



非编码保守元件在物种适应性演化中的作用

赵金金^{1,2}, 金红², 胡义波^{2*}

1. 河南师范大学生命科学学院, 新乡 453000

2. 中国科学院动物研究所, 中国科学院动物生态与保护生物学重点实验室, 北京 100101

* 联系人, E-mail: ybhu@ioz.ac.cn

收稿日期: 2024-08-14; 接受日期: 2024-11-15; 网络版发表日期: 2025-04-03

摘要 在生物体基因组中除编码序列外, 还存在大量非编码保守元件(conserved non-coding elements). 以往研究物种适应性演化的正选择和氨基酸趋同分析都聚焦在蛋白编码基因的编码区, 而非编码区的研究较少. 研究发现, 非编码保守元件在基因组分布广泛, 可作为增强子等顺式调控元件发挥作用, 其序列特征变化和快速演化在物种适应性演化和趋同演化中发挥重要作用, 影响着较多物种的表型和生理等适应性性状. 本文总结了非编码保守元件在物种适应性演化中的作用机制、主要特点及鉴定方法, 并进一步介绍了其在趋同演化中的作用机制及模式, 最后提出了非编码保守元件研究展望. 该综述可为从非编码区视角深入研究物种适应性演化和趋同演化的分子机制提供科学指导.

关键词 适应性演化, 趋同演化, 非编码保守元件, 比较基因组学

非编码保守元件(conserved non-coding elements, CNEs)是指在基因组中不直接编码蛋白质, 但在演化过程中发挥重要功能的非编码序列^[1]. 基因组中包含数千个非编码序列, 这些序列在物种内以及物种之间具有高度的保守性, 甚至超过了蛋白质编码外显子中发现的保守性^[2]. 在人类基因组中, 由于选择压力的作用, 大概有5%的碱基序列成为保守序列, 其中蛋白编码序列仅占1.5%, 而剩余的3.5%则是由CNEs构成^[3]. 随着多组学技术的不断发展, 较多研究发现CNEs通过直接或间接调控基因表达, 进而影响物种的适应性演化, 在物种演化中发挥着不可或缺的重要作用.

本文旨在总结CNEs在物种适应性演化和趋同演化中的研究进展, 从CNEs调控转录因子、CNEs序列

特征变化、CNEs的快速演化等方面详细介绍了非编码保守元件的作用机制, 这有助于深入理解物种多样化和表型多样化的分子基础, 为从非编码区视角深入研究物种适应性演化和趋同演化的分子机制提供科学指导.

1 非编码保守元件简介

1.1 CNEs在基因组中的分布

研究发现, CNEs遍布人类基因组的所有染色体, 分布十分密集但并不均匀^[4]. 尽管各个物种的CNEs序列同源性差异较大, 但它们在基因组中的分布格局却十分相似^[5]. 大多数CNEs倾向聚集在编码发育调节基

引用格式: 赵金金, 金红, 胡义波. 非编码保守元件在物种适应性演化中的作用. 中国科学: 生命科学, 2025, 55: 1409–1420

Zhao J J, Jin H, Hu Y B. Role of conserved non-coding elements in adaptive evolution of species (in Chinese). Sci Sin Vitae, 2025, 55: 1409–1420, doi: 10.1360/SSV-2024-0247

因的周围,尤其是转录因子和信号基因,并且在一定程度上与脊椎动物和无脊椎动物的细胞信号相关^[6]。这些CNEs是基因调控网络的关键构成元件,通过充当转录因子结合位点调节基因表达,具有高度保守性。同时,CNEs在基因荒漠区(gene desert)广泛存在,其数量高出其他区域约4倍,在荒漠区的边缘有许多与生物发育密切相关的基因^[7]。还有许多CNEs在基因组中嵌套在旁系同源基因区域中,被称为重复的保守非编码元件(duplicated conserved non-coding elements, dCNEs)。dCNEs与旁系同源基因之间有着紧密且复杂的相互关系,它们共同参与了转录调控、发育进程以及环境适应性应答等生物过程。在基因数量加倍增长的同时,一些重要的顺式调控元件也会被保留在复制产生的旁系同源区域内^[8]。

1.2 超保守元件

非编码序列中同样包含着具有关键调控作用的保守元件——超保守元件(ultra-conserved elements, UCEs)。UCEs是基因组中极为特殊的序列,它们在漫长的进化历程中也保持了高度的保守性。人、小鼠以及大鼠的直系同源区域发现了481个长度超过200 bp且序列完全一致的超保守元件。这些UCEs在哺乳动物、鸟类以及鱼类中也几乎保守,上述UCEs在鸡和狗的基因组中发现存在一系列相似度约达95%的同源区段^[9]。UCEs在基因组中的位置分布多样,11.3% UCEs位于外显子区域,31.2% UCEs位于内含子区域,37.8% UCEs散布于基因间区,8.9% UCEs居于UTR(untranslated region)区域,10.8%位于非编码RNA区域^[10]。CNEs与UCEs部分重叠,相互补充,两者都经常聚集在转录调控或与发育相关的基因附近,在发育调控中起作用;在不同发育阶段调控的关键基因周围,大部分CNEs与UCEs均展现出增强子的功能特性,CNEs虽然广泛分布在基因组中,但并不编码蛋白质;而部分UCEs存在于蛋白质编码区域,起着可变剪切调控的作用。本文着重关注于CNEs的生物学功能及其意义。

1.3 CNEs的生物学意义

CNEs虽不编码蛋白质,但具有重要的生物学功能。它们调控基因表达模式,维护基因组的整体稳定性,同时也推动物种适应性演化(图1)。如CNEs通过调

控基因的转录水平,影响生物体内各种生物学过程^[11]。同时,它们的保守性有助于维持基因组结构和功能的稳定性,为生物体适应环境提供了基因组水平的遗传基础^[12]。因此,通过研究CNEs,可以深入了解物种间的演化关系,推断演化过程中的变化和适应,为揭示演化发生的遗传学机制提供新的线索。

2 CNEs在物种适应性演化中发挥的作用

适应是生物通过形态结构和生理机能的改变,与所处的环境相适应并进行生存和繁衍的过程^[13],而适应性演化(adaptive evolution)是物种生存和延续的重要策略之一。由于生物的适应性演化,形成了生物的复杂性和多样性。CNEs既可以作为调控元件影响物种的适应性演化,也可以通过序列的变异或快速演化在物种适应性演化过程中发挥作用。CNEs是推动生物多样性形成的重要驱动力,从鱼类、两栖爬行类到哺乳动物都有研究案例(表1)^[14-23]。

2.1 CNEs作为顺式调控元件影响物种适应性演化

在早期研究中,普遍使用基因组中的蛋白编码序列来探索物种的适应性演化,对非编码序列关注较少^[24]。但越来越多的研究发现,非编码序列在基因组中发挥至关重要的作用,为生物适应性演化提供物质和结构基础,且在基因组非编码区域中探寻保守元件,正逐渐成为寻找功能元件的新途径^[25]。研究发现,CNEs作为顺式调控元件调节基因表达,影响表型的变化^[26]。例如,探索地下哺乳动物眼睛退化机制,鉴定出*Tdrd7*基因上游的CNEs与这一眼睛退化机制紧密相关。研究利用CRISPR-Cas9技术将小鼠*Tdrd7*基因上游调控元件替换为盲鼯鼠(*Talpa caeca*)序列,实验结果显示,小鼠晶状体内*Tdrd7*基因表达上调,证明该调控元件在调控与晶状体发育相关的*Tdrd7*基因表达上具有重要作用^[27]。通过对果蝇顺式调控元件的深入研究,揭示了果蝇翅膀上特定色素沉着模式的形成机制。以圆点果蝇(*Drosophila guttifera*)为例,其翅膀上呈现出独特的斑点图案,是由yellow蛋白精准定位所形成的色素沉着区域。其中,*wingless*基因在斑点中心表达,并通过异位表达实验证明该基因具有诱导色素沉着的作用。进一步研究发现,*wingless*基因的顺式调控区域在圆点果蝇中经历了独特的进化,形成了特有的表达

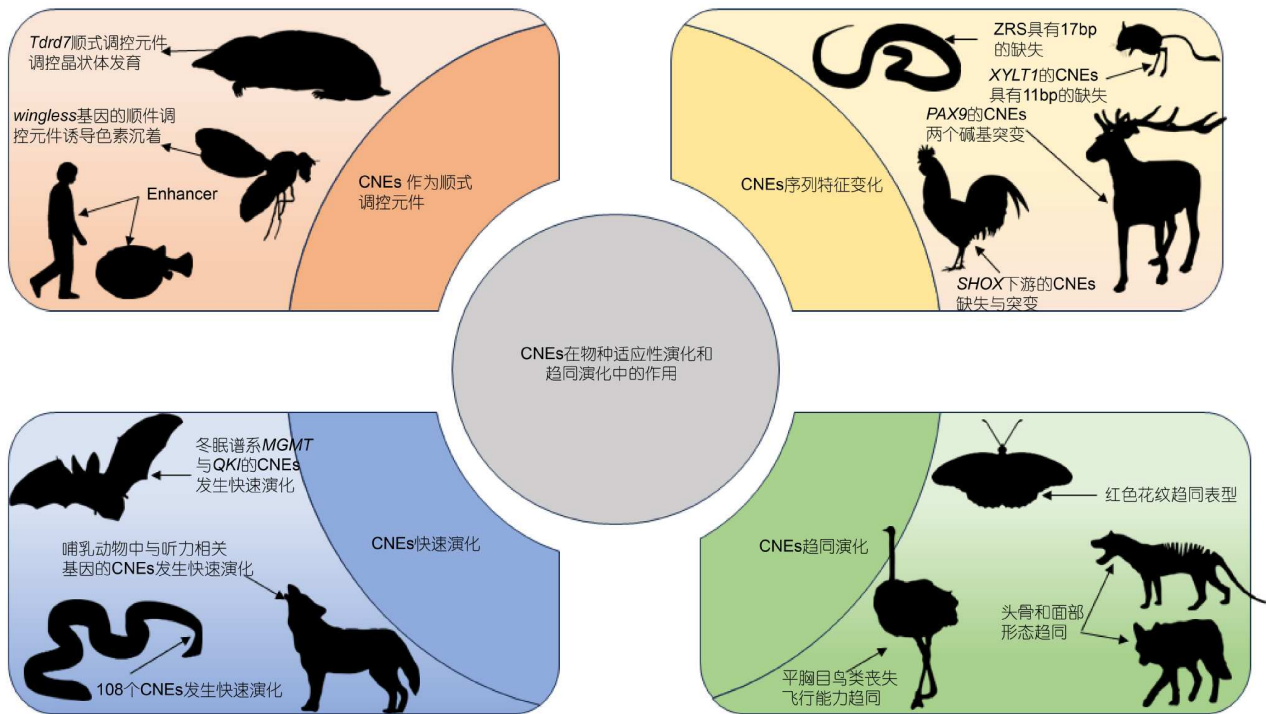


图 1 非编码保守元件在物种适应性演化和趋同演化中的作用

Figure 1 The role of conserved non-coding elements during adaptive evolution and convergent evolution of species

模式和增强子活性^[28]。

此外, CNEs和UCEs在物种适应性演化中发挥增强子作用^[29]。例如, 研究者在人与小鼠、人与鸡、人与河豚的非编码序列中发现CNEs可作为发育调控的增强子, 并且这些CNEs在斑马鱼中证实了其作为转录增强子的功能^[30]。在鱼类研究中也发现了CNEs发挥增强子的调控作用, 如花鲈鱼类中多个物种常栖息于硫化氢丰富的有毒泉域, 那些生活在硫化物富集环境下的种群被称作硫化群体, 对7个独立谱系硫化与非硫化群体比较分析, 发现*G6PD*与*RDH13*基因在增强子区域的变异对缓解氧化应激至关重要, 可增强这类物种对极端环境的适应能力^[31]。非编码序列中的UCEs同样可发挥增强子作用。Pennacchio等人^[32]对一个古老的保守增强子*Dc2*进行深入研究, 结果表明424 bp的*Dc2*保守元件对于大脑组织增强子活性十分重要, 实验验证结果表明, 45%的元件在胚胎发育阶段具有组织特异性增强子的作用。此外, 该研究完成了120个哺乳动物的全基因组比对, 首次对胎盘哺乳动物中UCEs的变异性进行量化, 估计有4.7%~15.6%的序列是可变的。研究鉴定了仅在胎盘哺乳动物中保守的增强子序

列, 并发现这些增强子与胎盘相关基因表达显著相关, 表明其中一些增强子可能参与了胎盘哺乳动物中特有胎盘的演化^[33]。

2.2 CNEs序列特征变化影响物种适应性演化

CNEs展现出高度保守性, 也表现出关键的调控作用, 但其仍可能发生缺失、插入、点突变以及序列重复等变异^[3]。这些变异并不总伴随显著的表型变异出现, 但能对基因的表达产生一定影响并可能成为许多性状变化的基础。一方面, CNEs的缺失能够引发关键的调控作用甚至引起表型的改变。研究显示在多种哺乳动物中, 数百种CNEs各自独立经历了缺失的过程^[34]。例如, 蛇类的四肢缺失是因为与其他物种相比, 只有在蛇类ZRS(zone of polarizing activity regulatory sequence)中具有17 bp的缺失, 这一缺失导致蛇体四肢的发育受到抑制, 此现象与调控四肢发育相关基因中CNEs的部分或全部独立丢失紧密相关^[35]。在研究海牛(*Trichechus*)和海豚(Delphinidae)时发现, *EGR2*基因附近CNEs发生了特殊的独立丢失现象, 推测这一变化是驱动两者前肢形态显著差异的关键因素^[36]。在淡水

表 1 非编码保守元件影响物种适应性演化的案例

Table 1 Examples of CNEs affecting the adaptive evolution of species

CNEs序列特征	物种	主要发现
缺失、插入	无肢蜥蜴	对来自不同谱系的无肢蜥蜴进行分析, 发现这些物种中CNEs的插入与缺失数量各不相同. 这些CNEs与肢体起始和生长所必需的基因(如 <i>Tbx4</i> , <i>Fgf10</i> 和 <i>Gli3</i>)紧密相关, 并且与肢体发育早期阶段的调节信号有所联系, 共同调节着表型的进化过程 ^[14] .
插入	青鳉鱼	对青鳉鱼(<i>Oryzias latipes</i>)的不同尾鳍进行比对研究后, 发现基因 <i>zic1</i> 与 <i>zic4</i> 的增强子中存在特异性插入, 这些插入在髓间充质的背侧表现出强烈的表达, 进而诱导野生型胚胎呈现出不对称的尾鳍发育特征 ^[15] .
缺失	两栖动物	对两栖动物CNEs的研究揭示, 位于Hox复合体3'-末端的特定元件ZNF503和ZNF703附近的CNEs呈现出增强子活性, 调控相关基因表达, 从而对神经与肢体发育产生重要影响 ^[16] .
插入、缺失	鲤科	对11种硬骨鱼进行比较分析后, 发现在两个特定物种中, <i>atg7</i> 基因附近CNE存在一个7 bp的特异性缺失, 而 <i>usp45</i> 基因附近的另一个CNE则具有7 bp的特异性插入. 这些CNEs均展现出更高的启动子活性, 可能是它们适应海水环境的关键机制 ^[17] .
缺失、突变	鸡	使用比较基因组学分析, <i>SHOX</i> 基因的非编码序列在鸡(<i>Gallus gallus</i>)的肢体发育中具有增强子作用, 多个位于 <i>SHOX</i> 基因下游的顺式调控元件参与了调控肢体发育过程中 <i>SHOX</i> 基因的表达, <i>SHOX</i> 下游的CNEs缺失与突变会导致软骨化障碍 ^[18] .
缺失	鱼类	通过对36种鱼类进行比较分析发现, 在野生鳍减少的物种中, 保守序列反复缺失的基因显著富集在与附肢发育相关通路. 进一步发现, <i>PelA</i> 增强子在远缘物种中发生了独立丢失, 这一现象可能影响了鳍的形态 ^[19] .
突变、插入	反刍动物	与有角类反刍动物多腔胃结构相关的 <i>SIM2</i> , <i>PAX9</i> 转录因子以及 <i>KCNK5</i> , <i>DENND2C</i> 两个基因, 紧邻它们上游/下游各10 kb区域内具有有角类动物特有的CNEs, 且与 <i>PAX9</i> 基因相关的CNEs具有有角类动物特有的两个碱基突变, 推测与瓣胃上皮和瘤胃不同有关 ^[20] .
缺失	跳鼠	<i>XYLT1</i> 基因相关的CNEs上存在一个11 bp的缺失, 这一缺失可能对 <i>XYLT1</i> 的表达、软骨形成以及跳鼠独特的肢干发育具有调控作用; 同时 <i>ROR2</i> 基因附近的CNEs缺失可能在跳鼠(<i>Allactaga sibirica</i>)适应干旱环境的过程中发挥关键作用 ^[21] .
缺失	长喙鱼	剑鱼(<i>Xiphias gladius</i>)和旗鱼(<i>Istiophorus platypterus</i>)缺失的CNEs中, 一个CNE(UCE-1582)位于 <i>chd9</i> 上游172 kb处, <i>chd9</i> 是 <i>runx2</i> 的潜在调节因子, 推测与喙再生有关; <i>Hoxd3a</i> 上游2 kb的CNE和 <i>Hoxd4a</i> 上游3 kb的CNE在剑鱼和旗鱼中缺失, 可能与鳍的变化有关 ^[22] .
缺失	驯鹿	驯鹿(<i>Rangifer tarandus</i>)77个CNEs发生缺失, 159个CNEs存在差异. 富集分析表明, 它们与神经系统发育有关, 其中一个位于 <i>FOXB1</i> 上游约58 kb处, 该CNE可能是 <i>FOXB1</i> 的增强子. <i>FOXB1</i> 相关增强子和其他CNE的缺失可能导致驯鹿神经系统响应速度减弱, 促进驯化 ^[23] .

环境中, 面对食物稀缺以及钙利用率下降的生存挑战, 刺鱼体内的*PITX1*基因作为发育调节关键因子, 研究表明与*PITX1*相关CNEs片段的缺失, 促使鱼体演化出骨盆缩小的适应性特征, 且降低了对钙的需求, 成为它们应对不利环境条件的生存策略. 这种变化在多个淡水刺鱼种群中均有发现^[37]. 在鸟类基因组中也识别出具有谱系特异性的CNEs, 这些CNEs多数包含转录因子结合位点并参与调控发育过程, 并且发现了在鸟类飞羽形成中CNEs对其靶基因*SIM1*的调控机制, CNEs的丢失直接导致*SIM1*基因表达受阻, 结果说明特定CNEs对靶基因*SIM1*的调控至关重要^[38]. 此外, 人们推测硬骨鱼类的表型多样性可能与祖先脊椎动物中CNEs的大规模丢失和替代有关^[39]. 通过解析非洲肺鱼(*Protopterus annectens*)的基因组, 发现*Foxp2*附近的CNEs影响该基因功能, 导致肺鱼呼吸能力增强, 且在*Hoxa11*上游发现四足动物特有的CNEs, 推测与五趾的起源有关; 两个四足动物*Hoxc10*上游特异性的CNEs被

认为是候选增强子, 可能与陆地运动有关, 这对于脊椎动物从水生到陆生的演化具有重要意义^[40]. 另一方面, CNEs序列的突变与重复与多种疾病的发生相关. 例如, *IRF6*基因增强子区域发生突变导致孟德尔唇裂障碍^[41]; 在具有发育增强子功能的区域中发现杂合点突变, 与皮埃尔·罗班综合征有关^[42]; *Six3*远程影响*Shh*前脑增强子, 与前脑无裂畸形发生相关^[43]; 在*RET*基因内含子中的保守增强子序列中发现非编码变异, 诱导先天性巨结肠发生^[44]. 此外, CNEs序列变异可能诱发多种疾病, 包括多动症^[45]、孤独症^[46]以及不宁腿综合征^[47]等. 而CNEs的重复现象也可能导致A2型短指^[48]、短指症无指甲以及并指和颅缝早闭等问题^[49].

2.3 CNEs的快速演化影响物种适应性演化

在某些物种中, CNEs表现出快速演化的信号, 这些信号可能与物种对特定环境的适应性演化有关. 例如, 海蛇的运动适应性演化研究中, 在平颏海蛇(*Hy-*

drosophila curtus)中发现了108个CNEs经历快速演化, 其相关基因的功能主要富集于骨骼、肌肉及神经系统发育相关的信号通路上, 而这些信号通路可能与平颚海蛇体型特征的变化有关, 如与身体横向压缩、尾巴扁平以及陆地到海洋运动方式的转变密切相关^[50]. 在哺乳动物的演化研究中, 由于哺乳动物拥有独特的听力系统, 研究显示, *LOXHD1*, *TMCI*, *OTOF*, *CDH23*和*PCDH15*等基因对哺乳动物内耳发育的正常功能至关重要, 显示出强烈正选择信号, 且其非编码区域经历了快速演化, 作为发育增强子发挥作用, 推测非编码区域的快速演化影响哺乳动物听力系统的适应性演化^[51]. 异温动物(heterothermy)分布在几个哺乳动物谱系中, 冬眠是异温动物适应环境温度变化的一种策略. 在动物冬眠适应性演化研究中, 对9种冬眠脊椎动物进行CNEs的快速演化研究, 在其中5个冬眠谱系中发现*MGMT*与*QKI*基因的CNEs发生快速演化, 这些基因参与DNA修复与缺血耐受性的形成, 可能与冬眠适应相关^[52]. 对灵长类基因组计划的大规模组学数据进行分析, 鉴定出支系特异快速演化的调控元件(conserved non-coding elements that exhibited rapid evolution in specific lineage, RECNEs), 即在特定支系的基因组中演化速率显著升高的非编码保守元件, 并鉴定注释出灵长类8个主要祖先节点上的RECNEs, 发现每个祖先节点上, 许多与大脑发育基因相关的调控序列都经历了快速演化, 尤其是类人猿下目祖先(Simiiiformes)富集了最多的RECNEs, 主要作用于神经发生、神经元发育与迁移等功能通路. RECNEs可能驱动灵长类神经细胞中某些基因的高表达, 影响神经元功能^[53].

3 CNEs在物种趋同演化中发挥的作用

趋同演化(convergent evolution)是指远缘物种在相似环境或共同选择压力下, 各自演化出形态或功能上相似的特征^[54]. 例如, 鸟类与蝙蝠各自独立地演化出相似的翅膀结构而使它们具备飞行的能力, 鲸类与鱼类在水中高效游动而独立演化出流线型的身体, 鼯鼠和金毛鼯适应地下挖掘而独立演化出适合掘土的附肢等^[55].

在动物基因组中, 仅1%~10%的区域为编码基因所占据, 而超过90%的部分则是由富含基因表达调控元件的非编码序列所构成. 大量研究表明, CNEs在物

种对环境的适应过程中具有重要作用^[56]. 趋同演化的研究目前主要集中于全基因组层面的筛选检测, 已经识别出大量编码基因的氨基酸趋同替换, 但非编码区变异对趋同演化的相对贡献仍缺乏深入了解. 通过对趋同演化中CNEs的序列变异、调控元件的功能特性以及它们与基因表达之间的关系开展深入研究, 可以了解CNEs在物种趋同演化中发挥的作用.

在非编码区水平研究远缘物种趋同演化的分子机制方面已经开展了较多研究. 例如, 在平胸总目(Ratitae)鸟类丧失飞行能力的演化研究中^[57], 通过分析包括已灭绝的恐鸟在内的11种古代鸟类基因组, 在蛋白编码区域未检测到与飞行能力丧失相关的氨基酸变化. 但在非编码区, 研究人员发现了2355个CNEs, 这些元件在平胸类动物中展现了快速演化的特征. 部分经历快速演化的CNEs邻近的基因, 如*TBX5*, *DACH1*, *PAX9*及*IRX*家族基因等, 均紧密关联肢体发育功能. 其中, 元件mCE1916069与哺乳动物内识别的肢体增强子重叠, 定位于*TEAD1*基因的内含子中, 在早期鸡前肢研究中证实具有增强子活性. 该元件在具有飞行能力的鸟类中表现出保守性, 但在平胸目类群中则显示出趋同快速演化的趋势, 进一步功能验证表明, 该特定元件在平胸目鸟类并不具有启动子活性. 基于这一发现, 推测该调控元件的快速演化与平胸目鸟类丧失飞行能力的趋同演化特征紧密相关^[57]. 在果蝇类群的研究中, Frankel等人^[58]发现两个在大约4000万年前分化的果蝇物种, 它们的幼虫相较于各自近缘物种都表现出表皮毛状体结构缺失的共同特征, 为探究此现象的遗传机制, 研究者对调控毛状体发育的关键基因*Shavenbaby (Svb)*的上游启动子区域进行了深入的结构与功能剖析. 研究显示, 所有被研究的果蝇物种的*Svb*基因上游表达调控区均包含7个增强子结构, 但两个目标物种在第三个增强子的功能上却表现出趋同的缺失. 这种缺失阻碍了它们有效启动下游基因的表达, 进而产生了与近缘物种不同的缺失毛状体的表型. 在蝙蝠听觉系统的演化研究中, 研究者检查了上千种动物类群中的CNEs, 并测试了与听觉系统相关基因附近的CNEs的演化速率变化, 结果显示, 与内耳发育有关的*SHH*和*TSHZ*在所有6种蝙蝠物种的替代率都显著更高, 反映不同基因在蝙蝠表现不同的功能, 也演化出不同的回声定位模式^[59]. 在地下哺乳动物演化研究中发现, 与眼睛发育紧密相关的基因*Pax6*的增强子区域显示

出演化速率加速的迹象. 这种快速演化表现为突变的累积, 这些突变影响了*Pax6*基因的表达调控和晶状体的正常发育. 进一步分析显示, 这种调控区层面的变化与地下哺乳动物因选择压力减少而导致的视觉功能衰退的趋同现象有着密切的关联^[60]. 越来越多的研究发现, CNEs趋同在物种趋同演化中发挥着重要作用(表2)^[6,61~64].

4 CNEs鉴定方法

比较基因组学可以对非编码保守序列进行鉴别和功能注释. CNEs通常作为特定组织及特定时间点附近基因的顺式调节元件发挥作用^[65]. 许多调控基因似乎总以相似的方式发挥功能, 这些功能和作用主要通过调控元件完成, 而这些调控基因相互作用共同参与不同物种的发育过程, 从而促进了物种多样性形成^[66]. 因此识别非编码保守元件的方法和工具, 对于深入挖掘基因组调控网络的结构和功能至关重要.

4.1 CNEs鉴定方法

目前现有的鉴定策略大致可以分为四类, 窗口聚类法(window clustering)、概率建模法(probabilistic modelling)、系统发育足迹法(phylogenetic footprinting)和判别建模法(discriminative modelling)^[67]. 窗口

聚类法主要是在一个序列窗口内高密度结合位点显著聚集; 概率模型法寻找目标序列区域, 这些区域构成了模式集群的统计模型, 而不是构成“背景DNA”模型; 系统发育足迹法是搜索进化关系相对密切物种之间保守结合位点的高密度区域; 判别建模法旨在识别一组可以最大化调节区域和非调节区域之间差异的信号.

在识别CNEs的方法中, MSCAN, MCAST, Cis-PlusFinder与EEL属于窗口聚类法; ClusterBuster, StubbMS, MorphMS, CisModule, PhylCRM, Regulatory Potential和EMMA属于概率建模法; MultiModule, Cis-PlusFinder, EEL, Regulatory Potential, PhastCons, PhyloP以及PhyloACC属于系统发育足迹法; 判别建模法包含Regulatory Potential和HexDiff等方法(表3)^[68~83].

4.2 CNEs鉴定方法的异同

不同的CNEs鉴定方法针对的基因组对象和采用的模型方法存在异同. MSCAN和MCAST都只针对单个基因组扫描. CisPlusFinder基于多物种保守性识别PLUSs的高密度区域. ClusterBuster, StubbMS, MorphMS, CisModule和MultiModule都以HMM为识别基础, 将CRM序列建模为由一组结合位点的组合产生; ClusterBuster和CisModule是基于单个基因组, 而StubbMS, MorphMS和MultiModule是以一对或多个同源基因组为基础; StubbMS与MorphMS的区别在于第

表2 非编码保守元件在物种趋同演化中的作用研究实例
Table 2 Examples of the role of CNEs in convergent evolution of species

物种	主要发现
蠕虫、苍蝇、脊椎动物	一组调节发育的核心基因与三个动物类群(蠕虫、苍蝇和脊椎动物)的CNEs相关. 这些CNEs揭示了在不同动物群体中, 经历了平行进化, 并共同调控着一组关键的发育调控基因 ^[6] .
红带袖蝶	在袖蝶属(<i>Heliconius</i>)中, 拟态物种各自独立地演化出了几乎相同的翅膀颜色图案, 这些图案由一组基因所控制. 研究发现, 基因 <i>optix</i> 在“红色”颜色模式中发挥着关键的调控作用, 而 <i>optix</i> 附近的顺式调控元件(CRE)是诱导红色表型的必要条件. 在红带袖蝶(<i>Heliconius melpomene</i>)中的 <i>optix</i> CRE序列在与袖蝶属关系更远的蝴蝶物种中完全缺失, 这进一步证明了 <i>optix</i> 及其CRE在形态趋同过程中, 对翅膀颜色模式基因的独特调整合作用. 该基因是构成红带袖蝶独特趋同表型的基础 ^[61] .
鲸、大象、人和裸鼹鼠	对62种哺乳动物进行全基因组扫描后, 发现了数百种与毛发发育相关的CNEs. 在这些CNEs附近, 具有快速演化特征的非编码区域周围的基因包括 <i>ELF3</i> , <i>FOXC1</i> , <i>KRT</i> 和 <i>KRTAP</i> 等, 它们均与毛发发育和生长紧密相关. 这一发现表明, 无毛性状在进化过程中经历了多次独立的趋同演化 ^[62] .
电鱼	研究聚焦于探讨南美和非洲弱电鱼肌肉中电压门控钠通道基因(<i>Scn4aa</i>)如何针对性地作用于脊椎动物的肌肉机制. 结果显示, 一个保守增强子CNE3在不同脊椎动物中驱动 <i>Scn4aa</i> 基因在肌肉中的表达. CNE3包含肌肉转录因子的结合位点, 与肌肉发育密切相关. 进一步发现, 在电鱼中, CNE3序列的退化与肌肉中 <i>Scn4aa</i> 表达的缺失存在关联 ^[63] .
袋狼、狼	袋狼与狼虽然亲缘关系较远, 却各自独立演化出了极其相似的头骨和面部形态. 研究发现, 袋狼和狼在颅面发育过程中活跃的CREs存在广泛的趋同演化, 在袋狼和狼中发现了许多快速演化的区域, 这些区域位于关键形态转录因子和保守发育途径成员的附近, 并在TGFB和BMP信号通路的基因附近富集. 这些信号通路在颅面组织的发育中起着关键作用. 此外, 与颅面神经发育相关基因的增强子也表现出趋同选择, 并涉及上述信号通路. 在两个物种中, 快速演化的区域还在脑发育相关基因附近有富集 ^[64] .

表 3 识别非编码保守元件的部分方法

Table 3 Selected methods for recognizing CNEs

方法	原理
MSCAN	一种结合位点聚类检测方法, 能够有效确定所观察位点的显著性, 并同时校正序列中存在的局部组成偏差 ^[68] .
MCAST	基于基序隐马尔可夫模型(HMM), 可以直接比较不同宽度和特异性的基序得分, 以搜索与查询主题不重叠且具有统计显著性的簇 ^[69] .
CisPlusFinder	基于多物种保守性分析, 可以识别出完全局部无重叠序列(PLUSs)的高密度区域. 这些序列在近缘物种间保守. 通过假设PLUSs中包含局部过度表达的核心转录因子结合位点(TFBSs)基序, 找到顺式调控模块(<i>cis-regulatory module</i> , CRM)的结构和进化中的预期特征 ^[70] .
ClusterBuster	使用线性时间向前-向后算法(<i>forward-backward algorithm</i>)进行聚类预测时, 首先执行向前算法, 对每个从核苷酸开始到结束的子序列进行计算, 得出其对数似然分数. 接着, 针对这些子序列, 再使用向后算法来调整最佳起点, 从而完成聚类预测的过程 ^[71] .
StubbMS	HMM框架作为序列比对的起始阶段, 与多物种比对策略相结合. 在双物种比对中, Lagan算法专注于寻找不间断保守区域的最佳联合布局. 利用基于HMM构建的系统发育模型来评估保守区块内部结合位点的匹配质量. 对于未对齐的序列, 它们被视作一个单一物种进行计算, 并作为同源窗口的最终分数来确定 ^[72] .
MorphMS	在给定两个同源序列的情况下, 该程序会使用固定长度的窗口来扫描其中一个序列. 对于每个这样的窗口, 程序会通过使用标准的比对工具来计算它在另一个序列中对应同源窗口的边界 ^[73] .
CisModule	这是一个两层次的混合模型. 在第一层次, 序列被视为由纯背景序列与CRM混合而成; 在第二层次, CRM则被建模为主题与模块内背景序列的混合物. 该模型基于模块、TFBSs和基序模式的联合后验分布, 进行贝叶斯推理, 以检测模块、TFBSs和基序模式 ^[74] .
MultiModule	采用HMM来模拟各物种中TFBSs的协同定位, 借助多次比对技术将TFBSs与调控模块的位置紧密关联起来. 随后, 运用马尔可夫链蒙特卡罗法, 根据联合后验分布同步采样CRM及其内部包含的TFBSs ^[75] .
EEL	结合转录因子(TF)结合的保守性与生物物理原理, 可以确定能量最为显著的元素位置. 这一过程中会考虑TF与其各自结合位点的结合亲和力, 以及相邻结合位点之间的距离 ^[76] .
Regulatory Potential	通过统计模拟调节区域和背景序列中短比对模式的频率来识别调节区域. 用5个符号来描述两种物种的配对: A和T的配对、C和G的配对、过渡、翻转和间隙, 对这些符号的一组 <i>k</i> -mers进行分类. 该序列由马尔可夫链建模, 参数来自实验确认的调控区域和对齐的祖先散布重复序列 ^[77] .
PhyloP	通过测量每个核苷酸位点的进化加速度(即比预期更快的进化)或保守性(即比预期更慢的进化)来评估其保守程度. 基于多序列比对数据, 考虑了物种之间的进化关系, 从而可以评估每个核苷酸位点在不同物种之间的保守性 ^[78] .
PhyloACC	通过分析一组保守元素检测基因组中DNA替换率的转变模式, 进而识别出在特定物种中呈现加速进化的基因组元素. 通过评估每个基因组元素在进化树上的相对替换率状态(这些状态可以是中性的、保守的或加速的), 来揭示基因组的动态性 ^[79] .
PhastCons	利用多物种比对结果来鉴别CNEs, 基于一个系统发育隐马尔可夫模型(<i>phylo-HMM</i>), 通过最大似然法将 <i>phylo-HMM</i> 与数据拟合, 从而对不同物种间的序列比较模型进行校准, 最后基于该模型预测出保守元件, 并给出每个碱基的保守分数 ^[80] .
EMMA	依据进化模型, 考量TFBSs的变异性与得失, 运用统计对齐技术及HKY模型估算核苷酸替换率. 通过切换控制TFBSs和非TFBSs进化的模型来模拟结合位点的功能增益和损失 ^[81] .
HexDiff	通过使用六聚体频率来区分CRM和非CRM序列来预测调控系统中的新CRM ^[82] .
PhylCRM	量化了MONKEY在多个对比识别出的模块(<i>motif</i>)的聚类和保守性来寻找候选CRMs ^[83] .

一步对输入的同源序列进行比对时, StubbMS使用Lagan进行序列比对, MorphMS通过对所有可能的序列与潜在结合位点的匹配进行概率相加来对序列进行比对. CisModule和MultiModule与其他方法不同, 它们在一部中预测结合位点和CRM, CisModule将结合位点和一个CRM组成一个分层混合模型, 利用贝叶斯推理来同时预测CRM以及CRM中结合位点的位置. MultiModule也使用相同的模型, 但与CisModule不同的是, 它加入了来自比较基因组的信息. 与MorphMS类似的还有EEL, PhylCRM和EMMA, 它们旨在通过

实施基于结合位点的比对方法, 更好地利用多个基因组信息. EMMA采用与MorphMS类似的方法, 与MorphMS不同的是, EMMA增加了结合位点的增益和损失, 这使得计算成本随着转录因子的数量呈指数增长, 并限制了EMMA聚焦更多问题, 而不是全基因组的研究. CisModule, MultiModule和CisPlusFinder是不依赖于基序库(*motif library*)先验信息的三种方法, 采用了一种完全基于序列保守的方法作为独立校准. PhyloP, PhyloACC和PhastCons都是用于分析基因组数据的重要工具. PhyloP关注于评估特定核苷酸或核

苷酸类别的选择特征, 而PhastCons则用于识别保守元件. PhyloACC则侧重于检测基因组区域DNA替换率的转变模式, 并识别在特定物种中加速的基因组元件. 实际分析时往往采用多种方法结合来达到鉴定CNEs的目的.

5 非编码保守元件研究展望

据估计, 人类基因组中的非编码保守DNA数量高达6万个, 相比之下, 直接参与蛋白质编码的基因数量仅为3万~4万个, 两者在数量上几乎呈现出两倍的差异^[84]. 据此推测, 非编码保守序列蕴藏着众多尚未揭示的重要功能.

随着三代测序技术以及T2T(telomere-to-telomere)全基因组测序技术的发展, 将有更多完整准确的物种基因组序列被获得, 这将促进CNEs的识别与分析, 使得我们能够更加精确地解析CNEs在基因组中的分布、保守性及其重要功能^[85]. 例如, 研究者采用Nanopore ultralong以及PacBio HiFi测序技术, 构建了一个包含所有染色体(W染色体除外)的T2T完整序列的新的鸡参考基因组; 鉴定了10条具有不同序列和表观遗传特征的小染色体(称为点染色体), 包含原基因组缺失的6条小染色体. T2T基因组测序能够克服着丝粒或高重复区域的组装困难问题, 提高了组装染色体的连续性

和完整性, 可为深入探索基因组中的特殊序列以及结构提供有力支持^[86].

为了揭示CNEs在生物体中的具体作用及其机制, 功能验证将是至关重要、不可或缺的一环. 随着CRISPR-Cas9基因编辑技术的快速发展, 可以更加容易地对CNEs进行敲除、突变替换等操作, 进而深入研究对基因表达、基因功能以及生物表型的影响. 例如在棘背鱼中, *PelB*驱动发育中的后肢后部的基因表达, 利用CRISPR-Cas9技术删除小鼠内源性位点中一个9425 bp区域, 该区域几乎涵盖了研究的整个*PelB*区域, 导致更明显的后肢表型, 包括髌骨的完全缺失. *PelB*序列在鱼类至哺乳动物进化过程中高度保守. 在缺乏骨盆的野生棘背鱼种群中, 观察到有插入、缺失突变, 导致*PelB*的结构和功能改变, 表明这种古老增强子的变化有助于自然界中骨盆附属物的进化修饰^[87]. 利用CRISPR-Cas9技术, 在非洲爪蟾(*Xenopus laevis*)中生成了针对CE14的突变体, 这些突变体表现出眼睛发育异常, 包括眼睛形态异常和晶状体缺陷, 这些调控元件之间可能存在复杂的相互作用, 共同调控*MAB21L2*基因的表达和眼睛发育^[88].

随着基因组研究的快速发展和功能实验验证技术的成熟, 对CNEs的作用机制将更加深入, 未来非编码序列与蛋白编码序列的整合研究将是演化发育生物学研究的主要范式.

参考文献

- 1 Rao E, Weiss B, Fukami M, et al. Pseudoautosomal deletions encompassing a novel homeobox gene cause growth failure in idiopathic short stature and Turner syndrome. *Nat Genet*, 1997, 16: 54–63
- 2 Mattick J S, Makunin I V. Non-coding RNA. *Hum Mol Genet*, 2006, 15: R17–R29
- 3 Glazko G V, Koonin E V, Rogozin I B, et al. A significant fraction of conserved noncoding DNA in human and mouse consists of predicted matrix attachment regions. *Trends Genet*, 2003, 19: 119–124
- 4 Leypold N A, Speicher M R. Evolutionary conservation in noncoding genomic regions. *Trends Genet*, 2021, 37: 903–918
- 5 Polychronopoulos D, King J W D, Nash A J, et al. Conserved non-coding elements: developmental gene regulation meets genome organization. *Nucleic Acids Res*, 2017, 45: 12611–12624
- 6 Woolfe A, Goodson M, Goode D K, et al. Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol*, 2005, 3: e7
- 7 Sanges R, Hadzhiev Y, Gueroult-Bellone M, et al. Highly conserved elements discovered in vertebrates are present in non-syntenic loci of tunicates, act as enhancers and can be transcribed during development. *Nucleic Acids Res*, 2013, 41: 3600–3618
- 8 Farré D, Bellora N, Mularoni L, et al. Housekeeping genes tend to show reduced upstream sequence conservation. *Genome Biol*, 2007, 8: R140
- 9 Tsang W H, Shek K F, Lee T Y, et al. An evolutionarily conserved nested gene pair—*Mab21* and *Lrba/Nbea* in metazoan. *Genomics*, 2009, 94: 177–187
- 10 Chen W H. Identification and analysis of ultraconservative components in birds (in Chinese). Dissertation for Master's Degree. Yangling:

Northwest A&F University, 2019 [陈伟煌. 鸟类超保守元件鉴定与分析. 硕士学位论文. 杨凌: 西北农林科技大学, 2019]

- 11 O'Loughlin S M, Forster A J, Fuchs S, et al. Ultra-conserved sequences in the genomes of highly diverse *Anopheles* mosquitoes, with implications for malaria vector control. *G3*, 2021, 11: jkab086
- 12 Sandelin A, Bailey P, Bruce S, et al. Arrays of ultraconserved non-coding regions span the loci of key developmental genes in vertebrate genomes. *BMC Genomics*, 2004, 5: 1–9
- 13 Drake J A, Bird C, Nemesh J, et al. Conserved noncoding sequences are selectively constrained and not mutation cold spots. *Nat Genet*, 2006, 38: 223–227
- 14 Wang Z, Peng C, Wu W, et al. Developmental regulation of conserved non-coding element evolution provides insights into limb loss in squamates. *Sci China Life Sci*, 2023, 66: 2399–2414
- 15 Moriyama Y, Kawanishi T, Nakamura R, et al. The medaka *zic1/zic4* mutant provides molecular insights into teleost caudal fin evolution. *Curr Biol*, 2012, 22: 601–607
- 16 Beaster-Jones L. *Cis*-regulation and conserved non-coding elements in amphioxus. *Brief Funct Genomics*, 2012, 11: 118–130
- 17 Wang Y, Zhang X, Wang J, et al. Genomic insights into the seawater adaptation in Cyprinidae. *BMC Biol*, 2024, 22: 87
- 18 Sabherwal N, Bangs F, Roth R, et al. Long-range conserved non-coding *SHOX* sequences regulate expression in developing chicken limb and are associated with short stature phenotypes in human patients. *Hum Mol Genet*, 2007, 16: 210–222
- 19 Chen H I, Turakhia Y, Bejerano G, et al. Whole-genome comparisons identify repeated regulatory changes underlying convergent appendage evolution in diverse fish lineages. *Mol Biol Evol*, 2023, 40: msad188
- 20 Chen L, Qiu Q, Jiang Y, et al. Large-scale ruminant genome sequencing provides insights into their evolution and distinct traits. *Science*, 2019, 364: eaav6202
- 21 Chai S, Chong Y, Yin D, et al. Genomic insights into adaptation to bipedal saltation and desert-like habitats of jerboas. *Sci China Life Sci*, 2024, 67: 2003–2015
- 22 Wu B, Feng C, Zhu C, et al. The genomes of two billfishes provide insights into the evolution of endothermy in teleosts. *Mol Biol Evol*, 2021, 38: 2413–2427
- 23 Wu B, Ren Q, Yan X, et al. Resequencing of reindeer genomes provides clues to their docile habits. *Evol Lett*, 2024, 8: 494–504
- 24 Orr H A. The genetic theory of adaptation: A brief history. *Nat Rev Genet*, 2005, 6: 119–127
- 25 Polychronopoulos D, Athanasopoulou L, Almirantis Y. Fractality and entropic scaling in the chromosomal distribution of conserved noncoding elements in the human genome. *Gene*, 2016, 584: 148–160
- 26 Aparicio S, Morrison A, Gould A, et al. Detecting conserved regulatory elements with the model genome of the Japanese puffer fish, *Fugu rubripes*. *Proc Natl Acad Sci USA*, 1995, 92: 1684–1688
- 27 Roscito J G, Subramanian K, Naumann R, et al. Recapitulating evolutionary divergence in a single *cis*-regulatory element is sufficient to cause expression changes of the lens gene *Tdrd7*. *Mol Biol Evol*, 2021, 38: 380–392
- 28 Koshikawa S. Evolution of wing pigmentation in *Drosophila*: diversity, physiological regulation, and *cis*-regulatory evolution. *Dev Growth Differ*, 2020, 62: 269–278
- 29 Xu X Q, Cui T, Zhang H, et al. Computational methods to predict Enhancer-target Gene Pairs (in Chinese). *Sci Sin Vitae*, 2023, 53: 1370–1382 [徐晓强, 崔婷, 张涵, 等. 增强子靶标基因的预测方法研究. 中国科学: 生命科学, 2023, 53: 1370–1382]
- 30 Visel A, Bristow J, Pennacchio L A. Enhancer identification through comparative genomics. *Semin Cell Dev Biol*, 2007, 18: 140–152
- 31 Allende M L, Manzanares M, Tena J J, et al. Cracking the genome's second code: enhancer detection by combined phylogenetic footprinting and transgenic fish and frog embryos. *Methods*, 2006, 39: 212–219
- 32 Pennacchio L, Ahituv N, Moses A, et al. *In vivo* enhancer analysis of human conserved non-coding sequences. *Nature*, 2006, 444: 499–502
- 33 Poulin F, Nobrega M A, Plajzer-Frick I, et al. *In vivo* characterization of a vertebrate ultraconserved enhancer. *Genomics*, 2005, 85: 774–781
- 34 Hecker N, Hiller M. A genome alignment of 120 mammals highlights ultraconserved element variability and placenta-associated enhancers. *GigaScience*, 2020, 9: giz159
- 35 Kvon E Z, Kamneva O K, Melo U S, et al. Progressive loss of function in a limb enhancer during snake evolution. *Cell*, 2016, 167: 633–642.e11
- 36 Sun L, Rong X, Liu X, et al. Evolutionary genetics of flipper forelimb and hindlimb loss from limb development-related genes in cetaceans. *BMC Genomics*, 2022, 23: 797
- 37 Chan Y F, Marks M E, Jones F C, et al. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. *Science*,

- 2010, 327: 302–305
- 38 Wray G A. The evolutionary significance of *cis*-regulatory mutations. *Nat Rev Genet*, 2007, 8: 206–216
- 39 Iliopoulou E, Papadogiannis V, Tsigenopoulos C S, et al. Extensive loss and gain of conserved noncoding elements during early teleost evolution. *Genome Biol Evol*, 2024, 16: evae061
- 40 Wang K, Wang J, Zhu C, et al. African lungfish genome sheds light on the vertebrate water-to-land transition. *Cell*, 2021, 184: 1362–1376.e18
- 41 Rahimov F, Marazita M L, Visel A, et al. Disruption of an AP-2 α binding site in an *IRF6* enhancer is associated with cleft lip. *Nat Genet*, 2008, 40: 1341–1347
- 42 Benko S, Fantes J A, Amiel J, et al. Highly conserved non-coding elements on either side of *SOX9* associated with Pierre Robin sequence. *Nat Genet*, 2009, 41: 359–364
- 43 Jeong Y, Leskow F C, El-Jaick K, et al. Regulation of a remote *Shh* forebrain enhancer by the Six3 homeoprotein. *Nat Genet*, 2008, 40: 1348–1353
- 44 Emison E S, McCallion A S, Kashuk C S, et al. A common sex-dependent mutation in a *RET* enhancer underlies Hirschsprung disease risk. *Nature*, 2005, 434: 857–863
- 45 Martinez A F, Abe Y, Hong S, et al. An ultraconserved brain-specific enhancer within *ADGRL3* (*LPHN3*) underpins attention-deficit/hyperactivity disorder susceptibility. *Biol Psychiatry*, 2016, 80: 943–954
- 46 Doan R N, Bae B I, Cubelos B, et al. Mutations in human accelerated regions disrupt cognition and social behavior. *Cell*, 2016, 167: 341–354.e12
- 47 Spieler D, Kaffé M, Knauf F, et al. Restless Legs Syndrome-associated intronic common variant in *Meis1* alters enhancer function in the developing telencephalon. *Genome Res*, 2014, 24: 592–603
- 48 Dathe K, Kjaer K W, Brehm A, et al. Duplications involving a conserved regulatory element downstream of *BMP2* are associated with brachydactyly type A2. *Am J Hum Genet*, 2009, 84: 483–492
- 49 Kurth I, Klopocki E, Stricker S, et al. Duplications of noncoding elements 5' of *SOX9* are associated with brachydactyly-anonychia. *Nat Genet*, 2009, 41: 862–863
- 50 Peng C, Ren J L, Deng C, et al. The genome of Shaw's sea snake (*Hydrophis curtus*) reveals secondary adaptation to its marine environment. *Mol Biol Evol*, 2020, 37: 1744–1760
- 51 Trigila A P, Pisciotto F, Franchini L F. Hearing loss genes reveal patterns of adaptive evolution at the coding and non-coding levels in mammals. *BMC Biol*, 2021, 19: 1–24
- 52 Nakayama D, Makino T. Convergent accelerated evolution of mammal-specific conserved non-coding elements in hibernators. *Sci Rep*, 2024, 14: 11754
- 53 Kuderna L F K, Ulirsch J C, Rashid S, et al. Identification of constrained sequence elements across 239 primate genomes. *Nature*, 2024, 625: 735–742
- 54 Duan H X, Li Y. Research progress and perspectives on the evolution of mammal diet (in Chinese). *Sci Sin Vitae*, 2025, 55: 15–29 [段红祥, 李艳. 哺乳动物食性进化的研究状况与展望. 中国科学: 生命科学, 2024, 55: 15–29]
- 55 Lee J H, Hiller M. Genome-wide screens for molecular convergent evolution in mammals. In: Pontarotti P, ed. *Evolutionary Biology: Self/Nonself Evolution, Species and Complex Traits Evolution, Methods and Concepts*. Cham: Springer, 2017. 297–312
- 56 Lin Q, Fan S, Zhang Y, et al. The seahorse genome and the evolution of its specialized morphology. *Nature*, 2016, 540: 395–399
- 57 Sackton T B, Grayson P, Cloutier A, et al. Convergent regulatory evolution and loss of flight in paleognathous birds. *Science*, 2019, 364: 74–78
- 58 Frankel N, Wang S, Stern D L. Conserved regulatory architecture underlies parallel genetic changes and convergent phenotypic evolution. *Proc Natl Acad Sci USA*, 2012, 109: 20975–20979
- 59 Jebb D, Huang Z, Pippel M, et al. Six reference-quality genomes reveal evolution of bat adaptations. *Nature*, 2020, 583: 578–584
- 60 Partha R, Chauhan B K, Ferreira Z, et al. Subterranean mammals show convergent regression in ocular genes and enhancers, along with adaptation to tunneling. *eLife*, 2017, 6: e25884
- 61 Van Belleghem S M, Ruggieri A A, Concha C, et al. High level of novelty under the hood of convergent evolution. *Science*, 2023, 379: 1043–1049
- 62 Kowalczyk A, Chikina M, Clark N. Complementary evolution of coding and noncoding sequence underlies mammalian hairlessness. *eLife*, 2022, 11: e76911
- 63 LaPotin S, Swartz M E, Luecke D M, et al. Divergent *cis*-regulatory evolution underlies the convergent loss of sodium channel expression in

- electric fish. *Sci Adv*, 2022, 8: eabm2970
- 64 Feigin C Y, Newton A H, Pask A J. Widespread *cis*-regulatory convergence between the extinct Tasmanian tiger and gray wolf. *Genome Res*, 2019, 29: 1648–1658
- 65 Song H, Wang Q, Zhang Z, et al. Identification of clade-wide putative *cis*-regulatory elements from conserved non-coding sequences in Cucurbitaceae genomes. *Horticulture Res*, 2023, 10: uhad038
- 66 Erwin D H, Davidson E H. The last common bilaterian ancestor. *Development*, 2002, 129: 3021–3032
- 67 Su J, Teichmann S A, Down T A, et al. Assessing computational methods of *cis*-regulatory module prediction. *PLoS Comput Biol*, 2010, 6: e1001020
- 68 Alkema W B L, Johansson O, Lagergren J, et al. MSCAN: identification of functional clusters of transcription factor binding sites. *Nucleic Acids Res*, 2004, 32: W195–W198
- 69 Bailey T L, Noble W S. Searching for statistically significant regulatory modules. *Bioinformatics*, 2003, 19: ii16–ii25
- 70 Pierstorff N, Bergman C M, Wiehe T. Identifying *cis*-regulatory modules by combining comparative and compositional analysis of DNA. *Bioinformatics*, 2006, 22: 2858–2864
- 71 Frith M C, Li M C, Weng Z. Cluster-Buster: finding dense clusters of motifs in DNA sequences. *Nucleic Acids Res*, 2003, 31: 3666–3668
- 72 Sinha S, Schroeder M D, Unnerstall U, et al. Cross-species comparison significantly improves genome-wide prediction of *cis*-regulatory modules in *Drosophila*. *BMC Bioinf*, 2004, 5: 129
- 73 Sinha S, He X, Pilpel Y. MORPH: probabilistic alignment combined with hidden Markov models of *cis*-regulatory modules. *PLoS Comput Biol*, 2007, 3: e216
- 74 Zhou Q, Wong W H. CisModule: *De novo* discovery of *cis*-regulatory modules by hierarchical mixture modeling. *Proc Natl Acad Sci USA*, 2004, 101: 12114–12119
- 75 Zhou Q, Wong W H. Coupling hidden Markov models for the discovery of *cis*-regulatory modules in multiple species. *Ann Appl Stat*, 2007, 1: 36–65
- 76 Hallikas O, Palin K, Sinjushina N, et al. Genome-wide prediction of mammalian enhancers based on analysis of transcription-factor binding affinity. *Cell*, 2006, 124: 47–59
- 77 Kolbe D, Taylor J, Elnitski L, et al. Regulatory potential scores from genome-wide three-way alignments of human, mouse, and rat. *Genome Res*, 2004, 14: 700–707
- 78 Pollard K S, Hubisz M J, Rosenbloom K R, et al. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res*, 2010, 20: 110–121
- 79 Yan H, Hu Z, Thomas G W C, et al. PhyloAcc-GT: a Bayesian method for inferring patterns of substitution rate shifts on targeted lineages accounting for gene tree discordance. *Mol Biol Evol*, 2023, 40: msad195
- 80 Siepel A, Haussler D. Combining phylogenetic and hidden Markov models in biosequence analysis. *J Comput Biol*, 2003, 11: 413–428
- 81 He X, Ling X, Sinha S, et al. Alignment and prediction of *cis*-regulatory modules based on a probabilistic model of evolution. *PLoS Comput Biol*, 2009, 5: e1000299
- 82 Chan B Y, Kibler D. Using hexamers to predict *cis*-regulatory motifs in *Drosophila*. *BMC Bioinf*, 2005, 6: 1–9
- 83 Warner J B, Philippakis A A, Jaeger S A, et al. Systematic identification of mammalian regulatory motifs' target genes and functions. *Nat Methods*, 2008, 5: 347–353
- 84 Freimer J W, Shaked O, Naqvi S, et al. Systematic discovery and perturbation of regulatory genes in human T cells reveals the architecture of immune networks. *Nat Genet*, 2022, 54: 1133–1144
- 85 Wang Y J, Wang X P, Yu L. Research progresses in revealing complex traits of animals based on genomic structural variations (in Chinese). *Sci Sin Vitae*, 2024, 54: 283–299 [汪亚军, 王晓萍, 于黎. 基于基因组结构变异解析动物复杂性状的研究进展. 中国科学: 生命科学, 2024, 54: 283–299]
- 86 Taher L, Ovcharenko I. Variable locus length in the human genome leads to ascertainment bias in functional inference for non-coding elements. *Bioinformatics*, 2009, 25: 578–584
- 87 Huang Z, Xu Z, Bai H, et al. Evolutionary analysis of a complete chicken genome. *Proc Natl Acad Sci USA*, 2023, 120: e2216641120
- 88 Ceroni F, Cicekdal M B, Holt R, et al. Deletion upstream of *MAB21L2* highlights the importance of evolutionarily conserved non-coding sequences for eye development. *Nat Commun*, 2024, 15: 9245

Role of conserved non-coding elements in adaptive evolution of species

ZHAO JinJin^{1,2}, JIN Hong² & HU YiBo^{2*}

¹ College of Life Sciences, Henan Normal University, Xinxiang 453000, China

² Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China

* Corresponding author; E-mail: ybhu@ioz.ac.cn

In addition to coding sequences, a large number of conserved non-coding elements exist in the genome of an organism. Previous studies focused on the roles of positive selection and amino acid convergence in species adaptation based on coding regions of protein-coding genes, whereas non-coding regions have been less studied. Comparative genomics has revealed that conserved non-coding elements are widely distributed in the genome and may function as *cis*-regulatory elements such as enhancers. The changes of their sequence characteristics and rapid evolution play an important role during adaptive evolution and convergent evolution of species, affecting the adaptive phenotype and physiological traits of many species. This review summarizes the functional mechanisms, main features and identification methods of conserved non-coding elements during adaptive evolution of species, and further summarizes the functional mechanisms and patterns of conserved non-coding elements during convergent evolution. Finally, we discuss the prospect of conserved non-coding element research. This review will provide scientific guidance for the study of molecular mechanisms of adaptive evolution and convergent evolution from the perspective of non-coding regions.

adaptive evolution, convergent evolution, conserved non-coding elements, comparative genomics

doi: [10.1360/SSV-2024-0247](https://doi.org/10.1360/SSV-2024-0247)