

doi: 10.7541/2023.2022.0465

基于全基因组的龟鳖目 *Hox* 基因序列特征及进化分析

黄保友^{1,2} 楼灵媛^{1,2} 樊嘉伟^{1,2} 孙伟^{2,3} 钱国英^{2,3} 葛楚天^{2,3} 王宗吉^{2,3}

(1. 上海海洋大学水产与生命学院, 上海 201306; 2. 浙江万里学院动物性别与发育重点研究所, 宁波 315100;
3. 浙江万里学院生物与环境学院, 宁波 315100)

摘要: 为对 *Hox* 基因在龟鳖目物种中进行系统地序列比较分析和进化研究, 文章对目前具有染色体水平的龟鳖目基因组进行了 *Hox* 基因的鉴定, 序列特征、进化和转录组分析。研究结果表明龟鳖物种的 *Hox* 基因簇是高度保守的。非重复序列的缺失导致鳖科 *HoxB9*—*HoxB13* 基因间区相对龟科短了 10 kb。大量 *Hox* 基因编码区发生了鳖科或龟科特异的序列替换、插入和缺失。胸部骨骼发育相关的 *Hox* 基因在鳖科祖先发生了快速进化和受到正选择。*Hox* 基因的表达具有组织、时期特异性, 主要在胚胎时期的顶端外胚层嵴、背甲嵴和性腺表达。研究为龟鳖目 *Hox* 基因不同胚胎时期的多组学及表达调控分析提供了靶标, 也为进一步厘清龟鳖物种演化创新提供了参考。

关键词: *Hox* 基因; 龟鳖目; 演化; 鳖

中图分类号: Q344⁺.1

文献标识码: A

文章编号: 1000-3207(2023)10-1617-11



Hox 基因最早在黑腹果蝇中发现^[1], 并在真核生物中广泛存在^[2-5]。*Hox* 基因由 2 个外显子和 1 个内含子组成, 第 2 个外显子编码的多肽链存在一段由 60 个氨基酸序列组成的高保守同源结构域 (Homeodomain), 因此也被称为同源异型盒基因^[6]。*Hox* 基因在染色体上排列成簇, 单个基因簇上最多具有 15 个不同的基因成员^[7]。脊椎动物两次全基因组复制导致 *Hox* 基因在高等脊椎动物(除硬骨鱼类外)中形成了 4 个基因簇, 每个簇由 13 个同源基因组成并线性排列在同一条染色体上^[8]。

Hox 基因作为一类转录因子, 调控着后生动物的体轴形态发育^[9]。在 *Hox* 基因的调控下, 体节分化为不同形态的椎骨(颈椎、胸椎、腰椎、骶椎和尾椎)和骨骼肌^[10]。*Hox* 基因簇的表达具有时空共线性: 在胚胎发育过程中, *Hox* 基因簇的基因自 3' 至 5' 依次激活, 并分别在胚胎前、中、后部表达^[11]。鉴于 *Hox* 基因在身体形态发育模式中的重要作用, *Hox* 基因、基因簇的改变被认为是塑造后生物形态多样性的重要来源^[12]。

哺乳类、鸟类及爬行类已鉴定出分布在 4 个簇上的 39 个同源 *Hox* 基因^[6, 13]。在爬行动物中, 蛇类特异性地丢失了 *HoxD12*^[14], *HoxC3* 仅存在于蜥蜴和蛇类中^[6]。不同于上述类群, 两栖类保留了 *HoxC3* 但丢失了 *HoxC4* 和 *HoxC5*^[15]。腔棘鱼具有 42 个 *Hox* 基因, 其中保留了哺乳类、鸟类丢失的 *HoxA14*、*HoxB10*、*HoxC1* 和 *HoxC3*, 但丢失了 *HoxD13*^[16]。在物种演化过程中, 一些功能冗余的基因可能会被随机丢失, 从而导致了不同物种之间的 *Hox* 基因的种类和数目差异^[17]。

龟鳖目 (Testudines) 隶属于脊索动物门爬行纲, 是爬行动物中古老而特化的类群。龟鳖祖先最早可以追溯到三叠纪时期的半甲齿龟 (*Odontochelys semitestacea*)^[18]。与其他爬行动物相比, 龟鳖颈部可以灵活伸缩, 并且具有独特、外部可见的甲壳(包括腹甲和背甲)。在龟鳖胚胎发育过程中, 位于背侧体节边界、前后肢中间的背甲嵴 (Carapacial ridge) 诱导胸椎骨相连的肋骨呈扇形排列, 最终形成了背甲^[19]。鳖科背甲由可见的肋板支撑革质软皮组成(“软壳”),

收稿日期: 2022-11-19; 修订日期: 2023-01-13

基金项目: 国家重点研发计划(2018YFD0900203); 国家自然科学基金(32102772); 2022 年浙江省“生物工程”一流学科(ZS2022008); 浙江省高校基础科研经费项目资助 [Supported by the National Key Research and Development Program (2018YFD0900203); the National Natural Science Foundation of China (32102772); Zhejiang Provincial Top Discipline of Biological Engineering (Level A), Zhejiang Wanli University (ZS2022008); Basic Scientific Research Foundation of Zhejiang Provincial Universities]

作者简介: 黄保友(1996—), 男, 硕士研究生; 主要从事龟鳖类性别决定研究。E-mail: 610579119@qq.com

通信作者: 王宗吉(1988—), 男, 博士; 主要从事多组学、进化发育学研究。E-mail: zongji.wang@zwu.edu.cn

不同于龟科坚硬并有纹理的角质盾片(“硬壳”)。*Hox*基因对脊椎动物的体轴和骨骼发育有着重要作用,因此发掘龟鳖*Hox*基因的序列、表达特征和进化有助于解析其特殊的外部形态特征。

目前,*Hox*基因家族仅在隶属于龟鳖目的中华鳖中进行了鉴定^[6],尚缺乏对龟鳖目*Hox*基因序列特征和进化水平的系统性研究。因此,本研究对目前具有染色体水平基因组(利于完整*Hox*基因簇预测)的龟鳖物种进行*Hox*基因鉴定、染色体定位、重复序列富集、序列特征、蛋白质空间结构、适应性进化及转录组学分析,为进一步研究*Hox*基因在龟鳖物种中的功能奠定了基础。

1 材料与方法

1.1 序列搜集和*Hox*基因同源预测

Ensembl (<https://www.ensembl.org/index.html>)获取小鼠(*Mus musculus*)、家鸡(*Gallus gallus*)、锦龟(*Chrysemys picta bellii*)、普通岩蜥(*Podarcis muralis*)、热带爪蟾(*Xenopus tropicalis*)和斑马鱼(*Danio rerio*)的*Hox*基因蛋白质序列(最长的同源异构体)作为参考序列,用于基因同源预测。

DNA Zoo(<https://www.dnazoo.org/>)获取密西西比鳄(*Alligator mississippiensis*, ASM28112v4)、Ensembl获取家鸡(*Gallus gallus*, GCA_016699485.1)的基因组序列。中华鳖(*Pelodiscus sinensis*, GCA_000230535.1)、斑鳖(*Rafetus swinhoei*, GCA_019425775.1)、刺鳖(*Apalone spinifera*, GCA_000385615.1)、黄喉拟水龟(*Mauremys mutica*, GCF_020497125.1)、草龟(*Mauremys reevesii*, GCF_016161935.1)、古兹沙漠陆龟(*Gopherus evgoodei*, GCF_007399415.2)、锦龟(*Chrysemys picta bellii*, GCA_000241765.5)、红耳龟(*Trachemys scripta elegans*, GCF_013100865.1)、绿海龟(*Chelonia mydas*, GCF_015237465.1)和棱皮龟(*Dermochelys coriacea*, GCF_009764565.3)的基因组序列从NCBI获得。刺鳖基因组(contig N50: 4693 bp)不能预测出完整*Hox*基因集,本研究仅对其部分*Hox*基因进行分析。

使用Blast^[20](v2.2.26)将搜集的蛋白质序列与基因组序列进行序列比对($E < 10^{-5}$);剔除比对率<30%,相似度<50%的序列;使用solar^[21](v0.9.6)将Blast结果连接,连接后的区域上下游各延伸5 kb;使用GENEWISE^[22](v2.4.1)对其进行基因结构精细比对,去冗余并得到最佳匹配结果;预测得到的同源基因集与NR数据库(v4)进行Blast比对,根据序列相似性(top three best hits)确定正确的同源预测基因。

1.2 基因染色体定位分析

*Hox*基因的染色体定位使用R包gggenes([https://](https://github.com/wilkox/gggenes)

github.com/wilkox/gggenes)进行可视化展示。

1.3 基因组重复序列注释

分别对9个龟鳖物种及密西西比鳄基因组进行重复序列注释:使用RepeatModeler(v2.0.3)进行重复序列*de novo*注释,将注释文库与Rebase^[23]数据库合并,RepeatMasker(v4.0.6)进行重复序列同源注释。

1.4 序列分析

使用FASTML^[24](<https://fastml.tau.ac.il/>)进行*Hox*基因祖先(一致)序列预测。通过Jalview^[25]的MUSCLE^[26]进行序列比对,识别类群特异的碱基突变、插入或缺失,并可视化展示。

AlphaFold^[27]进行氨基酸序列的空间结构预测, PYMOL (<http://pymol.Sourceforge.Net/>)用于蛋白质3d结构可视化及突变位点定位。SMART (<http://smart.embl-heidelberg.de/>)预测蛋白质的结构域。

1.5 进化分析

已发表的龟鳖系统发生树作为物种树^[28],利用PAML^[29](v4.9)codeml模块的分支模型(Branch model)计算*Hox*基因进化速率。自由比率模型(Free ratio model)检测各个分支的进化速率。二比率模型(One-ratio vs. two ratio, two ratio vs. two ratio with the constraint $\omega \leq 1$)进行最大似然率检验(Likelihood ratio test)检测*Hox*基因是否在特定分支中发生快速进化和正选择。

1.6 *Hox*基因的表达分析

NCBI获取龟鳖转录组数据,存在3个及以上生物学重复的样品用于后续分析。使用fastp^[30](v0.23.2)进行数据质控和过滤。利用new Tuxedo^[31]流程进行转录组分析: HISAT2^[32](v2.2.1)将reads比对回基因组序列, StringTie^[33](v2.2.1)计算转录本表达量。使用R包tximport^[34]和edgeR^[35](v3.40.0)获得样品间校正(TMM)后基因水平的FPKM值。

2 结果

2.1 龟鳖目*Hox*基因簇是高度保守的

为了厘清龟鳖目*Hox*基因和基因簇是否保守,我们选取脊椎动物各类群(鱼类、两栖类、爬行类、鸟类和哺乳类)代表物种的*Hox*蛋白质序列作为参考,分别对9个龟鳖物种(隶属于曲颈龟亚目的不同科),家鸡和密西西比鳄基因组序列进行*Hox*基因同源预测。所研究的8个龟鳖物种、家鸡和密西西比鳄都具有同源的39个*Hox*基因,仅红耳龟未预测出*HoxB13*和*HoxC5*。这些*Hox*基因分布在4个基因簇上,其中A簇包括11个基因(A1—A7、A9—A11和A13),B簇10个基因(B1—B9和B13),C簇9个基因(C4—C6和C8—C13)和D簇9个基因(D1、

D3、D4和D8—D13; 图1)。新组装的红耳龟基因组序列(未发表)预测出了*HoxB13*和*HoxC5*, 这说明红耳龟并未丢失这两个基因。值得一提的是, 相比羊膜动物祖先的*Hox*基因簇, 龟鳖、鳄鱼和鸟类都丢失了*HoxC3*。上述结果提示龟鳖特异形态(例如背甲等)的产生并不来源于*Hox*基因数目和基因簇的剧烈变化。

2.2 鳖科HoxB基因簇末端基因间区存在特异性序列变化

通过比较龟鳖*Hox*基因和基因间区的序列特

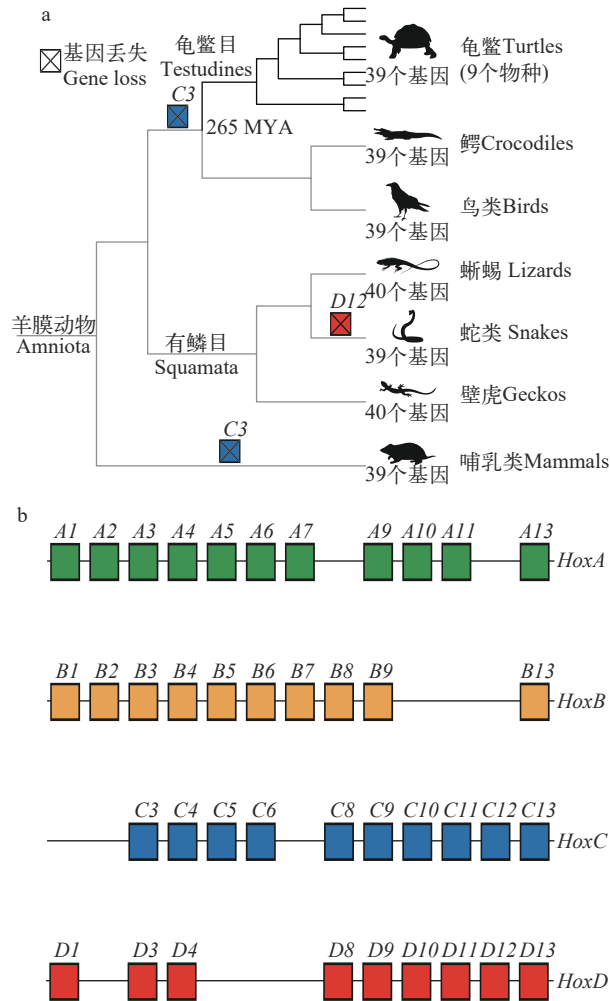


图1 羊膜动物的*Hox*基因簇

Fig. 1 Schematic diagram of *Hox* clusters in amniotes

a. *Hox*基因在羊膜动物中的演化历史。系统发育树的黑色分支代表本项目研究的龟鳖物种。基因丢失时间标记在对应的系统发育节点。265MYA. 265百万年前(million years ago)^[36]。
b. 羊膜动物祖先的*Hox*基因簇

The left part(a) shows the evolutionary history of *Hox* genes across amniotic lineages. The black branches of the phylogenetic tree highlight the studied turtles in this study. Gene losses are labelled at corresponding phylogenetic nodes. 265MYA. 265 million years ago. The right part(b) shows the ancestral *Hox* clusters in amniotes

征, 我们发现龟鳖*Hox*基因都是由2个外显子和1个内含子组成, 二者的长度在龟鳖物种间没有明显差异。这说明龟鳖目*Hox*基因结构高度保守, 并不存在外显子、内含子长度的特异性变化。龟鳖*Hox*基因簇(A簇、B簇B1—B9、C簇和D簇)的基因间区也不存在明显差异(图2, S2—S4)。同龟科相比, 鳖科物种的*HoxB9*—*HoxB13*基因间区的长度明显短

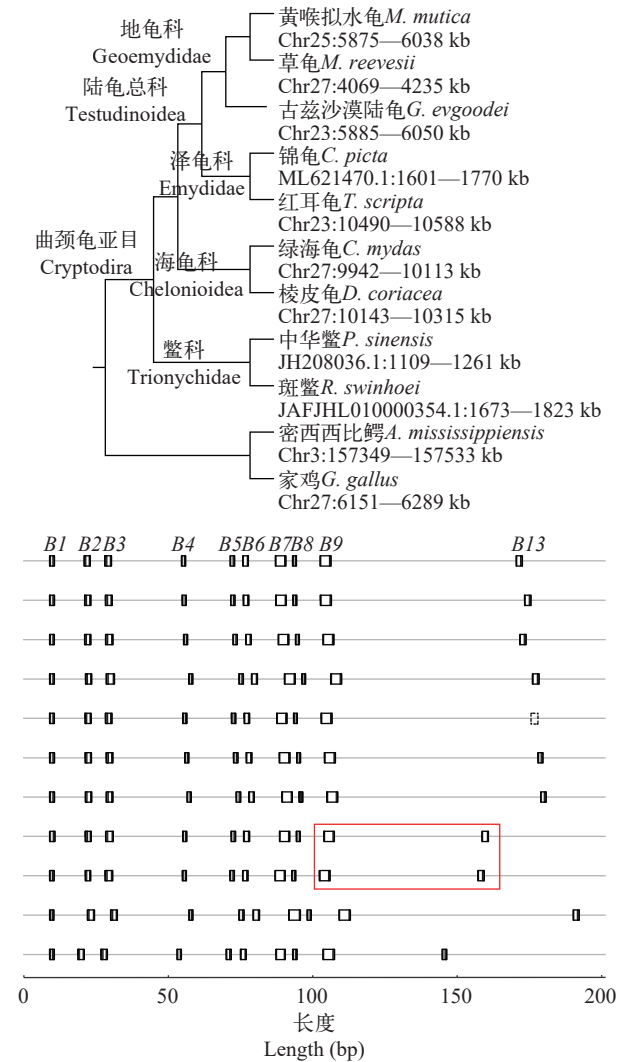


图2 *HoxB*基因簇在11个物种中的染色体定位

Fig. 2 Genomic organization of *HoxB* clusters in 11 species
左边为物种间的系统发生树(亲缘关系)、分类、物种名及*Hox*基因簇的坐标(染色体: 起始坐标-终止坐标)。基因名字标记在第一行。横轴标记出预测的*Hox*基因外显子(方框)、内含子(白框)和基因间区域(粗线)的长度大小

The left part shows the phylogenetic relationships, lineages, species name and the coordinates of the *Hox* gene cluster (chromosome: start- end) in studied species. Gene names are shown above each box on the top cluster. The annotated sizes of predicted exons (black boxes), introns (white boxes), and intergenic regions (thick lines) for *Hox* genes allow for direct size comparisons between species

了10 kb左右(t 检验: p 值=2.872e-06, 图2), 这提示鳖科在该基因间区存在特异的序列变化。

我们对研究的9个龟鳖目物种和密西西比鳄进行了重复序列注释(从头预测和同源预测), 重点比较了不同物种 $HoxB9$ — $HoxB13$ 基因间区的重复序列。鳖科和龟科在该基因间区的长度差异并不是重复序列总量导致, 而是主要来源于非重复序列的差异(鳖科长度恰好比龟科短10 kb左右; 图3a)。此外, 亲缘关系较近的物种(锦龟和红耳龟; 绿海龟和棱皮龟; 中华鳖和斑鳖)在该基因间区具有相似的重复序列和非重复序列含量, 这说明该基因间区的序列含量具有科内保守性和科间特异性。

通过比较物种间该基因间区的重复序列元件含量, 我们发现相比龟科物种, 鳖科的DNA转座子和逆转座子[长末端重复序列(Long terminal repeat, LTR)和非LTR逆转座子]的含量较低(图3b)。然而其他重复序列[(剔除前述两类序列: 包括低复杂度

序列(Low complexity)、微卫星序列(Satellite)、简单重复序列(Simple repeat)及未知序列(Unknown)]在鳖科的含量接近20%。结合外群密西西比鳄的序列特征, 证实鳖科物种在 $HoxB9$ — $HoxB13$ 基因间区发生了特异的重复序列扩张(其他重复序列)和收缩(DNA转座子和逆转座子)。对重复序列元件类型的进一步分析发现鳖科的长散布重复元件(Long interspersed repetitive element, LINE/L2), 短散布重复元件/哺乳动物广泛散布重复元件(Short interspersed nuclear element/mammalian-Wide Interspersed Repeats, SINE/MIR)及DNA转座子/P不稳定因子-Harbinger超家族(DNA transposon/P Instability Factor-Harbinger superfamily, DNA/PIF-Harbinger)的含量很低(<1%), 未知序列(Unknown)的含量接近15%(图3c), 这说明鳖科物种在该基因间区发生了特异重复序列扩张, 进而呈现出不同于龟科物种的重复序列特征。

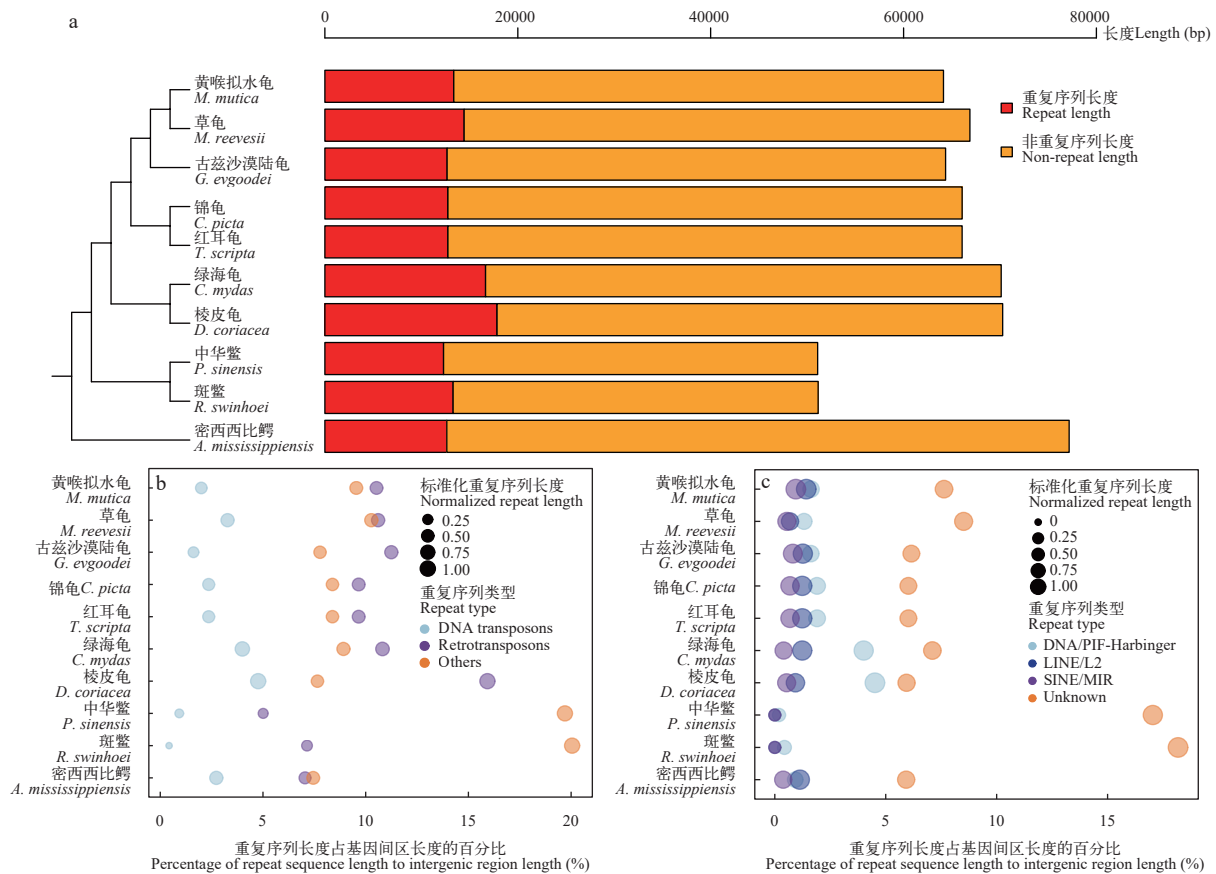


图3 龟鳖物种及短吻鳄 $HoxB9$ — $HoxB13$ 基因间区的序列组成比较

Fig. 3 Comparison of the sequence content within the intergenic region between $HoxB9$ and $HoxB13$ in turtle and alligator genomes

a. $HoxB9$ — $HoxB13$ 基因间区的序列组成比较; b和c分别按照转座元件的大类和子类统计。横轴表示重复序列长度占基因间区长度的百分比。不同颜色表示重复序列类型, 圆的大小代表基因间区长度相对物种间最长基因间区长度的比值

a. shows the comparison of the sequence content within the intergenic region between $HoxB9$ and $HoxB13$; The transposable element content is analyzed at type (b) and subtype (c) levels. The x axis represents the percentage of repeat sequence length to intergenic region length. The colors of the circles indicate the types of repetitive sequences, the areas of the circles represent the ratio of the intergenic region length to the longest intergenic region length across species

2.3 Hox基因编码区存在大量鳖科特异性序列替换、插入和缺失

通过构建Hox基因祖先序列和物种间序列比较,我们发现39个Hox基因有16个基因(41%)的编码区发生了类群特异性的序列插入缺失(Insertion and deletion, indel)和替换(表1和图4)。类群特异性的序列变化主要来源于鳖科Hox基因(14/16, 87.5%),其中10个Hox基因(62.5%)发生了鳖科特异的序列缺失,包括B1—B4、B9、B13、C9、C10、C12和D8。5个Hox基因(31.3%)发生了鳖科特异的序列插入,包括A5、B5、B13、C13和D9。其中A1、B4和B13,以及D1和D9基因分别发生了龟科特异的序列插入和缺失。虽然龟鳖物种广泛存在类群特异的序列插入和缺失,但核苷酸序列长度变化都是3的倍数,并不会大范围改变基因的读码框和氨基酸编码(表1)。

为了厘清上述类群特异的序列插入和缺失是否改变蛋白质的功能,我们使用AlphaFold^[27]对突变前后的氨基酸序列进行了空间结构预测。预测结果表明这些类群特异的序列插入和缺失并未导致其空间结构发生明显变化(图4c)。Hox基因B4、B13和D9发生了多种类型的插入/缺失突变,但这些突变都未明显改变其空间结构,也并不位于结构域上(图4a)。这些结果表明龟鳖目Hox基因序列在演化过程中存在类群特异的突变,但并不改变基因的

结构/功能。此外,Hox基因存在大量鳖科特异的序列突变暗示着鳖科具有更高的进化速率。

2.4 鳖科Hox基因的适应性进化

为了检验上述鳖科序列特异突变的Hox基因是否具有更高的进化速率,我们选取家鸡为外群计算基因序列在不同分支的进化速率。Hox基因在龟鳖物种间的进化速率呈现出明显的差异性,多数鳖科Hox基因相对龟科具有更高进化速率(图5a—d),这和上述鳖科具有更多的序列突变相一致。

基于鳖科物种不同于龟科的特异表型,我们检测了鳖科Hox基因的自然选择压力。Hox基因A1、D8和D9基因在鳖科物种最近祖先发生了快速进化,而A9、B5、C6和D12受到正选择(图5e)。

2.5 龟鳖Hox基因的组织、时期特异性表达

为了探明Hox基因在龟鳖生长发育过程中的表达特征,我们搜集、分析了已发表的龟鳖转录组数据。分析结果表明Hox基因在中华鳖、锦龟和绿海龟孵化后/成体的肝脏、大脑和心脏几乎不表达,但在胚胎时期的背甲嵴(发育成背甲)、顶端外胚层嵴(Apical ectodermal ridge,发育成四肢)、性腺及孵化后的肾脏高表达(图6)。Hox基因B8、B9和C10分别在中华鳖背甲嵴(14—16期),性腺(9d、16d,雌雄)持续高表达。此外,C4、C5、C8、D1、D3和D8在中华鳖背甲嵴和性腺也具有高表达量。HoxD8在中华鳖雌、雄性腺(16d)都特异高表达。Hox基因在锦龟相同胚胎时期(16期)背甲嵴(A7、A9、A10、B7、C8—C11和D9)和顶端外胚层嵴(A9—A13、C10、C13、D12和D13)存在组织特异性高表达。值得一提的是,锦龟顶端外胚层嵴特异高表达的基因都位于Hox基因簇的5'端(A9—A13、C10、C13、D12和D13)。然而,在绿海龟胚胎早期(产卵后0、36h)首先出现表达的Hox基因(A1、A2、B1和D1)均位于Hox基因簇3'端。上述结果一定程度上佐证了Hox基因簇的表达具有时间共线性。此外,Hox基因(A7、A9、B6—B9、C4、C10和D8—D10)在绿海龟孵化后肾脏中高表达。综上,龟鳖Hox基因的表达具有组织、时期特异性。

3 讨论

龟鳖目因其独特的甲壳(背甲、腹甲)在四足动物中独树一帜。龟、鳖在背甲形成过程中,骨骼的发育存在差异。龟科(“硬壳”)物种的肋骨随着背甲而发育,最终轴向延伸至背甲的边缘,然而鳖科(“软壳”)的肋骨会提前骨化,不再沿着背甲延伸,大多数物种的背甲边缘都没有骨板^[37]。此外,鳖科的腹甲存在特异的新腹骨板(Neoplastron)结构^[38],而

表1 Hox基因的序列插入缺失统计

Tab. 1 Statistics of insertions and deletions in Hox genes

基因 Gene	类型 Type	坐标 Position (bp)	长度 Length (bp)
HoxA1	龟科插入	81—82, 247—248	3, 6
HoxA5	鳖科插入	468—469	9
HoxB1	鳖科缺失	121—129	9
HoxB2	鳖科缺失	756—761, 792—836	6, 45
HoxB3	鳖科缺失	105—107, 951—959	3, 9
HoxB4	鳖科缺失; 龟科插入	291—293, 364—366; 315—316	3, 3; 3
HoxB5	鳖科插入	229—230	3
HoxB9	鳖科缺失	516—524	9
HoxB13	龟科插入; 鳖科插入; 鳖科缺失	41—42; 81—82; 100—111, 160—174, 230—244	9; 3; 12, 15, 15
HoxC9	鳖科缺失	298—309	12
HoxC10	鳖科缺失	574—576	3
HoxC12	鳖科缺失	556—558	3
HoxC13	鳖科插入	152—153, 906(S端)	9, 3
HoxD1	龟科缺失	521—532	12
HoxD8	鳖科缺失	161—163	3
HoxD9	龟科缺失; 鳖科插入	324—329; 394—395	6; 9

内腹骨板是龟科特有的^[39]。与龟科相比,鳖科背甲的骨化程度比较低^[40],其甲壳边缘骨骼、肋骨远端部分及大部分塑形骨骼的数量较少^[9]。这些特异的表型暗示着鳖科在体轴形态发育、骨骼发育相关通路基因序列及表达调控上不同于龟科。然而,以往的研究并未对龟、鳖类群表型差异的遗传分子基础进行系统的研究。

本研究发现龟鳖目Hox基因簇并不存在基因数目、基因元件长度的特异性变化。这说明在龟鳖目长达2.65亿年的演化中^[36],Hox基因簇是高度保守的(图1)。与有鳞目相比,龟鳖目、鳄鱼和鸟类的最近共同祖先在演化过程中,特异性地丢失了HoxC3基因^[6]。除了蛇类特异性丢失HoxD12基因

外,Hox基因在龟鳖目和整个羊膜类其他类群都是高度保守的(图1)。两次全基因组复制事件使得高等脊椎动物获得了4个Hox基因簇^[8,41],某些旁系同源基因可能由于表型重塑(例如蛇类HoxD12^[6,42])或功能冗余导致基因丢失^[43]。因此哺乳类、鸟类和鳄鱼HoxC3基因的丢失很可能是由于与其他基因存在功能上的冗余。

非重复序列的含量差异导致鳖科HoxB9—HoxB13基因间区长度比龟科短10 kb。此外,该基因间区的转座元件类型和含量也存在龟鳖类群差异。转座子可以调控基因的转录表达^[44],影响异染色质的表观遗传修饰,也可能会破坏Hox基因的调控元件^[45]。基因间区的非重复序列主要包括启动子、终止子

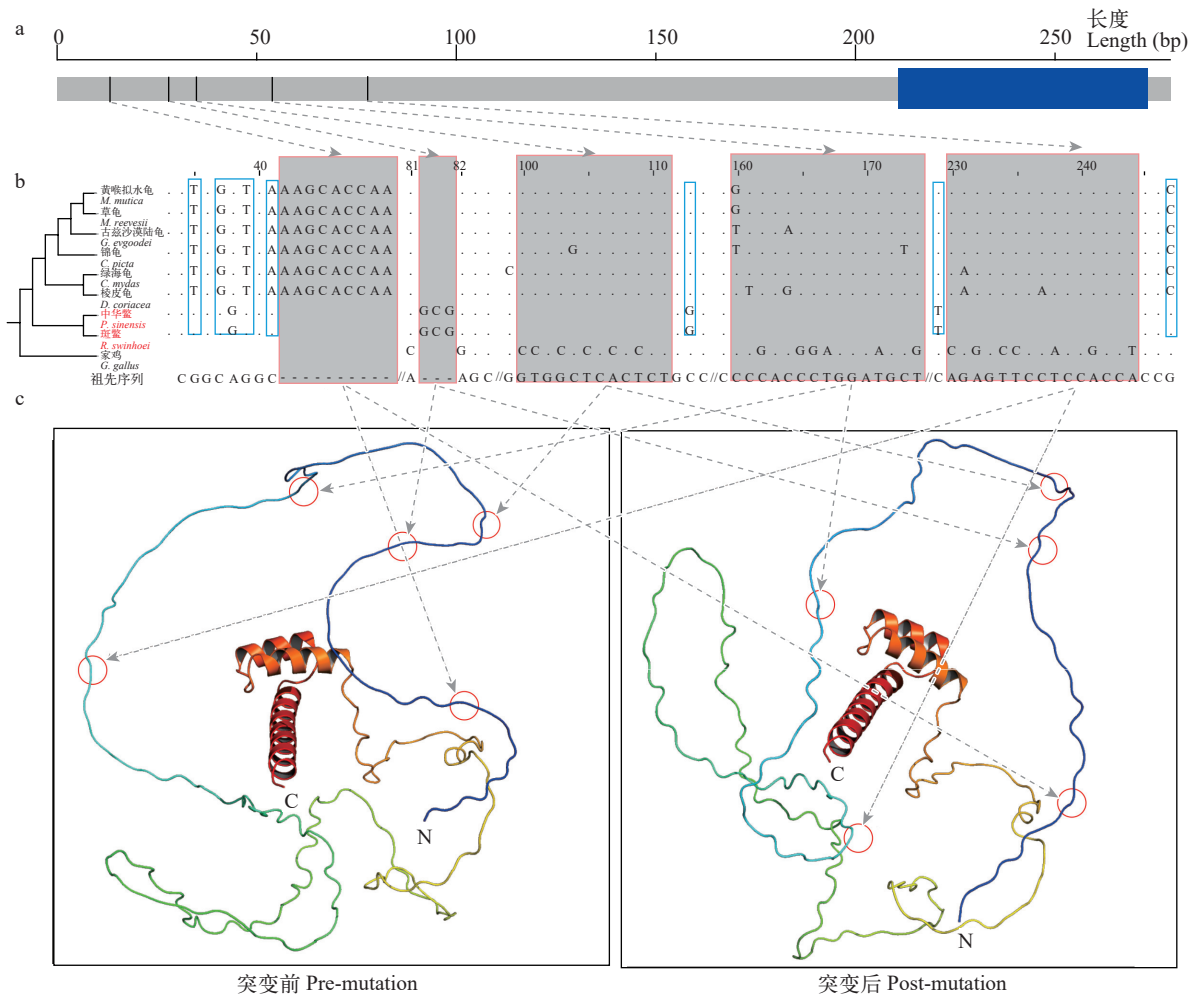


图4 龟鳖目HoxB13基因序列及3D结构预测的比对

Fig. 4 Comparisons of HoxB13 gene sequences and its predicted 3D structure in turtles

a. Hox基因编码氨基酸长度,黑色线条标记发生序列插入缺失,蓝色矩形为结构域。b. 物种间的亲缘关系及Hox基因序列比对。红色字体标记鳖科物种,以祖先序列为参照,灰色高亮序列插入或缺失,蓝色矩形标记类群特异的碱基替换。c. Hox基因的预测空间结构。红色圆圈为发生突变的位置

a. The upper part shows the length of amino acid encoded by the Hox gene. The black bars and blue rectangle indicate the positions of indels (Insertion and deletion) and the structural domain, respectively; b. The evolutionary relationships of studied species and their Hox gene sequence alignments. The soft-shell turtles are marked in red fonts. Compared with ancestral sequence, indels are highlighted in gray, substitutions are marked by blue rectangles; c. Predicted 3D structure of Hox gene. The red circles indicate the positions of indels

及增强子等调控元件。蛇类在进化过程中, 极性活性区调控序列(Zone of Polarizing Activity Regulatory Sequence)缺失一段17 bp序列会使该增强子失去功能, 从而导致四肢退化^[46]。值得一提的, *HoxB9*和*HoxB13*在相同胚胎发育时期(16期)的背甲嵴, 相比锦龟, 中华鳖分别呈现特异的高表达和不表达(图6), 这暗示着*HoxB9*和*HoxB13*在龟、鳖背甲发育过程中的表达调控是不同的。*HoxB9*—*HoxB13*基因间区不同的转座子类型和序列变化很可能调控着*Hox*基因表达, 和鳖科、龟科特异的表型特征相关联。为了进一步厘清该基因间区的调控作用, 还需要针对胚胎时期的多组学分析和相关功能验证。

大量*Hox*基因发生了鳖科或龟科特异的序列替换、插入和丢失, 这说明*Hox*基因在鳖科和龟科物种之间具有明显的序列差异。然而, 这些序列突变并不在其结构域上, 也未导致其蛋白质空间结构发生变化, 不同于基因序列突变会对基因的功能乃至表型造成重要影响的定论。*Hox4*—*Hox9*控制着动物胸部区域骨骼的形态结构^[47], 发生序列突变的16个基因中, 有7个基因在这个区间之内, 这暗示

着鳖科和龟科胸部骨骼的形态结构差异可能和*Hox*基因序列的特异性变化相关。

*Hox*基因在原始鱼类和陆生脊椎动物进化速率是不同的^[48]。不同物种之间*Hox*基因簇数目及每个簇所包含的基因数目也不尽相同^[6, 49], 推测*Hox*基因可能在不同的物种中经历了不同选择压力。*HoxA1*、*HoxD8*和*HoxD9*在鳖科祖先发生了快速进化, 而*HoxA9*、*HoxB5*、*HoxC6*和*HoxD12*在鳖科共同祖先受到了正选择。*HoxC6*在中华鳖的背甲真皮层中表达, 这可能与其腹侧肋骨缺失相关^[50]。与小鼠、家鸡胚胎的*Hox*基因表达相比, *HoxA5*、*HoxA7*、*HoxB5*、*HoxC6*和*HoxC8*在中华鳖的表达位置前移, 这可能与中华鳖特有身体结构有关^[50, 51]。我们推测在鳖科祖先发生正选择的*HoxB5*和*HoxC6*很可能和鳖科特异的形态相关联。

绿海龟孵化后肾脏和中华鳖胚胎性腺中检测到多数*Hox*基因的表达(图6), 说明*Hox*基因在龟鳖目肾脏功能和性腺发育过程中起着重要作用。有研究报道*HoxA13*和*HoxD11*在太平洋丽龟泌尿生殖嵴呈现产雌/雄温度的差异表达^[52]。上述两个基因

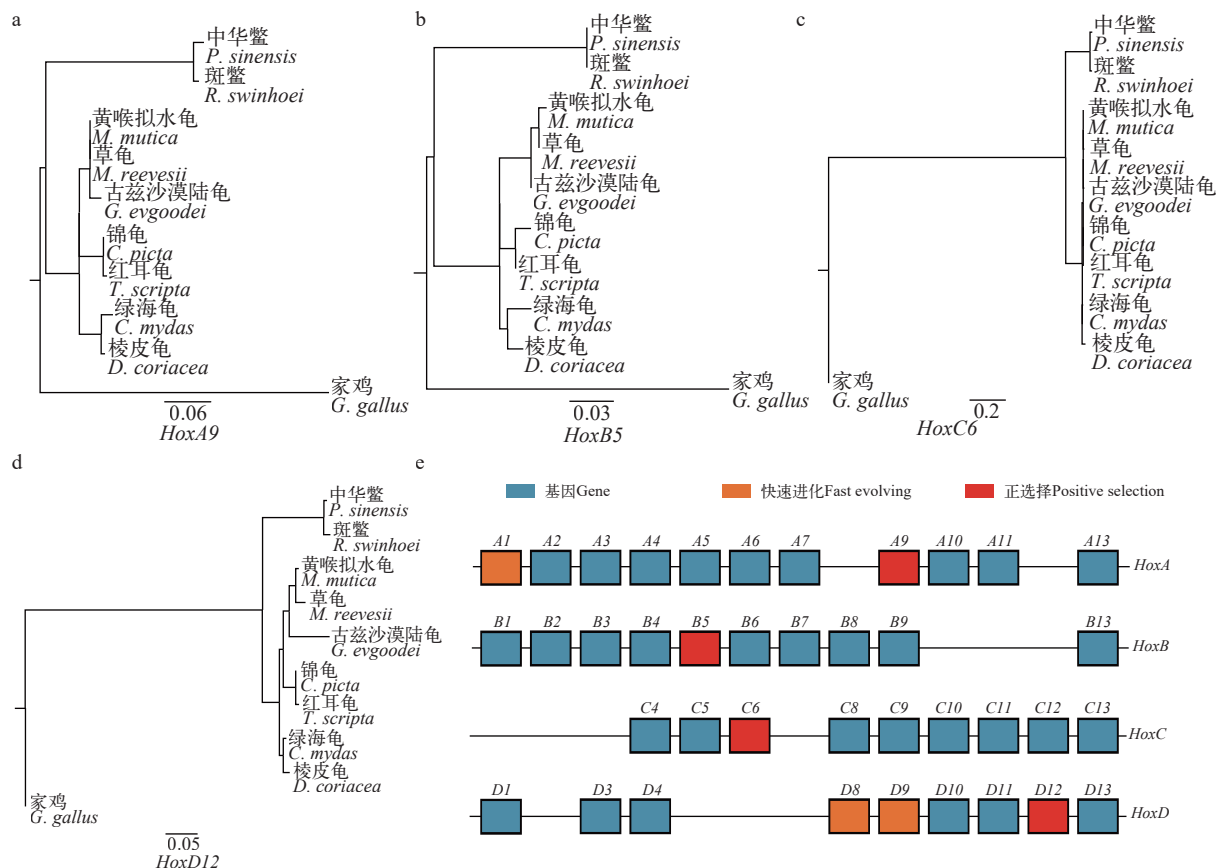
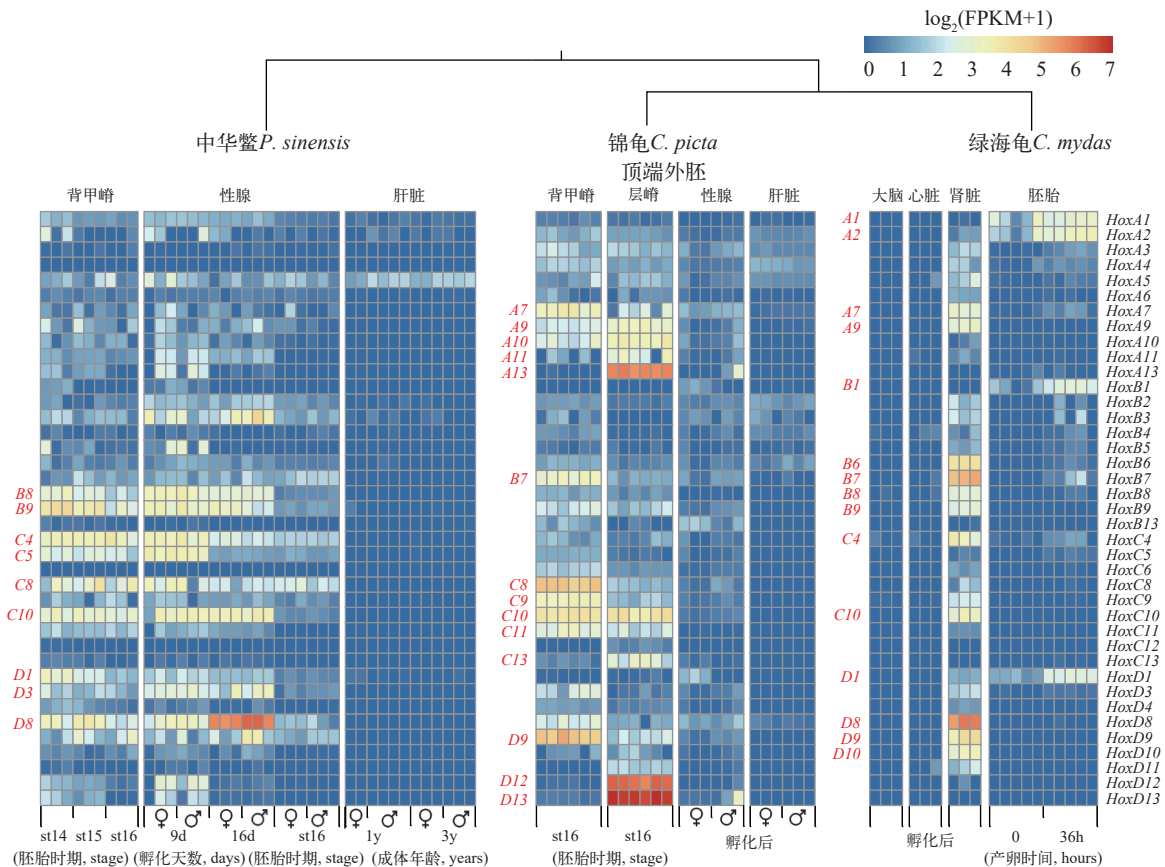


图5 龟鳖Hox基因的进化

Fig. 5 Evolution of *Hox* genes in turtles

a—d. *Hox*基因的同义置换速率(dS, free-ratio model); e. 鳖科祖先*Hox*基因的进化

a—d. Synonymous substitution rates(dS) of *Hox* genes(free-ratio model); e. evolution of *Hox* genes in the ancestor of soft-shell turtles

图6 *Hox*基因龟鳖组织中的表达Fig. 6 Expression patterns of *Hox* genes in turtle tissues

在中华鳖胚胎雌雄性腺(9d、16d和16期)几乎不表达,这可能与取样时期及性别决定机制(中华鳖为遗传决定型)相关。超过25个*Hox*基因在哺乳动物的肾脏表达^[53],*Hox11*基因对于肾脏发育至关重要^[54]。*Hox*基因在绿海龟孵化后肾脏也广泛表达(图6)暗示着*Hox*基因对肾脏发育作用的保守性。

顶端外胚层嵴位于胚胎肢芽的顶端,是调节肢体生长和发育的信号中心^[55,56]。*HoxA11*和*HoxA13*能够调控肢芽近端-远端(Proximodistal)的生长模式^[56]。鉴于与顶端外胚层嵴的相似性^[57],背甲嵴被认为是调控背甲发育的信号源。在龟胚胎中进行的背甲嵴消融实验导致肋骨在轴向域中扇形方向发生了改变,从而佐证了背甲嵴调控背甲形成的假设^[58]。转录组分析结果表明多数*Hox*基因在中华鳖、锦龟胚胎时期的背甲嵴、顶端外胚层嵴高表达(图6),可见*Hox*基因对于龟鳖的四肢、背甲发育的重要性。中华鳖和锦龟在相同胚胎时期(16期)的背甲嵴中招募了不同的*Hox*基因表达,这暗示着龟鳖物种间背甲嵴发育存在时空差异。相比背甲嵴,*HoxA13*、*HoxD12*和*HoxD13*在锦龟相同胚胎时期(16期)的顶端外胚层嵴显著高表达(图6),这和之前对二者转录组比较分析相一致^[59]。厘清龟鳖背甲

嵴形成、发育机制需要更系统地取样(更早时期)和多维度的比较分析。

综上所述,本研究从经典*Hox*基因视角,结合多组学手段,较为全面地分析了龟、鳖两个类群的序列、基因表达和进化特征。本研究推测龟鳖类群表型差异的遗传基础可能来源于*HoxB9—HoxB13*基因间区存在特异调控元件调控基因的表达,以及鳖科祖先受到正选择的*HoxB5*和*HoxC6*基因。本研究为后续龟鳖目*Hox*基因在不同胚胎时期的多组学(包括单细胞多组学)及表达调控分析提供了靶标,为进一步厘清龟鳖演化创新(背甲和腹甲等)提供了参考。

参考文献:

- [1] Mallo M, Alonso C R. The regulation of Hox gene expression during animal development [J]. *Development*, 2013, **140**(19): 3951-3963.
- [2] Li R, Ge H, Dai Y, et al. Genomewide analysis of homeobox gene family in apple (*Malus domestica* Borkh.) and their response to abiotic stress [J]. *Journal of Genetics*, 2019, **98**(1): 1-11.
- [3] Kappen C. The homeodomain: an ancient evolutionary motif in animals and plants [J]. *Computers & Chemistry*,

- 2000, **24**(1): 95-103.
- [4] Afzal Z, Krumlauf R. Transcriptional regulation and implications for controlling *Hox* gene expression [J]. *Journal of Developmental Biology*, 2022, **10**(1): 4.
 - [5] Derelle R, Lopez P, Le Guyader H, *et al.* Homeodomain proteins belong to the ancestral molecular toolkit of eukaryotes [J]. *Evolution & Development*, 2007, **9**(3): 212-219.
 - [6] Liang D, Wu R, Geng J, *et al.* A general scenario of *Hox*-gene inventory variation among major sarcopterygian lineages [J]. *BMC Evolutionary Biology*, 2011, **11**(1): 25.
 - [7] Holland L Z, Albalat R, Azumi K, *et al.* The amphioxus genome illuminates vertebrate origins and cephalochordate biology [J]. *Genome Research*, 2008, **18**(7): 1100-1111.
 - [8] Duboule D. The rise and fall of *Hox* gene clusters [J]. *Development*, 2007, **134**(14): 2549-2560.
 - [9] Alexander T, Nolte C, Krumlauf R. *Hox* genes and segmentation of the hindbrain and axial skeleton [J]. *Annual Review of Cell and Developmental Biology*, 2009(25): 431-456.
 - [10] Christ B, Huang R, Scaal M. Amniote somite derivatives [J]. *Developmental Dynamics*, 2007, **236**(9): 2382-2396.
 - [11] Shah N, Sukumar S. The *Hox* genes and their roles in oncogenesis [J]. *Nature Reviews Cancer*, 2010, **10**(5): 361-371.
 - [12] Carroll S B. Chance and necessity: the evolution of morphological complexity and diversity [J]. *Nature*, 2001, **409**(6823): 1102-1109.
 - [13] Richardson M K, Crooijmans R P, Groenen M A. Sequencing and genomic annotation of the chicken (*Gallus gallus*) *Hox* clusters, and mapping of evolutionarily conserved regions [J]. *Cytogenetic and Genome Research*, 2007, **117**(1/2/3/4): 110-119.
 - [14] Martín-Del-Campo R, Sifuentes-Romero I, García-Gasca A. *Hox* genes in reptile development, epigenetic regulation, and teratogenesis [J]. *Cytogenetic and Genome Research*, 2019, **157**(1/2): 34-45.
 - [15] Hoegg S, Meyer A. *Hox* clusters as models for vertebrate genome evolution [J]. *Trends in Genetics*, 2005, **21**(8): 421-424.
 - [16] Amemiya C T, Powers T P, Prohaska S J, *et al.* Complete HOX cluster characterization of the coelacanth provides further evidence for slow evolution of its genome [J]. *Proceedings of the National Academy of Sciences*, 2010, **107**(8): 3622-3627.
 - [17] Di-Poï N, Montoya-Burgos J I, Miller H, *et al.* Changes in *Hox* genes' structure and function during the evolution of the squamate body plan [J]. *Nature*, 2010, **464**(7285): 99-103.
 - [18] Li C, Wu X C, Rieppel O, *et al.* An ancestral turtle from the late triassic of southwestern China [J]. *Nature*, 2008, **456**(7221): 497-501.
 - [19] Böhmer C, Werneburg I. Deep time perspective on turtle neck evolution: chasing the *Hox* code by vertebral morphology [J]. *Scientific Reports*, 2017(7): 8939.
 - [20] Altschul S F, Gish W, Miller W, *et al.* Basic local alignment search tool [J]. *Journal of Molecular Biology*, 1990, **215**(3): 403-410.
 - [21] Yu X J, Zheng H K, Wang J, *et al.* Detecting lineage-specific adaptive evolution of brain-expressed genes in human using rhesus macaque as outgroup [J]. *Genomics*, 2006, **88**(6): 745-751.
 - [22] Birney E, Clamp M, Durbin R. GeneWise and genomewise [J]. *Genome Research*, 2004, **14**(5): 988-995.
 - [23] Jurka J. Repbase update: a database and an electronic journal of repetitive elements [J]. *Trends in Genetics*, 2000, **16**(9): 418-420.
 - [24] Ashkenazy H, Penn O, Doron-Faigenboim A, *et al.* FastML: a web server for probabilistic reconstruction of ancestral sequences [J]. *Nucleic Acids Research*, 2012, **40**(W1): W580-W584.
 - [25] Clamp M, Cuff J, Searle S M, *et al.* The Jalview Java alignment editor [J]. *Bioinformatics*, 2004, **20**(3): 426-427.
 - [26] Edgar R C. MUSCLE: multiple sequence alignment with high accuracy and high throughput [J]. *Nucleic acids research*, 2004, **32**(5): 1792-1797.
 - [27] Baek M, DiMaio F, Anishchenko I, *et al.* Accurate prediction of protein structures and interactions using a three-track neural network [J]. *Science*, 2021, **373**(6557): 871-876.
 - [28] Thomson R C, Spinks P Q, Shaffer H B. A global phylogeny of turtles reveals a burst of climate-associated diversification on continental margins [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2021, **118**(7): e2012215118.
 - [29] Yang Z. PAML 4: phylogenetic analysis by maximum likelihood [J]. *Molecular Biology and Evolution*, 2007, **24**(8): 1586-1591.
 - [30] Chen S, Zhou Y, Chen Y, *et al.* Fastp: an ultra-fast all-in-one FASTQ preprocessor [J]. *Bioinformatics*, 2018, **34**(17): i884-i890.
 - [31] Pertea M, Kim D, Pertea G M, *et al.* Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown [J]. *Nature Protocols*, 2016, **11**(9): 1650-1667.
 - [32] Kim D, Langmead B, Salzberg S L. HISAT: a fast spliced aligner with low memory requirements [J]. *Nature Methods*, 2015, **12**(4): 357-360.
 - [33] Pertea M, Pertea G M, Antonescu C M, *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads [J]. *Nature Biotechnology*, 2015, **33**(3): 290-295.
 - [34] Soneson C, Love M I, Robinson M D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences [J]. *F1000Research*, 2015(4): 1521.
 - [35] Robinson M D, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data [J]. *Genome Biology*, 2010, **11**(3): 1-9.

- [36] Zheng Y, Wiens J J. Combining phylogenomic and supermatrix approaches, and a time-calibrated phylogeny for squamate reptiles (lizards and snakes) based on 52 genes and 4162 species [J]. *Molecular Phylogenetics and Evolution*, 2016(94): 537-547.
- [37] Rice R, Riccio P, Gilbert S F, *et al.* Emerging from the rib: resolving the turtle controversies [J]. *Journal of Experimental Zoology Part B, Molecular and Developmental Evolution*, 2015, **324**(3): 208-220.
- [38] Xu S K, Liu Z X. The anatomy of the skeletal of Testudo horsfieldi gray [J]. *Journal of Xinjiang University* (Natural Science Edition), 1993(1): 93-102. [许设科, 刘志霄. 四爪陆龟骨骼系统的解剖 [J]. 新疆大学学报(自然科学版), 1993(1): 93-102.]
- [39] Xie M M, Hong X Y, Li W, *et al.* Skeletal system analysis of assign giant soft-shelled turtles and comparison with chinese soft-shell turtle [J]. *Acta Hydrobiologica Sinica*, 2022, **46**(5): 654-663. [谢敏敏, 洪孝友, 李伟, 等. 鼋骨骼系统与中华鳖比较 [J]. 水生生物学报, 2022, **46**(5): 654-663.]
- [40] Scheyer T M, Martin Sander P, Joyce W G, *et al.* A plywood structure in the shell of fossil and living soft-shelled turtles (Trionychidae) and its evolutionary implications [J]. *Organisms Diversity & Evolution*, 2007, **7**(2): 136-144.
- [41] Matsunami M, Sumiyama K, Saitou N. Evolution of conserved non-coding sequences within the vertebrate Hox clusters through the two-round whole genome duplications revealed by phylogenetic footprinting analysis [J]. *Journal of Molecular Evolution*, 2010, **71**(5): 427-436.
- [42] Mannaert A, Roelants K, Bossuyt F, *et al.* A PCR survey for posterior Hox genes in amphibians [J]. *Molecular Phylogenetics and Evolution*, 2006, **38**(2): 449-458.
- [43] Soshnikova N, Dewaele R, Janvier P, *et al.* Duplications of hox gene clusters and the emergence of vertebrates [J]. *Developmental Biology*, 2013, **378**(2): 194-199.
- [44] Feschotte C, Pritham E J. DNA transposons and the evolution of eukaryotic genomes [J]. *Annual Review of Genetics*, 2007(41): 331-368.
- [45] Feiner N. Accumulation of transposable elements in Hox gene clusters during adaptive radiation of Anolis lizards [J]. *Proceedings of the Royal Society B: Biological Sciences*, 2016, **283**(1840): 20161555.
- [46] Kvon E Z, Kamneva O K, Melo U S, *et al.* Progressive loss of function in a limb enhancer during snake evolution [J]. *Cell*, 2016, **167**(3): 633-642.
- [47] Mallo M, Wellik D M, Deschamps J. Hox genes and regional patterning of the vertebrate body plan [J]. *Developmental Biology*, 2010, **344**(1): 7-15.
- [48] Málaga-Trillo E, Meyer A. Genome duplications and accelerated evolution of Hox genes and cluster architecture in teleost fishes [J]. *Integrative and Comparative Biology*, 2001, **41**(3): 676-686.
- [49] Zou S M, Jiang X Y, He Z Z, *et al.* Hox gene clusters in blunt snout bream, *Megalobrama amblycephala* and comparison with those of zebrafish, Fugu and medaka genomes [J]. *Gene*, 2007, **400**(1/2): 60-70.
- [50] Ohya Y K, Kuraku S, Kuratani S. Hox code in embryos of Chinese soft-shelled turtle *Pelodiscus sinensis* correlates with the evolutionary innovation in the turtle [J]. *Journal of Experimental Zoology Part B, Molecular and Developmental Evolution*, 2005, **304**(2): 107-118.
- [51] Böhmer C, Rauhut O W M, Wörheide G. New insights into the vertebral Hox code of archosaurs [J]. *Evolution & Development*, 2015, **17**(5): 258-269.
- [52] Sifuentes-Romero I, Merchant-Larios H, García-Gasca A. Hox gene expression in the embryonic genital system of the sea turtle *Lepidochelys olivacea* (Eschscholtz, 1829), a species with temperature-dependent sex determination [J]. *Gene Expression Patterns*, 2010, **10**(6): 290-298.
- [53] Patterson L T, Potter S S. Atlas of Hox gene expression in the developing kidney [J]. *Developmental Dynamics*, 2004, **229**(4): 771-779.
- [54] Wellik D M, Hawkes P J, Capecchi M R. Hox11 paralogous genes are essential for metanephric kidney induction [J]. *Genes & Development*, 2002, **16**(11): 1423-1432.
- [55] Mallick A. The function of apical ectodermal ridge in the formation of limb [J]. *Bangladesh Journal of Scientific Research*, 2014, **26**(1/2): 95-99.
- [56] Lin G H, Zhang L. Apical ectodermal ridge regulates three principal axes of the developing limb [J]. *Journal of Zhejiang University-Science B*, 2020, **21**(10): 757-766.
- [57] Yang J, Xia Y, Li S, *et al.* Gene regulation during carapacial ridge development of *Mauremys reevesii*: the development of carapacial ridge, ribs and scutes [J]. *Genes*, 2022, **13**(9): 1676.
- [58] Nagashima H, Kuraku S, Uchida K, *et al.* On the carapacial ridge in turtle embryos: its developmental origin, function and the chelonian body plan [J]. *Development*, 2007, **134**(12): 2219-2226.
- [59] Cordero G A. Transcriptomic similarities and differences between the limb bud AER and unique carapacial ridge of turtle embryos [J]. *Evolution & Development*, 2020, **22**(5): 370-383.

EVOLUTIONARY ANALYSES AND SEQUENCE FEATURE OF *HOX* GENES IN TURTLES

HUANG Bao-You^{1,2}, LOU Ling-Yuan^{1,2}, FAN Jia-Wei^{1,2}, SUN Wei^{2,3}, QIAN Guo-Ying^{2,3},
GE Chu-Tian^{2,3} and WANG Zong-Ji^{2,3}

(1. College of Fisheries and Life Sciences, Shanghai Ocean University, Shanghai 201306, China; 2. Institute of Animal Sex and Development, Zhejiang Wanli University, Ningbo 315100, China; 3. College of Biological and Environmental Sciences, Zhejiang Wanli University, Ningbo 315100, China)

Abstract: *Hox* genes, as a family of transcription factors, play a central role in the specification of regional identities along the anterior-posterior body axis, skeletal development, and limb morphogenesis in vertebrates. Previous analyses of *Hox* genes revealed an unexpected diversity in gene number, genomic organization, and expression patterns in different vertebrate groups. However, the *Hox* genes have not been systematically analyzed in terms of sequence and evolutionary features in turtles. In this study, we performed *Hox* gene identification, sequence characterization and evolutionary analyses in turtles with the publicly available chromosomal-level genome sequences and RNA-seq data, and found that the *Hox* gene clusters in turtles are highly conserved. Turtles possess the same *Hox* inventory (39 genes, lost *HoxC3*) as birds, crocodiles, and placental mammals. The *HoxB9-HoxB13* intergenic region in soft-shell turtles is about 10 kb shorter than that in hard-shell turtles, which resulted from the deletion of the non-repetitive sequences. There are hard-shell/soft-shell turtle specific nucleotide substitutions, insertions, and deletions located in the coding regions of tens of *Hox* genes. However, these sequence mutations are not in functional domains, nor do they lead to changes in protein spatial structure. The *Hox* genes related to thoracic skeletal development have been undergone rapid evolution and positive selection in the ancestor of soft-shell turtles. The *Hox* gene expression is tissue- and stage- specific, and mainly expressed in the apical ectoderm ridge, carapace ridge and gonad during the embryonic stages in turtles. We speculate that the phenotypic differences between soft-shell and hard-shell turtles might be related to non-coding regulatory elements located in the *HoxB9-HoxB13* intergenic region, and the *HoxB5* and *HoxC6*, which are positively selected in the ancestors of soft-shell turtles. This study provides a target for the analyses of multi-omics, gene expression and regulatory mechanism of *Hox* genes across different embryonic stages in turtles, also provides a reference for further clarifying the evolutionary innovation in turtles.

Key words: *Hox* genes; Testudines; Evolution; Soft-shell turtles