *HoxA3a* and *HoxC10a* genes were detected in 4nAT, which also originated from RCC and CC. *HoxA3a*, *HoxA4a* and *HoxB4a* from zebrafish were found in *Oncorhynchus mykiss*, and these genes in rainbow trout had two specific homologous zebrafish genes [31]. Two specific homologous zebrafish genes of the *Hox* gene in RCC and CC were found, and there were more subtypes in 4nAT. These results suggested that genome duplication had occurred in other cyprinid fish. A comparison of the homologous parts of all subtypes of the *Hox* genes in the 4nAT, RCC and CC, showed that the homologous frame parts of the *Hox* genes in the 4nAT did not exhibit large-scale base mutations after doubling. The sequence of this segment was stable. The *Hox* genes in the 4nAT were more complex and abundant than those in RCC and CC.

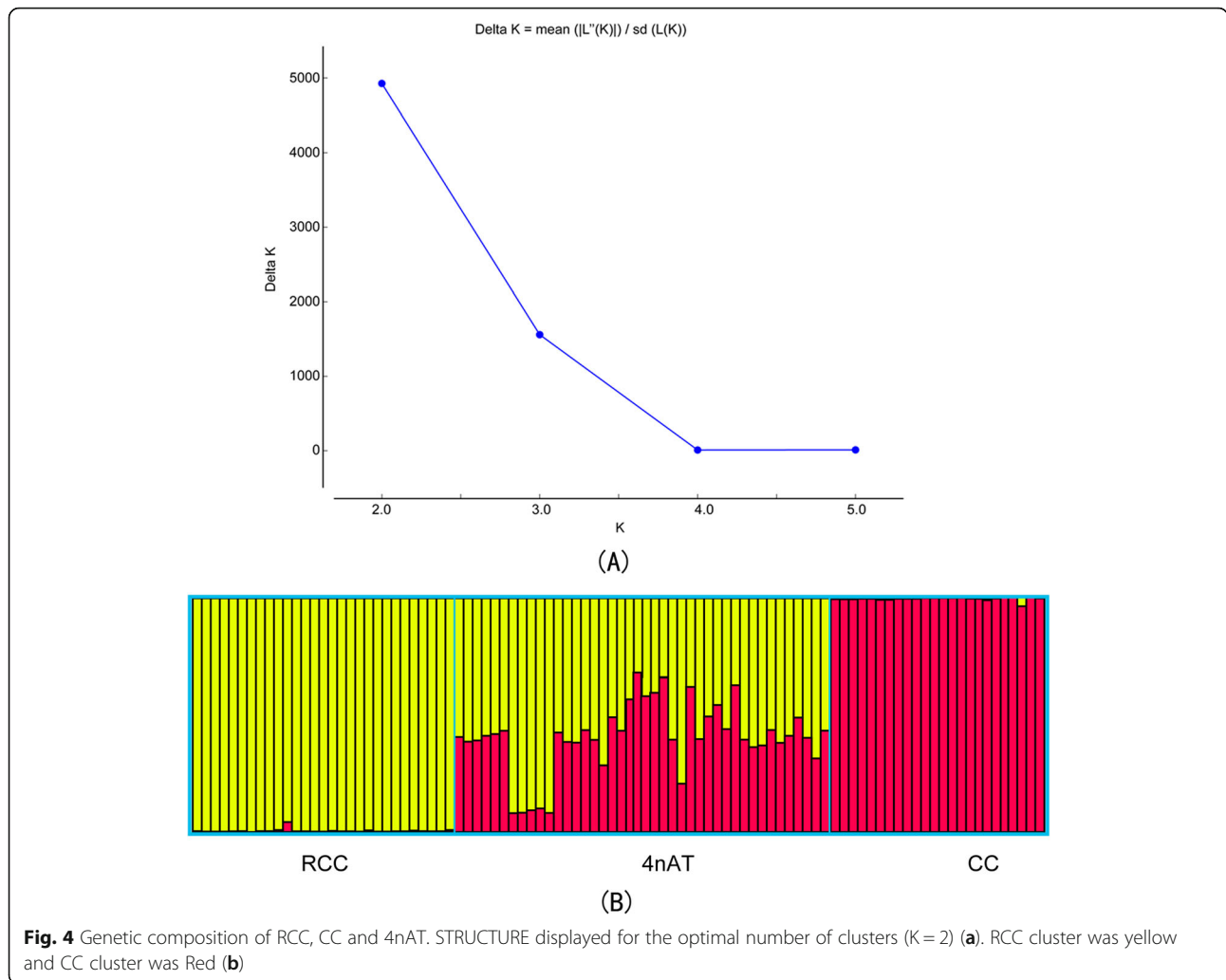
After genome doubling, the new genome could face selection pressure, which could result in the evolution of new features. Genome duplication produced abundant genomic DNA, so the hybrid maintained the dosage balance or rapidly stabilized the duplicated genomes via retention/exclusion of redundancy. Lynch et al. [32] suggested three possible outcomes from the evolution of duplicate genes: non-functionalization, neofunctionalization and subfunctionalization. Duplicated genes were either retained without changes, mutated into other genes with new functions or degenerated into nonfunctional pseudogenes. We found that the *HoxA10b* gene in 4nAT exhibited two base deletions, while the *HoxC13a* gene had two base insertions. In the *Hox* gene sequence of RCC, a base deletion was also found in the *HoxA11a* gene sequence. However,



**Fig. 3** Comparison of recombinant *HoxC9a* gene in 4nAT with *HoxC9a* gene in RCC and CC. Except for one nucleotide marked with the double underscore "=", which was inconsistent with both RCC and CC, the former part of nucleotide sequence of the recombinant *HoxC9a* gene that occurred in 4nAT was consistent with CC (position: 1–1026); the latter half was consistent with RCC (position: 1027–1825). The boundary bases of introns were underlined with "GT-AG"; "\*" meant that the nucleotide was identical at this site

whether the *Hox* gene in CC also had this base variation remained unknown. The occurrence of these pseudogenes (*HoxA10b*, *HoxA11a* and *HoxC13a*) helped to reduce the pressure created by genome doubling. This pressure was consistent with the expectation that there were *Hox* clusters in the 4nAT

genome that had lost functional *Hox* genes due to the reduction of redundancy of the *Hox* genes due to the polyploidization event. In the hybridization process, 4nAT required genetic recombination, mutation, and pseudogenization to reduce the amount of incompatible genetic material and improve fertility [13]. Thus,



**Fig. 4** Genetic composition of RCC, CC and 4nAT. STRUCTURE displayed for the optimal number of clusters (K = 2) (a). RCC cluster was yellow and CC cluster was Red (b)

we established an allotetraploid fish lineages [7]. The results indicated that the genome after doubling could reduce redundant genes by formation of pseudogenes, as indicated by the *Hox* gene in 4nAT. Characterization of the *Hox* gene clusters in allotetraploid hybrids increased understanding of the evolutionary process which occurred after *Hox* gene doubling.

We cloned the full-length sequence of the *HoxC9a* gene, and its structure was divided into three parts: exon1, intron and exon2. This result was consistent with the *Hox* structure described in previous reports [33, 34], in which the bases of the introns conformed to the GT-AG rule. Analysis of the *HoxC9a* genes in RCC, CC and 4nAT, showed that genetic recombination occurred in 4nAT, and the intron region occurred in the recombination region. The former part was consistent with CC, and the latter part was consistent with RCC. This recombinant gene indicated that 4nAT underwent meiosis, alien chromosome synapsis and exchange recombination. The

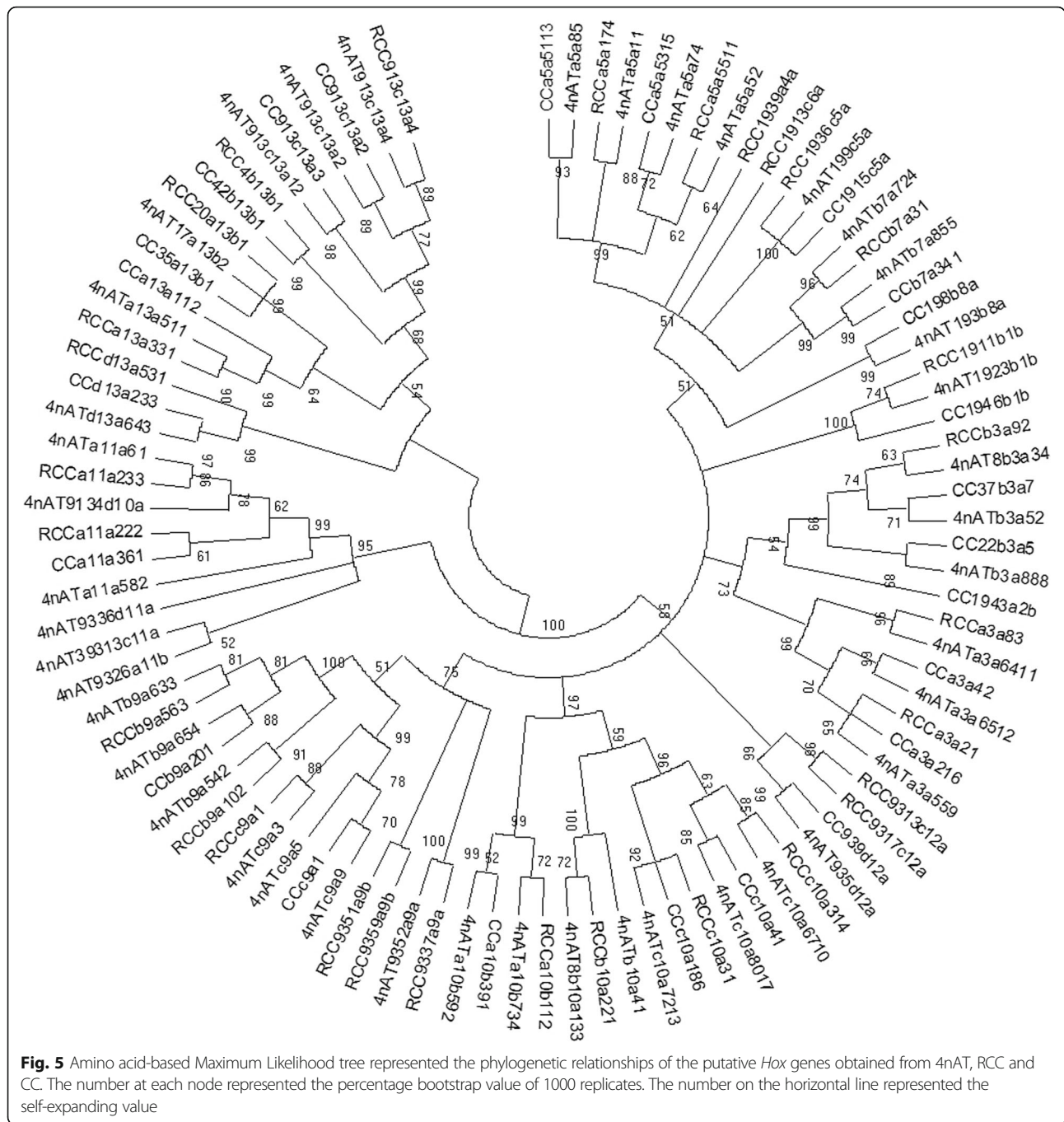
exchange recombination provided a genetic basis for the variation and diversity of the 4nAT and provided raw materials which might facilitate species adaptation and evolution.

In summary, we investigated the gene organization and structures of *Hox* in 4nAT and its parents. There was significant variation in *Hox* genes, which enabled rapid genomic evolution in 4nAT after hybridization and polyploidization. In fish, hybridization and polyploidization could help drive speciation by changing genome structures and generating new genes.

### Conclusions

We found four copies of *Hox* genes in the 4nAT, two copies in RCC and two copies in CC. Obvious variation and pseudogenization were found in some *Hox* genes of the 4nAT. These results revealed the effects of polyploidization on the organization and evolution of *Hox* gene clusters in





fish, and helped to clarify aspects of vertebrate genome evolution.

**Methods**

**Samples and ethics statement**

RCC, CC and 4nAT were obtained from State Key Laboratory of Developmental Biology of Freshwater Fish, Hunan Normal University, Changsha, Hunan, China. The procedures were conducted in accordance with approved guidelines. Fish were housed in open pools (0.067 ha) with

a suitable pH (7.0–8.5), water temperature (22–24 °C), and dissolved oxygen content (5.0–8.0 mg/L) and with adequate food. The fish used for blood samples were anesthetized with 100 mg/L MS-222 (Sigma-Aldrich, St. Louis, MO, United States).

**DNA extraction**

One-year-old 4nAT, one-year-old RCC, and one-year-old CC, were randomly selected. Peripheral blood was collected from the caudal vein of each fish (for total genomic

```

HoxA10b-4nAT
      1      11      21      31      41      51      61      71
1 CGAAAGAAGCGTGTTCCTTATACCAAGATTCAACTACTTGAGCTAGAGAAGGAGTTCCTTGTTCGCATGTTAATGACTCA
1(+1) R K K R V P Y T K I Q L L E L E K E F L F A C * L T Q
1(+2) E R S V F L I P R F N Y L S * R R S S C S H V N * L
1(+3) K E A C S L Y Q D S T T * A R E G V L V R M L I D S

      81
28(+1) AACGCGACGTCTGGAGATCCTGCGAAGTATCACCTCTCCAAACAGACAGGTCATGTGGGTTAAAAACCGCAGGATG
27(+2) T R R L E I L R S I T S P N R Q V N V G * K P Q D
27(+3) K R D V W R S C E V S P L Q T D R S M W V K N R R M
      N A T S G D P A K Y H L S K Q T G Q C G L K T A G
    
```

A

```

HoxA11a1 RKKSVPYTKFQIRELEREFFFSVYINKEKRVQLSRMLNLTDRQVKIWFQNRMR-
HoxA11a2 RKKSVRIQNFKSGSSRGNFSSACTLTKRNEFSSQGCL---TSPIAKKCGSKTAG
*****  :*  . . :*  :  .*. . . . *  :  .:
    
```

B

```

HoxC13a1 RKKRVPYTKIQLKELEKEYAASKFITKDKRRRISATTNLSERQVTIWFQNRMR-
HoxC13a2 RKKRVPYTPRYNRNWKKSMQQASLLPKTRDD--ASLRQPTSQNAKLLCGSKTAG
*****  :: :*  . : . . . * :  :  : : . . . . . :  .:
    
```

C

**Fig. 6** The pseudogenes *HoxA10b* and *HoxC13a2* in 4nAT and the pseudogene *HoxA11a2* in RCC. **a**: listed all the possible amino acid sequences of *HoxA10b* gene in 4nAT which had base deletions. "\*" meant that this site was a terminator when a nucleotide was translated into an amino acid. **b** and **c** reflected transcoding mutations in *HoxA11a* and *HoxC13a* genes. The amino acid sequences of *HoxA11a* and *HoxC13a* were expressed as *HoxA11a1* and *HoxC13a1*, respectively. *HoxA11a2* meant that the amino acid sequence of *HoxA11a* gene with the deletion of base, and *HoxC13a2* was on behalf of the amino acid sequence of *HoxC13a* gene with the insertion of base. "\*" represented the same amino acid at this site; ":" meant that this site was a conservative mutation; "." indicated that the position was a semi-conservative mutation; The blank space represented a large variation in the position

DNA extraction). Total genomic DNA was isolated from blood cells following the manufacturer's instructions (Sangon, Shanghai, China). The concentration and quality of DNA was assessed using agarose gel electrophoresis.

**PCR amplification, cloning, and sequencing**

PCR amplification of 108 bp of the highly conserved homeobox PG1–9 was performed using a degenerate homeobox primer pair: forward primer [5'-GAA TTC CAC TTC AAC (C/A)(G/A)(C/G) TAC CT-3'] and reverse primer [5'-CAT CCT GCG GTT TTG GAA CCA NAT-3']. PCR amplification of 159 bp of the highly

conserved homeobox PG9–13 was performed using a degenerate homeobox primer pair: forward primer [5'-CGA AAG (C/A)G(N/C) GT(N/C) CC(N/C) TA(T/C) AC-3'] and reverse primer [5'-CAT CCT GCG GTT TTG GAA CCA NAT-3'] as described by Amores et al. (2004) [35]. We used Primer Premier 5.0 to design a pair of degenerate primers [HC9aF, 5'-ATT ATT ATG TGG A(C/T) T C(C/T) T TGAT-3'; HC9aR, 5'-TA(C/T) TGT TC(C/T) TT(A/G) CTG TCG TT(C/T) T-3'] based on nucleotide and amino acid sequences of *HoxC9a* in zebrafish (*Danio rerio*), pufferfish species (*Takifugu rubripes* and *Tetraodon nigroviridis*) and



combined the 20 runs at  $K = 2$  using CLUMPP [38] and conducted a graphical display by DISTRUCT [39].

### Phylogenetic analysis

The *Hox* gene sequences were aligned using ClustalX software [40] and a Maximum Likelihood (ML) phylogenetic tree was constructed using MEGA 5.0 software [41]. We assigned PCR fragments based on the identity of the subtree in which they were located. Phylogenetic analysis was performed by ML methods and the best-fitting nucleotide substitution model with the lowest BIC score was determined by using MEGA5 [42]. Conserved regions were determined using the Gblocks program. ML analyses were performed using the GTR + G model, and the robustness of the tree topology was assessed with 1000 bootstrap replicates [43].

### Abbreviations

2nRCF<sub>1</sub>: The first-generation diploid hybrids of RCC (♀) × CC (♂); 2nRCF<sub>2</sub>: The self-mating of 2nRCF<sub>1</sub>; 4nAT: Allotetraploid hybrids; CC: Common carp; RCC: Red crucian carp

### Acknowledgements

We would like to sincerely thank the researchers who helped to complete this manuscript, including professor Zongzhao Zhai, associate professor Yi Zhou, Dr Shi Wang.

### Authors' contributions

RRZ, YD. W, and SJ. L designed the study, carried out the analyses, performed the technical discussions, prepared and draft the manuscript. L. Z, YX. L and HF. T participated in data collection and discussions. JJ. Y and MH. Z were involved in the statistical analysis. All the authors read and approved the final manuscript.

### Funding

This work was supported by the National Natural Science Foundation of China (Grant No. 31430088, 31730098, 31702328), the Key Research and Development Project of Hunan Province (Grant No. 2016NK2130), the earmarked fund for China Agriculture Research System (Grant No.CARS-45), Hunan Provincial Natural Science and Technology Major Project (Grant No.2017NK1031), the China Postdoctoral Science Foundation (Grant No. 2019M662788), the Cooperative Innovation Center of Engineering and New Products for Developmental Biology of Hunan Province (Grant No.20134486). Funds for this project provided funding for reagents, consumables, and laboratory apparatus for this study. The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

### Availability of data and materials

The datasets supporting the conclusions of this article were available in the GenBank repository with access No. MN584925-MN584927.

### Ethics approval and consent to participate

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 fish used as the samples were anesthetized with 100 mg/L MS-222 (Sigma-Aldrich, St. Louis, MO, United States) before dissection and all efforts were made to minimize suffering.

### Consent for publication

Not applicable.

### Competing interests

The authors declared that they had no competing interests.

### Author details

<sup>1</sup>State Key Laboratory of Developmental Biology of Freshwater Fish, Hunan Normal University, Changsha 410081, Hunan, People's Republic of China.

<sup>2</sup>College of Life Sciences, Hunan Normal University, Changsha 410081, Hunan, People's Republic of China. <sup>3</sup>Fisheries Research Institute of Hunan Province, Changsha 410153, People's Republic of China.

Received: 28 October 2019 Accepted: 4 February 2020

Published online: 04 March 2020

### References

- Barker MS, Husband BC, Pires JC. Spreading Winge and flying high: the evolutionary importance of polyploidy after a century of study. *Am J Bot.* 2016;103(7):1139.
- dPY V, Mizrachi E, Marchal K. The evolutionary significance of polyploidy. *Nat Rev Gene.* 2017;18(7):411–24.
- Otto SP. The evolutionary consequences of polyploidy. *Cell.* 2007;131(3):452–62.
- McClintock B: The significance of responses of the genome to challenge. 1983.
- Cox MP, Dong T, Shen G, Dalvi Y, Scott DB, Ganley AR. An interspecific fungal hybrid reveals cross-kingdom rules for allopolyploid gene expression patterns. *PLoS Genet.* 2014;10(3):e1004180.
- Gerstein AC, Chun H-JE, Grant A, Otto SP. Genomic convergence toward diploidy in *Saccharomyces cerevisiae*. *PLoS Genet.* 2006;2(9):e145.
- Liu S, Luo J, Chai J, Ren L, Zhou Y, Huang F, Liu X, Chen Y, Zhang C, Tao M. Genomic incompatibilities in the diploid and tetraploid offspring of the goldfish × common carp cross. *Proc Natl Acad Sci.* 2016;113(5):1327–32.
- Moshe F, Levy AA. Genome evolution due to allopolyploidization in wheat. *Genetics.* 2012;192(3):763–74.
- Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJ, Bierne N, Boughman J, Brelsford A, Buerkle CA, Buggs R. Hybridization and speciation. *J Evol Biol.* 2013;26(2):229–46.
- 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(2–4):171–86.
- Liu S. Distant hybridization leads to different ploidy fishes. *Sci China Life Sci.* 2010;53(4):416–25.
- Song C, Wang J, Liu S, Qin Q, Xiao J, Duan W, Luo K, Liu J, Liu Y. Biological characteristics of an improved triploid crucian carp. *Sci China.* 2009;52(8):733–8.
- Wang J, Ye L, Liu Q, Peng L, Liu W, Yi X, Wang Y, Xiao J, Xu K, Hu F, Ren L, Tao M, Zhang C, Liu Y, Hong Y, Liu S. Rapid genomic DNA changes in allotetraploid fish hybrids. *Heredity.* 2015;114(6):601–9.
- Santini S, Bernardi G. Organization and base composition of tilapia *Hox* genes: implications for the evolution of *Hox* clusters in fish. *Gene.* 2005;346(346):51–61.
- Koh EGL, Lam K, Christoffels A, Erdmann MV, Brenner S, Venkatesh B. *Hox* gene clusters in the Indonesian coelacanth, *Latimeria menadoensis*. *Proc Natl Acad Sci U S A.* 2003;100(3):1084–8.
- Mallo M, Alonso CR. The regulation of *Hox* gene expression during animal development. *Development.* 2013;140(19):3951–63.
- Lemons D, McGinnis W. Genomic evolution of *Hox* gene clusters. *Science.* 2006;313(5795):1918–22.
- Pruett ND, Visconti RP, Jacobs DF, Scholz D, McQuinn T, Sundberg JP, Awgulewitsch A. Evidence for *Hox*-specified positional identities in adult vasculature. *BMC Dev Biol.* 2008;8(1):93.
- Thisse B, Thisse C. Fast release clones: a high throughput expression analysis. *ZFIN direct data submission 2004.*
- Chung N, Jee BK, Chae SW, Jeon Y-W, Lee KH, Rha HK. *HOX* gene analysis of endothelial cell differentiation in human bone marrow-derived mesenchymal stem cells. *Mol Biol Rep.* 2009;36(2):227–35.
- McGinnis W, Krumlauf R. Homeobox genes and axial patterning. *Cell.* 1992;68(2):283–302.
- Holland PWH, Garcia-Fernández J. *Hox* Genes and chordate evolution. *Dev Biol.* 1996;173(2):382–95.
- Ohno K, Ji T. Masuda a: rules and tools to predict the splicing effects of exonic and intronic mutations. *Wiley Interdiscip Rev RNA.* 2018;9(1):e1451.
- Harada N, Yamada K, Saito K, Kibe N, Dohmae S, Takagi Y. Structural characterization of the human estrogen synthetase (aromatase) gene. *Biochem Biophys Res Commun.* 1990;166(1):365–72.

25. Postlethwait JH, Yan Y-L, Gates MA, Horne S, Amores A, Brownlie A, Donovan A, Egan ES, Force A, Gong Z, et al. Vertebrate genome evolution and the zebrafish gene map. *Nat Genet.* 1998;18(4):345–9.
26. Luo J, Stadler PF, He S, Meyer A, Luo J, Stadler PF, He S, Meyer A. PCR survey of Hox genes in the goldfish *Carassius auratus auratus*. *J Exp Zool B Mol Dev Evol.* 2010;308B(3):250–8.
27. Jaillon O, Aury J-M, Brunet F, Petit J-L, Stange-Thomann N, Mauceci E, Bouneau L, Fischer C, Ozouf-Costaz C, Bernot A. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature.* 2004;431(7011):946.
28. Aparicio S. Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science.* 2002;297(5585):1301–10.
29. Naruse K, Fukamachi S, Mitani H, Kondo M, Matsuoka T, Shu K, Hanamura N, Morita Y, Hasegawa K, Nishigaki R. A detailed linkage map of Medaka, *Oryzias latipes*: comparative genomics and genome evolution. *Genetics.* 2000;154(4):1773–84.
30. Hoegg S, Meyer A. Hox clusters as models for vertebrate genome evolution. *Trends Genet.* 2005;21(8):421–4.
31. Moghadam HK, Ferguson MM, Danzmann RG. Evidence for Hox gene duplication in rainbow trout (*Oncorhynchus mykiss*): a Tetraploid model species. *J Mol Evol.* 2005;61(6):804–18.
32. Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. *Science.* 2000;290(5494):1151–5.
33. Zou S-M, Jiang X-Y, He Z-Z, Yuan J, Yuan X-N, Li S-F. Hox gene clusters in blunt snout bream, *Megalobrama amblycephala* and comparison with those of zebrafish, fugu and medaka genomes. *Gene.* 2007;400(1–2):60–70.
34. Xue L, Qian K, Qian H, Li L, Yang Q, Li M. Molecular cloning and characterization of the Myostatin gene in Croceine croaker, *Pseudosciaena crocea*. *Mol Biol Rep.* 2006;33(2):129–35.
35. 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.
36. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics.* 2000;155(2):945–59.
37. Earl DA. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour.* 2012;4(2):359–61.
38. Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics.* 2007;23(14):1801–6.
39. Rosenberg NA. DISTRUCT: a program for the graphical display of population structure. *Mol Ecol Notes.* 2004;4(1):137–8.
40. Thompson JD, Gibson TJ, Higgins DG. Multiple Sequence Alignment Using ClustalW and ClustalX. *Curr Protoc Bioinform.* 2002;Chapter 2(Unit 2):Unit 2.3.
41. Cummings MP: MEGA (molecular evolutionary genetics analysis); 2004.
42. Gao F, Lin W, Shen J, Liao F. Genetic diversity and molecular evolution of arabis mosaic virus based on the CP gene sequence. *Arch Virol.* 2016;161(4):1047–51.
43. Yang Z, Rannala B. Molecular phylogenetics: principles and practice. *Nat Rev Genet.* 2012;13(5):303.

## Publisher's Note

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

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

