Integrative Zoology 2018; 13: 21–35 doi: 10.1111/1749-4877.12271

ORIGINAL ARTICLE

Evolutionary changes of *Hox* genes and relevant regulatory factors provide novel insights into mammalian morphological modifications

Kui LI, Xiaohui SUN, Meixiu CHEN, Yingying SUN, Ran TIAN, Zhengfei WANG, Shixia XU and Guang YANG

Jiangsu Key Laboratory for Biodiversity and Biotechnology, College of Life Sciences, Nanjing Normal University, Nanjing 210023, China

Abstract

The diversity of body plans of mammals accelerates the innovation of lifestyles and the extensive adaptation to different habitats, including terrestrial, aerial and aquatic habitats. However, the genetic basis of those phenotypic modifications, which have occurred during mammalian evolution, remains poorly explored. In the present study, we synthetically surveyed the evolutionary pattern of *Hox* clusters that played a powerful role in the morphogenesis along the head–tail axis of animal embryos and the main regulatory factors (*Mll*, *Bmi1* and *E2f6*) that control the expression of *Hox* genes. A deflected density of repetitive elements and lineage-specific radical mutations of *Mll* have been determined in marine mammals with morphological changes, suggesting that evolutionary changes may alter *Hox* gene expression in these lineages, leading to the morphological modification of these lineages. Although no positive selection was detected at certain ancestor nodes of lineages, the increased ω values of *Hox* genes implied the relaxation of functional constraints of these genes during the mammalian evolutionary process. More importantly, 49 positively-selected sites were identified in mammalian lineages with phenotypic modifications, indicating adaptive evolution acting on *Hox* genes and regulatory factors. In addition, 3 parallel amino acid substitutions in some *Hox* genes were examined in marine mammals, which might be responsible for their streamlined body.

Key words: evolutionary changes, *Hox* genes, mammals, morphological modifications

INTRODUCTION

As an evolutionarily superlative animal group, mammals present morphological diversity, which facilitates the complexity of mammalian body structure, flexibility of locomotion, and the extensive invasion of various habitats, including terrestrial, aerial and aquatic niches. Adaptation refers to the evolutionary changes of organisms which are better able to live in their habitats

Correspondence: Guang Yang and Shixia Xu, Jiangsu Key Laboratory for Biodiversity and Biotechnology, College of Life Sciences, Nanjing Normal University, Nanjing 210023, China. Email: gyang@njnu.edu.cn; xushixia78@163.com.

(Dobzhansky 1968). The morphological modification of axial skeleton and limbs in the tetrapods have been broadly recognized as the adaptive response of locomotion to the change of lifestyles and habitats (Crompton & Jenkins 1973; Shubin *et al*. 1997). Despite the increased number of thoracolumbar in early mammals, it has been recognized that modern mammals have a highly conserved vertebral formulae with 7 cervical vertebrae and 19 combined thoracolumbar vertebrae (Narita & Kuratani 2005; Luo *et al*. 2007; Hirasawa & Kuratani 2013). However, apparent deviations of cervical counts in the sloth (with unequal 5–10) and the manatee (fixed at 6) have broken the cervical constraint (Narita & Kuratani 2005; Hautier 2010). Furthermore, the number of thoracolumbar vertebrae tends to vary in different lineages, such as carnivore, anthropoids, afrotherians and perissodactyls, which unswervingly have escaped from typical vertebral counts (Narita & Kuratani 2005; Asher *et al.* 2011). In addition, morphological transformations are also important to the diversity of lifestyles and habitats. For example, a streamlined body shape is essential for marine mammals (whales, pinnipeds and manatee) to conquer aquatic lifestyles (Foote *et al*. 2015). Cetaceans, originated from the lineage of artiodactyls, have been reconfigured discretely with exceptionally foreshortened and variably coalesced cervical vertebrae as well as numerous highly uniform vertebral units in the torso extending from the chest to the fluke (Buchholtz 2001; Buchholtz & Schur 2004). Interestingly, the manatee, split from Afrotheria, shares some similar phenotypes with cetaceans, such as short centrum length and quite variable fusion in the neck area (Buchholtz *et al.* 2007). Such morphological convergence between cetaceans and the manatee, 2 highly diverged groups, has been suggested to facilitate their adaption to the aquatic environment (Scotland 2011; Foote *et al*. 2015). Except for the various fusions of anterior ribs in cetaceans, the syncretic sternums and increased floating ribs in cetaceans and the manatee, as well as the robust chest region in pinnipeds, also contribute to the morphological adaptation for aquatic lifestyles (Buchholtz 2001; Zhou 2004; DeLynn *et al.* 2011; Pierce *et al.* 2011). In the case of bats, they have evolved a well-developed keel on the sternum, where the mesosterni and xiphoid are fused to be a solid structure, as an adaptation to the powered flight (Walton & Walton 1970). A similar sternal keel evolved in the birds, thereby representing an example of convergent evolution. Moreover, in spite of the conservative organization of limbs in mammals, the paddle-shaped flippers of marine mammals and bats' wings are extremely specialized forelimbs, both of which are of benefit to their locomotory changes and the adaptation to new habitats (Zhou 2004; Liang *et al.* 2013). Taken together, such remarkable morphological changes provide us an ideal opportunity to trace the molecular mechanism propelling the mammalian morphological alterations.

Hox genes, encoding homeodomain-containing transcription factors, are vital for the evolutionary diversification due to their powerful functions in diversifying morphology on the head–tail axis of animal embryos (Lemons & McGinnis 2006; Khaner 2007). These genes have been widely considered to be collinearly organized in clusters, which are consecutively activated from "head" to "tail" during the establishment of body axis (Liang *et al.* 2013). In a mammalian genome, 39 *Hox* gene members were subdivided into 13 paralogous groups (PGs) and closely clustered at 4 loci: *Hox a*, *b*, *c* and *d* (Ruddle *et al.* 1994). It has long been known that these genes are central players in patterning the vertebrate axial skeleton and limbs through changing protein sequences and expression patterns (Pearson *et al.* 2005; Casaca *et al.* 2014). Malformations of the axial and appendicular skeleton in single and compound *Hox* mutants have been studied extensively, which uncovered the functional cooperation and redundancy of *H*ox genes during the establishment of axial skeleton and limbs (Wellik & Capecchi 2003; Mallo *et al.* 2010; Xu & Wellik 2011; Casaca *et al.* 2014). Numerous researchers have clarified that *Hox* genes are expressed along the head–tail axis and regulate the establishment of axial skeleton morphology. The functional connection between *Hox* gene expression and axial patterning has been demonstrated in transgenic mice, which display different types of homeotic transformations of axial skeleton (Casaca *et al.* 2014). In addition, *Hoxd12* misexpression in transgenic mice by gain-of-function has engendered apparent transformations from anterior digits to posterior morphology and digit duplications (Knezevic *et al.* 1997). Furthermore, it has been recognized that *PcG* and *trxG* genes are essential for the maintenance of appropriate expression of *Hox* genes after their initial activation during the establishment of the axial skeleton morphology (Casaca *et al.* 2014). *Bmi1* and *Mll* are pivotal members of *PcG* and *TrxG* (Hanson *et al.* 1999). Mutations of the 2 genes have given rise to abundant morphological changes of the axial skeleton through the alteration of *Hox* gene expression (Benjamin *et al.* 1995; Hanson *et al.* 1999; Cao *et al.* 2005). Moreover, cofactor *E2f6* and *Bmi1* compound mutations lead

to an increase in skeletal defects (Courel *et al.* 2008). Recently, the role of *Hox* evolutionary changes during the mammalian morphological modifications has been receiving increasing attention. Furthermore, adaptive evolution of certain *Hox* genes has partially revealed its function on cetacean phenotypic alternation (Wang *et al.* 2009; Liang *et al.* 2013; Yim *et al.* 2014). Changes in *Hox* structures have also demonstrated indispensable roles for the diversity of body plans in reptiles (Di-Poï *et al.* 2010; Wu *et al.* 2015). However, whether the modifications of *Hox* cluster organization, *Hox* mutations and *Hox* expression patterns impact morphological changes of axial skeletons and limbs have not been comprehensively addressed in mammals.

Here, we synthetically surveyed mammalian *Hox* genes and relevant regulatory factors to evaluate their evolutionary roles in skeletal modifications. Our results uncovered the adaptive evolution of *Hox* genes and relative regulatory factors in mammalian lineages with morphological changes, which may have contributed to their morphological alteration. In addition, we revealed the indispensable roles of *Hox* evolutionary changes in the diversities of different mammals and even in the same species.

MATERIALS AND METHODS

Detection of repetitive elements in *Hox* clusters

To determine the repetitive elements and their distribution in mammalian *Hox* clusters, we first obtained 4 *Hox* clusters from Human (*Homo sapiens*), mouse (*Mus musculus*), giant panda (*Ailuropoda melanoleuca*), seal (*Leptonychotes weddellii*), walrus (*Odobenus rosmarus*), horse (*Equus caballus*), minke whale (*Balaenoptera acutorostrata*), killer whale (*Orcinus orca*), ying fox (P*teropus vampyrus*), little brown bat (*Myotis lucifugus*), armadillo (*Dasypus novemcinctus*), manatee (*Trichechus manatus*) and elephant (*Loxodonta Africana*) from the National Center for Biotechnology Information (NCBI: http://www.ncbi.nlm.nih.gov/). Secondly, the location of each gene has been extracted by GenScan (http://genes.mit.edu/genomescan.html) and Blastn. Finally, the interspersed repetitive elements were detected using CENSOR (http://www.girinst.org/censor/) with default parameters against the Repbase library of mammalian repeat sequences. Simple repeats, low-complexity sequences and poor sequence coverage were excluded in the comparative analysis, as they have been studied previously (Wu *et al.* 2015). Meanwhile, the density of repetitive elements was also estimated using the criteria of Wu *et al*. (2015) to explore the structural changes of mammalian *Hox* clusters.

Sequences retrieval and alignment

To assess the molecular evolution of *Hox* genes, 39 protein-coding sequences (i.e. *Hoxa1*, *2*, *3*, *4*, *5*, *6*, *7*, *9*, *10*, *11*, *13*; *Hoxb1*, *2*, *3*, *4*, *5*, *6*, *7*, *8*, *9*, *13; Hoxc4*, *5*, *6*, *8*, *9*, *10*, *11*, *12*, *13*; *Hoxd1*, *3*, *4*, *8*, *9*, *10*, *11*,*12*, *13*) and main regulatory factors (*Mll*, *Bmi1* and *E2f6*) were acquired from 27 mammalian species representing all major lineages of eutheria and kangaroo as an extra group. All human and cow (*Bos taurus*) genes were collected from the Ensembl public database (http://www.ensembl.org/index.html) and NCBI. Corresponding orthologs were obtained from the Ensembl ortholog_one2one gene database and NCBI for Primates: chimpanzee (*Pan troglodytes*) and bushbaby (*Otolemur garnettii*); Rodentia: mouse (*M. musculus*) and rabbit (*Oryctolagus cuniculus*); Carnivora: cat (*Felis catus*), giant panda (*A. melanoleuca*), seal (*L. weddellii*) and walrus (*O. rosmarus*); Perissodactyla: rhinoceros (*Ceratotherium simum*) and horse (*E. caballus*); Cetartiodactyla: pig (*Sus scrofa*), minke whale (*B. acutorostrata*), sperm whale (*Physeter catodon*), baiji (*Lipotes vexillifer*), killer whale (*O. orca*) and common bottlenose dolphin (*Tursiops truncatus*), Chiroptera: ying fox (*P. vampyrus*) and little brown bat (*M. lucifugus*); Insectivora: hedgehog (*Erinaceus europaeus*); Xenarthra: sloth (*Choloepus hoffmanni*) and armadillo (*D. novemcinctus*); Afrotheria: manatee (*T. manatus*) and elephant (*L. Africana*); and Marsupialia: kangaroo (*Macropus eugenii*). To enhance the comprehensiveness of our searches, we next used the local BLAST program (E-value cut-off 10) against the genome of bowhead whale (*Balaena mysticetus*) (Keane *et al.* 2015), finless porpoise (*Neophocaena phocaenoides*) (unpublished database in our laboratory) and other genomes for missing data. If multiple transcriptional forms for a gene were encountered, the longest transcript was chosen for subsequent analyses. The accession numbers of all genes utilized in our study are listed in Table S1.

The nucleotide sequences and putative amino acid sequences were aligned using MEGA 6.0 (Tamura *et al.* 2013), further manually inspected and edited by eye. In addition, specific amino acid changes among lineages were determined through MEGA 6.0, and the radical changes in amino acid properties at these sites were confirmed by the TreeSAAP 3.2 software package (Woolley *et al.* 2003). This method classified non-synonymous changes into 8 categories according to the change in specific physicochemical properties from conserva-

tive (1–3) to very radical substitutions (6–8) (McClellan & Ellison 2010). For each category, a *z*-score was generated. Significantly positive *z*-scores mean that a given region is under the influence of positive selection. For our purpose, only amino acid properties in categories 6–8 were considered.

Selection detection

To investigate the possible evolutionary pattern of *Hox* genes and relevant regulatory factors in eutherian, we estimated the nonsynonymous (d_N)/synonymous (d_S) substitution ratios (omega, $\omega = d_N/d_S$) using the codon-based maximum likelihood models implemented in the CODEML program from PAML v4.4 (Yang 2007). The values of $\omega < 1$, $= 1$, and > 1 indicate purifying, neutral and positive selection, respectively. The well-supported phylogeny of mammals based on published literature was used as the input tree in all analyses (Murphy *et al.* 2001; Nishihara *et al.* 2005; McGowen *et al.* 2009; Zhou *et al.* 2012; Luan *et al.* 2013).

A pair of site models (M8 *vs* M8a) were applied to detect sites under positive selection. In these nested models, positive selection was determined by a likelihood ratio test (LRT) with ω significantly greater than 1 in alternative model M8. To further evaluate positive selection, the random effect likelihood (REL) (Bayes factor > 50) model in the Datamonkey web server (http://www.datamonkey.org/) was then used. REL allows the synonymous substitution rates to vary among codon sites and then infers positive selection by calculating the substitution rate for individual sites. The sites identified to be under positive selection using both methods were considered as candidates for selection.

To evaluate whether the selection was limited to specific lineages, 2 branch models, namely free-ratio (M1) and branch-site models, were adopted in our study. The free-ratio model (M1), which assumes separate ω ratios for each lineage in the entire tree, was compared with the one ratio model (M0) that allows a single ω for all branches in the phylogenetic tree. For M1, the branches and nodes of $S*d_S = 0$ were discarded in our study due to inaccurate estimations of this model (Yang 2000). Positive selection was further evaluated by modified branch-site model A (test 2) as lower rates of false positive results (Zhang *et al.* 2005). This algorithm allows variation of the ω ratios among amino acid sites and different lineages. In this case, terminal branches and ancestral nodes among eutherian lineages were labeled as foreground branches. Candidate positively-selected sites were identified using Bayes Empirical Bayes (BEB) analysis with posterior probabilities >0.8 (Yang 2007). Moreover, likelihood ratio test *P*-values were corrected for multiple testing with Benjamini and Hochberg's procedure with the threshold value 0.05 (Benjamini & Hochberg 1995; Anisimova & Yang 2007). To further support the PAML results, the complementary protein-level approach was also conducted using the TreeSAAP 3.2 software package.

Spatial distribution of positively selected sites in 3-D structures

The 3-D structures of proteins identified as being under positive selection were predicted using the homology modeling I-TASSER workspace (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) (Roy *et al*. 2010). The protein sequences being used to simulate 3-D structures were obtained from the Ensembl public database, and sequences of human or bottlenose dolphin were used as templates to assure the integrality of modeled structures. Corresponding protein domains were determined by the online Simple Modular Architecture Research Tool (SMART: http://smart.embl-heidelberg.de/). To provide further insight into the functional significance of these sites under positive selection, such sites were mapped onto the 3-D structure using PYMOL (http://pymol.sourceforge.net/). We conducted similar analysis on lineage-specific mutated sites with radical amino acid property changes to determine their functional significance.

Identification of parallel/convergent sites among mammals

The parallel/convergent sites among different mammalian lineages with similar axial and limb changes were identified according to the methods previously described (Foote *et al*. 2015). Ancestral sequences of each node across mammalian phylogeny of each *Hox* gene were reconstructed using the Codeml program in PAML v4.4. For each lineage of marine mammals (cetaceans, pinnipeds and the manatee), the amino acid changes at each position were compared between extant species and the most recent ancestor. Similar analyses were also conducted among terrestrial lineages with morphological convergence. Parallel amino acid changes are required to have the same descendant amino acid from the same ancestral amino acid, while convergent evolution occurs from a different ancestral amino acid. Then, we further tested whether the observed parallel/convergent substitutions were fixed randomly or by natural selection using CONVERG 2 (Zhang & Kumar 1997).

RESULTS

Organization of mammalian *Hox* clusters

All 39 *Hox* genes were obtained from species sampled in the present study, except for the *Hoxc* cluster in baiji and several *Hox* genes in certain species likely due to the gaps or the low quality of genome assembly. Our results showed that the size and organization of mammalian *Hox* clusters are relatively consistent, which reflected the conservation of this gene family. We aimed to detect repetitive elements in mammalian *Hox* clusters, but the number of interspersed repeats was likely underestimated because of the gaps and poor sequence coverage (e.g. *Hoxa* and *Hoxb* in seal, *Hoxc* in armadillo). Interestedly, we determined that ERV, non-LTR (CR1/LINE/SINE) and DNA/hAT transposons generally repeated in mammals. Moreover, in a given cluster, these transposons concentrated on similar regions among different mammals; that is, repeats detected from *Hoxa* cluster located in 5′terminal *Evx1-Hoxa13*, *Hoxb* in *Hoxb13-Hoxb9* and *Hoxd* in the region of *Hoxd9* to 3′terminal. Meanwhile, repeats in *Hoxc* show a symmetrical distribution in all selected mammals (Fig. S1). Notably, both bat *Hoxa* and afrotheria *Hoxb* present ample LINEs among these detected repeats. Excluding seal *Hoxa*/*Hoxb* and armadillo *Hoxc*, *Hoxb* showed the highest density of repetitive elements, and *Hoxa* was the second to the former (Fig. 1), which might be related to the different size of *Hox* clusters (*Hoxb* > *Hoxa* > *Hoxc* > *Hoxd*) (Ruddle *et al.* 1994). Although there is no significant difference in the repetitive elements and their distribution in the same *Hox* cluster among different mammals, subtle imparities of the repetitive element density (the length of all repeats to the total length of the given *Hox* cluster) still can be discovered in marine mammals when compared with their terrestrial relatives (Fig. 1). An increased density of repetitive elements was determined in pinniped *Hoxc*, cetacean *Hoxb* and *Hoxc*, and manatee *Hoxd* clusters. In contrast, the decreased density only presented in cetacean *Hoxd* clusters.

Lineage-specific amino acids changes of *Hox* proteins and relevant regulatory factors

Lineage-specific radical amino acid changes that occurred in Hox proteins and 3 main regulatory factors were counted and mapped onto each branch in the eutherian phylogeny (Fig. 2 and Table S2). Whales were found to have the most specific amino acid sites that

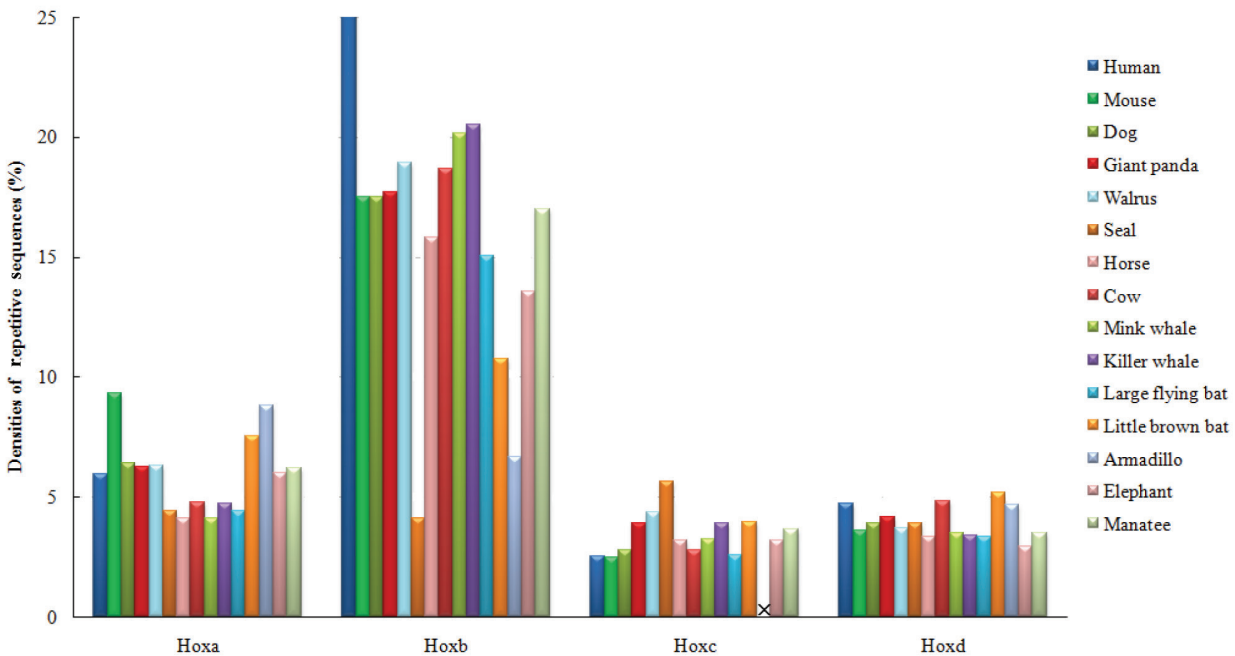

Figure 1 Densities of repetitive elements of mammalian *Hox* clusters. The density of repetitive elements for a specific *Hox* cluster was calculated by the total length of all repetitive sequences to the total length of the cluster.

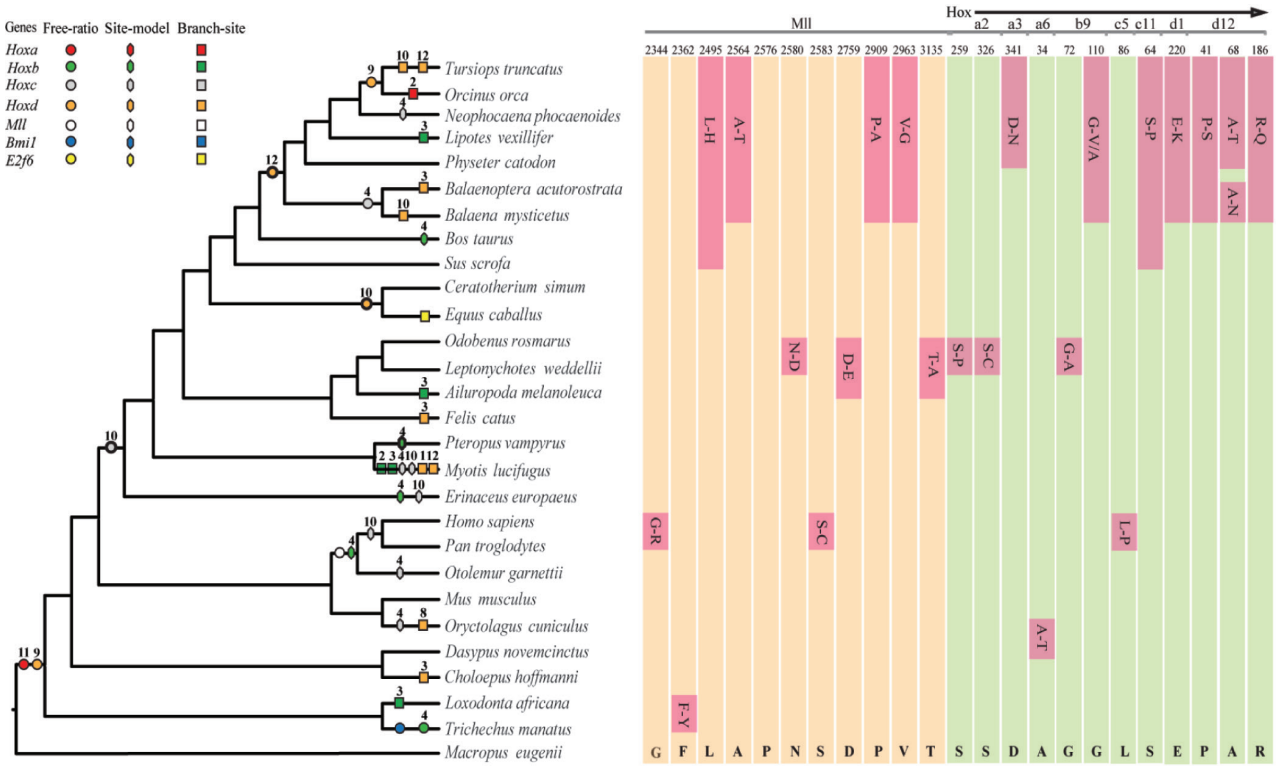

Figure 2 The distribution of genes with ω > 0.5, positive selection and lineage-specific mutations with radical amino acid property changes in the phylogenetic tree. Hox genes and regulatory factors are marked by different colors of various geometric shape with corresponding numbers above. Lineage-specific mutations with radical amino acid property changes among Mll (pink bars) and Hox proteins (aqua bars) are displayed by purple boxes in corresponding lineages and consistent amino acids are labeled at the bottom.

were identified as having undergone radical changes in Hox proteins (Hoxa3: 341, Hoxb9: 110, Hoxd1: 220, and Hoxd12: 68 and 186). Besides, 1 or 2 lineage-specific mutations with radical amino acid property changes were also detected in the Hox proteins of Xenarthra (Hoxa6: 34), Cetartiodactyla (Hoxc11: 64), Pinnipedia (Hoxa2: 295/326; Hoxb9: 72) and Anthropoid (Hoxc5: 86), respectively. When regulatory factors were considered, Mll also presented lineage-specific mutations with radical amino acid property changes, such as Cetacean: 2564, 2576, 2909 and 2963, Anthropoid: 2344 and 2583, Carnivora: 2759 and 3135, Afrotheria: 2362, Pinnipedia: 2580, and Cetartiodactyla: 2495. It has been recognized that radical changes in amino acid properties reflect the strength of natural selection (Sunagar *et al.* 2012). Therefore, these lineage-specific mutations with radical amino acid property changes suggested the independent evolutionary changes of *Hox* genes during species evolution.

Survey for positive selection in *Hox* genes and relevant regulatory factors

The average ω values from M0 for each *Hox* gene and regulatory factor ranged from 0.0145 to 0.2009 (Table S3), indicating powerful purifying selection imposed on these genes during the mammalian evolution. When compared with M1, some members of *Hoxa-d* cluster and 3 regulatory factors significantly fitted the data better than M0 (Table S3), suggesting different selective pressures acting on these genes. Although all ω estimates were less than 1, the increased ω (ω > 0.5) values were detected at certain ancestor nodes (Eutherian: *Hoxa11* & *Hoxd9*; Laurasiatheria: *Hoxc10*; Afrotheria: *Hoxb7*; Manatee: *Hoxb4* and *Bmi1*; Cetaceans: *Hoxd12*; Dolphins: *Hoxd9*; Baleens: *Hoxc4*; Perissodactyla: *Hoxd10*; and Primates: *Mll*) (Fig. 2). These markedly increased ω values may imply the relaxation of functional constraints or positive selection of these genes during

the mammalian evolutionary process.

In mammals, site model M8 significantly fitted the data better than M8a with $P < 0.05$ and $\omega > 1$ in *Hoxb4*, *Hoxc4* and *Hoxc10* (Table 1). Moreover, positive selection acting on *Hoxb4* and *Hoxc4* was further confirmed by REL and radical amino acid property changes of candidate positively selected sites (PP > 0.8) (Table 2). No positive selection was detected in *Hoxc10* by REL, but the 237 site with 6 radical amino acid property changes can be also viewed as positive selection by the criteria used by Sunagar *et al*. (2012). The distribution of these positively selected sites revealed that site 237 in *Hoxc10* and 236 in *Hoxb4* were located in the ancestors of anthropoids and primates, respectively, while the remaining residues were situated in terminal branches (Fig. 2). However, the branch-site model did not detect any molecular adaptive evidence of *Hox* genes and regulatory factors on ancestor nodes, which suggested a conservation or extensive purifying selection of these genes during the speciation and early adaptive radiation of mammals. Surprisingly, faithful positive selection ($\omega > 1$, $P < 0.05$) still can be determined in terminal branches with 46 positively selected sites (PP > 0.8) after correcting for multiple testing, such as in whales, bats and edentates (Fig. 2 and Tables S4 and S5). In addition, radical amino acid property changes were also identified in a few positively-selected sites, which further hinted the positive selection of these genes (Table S5).

Table1 Positive selection detected by site models among genes studied in the present study

Gene	Models	-lnL	LRT	*P*-value	**ω**
Hoxb4	M8a	2765.8184			$\omega = 1$
	M8	2762.0846	7.4677	0.0063	$\omega = 7.905$
Hoxc4	M8a	2174.3700			$\omega = 1$
	M8	2172.1354	4.4693	0.0345	$\omega = 5.775$
Hoxc10	M8a	3754.1411			$\omega = 1$
	M8	3750.9112	6.4598	0.0110	$\omega = 2.795$

Distribution of positively selected sites and specific radical AA changes in 3-D protein structures

Functional significance of the positively selected sites and specific sites with radical amino acid property changes was further estimated through mapping them onto 3-D structures of their corresponding proteins (Fig. S2). All specific sites that underwent radical amino acid property changes of *Hox* genes and positively selected sites of *Hoxb4* and *Hoxc1*0 were localized in the non-homeodomain. Meanwhile, most positively selected sites detected in terminal branches were also located in the non-homeodomain. In Mll, all lineage-specific mutations with radical amino acid property changes were situated in the transcription-activated domains (amino acids 2340-2771 and 2772-3123) (regions identified by Zeleznik-Le *et al*. [1994]) or close to the function domain.

Identification of parallel/convergent amino acid substitutions

To determine whether there is similar molecular mechanism propelling convergent phenotypic traits, we surveyed the parallel/convergent substitutions of Hox

Table 2 Positively selected sites detected by site-model and datamonkey (REL)

Gene	Site	Amino acid	PP (site-model)†	REL	Qualitative change‡	Total
Hoxb4	236	A	0.984*	114.043	P_α P_c P_t	3
Hoxc4	256	Q	0.981*	119.242	P_c c α_m H_t	4
Hoxc10	179	S	0.987*	—	—	—
	212	N	0.924	—	—	—
	237	T	0.820	—	N_s B_r pK′ R_a H_p H_t	6

†Amino acid sites detected by branch-site models with PP > 0.8 are regard as candidates for selection. *Significant. ‡Radical changes in amino acid properties under category 6–8 derived from TreeSAAP. Amino acid property symbols used: surrounding hydrophobicity (H_p), coil tendency (P_c), turn tendency (P_t), solvent accessible reduction ratio (R_a), power to be – middle α-helix (α_m), buriedness (B_r), composition (c), thermodynamic transfer hydrophobicity (H_t), average number of surrounding residues (N_s), equilibrium constant (ionization COOH) (pK′) and alpha-helical tendency (P_α).

proteins in mammals (Table 3). Two parallel substitutions were identified in marine mammals: for example, one was shared between pinnipeds and the manatee at Hoxa2 (124) and Hoxc10 (212), another one was determined between whales and the manatee at Hoxd12 (160). Interestingly, more parallel sites were also detected between marine mammals and other lineages: for example, pinniped and bats at Hoxb13 (212), manatee and xenarthra at Hoxd3 (10), and pinniped and xenarthra at Hoxd12 (164). Among the groups with increased thoracolumbar, no parallel substitution was determined. Besides, perissodactyla Hoxa10 (283) and carnivora Hoxd10 (205), respectively, shared a parallel site with anthropoid. The probabilities that parallel substitutions occurred at random were significantly rejected ($P < 0.01$) by CONVERG 2.

DISCUSSION

The molecular mechanisms underlying morphological specializations of mammalian axial skeleton and limbs have received increasing scientific interest science since 2009. *Hox* genes, as important regulators of embryonic development in animals, have earned much attention for their unique ability to regulate morphologies along the anterior–posterior (A–P) axis (Alexander *et al.* 2009; Mallo *et al.* 2010). Many researchers have provided powerful evidence of *Hox* genes' roles during the development of body plan, skeleton pattern and the formation of limbs (Pearson *et al.* 2005). Our analysis revealed widespread purifying selection in *Hox* clusters and 3 main regulatory factors during the mammalian evolution, which illuminated the strong conservation of these genes over a large evolutionary time frame. Nevertheless, the clue of molecular adaption related to the phenotypic decoration can still be determined by bioinformatics and comparative genomics analysis from *Hox* organization, main regulator factors and Hox protein coding sequences.

Lineage-specific mutations, adaptive evolution of *Hox* genes and regulatory factors responsible for mammalian morphological modifications

Large numbers of studies, mainly in laboratory mice, have shown that appropriate expression of *Hox* genes along head–tail axis specifies the morphogenesis of spine, ribcage and limbs (Zákány & Duboule 1999; Mallo *et al.* 2010). Moreover, changes in *Hox* gene spatio-temporal expression usually lead to various morphological modifications in corresponding regions (Casaca *et al.* 2014). For example, in comparison with other squamates, the delayed expression of *Hox10* and *Hox11* in the snake caused the expansion of caudal regions (DiPoi *et al.* 2010). In contrast, the combinational expression of various *Hox* genes (Hox code) specify distinct regional properties along embryonic axes (Kessel & Gruss 1991; Alexander *et al*. 2009). Furthermore, comparative analyses of *Hox* expression and axial skeleton formation between embryonic chickens and mice have shown that different patterns of Hox code exhibited different patterns of axial skeleton (Burke *et al*. 1995). It has long been recognized that regulatory factors of *Mll*, *Bmi1* and *E2f6* play essential roles in maintaining *Hox* expression after their initiation (Soshnikova & Duboule 2008; Courel *et al.* 2008). Thus, evolutionary changes in these genes may inevitably result in the alteration of *Hox* expression or Hox code. Our results have detected certain lineage-specific mutations with more than one radical amino acid property change in the Mll pro-

Table 3 *Hox* genes that encode parallel amino acid substitution between lineages

gene	Comparative lineage	Position†	Parallel amino acid substitution	*P*-value
Hoxa2	Pinnipeds/manatee	124	S(Ser)→G(Gly)	<0.01
Hoxa10	Perissodactyla/anthropoid	283	G(Gly)→S(Ser)	<0.01
Hoxb13	Pinniped/bats	212	A(Ala)→T(Thr)	<0.01
Hoxc10	Pinnipeds/manatee	212	N(Asn)→S(Ser)	<0.01
Hoxd3	Manatee/xenarthra	10	L(Leu)→P(Pro)	<0.01
Hoxd10	Carnivora/anthropoid	205	A(Ala)→V(Val)	<0.01
Hoxd12	Whales/manatee	160	A(Ala)→S(Ser)	<0.01
	Pinniped/xenarthra	164	D(Asp)→E(Glu)	<0.01

†Position of the amino acid substitution within the encoded product of *Hox* genes.

tein of marine mammals (especially in whales), anthropoid, carnivore and afrotheria (Fig. 2 and Table S2). The radical amino acid property changes can be viewed as a signal of natural selection and more radical changes of these properties might indicate adaptive evolution (Sunagar *et al*. 2012). Therefore, these radical mutations present us an obvious hint about the evolutionary changes of *Mll* gene in these lineages. Interestingly, all of these lineage-specific radical mutations were situated in the transcription activated domains. The expression of *Mll* in HeLa cells has confirmed that these domains play significant roles in the expression of downstream target genes, which suggests the functional importance of these radical mutations (Zeleznik-Le *et al*. 1994). As a methyl transferase, Mll can recuit many proteins to relevant *Hox* genes and further regulate *Hox* expression by the modification and remodeling of chromatins (Terranova *et al.* 2006; Krivtsov & Armstrong 2007). Moreover, heterozygous or homozygous mutations of *Mll* have presented the modification of many *Hox* gene expression patterns (*Hoxa3*, *4*, *5*, *7*, *9*, *10*; *Hoxb3*, *6*, *8*; *Hoxc4*, *5*, *6*, *8*, *9*) and further lead to improper transition of axial skeleton and malformed limbs (Benjamin *et al.* 1995; Hanson *et al.* 1999; Ernst *et al.* 2004; Terranova *et al.* 2006). Therefore, the lineage-specific radical mutations of Mll in the transcription-activated regions imply the changes of relevant *Hox* gene expression or the Hox code, which may contribute to the remarkable morphological changes of marine mammals and the deviation of thoracolumbar in anthropoid, carnivore and afrotheria. *Bmi1*, another indispensable regulatory factor, exhibits an obviously increased ω value (0.6227) in the manatee. The knockout of *Bmi1* in laboratory mice has presented significantly upregulated expression of 5′ *Hox* genes and downregulated expression of 3′ *Hox* members (Cao *et al.* 2005). Most importantly, both the heterozygous and homozygous mutations of *Bmi1* can cause morphological modification of the axial skeleton of laboratory mice along the head–hail axis, particularly at the cervical–thoracic and thoracic–lumbar boundaries (Courel *et al*. 2008). However, no positive selection was detected in *Bmi1*; thus, according to the suggestion of Di-Poi *et al*. (2010) and Sunagar *et al*. (2012), the increased ω value observed in *Bmi1* of the manatee likely represents the relaxation of functional constraints. This result may have contributed to the formation of streamlined body shape in manatees, including the morphological changes in axial skeleton and limbs.

Furthermore, the repetitive elements density of *Hox* genes were found to increase (*Hoxb*: whales; *Hoxc*: whales and pinnipeds; *Hoxd*: the manatee) and decrease (*Hoxd*: whales) in marine mammals, as well as increase the number of LINEs in bats (*Hoxa*) and afrotheria (*Hoxb*), implying the *Hox* clusters' structures have subtle imparities from the otherwise strong constraint in these lineages. It has been suggested that LINEs impact on gene expression (Feschotte & Pritham 2007). In addition, evidence has shown that the structural changes of *Hox* clusters caused by the insertion/deletion of repeats may influence the expression of *Hox* genes, such as in Banna caecilian (*Ichthyophis bannanicus*) and squamates (Di-Poi *et al*. 2010; Wu *et al*. 2015). Therefore, the *Hox* clusters' structural deviation lying in the abovementioned lineages *Hox* clusters may be associated with their morphological changes to respond to the modification of the locomotory pattern and different habitats.

Apart from changes in *Hox* expression, the crucial roles in the morphological modifications caused by the single and combined *Hox* mutation have also been widely discussed (Pearson *et al*. 2005; Alexander *et al*. 2009; Mallo *et al*. 2010). In the present study, we only detected the increased ω values ($\omega > 0.5$) of *Hoxa11* and *Hoxd9* at the ancestor of eutherians, and *Hoxc10* at the ancestor of laurasiatheres, respectively. These increased ω values may also indicate a process of the relaxed functional constraints in these *Hox* genes during the early speciation of different clades. Moreover, it has been clarified that mutated *Hox9*, *Hox10* and *Hox11* in mice have caused the thoracolumbar spine morphological transformation from posterior thoracic to sacral vertebrae and, thus, the phenotypic and numerical abnormality of corresponding vertebrae (Wellik & Capecchi 2003; McIntyre *et al*. 2007). Morphological analyses have shown that laurasiatheres and eutherians have exhibited frequent vertebral deviations from the highly conserved pattern presenting in modern mammals (Narita & Kuratani 2005). Therefore, we propose that in the early evolution of mammals, thoracolumbar vertebrae counts had been diversified concomitant with the relaxation of functional constraints of these genes in the eutherians (*Hoxa11* and *Hoxd9*) and laurasiatheres (*Hoxc10*).

The lineage-specific mutations that underwent radical amino acid property changes of Hox proteins are concentrated in cetaceans (Hoxa3, Hoxb9, Hoxd1 and Hoxd12) and pinnipeds (Hoxa2, Hoxb9). As far as we known, *Hox* genes are activated in a time sequence that follows their physical order within the cluster. In addition, mutations in the *Hoxa2*, *Hoxa3* and *Hoxd1* genes in mice have been reported to cause defects in cervical vertebra, whereas mutations in the *Hoxb9* and

Hoxd12 result in defects in posterior spine and ribcage and limbs, respectively (Alexander *et al*. 2009; Mallo *et al.* 2010). Thus, these mutations of Hox proteins in whales and pinnipeds may contribute to their convergent streamlined shape in order to adapt to the aquatic environment.

Functional redundancy is recognized as a common phenomenon among *Hox* paralogous genes, which was evidenced by the fact that paralogous mutants of *Hox5*, *6*, *7*, *8*, *9* and *11* would produce more severe axial phenotypes than in single or double mutant combinations (Wellik 2007). Thus, a few lineage-specific mutated sites with radical amino acid property changes of certain Hox proteins were identified to sporadically distribute in other lineages but without relevant phenotype changes. Taking the functional redundancy among *Hox* paralogous genes into account, these sporadic mutated loci may not be powerful enough to drive obvious phenotypic changes.

In addition, increased ω values (ω > 0.5) of certain *Hox* genes of the ancestors of whales (*Hoxd12*), dolphins (*Hoxd9*), baleen whales (*Hoxc4*) and the manatee (*Hoxb4*) were also detected, which hinted at a process of the relaxation of function constraints. Both single and compound mutations in mice have illuminated that *Hox4* exhibit a similar function to *Hox3* during the morphological assignation of cervical vertebrae (Horan *et al*. 1995). Mutated *Hoxd9* was found to cause morphological malformation of mice axial skeletons, ribcages and forelimbs (Alexander *et al*. 2009; Mallo *et al*. 2010; Xu & Wellik 2011). Thus, the relaxation of function constraints of these genes may also contribute to the morphological modifications of marine mammals. Interestingly, the anatomical evidence shows that cervical fusion occurred in marine mammals for their adaptation to aquatic habitat, especially for whales whose cervical vertebrae can form a single block in some genera (*Kogia*, *Pseudorca*, *Phocoenoides*) (Buchholtz & Schur 2004; Buchholtz 2001, 2007). Previous studies have shown that mutations of *Hox3* and *Hox4* genes in mice led to phenotypic malformation of cervical vertebra and fusion between adjacent vertebra (Capecchi 1996; Manley & Capecchi 1997). More importantly, the evolutionary changes of *Hox3* and *Hox4* were mainly found on whales, which may be associated with cervical fusion in this group. In addition, positive selection of *Hoxc10* and increased ω values (ω > 0.5) of *Hoxd10* have also been found in the ancestors of anthropoid and perissodactyla, respectively, which may be due to the number of the thoracolumbar vertebrae in both groups' (anthropoid: 17–18, perissodactyla: 22–24) deviations from the conservative pattern (19 vertebrae) in modern mammals (Narita & Kuratani 2005).

Unexpectedly, among the total 49 positively selected sites detected by PAML and REL in mammals (Table 2 and Table S5), almost all positive selection signals of *Hox* genes conformably distribute to terminal branches, especially in whales and bats. Clearly, this molecular evidence cannot provide a reliable explanation for the phenotypic changes of different lineages. According to Gellon and McGinnis (1998), *Hox* mutation might be an ubiquity strategy during *Hox* evolution, involving not only variations of different animals but also the polymorphism in a species. Moreover, the variation of axial skeletons among or even in the same species in mammals has been reported in certain morphological studies (Walton & Walton 1970; Yang & Huang 1982; Buchholtz 2001; Viertel & Trieb 2003; Narita & Kuratani 2005; Newitt *et al.* 2008). Therefore, the positive selection of *Hox* genes occurring in terminal branches may contribute to the diversity of mammalian morphologies.

Parallel amino acid substitutions drive the morphological convergence?

It has been reported that convergent morphologies have occurred among mammals during the interaction of similar habitats, such as marine mammals sharing numerous phenotypic adaptations to aquatic environment (Foote *et al.* 2015). Parallel amino acid substitutions of Hox proteins have been identified in marine mammals, such as pinnipeds and manatee sharing one site separately at Hoxa2 and Hoxc10, which might be considered as molecular evidence for their convergent streamlined shape. Furthermore, it has been reported that similar gene function may also lead to convergent phenotypes in unrelated taxa (Nery *et al*. 2016). Our results have detected the adaptive molecular signatures of *Hoxb9* in whales and pinnipeds, and *Hoxd12* in whales. Multiple mutations in mice have shown that the function of these 2 genes is necessary for the development of limbs (Zákány & Duboule 1999; Xu & Wellik 2011). Taking the functional overlap of different *Hox* genes into account, the adaptive molecular signatures of these 2 genes in whales and pinnipeds may also suggest a different evolutionary route to the similar limb morphologies in whales and pinnipeds. This seems to conform to the assumption that the phenotypic convergence of marine mammals may be achieved through the selection of different *Hox* genes with equivalent functions proposed by Nery *et al*. (2016). However, Nery *et al*. (2016) merely detected a small mi-

nority of positive selection of *Hox* genes with overlapping functions in marine mammals, which seemed to be insufficiently cogent case with respect to the remarkable convergent phenotypic traits of marine mammals.

In addition, the parallel amino acid substitution of Hoxd12 was identified between the whales and the manatee. More importantly, evolutionary analyses have clarified that adaptive evolution of *Hoxd12* contributed to the origin and diversification of the cetacean flipper (Wang *et al*. 2009). Thus, the parallel amino acid substitution of Hoxd12 may link to the origin of flippers in whales and the manatee. Unexpectedly, a parallel amino acid substitution (site: 164) of Hoxd12 was examined in pinnipeds and the xenarthra. Pinnipeds' limbs have specialized into short and flat flippers to produce propulsion while they swim (Pierce *et al*. 2011). In contrast, living armadillos have modified their limbs for scratch-digging, with big claws and long lever arms to faciliate the line of action of the principal muscles (Vizcaíno *et al*. 2006). Moreover, sloths have evolved a hook-like shape with long and slender limbs equipped with long, curved and pointed claws (Britton 1941). Thus, the parallel amino acid substitution may be associated with their adaptation to a different lifestyle in spite of no morphological convergence of limbs between these 2 lineages. Parallel amino acid substitutions of some *Hox* genes were also discovered between marine mammals and terrestrial contemporaries. These results were consistent with the inference that convergent molecular evolution is relatively common among mammals (Foote *et al*. 2015). As proposed by Foote *et al*. (2015) and Zhou *et al*. (2016), these similar morphologies may be derived by independent molecular evolution. Therefore, more detailed investigations from different aspects (i.e. parallel amino acid substitution, positive selection and functional changes of different genes with similar functions) should be synthetically employed to explore the inherent impetus driving phenotypic convergences.

CONCLUSIONS

The present study has systematically explored the evolutionary changes at *Hox* genes and its main regulatory factors in mammals. Our results have illuminated the organizational difference of *Hox* clusters and the radical lineage-specific mutations of Mll in marine mammals and other lineages with morphological modifications during the process of evolution. The evolutionary change of regulatory factors hints at the alteration of relevant *Hox* gene expression or Hox codes in these lineages. Furthermore, increased ω values of some *Hox* genes were detected at certain ancestor nodes of lineages, implying the relaxation of functional constraints of these genes during mammalian evolutionary process. Meanwhile, a total of 49 positive selected sites were determined by PAML and REL in mammalian lineages with phenotypic modifications. The increased ω values and the positive selection have indicated evolutionary changes of *Hox* genes, which may link to the morphological changes of relevant mammals. Nevertheless, almost all positively selected sites (46 sites) lie in terminal branches, especially in whales and bats. It seems that the *Hox* gene mutations may also play a significant role in the diversities of different mammals and even in the same species. However, the roles of *Hox* genes along with cooperation of related regulatory networks in the phenotypic formation of animals are complicated (Gellon & McGinnis 1998; Pearson *et al.* 2005; Casaca *et al.* 2014). Although the evolutionary changes of regulatory factors (mainly in *Mll*) and *Hox* structures had influenced the modifications of *Hox* expression, we still cannot know exactly how these changes influence the Hox code for complicated regulatory networks because of the limitation of fresh samples for functional verification. Moreover, it has been clarified that functional redundancy among *Hox* paralogous genes is one of the major hurdles for understanding the roles of mutation in *Hox* genes in the modification of axial morphology (Wellik 2007). Therefore, the deletion of this redundancy may contribute to a better understanding of genetic mechanism for the formation of axial skeletons. In addition, despite the synergetic function of paralogous *Hox* genes, it has also been shown that different members of paralogs present non-equivalent efficacy during the specification of corresponding axial regions (Manley & Capecchi 1997; Casaca *et al*. 2014). This also adds difficulty to our understanding of the roles of evolutionary changes of a single *Hox* gene in the morphological modification of mammals. Thus, more in-depth investigations should be conducted to explore the roles of changes in *Hox* expression patterns (Hox code), functional redundancy and the reciprocity of paralogous *Hox* genes and other regulatory factors in the phenotypic changes of mammals. The molecular mechanisms of convergent phenotypes also need more detailed discussion (including parallel amino acid substitution, positive selection and functional changes of different genes with similar functions) to attain a better understanding of the morphological modification of mammals.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Funds for Distinguished Young Scholars to G. Y. (grant number 31325025), the State Key Program of National Natural Science Foundation of China (NSFC) (grant number 31630071), the NSFC (grant numbers 31570379 to S. X., 31370401 to W. R.), the Priority Academic Program Development of Jiangsu Higher Education Institutions to G. Y. and S. X., and the Natural Science Foundation of Jiangsu Province of China (grant number BK20141449) to S.X.

REFERENCES

Alexander T, Nolte C, Krumlauf R (2009). Hox genes and segmentation of the hindbrain and axial skeleton. *Annual Review of Cell and Developmental* **25**, 431–56.

Anisimova M, Yang Z (2007). Multiple hypothesis testing to detect lineages under positive selection that affects only a few sites. *Molecular Biology and Evolution* **24**, 1219–28.

Asher RJ, Lin KH, Kardjilov N, Hautier L (2011). Variability and constraint in the mammalian vertebral column. *Journal of Evolutionary Biology* **24**, 1080–90.

Benjamini Y, Hochberg Y (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289–300.

Britton SW (1941). Form and function in the sloth. *The Quarterly Review of Biology* **16**, 13–34.

Buchholtz EA (2001). Vertebral osteology and swimming style in living and fossil whales (Order: Cetacea). *Journal of Zoology* **253**, 175–90.

Buchholtz EA, Schur SA (2004). Vertebral osteology in Delphinidae (Cetacea). *Zoological Journal of the Linnean Society* **140**, 383–401.

Buchholtz EA, Booth AC, Webbink KE (2007). Vertebral anatomy in the Florida manatee, *Trichechus manatus latirostris:* a developmental and evolutionary analysis. *The Anatomical Record* **290**, 624–37.

Burke AC, Nelson CE, Morgan BA, Tabin C (1995). Hox genes and the evolution of vertebrate axial morphology. *Development* **121**, 333–46.

Cao R, Tsukada YI, Zhang Y (2005). Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing. *Molecular Cell* **20**, 845–54.

Casaca A, Santos AC, Mallo M (2014). Controlling Hox gene expression and activity to build the vertebrate axial skeleton. *Developmental Dynamics* **243**, 24–36.

Courel M, Friesenhahn L, Lees JA (2008). E2f6 and Bmi1 cooperate in axial skeletal development. *Developmental Dynamics* **237**, 1232–42.

Crompton AW, Jenkins FA Jr. (1973). Mammals from reptiles: A review of mammalian origins. *Annual Review of Earth and Planetary Sciences* **1**, 131–55.

DeLynn R, Lovewell G, Wells RS, Early G (2011). Congenital scoliosis of a bottlenose dolphin. *Journal of Wildlife Diseases* **47**, 979–83.

Di-Poi N, Montoya-Burgos JI, Miller H, Pourquié O, Milinkovitch MC, Duboule D (2010). Changes in Hox genes' structure and function during the evolution of the squamate body plan. *Nature* **464**, 99–103.

Dobzhansky T (1968). On some fundamental concepts of Darwinian biology. In: Dobzhansky T, Hecht MK, Steere WC eds. *Evolutionary Biology*, vol II. Springer US, New York, pp. 1–34.

Ernst P, Mabon M, Davidson AJ, Zon LI, Korsmeyer SJ (2004). An Mll-dependent Hox program drives hematopoietic progenitor expansion. *Current Biology* **14**, 2063–9.

Feschotte C, Pritham EJ. (2007). DNA transposons and the evolution of eukaryotic genomes[J]. *Annual Review of Genetics* **41**, 331–68.

Foote AD, Liu Y, Thomas GW *et al.* (2015). Convergent evolution of the genomes of marine mammals. *Nature Genetics* **47**, 272–5.

Gellon G, McGinnis W (1998). Shaping animal body plans in development and evolution by modulation of Hox expression patterns. *BioEssays* **20**, 116–25.

Hanson RD, Hess JL, Benjamin DY *et al*. (1999). Mammalian Trithorax and polycomb-group homologues are antagonistic regulators of homeotic development. *Proceedings of the National Academy of Sciences* **96**, 14372–7.

Hautier L, Weisbecker V, Sánchez-Villagra MR, Goswami A, Asher RJ (2010). Skeletal development in sloths and the evolution of mammalian vertebral patterning. *PANS* **107**, 18903–8.

Hirasawa T, Kuratani S (2013). A new scenario of the evolutionary derivation of the mammalian diaphragm from shoulder muscles. *Journal of Anatomy* **222**, 504–517.

Horan GS, Ramírez-Solis R, Featherstone MS *et al.* (1995). Compound mutants for the paralogous hoxa-

4, hoxb-4, and hoxd-4 genes show more complete homeotic transformations and a dose-dependent increase in the number of vertebrae transformed. *Genes & Development* **9**, 1667–77.

Keane M, Semeiks J, Webb AE *et al*. (2015). Insights into the evolution of longevity from the bowhead whale genome. *Cell Reports* **10**, 112–22.

Kessel M, Gruss P (1991). Homeotic transformations of murine vertebrae and concomitant alteration of Hox codes induced by retinoic acid. *Cell* **67**, 89–104.

Khaner O (2007). Evolutionary innovations of the vertebrates. *Integrative Zoology* **2**, 60–7.

Knezevic V, De-Santo R, Schughart K *et al.* (1997). Hoxd-12 differentially affects preaxial and postaxial chondrogenic branches in the limb and regulates Sonic hedgehog in a positive feedback loop. *Development* **124**, 4523–36.

Krivtsov AV, Armstrong SA (2007). MLL translocations, histone modifications and leukaemia stem-cell development. *Nature Reviews Cancer* **7**, 823–33.

Lemons D, McGinnis W (2006). Genomic evolution of Hox gene clusters. *Science* **313**, 1918–22.

Liang L, Shen YY, Pan XW *et al.* (2013). Adaptive evolution of the Hox gene family for development in bats and dolphins. *PLoS ONE* **8**, e65944.

Luan PT, Ryder OA, Davis H, Zhang YP, Yu L (2013). Incorporating indels as phylogenetic characters: impact for interfamilial relationships within Arctoidea (Mammalia: Carnivora). *Molecular Phylogenetics and Evolution* **66**, 748–56.

Luo ZX, Chen P, Li G, Chen M (2007). A new eutriconodont mammal and evolutionary development in early mammals. *Nature* **446**, 288–93.

Mallo M, Wellik DM, Deschamps J (2010). Hox genes and regional patterning of the vertebrate body plan. *Developmental Biology* **344**, 7–15.

Manley NR, Capecchi MR (1997). Hox group 3 paralogous genes act synergistically in the formation of somitic and neural crest-derived structures. *Developmental Biology* **192**, 274–88.

Mann RS, Lelli KM, Joshi R (2009). Hox specificity: unique roles for cofactors and collaborators. *Current Topics in Developmental Biology* **88**, 63–101.

McClellan DA, Ellison DD (2010). Assessing and improving the accuracy of detecting protein adaptation with the TreeSAAP analytical software. *International Journal of Bioinformatics Research and Applications* **6**, 120–33.

McGowen MR, Spaulding M, Gatesy J (2009). Divergence date estimation and a comprehensive molecular tree of extant cetaceans. *Molecular Phylogenetics and Evolution* **53**, 891–906.

McIntyre DC, Rakshit S, Yallowitz AR *et al.* (2007). Hox patterning of the vertebrate rib cage. *Development* **134**, 2981–9.

Murphy WJ, Eizirik E, Johnson WE *et al.* (2001). Molecular phylogenetics and the origins of placental mammals. *Nature* **409**, 614–8.

Narita Y, Kuratani S (2005). Evolution of the vertebral formulae in mammals: a perspective on developmental constraints. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* **304**, 91–106.

Nery MF, Borges B, Dragalzew AC, Kohlsdorf T (2016). Selection on different genes with equivalent functions: the convergence story told by Hox genes along the evolution of aquatic mammalian lineages. *BMC Evolutionary Biology*, DOI: 10.1186/s12862-016-0682-4.

Newitt A, German AJ, Barr FJ (2008). Congenital abnormalities of the feline vertebral column. *Veterinary Radiology & Ultrasound* **49**, 35–41.

Nishihara H, Satta Y, Nikaido M *et al.* (2005). A retroposon analysis of Afrotherian phylogeny. *Molecular Biology and Evolution* **22**, 1823–33.

Pearson JC, Lemons D, McGinnis W (2005). Modulating Hox gene functions during animal body patterning. *Nature Reviews Genetics* **6**, 893–904.

Pick L, Heffer A (2012). Hox gene evolution: multiple mechanisms contributing to evolutionary novelties. *Annals of the New York Academy of Sciences* **1256**, 15–32.

Pierce SE, Clack JA, Hutchinson JR (2011). Comparative axial morphology in pinnipeds and its correlation with aquatic locomotory behaviour. *Journal of Anatomy* **219**, 502–14.

Roy A, Kucukural A, Zhang Y (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols* **5**, 725–38.

Ruddle FH, Bartels JL, Bentley KL, Kappen C, Murtha MT, Pendleton JW (1994). Evolution of Hox genes. *Annual Review of Genetics* **28**, 423–42.

Scotland RW (2011). What is parallelism? *Evolution & Development* **13**, 214–27.

Shubin N, Tabin C, Carroll S (1997). Fossils, genes and the evolution of animal limbs. *Nature* **388**, 639–48.

Soshnikova N, Duboule D (2008). Epigenetic regulation of Hox gene activation: the waltz of methyls. *Bioessays* **30**, 199–202.

Sunagar K, Johnson WE, O'Brien SJ, Vasconcelos V, Antunes A (2012). Evolution of CRISPs associated with toxicoferan-reptilian venom and mammalian reproduction. *Molecular Biology and Evolution* **29**, 1807–22.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* **30**, 2725–9.

Terranova R, Agherbi H, Boned A, Meresse S, Djabali M (2006). Histone and DNA methylation defects at Hox genes in mice expressing a SET domain-truncated form of Mll. *Proceedings of the National Academy of Sciences* **103**, 6629–34.

Viertel B, Trieb G (2003). The Himalayan rabbit (Oryctolagus cuniculus L.): spontaneous incidences of endpoints from prenatal developmental toxicity studies. *Laboratory Animals* **37**, 19–36.

Vizcaíno SF, Bargo MS, Kay RF, Milne N (2006). The armadillos (Mammalia, Xenarthra, Dasypodidae) of the Santa Cruz Formation (early–middle Miocene): an approach to their paleobiology. *Palaeogeography, Palaeoclimatology, Palaeoecology* **237**, 255–69.

Walton DW, Walton GM (1970). Post-cranial osteology of bats. In *About Bats: A Chiropteran Symposium*, pp. 93–126.

Wang Z, Yuan L, Rossiter SJ *et al*. (2009). Adaptive evolution of 5′ HoxD genes in the origin and diversification of the cetacean flipper. *Molecular Biology and Evolution* **26**, 613–22.

Wellik DM, Capecchi MR (2003). Hox10 and Hox11 genes are required to globally pattern the mammalian skeleton. *Science* **301**, 363–7.

Woolley S, Johnson J, Smith MJ, Crandall KA, McClellan DA (2003). TreeSAAP: Selection on amino acid properties using phylogenetic trees. *Bioinformatics* **19**, 671–2.

Wu R, Liu Q, Meng S, Zhang P, Liang D (2015). Hox cluster characterization of Banna caecilian (*Ichthyophis bannanicus*) provides hints for slow evolution of its genome. *BMC Genomics* **16**, DOI: 10.1186/s12864-015-1684-0.

Xu B, Wellik DM (2011). Axial Hox9 activity establishes the posterior field in the developing forelimb. *PNAS* **108,** 4888–91.

Yang Z (2000). Maximum likelihood estimation on large phylogenies and analysis of adaptive evolution in human influenza virus A. *Journal of Molecular Evolution* **51**, 423–32.

Yang Z (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* **24**, 1586–91.

Yang AF, Fang LX. (1982). The vertebral column and the thorax of the giant panda, ailuropoda melanoleuca. *Acta Theriologica Sinica* **2**, 1–7.

Yim HS, Cho YS, Guang X *et al.* (2014). Minke whale genome and aquatic adaptation in cetaceans. *Nature Genetics* **46**, 88–92.

Benjamin DY, Hess JL, Horning SE *et al.* (1995). Altered Hox expression and segmental identity in Mll-mutant mice. *Nature* **378**, 505–8.

Zákány J, Duboule D (1999). Hox genes in digit development and evolution. *Cell and Tissue Research* **296**, 19–25.

Zeleznik-Le NJ, Harden AM, Rowley JD (1994). 11q23 translocations split the "AT-hook" cruciform DNA-binding region and the transcriptional repression domain from the activation domain of the mixed-lineage leukemia (MLL) gene. *PNAS* **91**, 10610–4.

Zhang J, Nielsen R, Yang Z (2005). Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Molecular Biology and Evolution* **22**, 2472–9.

Zhang J, Kumar S (1997). Detection of convergent and parallel evolution at the amino acid sequence level. *Molecular Biology and Evolution* **14**, 527–36.

Zhou KY (2004). Cetacea, Carnivora: Phocoidea, Sirenia. In: Chen YY, ed. *Fauna Sinica Mammalia*. Science Press, Beijing, China, pp. 1–35.

Zhou X, Seim I, Gladyshev VN (2015). Convergent evolution of marine mammals is associated with distinct substitutions in common genes. *Scientific Reports* **5**, DOI: 10.1038/srep16550.

Zhou X, Xu S, Xu J *et al*. (2012). Phylogenomic analysis resolves the interordinal relationships and rapid diversification of the Laurasiatherian mammals. *Systematic Biology* **61**, 150–64.

SUPPLEMENTARY MATERIALS

Additional supporting information may be found in the online version of this article.

Table S1 Ensembl transcript IDs, GenBank accession numbers and genomic location information of all Hox genes and relevant regulatory factors used in the present research

Table S2 Lineage-specific amino acids changes in regulatory factors and *Hox* genes among mammals

Table S3 Parameter estimates under branch models among Hox genes and regulatory factors.

Table S4 The parameters of selective pressure for Hox genes and E2f6 among mammals by branch-site models

Table S5 Positively selected sites and quality changes in Hox genes and regulatory factors among mammals

Figure S1 The proportion of repetitive elements numbers to total repeats in different gene regions among species

Figure S2 3-D structure distribution of positively selected sites (red balls) and lineage-specific radical amino acids change sites (brown balls) in corresponding proteins. The homeodomain of Hox proteins, DNA binding region of E2f6, and transcriptional activation region of Mll are colored yellow. N-terminal and C-terminal regions are ocean blue and violet, respectively.

Cite this article as:

Li K, Sun X, Chen M *et al*. (2018). Evolutionary changes of *Hox* genes and relevant regulatory factors provide novel insights into mammalian morphological modifications. *Integrative Zoology* **13**, 21–35.