

龟鳖目表皮分化复合体基因的比较基因组学鉴定与分析

陈孟饶 李芳 陈其然 黄保友 孙伟 钱国英 葛楚天 王宗吉

COMPARATIVE GENOMIC IDENTIFICATION AND ANALYSES OF EPIDERMAL DIFFERENTIATION COMPLEX GENES IN TURTLES

CHEN Meng-Rao, LI Fang, CHEN Qi-Ran, HUANG Bao-You, SUN Wei, QIAN Guo-Ying, GE Chu-Tian, WANG Zong-Ji

在线阅读 View online: <https://doi.org/10.7541/2025.2024.0450>

您可能感兴趣的其他文章

Articles you may be interested in

基于全基因组的龟鳖目*Hox*基因序列特征及进化分析

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

水生生物学报. 2023, 47(10): 1617–1627 <https://doi.org/10.7541/2023.2022.0465>

一株雷氏普罗维登斯菌的分离鉴定、生物学特性及其基因组学研究

ISOLATION, IDENTIFICATION, BIOLOGICAL CHARACTERISTICS, AND WHOLE-GENOME SEQUENCING OF A *PROVIDENCIA RETTGERI* STRAIN

水生生物学报. 2024, 48(12): 2109–2121 <https://doi.org/10.7541/2024.2024.0085>

低蛋白饲料补充必需氨基酸对异育银鲫生长、氨基酸转运和mTOR信号通路相关基因的影响

LOW-PROTEIN DIETS SUPPLEMENTED WITH ESSENTIAL AMINO ACIDS ON GROWTH, AMINO ACID TRANSPORTERS AND GENE EXPRESSION OF THE MTOR SIGNALING PATHWAY IN GIBEL CARP (*CARASSIUS GIBELIO*)

水生生物学报. 2024, 48(1): 1–9 <https://doi.org/10.7541/2023.2023.0093>

拟柱孢藻N8全基因组测序及磷吸收转运通路的比较分析

WHOLE-GENOME SEQUENCING AND PHOSPHORUS UPTAKE AND TRANSPORT PATHWAY COMPARATIVE ANALYSIS OF *CYLINDROSPERMOPSIS RACIBORSKII* N8

水生生物学报. 2022, 46(8): 1130–1141 <https://doi.org/10.7541/2022.2021.0197>

基于全基因组InDel分析我国细鳞鲑遗传分化及本地适应性

GENETIC DIFFERENTIATION AND LOCAL ADAPTATION OF CHINESE LENOK (*BRACHYMYSTAX*) POPULATIONS BASED ON WHOLE GENOME INDELS

水生生物学报. 2024, 48(10): 1724–1735 <https://doi.org/10.7541/2024.2023.0424>



关注微信公众号，获得更多资讯信息

doi: 10.7541/2025.2024.0450

CSTR: 32229.14.SSSWXB.2024.0450

龟鳖目表皮分化复合体基因的比较基因组学鉴定与分析

陈孟饶^{1,2} 李芳^{1,2} 陈其然^{1,2} 黄保友^{1,2} 孙伟^{1,2} 钱国英^{1,2}
葛楚天^{1,2} 王宗吉^{1,2}

(1. 浙江万里学院动物性别与发育重点研究所, 宁波 315100; 2. 浙江万里学院生物与环境学院, 宁波 315100)

摘要: 发掘龟鳖表皮分化复合体(Epidermal Differentiation Complex, EDC)基因的序列特征和进化有助于解析其甲壳表皮特征的遗传基础和演化历史。文章对28个龟鳖的EDC基因进行了比较基因组学鉴定, 染色体定位、氨基酸含量、结构域、蛋白空间结构分析和演化历史的重塑。研究结果表明龟鳖EDC基因具有和其他爬行动物相同的四类基因家族, 其中简单表皮复合体(Simple epidermal differentiation complex, SEDC)基因家族成员通过多次串联重复事件, 实现基因扩张。龟鳖SEDC蛋白氨基酸组成高度多样化, SEDC和S100融合蛋白(S100-fused type proteins, SFTP)蛋白氨基酸含量存在类群(硬壳、软壳、棱皮)特异性, 这支持了龟鳖甲壳表皮结构的多样性。龟鳖S100钙结合蛋白A (S100 calcium binding protein A, S100A)和SFTP功能域和蛋白空间结构高度保守, 仅肽聚糖识别蛋白(Peptidolycan recognition protein3, PGLYRP3)的motif3存在鳖科特异性丢失。龟鳖特有的含有脯氨酸、半胱氨酸和缬氨酸的表皮分化蛋白(Epidermal Differentiation protein rich in Proline, Cysteine and Valine, EDPCV)来源于富含脯氨酸的表皮分化蛋白(Epidermal differentiation protein rich in proline 1, EDP1), 富含半胱氨酸和缬氨酸, 通过串联复制实现基因扩张。研究为进一步研究EDPCV在龟鳖中的作用提供了数据基础和理论依据。

关键词: 表皮分化复合体; 氨基酸组成; 基因扩张; EDPCV基因; 龟鳖

中图分类号: Q344⁺.1 文献标识码: A 文章编号: 1000-3207(2025)05-052510-10



表皮分化复合体(Epidermal differentiation complex, 缩写为EDC)基因最初在人类基因组上发现, 呈簇状分布于染色体1q21上一个2 Mb的区域, 包含多个编码结构、调控蛋白的基因^[1-3]。EDC基因对于角质形成细胞分化和角质层特性至关重要^[4], 主要包括四类基因家族: 编码S100钙结合蛋白A (S100 calcium binding protein A, S100A)、肽聚糖识别蛋白(Peptidolycan recognition protein, PGLYRP)、简单表皮分化复合体(Simple epidermal differentiation complex, SEDC)和S100融合蛋白(S100-fused type proteins, SFTP)的基因家族。SEDC是单外显子基因, 编码小脯氨酸富含蛋白(Small proline-rich protein,

SPRR)、外皮蛋白(Involucrin, IVL)和兜甲蛋白(loricrin, LOR)等。SFTP基因具有多个编码蛋白的外显子, 编码角质保护蛋白(cornulin, CRNN)、丝聚蛋白(Filaggrin, FLG)、毛角质透明蛋白(Trichohyalin, TCHH)和晚期角化包膜蛋白(Late cornified envelope, LCE)等^[5-7]。

EDC基因相继在脊椎动物不同分支中(哺乳类^[3,8,9]、鸟类^[10,11]、爬行^[11-15]、两栖^[16])被发现。哺乳类、鸟类、爬行类、两栖类的S100A家族高度保守, 鸟类不存在PGLYRP3基因。对于SEDC家族, 不同于其他脊椎动物, 哺乳动物缺乏 β 角蛋白基因(Beta-keratin)^[3,8], 哺乳类、鸟类及爬行动物都含有

收稿日期: 2024-11-21; 修订日期: 2025-01-07

基金项目: 国家自然科学基金(32102772); 宁波市自然科学基金(2023J043); 浙江省高校基础科研经费项目; 2022年浙江省“生物工程”一流学科(ZS2022008)资助 [Supported by the Natural Science Foundation of China (32102772); the Ningbo Natural Science Foundation (2023J043); the Basic Scientific Research Foundation of Zhejiang Provincial Universities; the Zhejiang Provincial Top Discipline of Biological Engineering (Level A), Zhejiang Wanli University (ZS2022008)]

作者简介: 陈孟饶(1999—), 硕士研究生; 主要从事水产动物遗传学研究。E-mail: 1300684186@qq.com

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

谷氨酰胺基序的表皮分化蛋白基因(Epidermal differentiation protein containing a glutamine (Q) Motif, *EDQM*)。爬行动物的*SEDC*家族较为保守, 蜥蜴、鳄鱼、蛇都不存在龟鳖特异的富含脯氨酸、半胱氨酸和缬氨酸的表皮分化蛋白基因(Epidermal differentiation protein rich in Proline, Cysteine and Valine, *EDPCV*)^[13-15]。与龟鳖、鳄鱼相比, 蜥蜴和蛇缺失了富含芳香氨基酸的表皮分化蛋白基因(Epidermal differentiation protein rich in Aromatic Amino, *EDAA*)^[11, 12]。两栖动物的*SEDC*家族仅由几个*SEDC1*、*SEDC2*基因组成^[16]。*SFTP*家族在脊椎动物中存在物种特异性, 哺乳动物中除含有*CRNN*基因外, 还有*FLG*、*TCHH*、*LCE*等基因^[3, 8, 9]。绿蜥蜴中缺失了*CRNN*^[11], 蛇类编码支架蛋白的基因(Scaffoldin, *SCFN*)出现了两个拷贝^[13]。

*EDC*基因调控羊膜动物角质形成细胞的分化, 对其鳞片、喙、爪子、皮肤等组织的发育发挥着至关重要的作用^[17, 18]。研究表明*EDC*基因编码蛋白的氨基酸组成会造成不同皮肤附属物物理特性的差异^[12, 13, 19]。在羽毛、毛发和爪子中, 富含半胱氨酸的蛋白质会形成更多的二硫键从而增加其硬度和强度^[20, 21]。甘氨酸和脯氨酸有助于形成柔韧的螺旋结构, 提高皮肤的柔韧性和弹性^[17, 21]。酪氨酸和苯丙氨酸等芳香族氨基酸在角质层蛋白中含量较高, 通过形成复杂的蛋白质网络, 增强外表皮的耐磨性和抵抗力^[21, 22]。丝氨酸和丙氨酸亲水性有助于蛋白质保持水分, 从而提高表皮和角质层的保湿能力^[7, 11, 21-24]。因此分析*EDC*基因编码蛋白的氨基酸组成有助于解析其功能及特性。

甲壳(背甲、腹甲)使龟鳖在脊椎动物中独树一帜, 然而龟鳖的甲壳由基本的骨骼、真皮和表皮三个部分组成, 这暗示其遗传基础是保守的。龟鳖甲壳骨骼相关的进化和发育已经被广泛、深入地研究, 然而其甲壳表皮结构进化的遗传分子基础研究才开始^[12, 25-29]。龟鳖甲壳的最外层表皮具有角质化纹理结构, 根据其骨化、角质化程度高低可分为硬壳、软壳和棱皮^[30-32]。尽管龟鳖表皮分化的分子机制尚未完全厘清, 但可以通过与其他羊膜动物比较分析推断出龟鳖甲壳特征是角质形成细胞分化过程特定适应的结果^[11]。表皮的主要细胞类型是角质形成细胞, 它不仅形成角质层, 还形成各种皮肤附属物^[33]。表皮组织的发育和分化与*EDC*基因的演化和特定的表达调控模式密切相关^[23, 33-35]。因此发掘龟鳖*EDC*基因的序列特征和进化有助于解析其甲壳表皮特征的遗传基础和演化历史。

目前, *EDC*基因仅在少数龟鳖物种中进行了鉴

定、分析^[12], 尚缺乏对*EDC*基因序列特征和进化水平在不同龟鳖类群的系统性比较分析。本研究对28个龟鳖*EDC*基因进行了比较基因组学鉴定, 染色体定位、氨基酸含量、结构域、蛋白空间结构分析和演化历史的重塑, 为进一步研究*EDPCV*基因在龟鳖中的作用提供了数据基础和理论依据。

1 材料与方法

1.1 *EDC*蛋白参考序列收集和同源预测

搜集人(*Homo sapiens*)、家鸡(*Gallus gallus*)、锦龟(*Chrysemys picta bellii*)、绿蜥蜴(*Anolis carolinensis*)、缅甸蟒(*Python bivittatus*)、短吻鳄(*Alligator mississippiensis*)*EDC*基因蛋白序列作为参考序列, 用于基因同源预测^[11-15]。NCBI数据库下载已公布的28个龟鳖物种基因组序列。这些物种覆盖鳖科(*Trionychia*, $n=4$)、地龟科(*Geoemydidae*, $n=4$)、陆龟科(*Testudinidae*, $n=4$)、泽龟科(*Emydidae*, $n=6$)、平胸龟科(*Platysternidae*, $n=1$)、海龟科(*Cheloniidae*, $n=3$)、鳄龟科(*Chelydridae*, $n=2$)、侧颈龟科(*Pelomedusoides*, $n=2$)和蛇颈龟科(*Chelidae*, $n=2$)。

将参考蛋白序列与龟鳖基因组进行比对^[36]($E < 10^{-5}$), 过滤掉低质量序列($\text{alignRate} < 50\%$, $\text{identity} < 40\%$); solar^[37](v0.9.6)连接比对结果, 对其上下游各延伸5 kb, 进行GENEWISE^[38]精细比对, 最终通过手动检查、筛选获得正确的同源基因序列。

对28个龟鳖鉴定的*EDC*蛋白序列进行统一命名, 其命名方式为由表皮分化(Epidermal differentiation, ED)和描述氨基酸组成或编码蛋白质中特定氨基酸序列基序组成^[11]。

1.2 基因染色体定位

分别使用R包gggenes^[39]和pheatmap^[40]对*EDC*基因进行可视化展示。

1.3 序列分析

龟鳖物种间亲缘关系来源于已发表的系统发生树^[41], 利用python脚本对蛋白序列进行氨基酸组成计算, 利用R的wilcox.test函数进行秩和检验, 并筛选出*EDC*蛋白序列具有显著差异的类群和氨基酸。并用R包reshape2^[42]和ggplot2^[43]对每个*EDC*蛋白序列氨基酸进行可视化展示。

利用AlphaFold2^[44]进行蛋白空间模型预测, PYMOL^[45]进行可视化展示。通过MEME^[46]软件对蛋白序列进行motif分析, 利用rpsblast^[47]进行蛋白结构域预测, 并利用tbtools^[48]进行可视化展示。

1.4 进化分析

使用mafft^[49]软件进行*EDC*蛋白序列的多序列比对, 并使用Jalview^[50]进行序列比对结果可视化。

利用iqtree^[51]设置参数(Ultrafast bootstrap=1000), 以鸡的*S100A11*基因作为外群并且使用最佳模型VT+F+R7对*EDC*基因进化树进行构建, 并利用ITOL^[52]进行可视化展示。

2 结果

2.1 龟鳖*EDC*基因呈现拷贝多样性并存在特有的*EDPCV*基因

为了解析龟鳖*EDC*基因是否具有保守性和类群特异性, 利用已发表的*EDC*基因蛋白序列(人、家鸡、锦龟、绿蜥蜴、缅甸蟒、短吻鳄)作为参考, 对囊括9个科的28个龟鳖基因组序列进行基因同源预测, 鉴定*S100A12*和*S100A11*基因间的*EDC*基因集。龟鳖*EDC*基因具有和外群物种(家鸡、鳄鱼、蜥蜴、蛇)相同的四类基因家族(*S100A*、*PGLYRP3*、*SEDC*、*SFTP*)且成簇分布(图1)。龟鳖的*S100A*、*PGLYRP3*和*SFTP*基因家族成员高度保守, 并未出现基因的复制和丢失。家鸡和绿蜥蜴分别丢失了*PGLYRP3*和*CRNN*基因, 这和之前的报道一致^[11]。

龟鳖*EDC*基因的多样性主要来源于*SEDC*基因家族, 其中五种不同类型的基因组成占据了主体, 包括*EDQM*、*EDAA*、*EDP*、*EDPCV*和β角蛋白基因(图1)。*EDQM*、*EDAA*和*EDP*基因在龟鳖物种间具有相似的拷贝数, 其中*EDAA*($n \approx 9$)和*EDP*($n = 3$)基因的拷贝数高于家鸡($n = 5, 1$), 表明这两类基因在龟鳖祖先发生了扩张。值得一提的是, 仅有龟鳖*EDPCV*在外群不存在对应直系同源基因, 这说明*EDPCV*基因起源于龟鳖的共同祖先。龟鳖*SEDC*基因家族最多拷贝来源于β角蛋白基因($n = 2—125$), 并呈现物种特异性。虽然存在零星的物种特异性基因丢失, 龟鳖*SEDC*基因家族的其他基因成员(例如*EDKM*、*EDQL*、*EDYMI*和*EDPE*)均高度保守(图1)。综上, 龟鳖*EDC*基因具有和其他爬行动物相同的四类基因家族, *SEDC*基因家族成员通过多次串联重复事件, 实现基因扩张。此外, 龟鳖*EDC*基因存在龟鳖特有的*EDPCV*基因。

2.2 龟鳖*SEDC*和*SFTP*蛋白氨基酸组成具有类群特异性

为了厘清龟鳖*EDC*蛋白的氨基酸组成及特征,

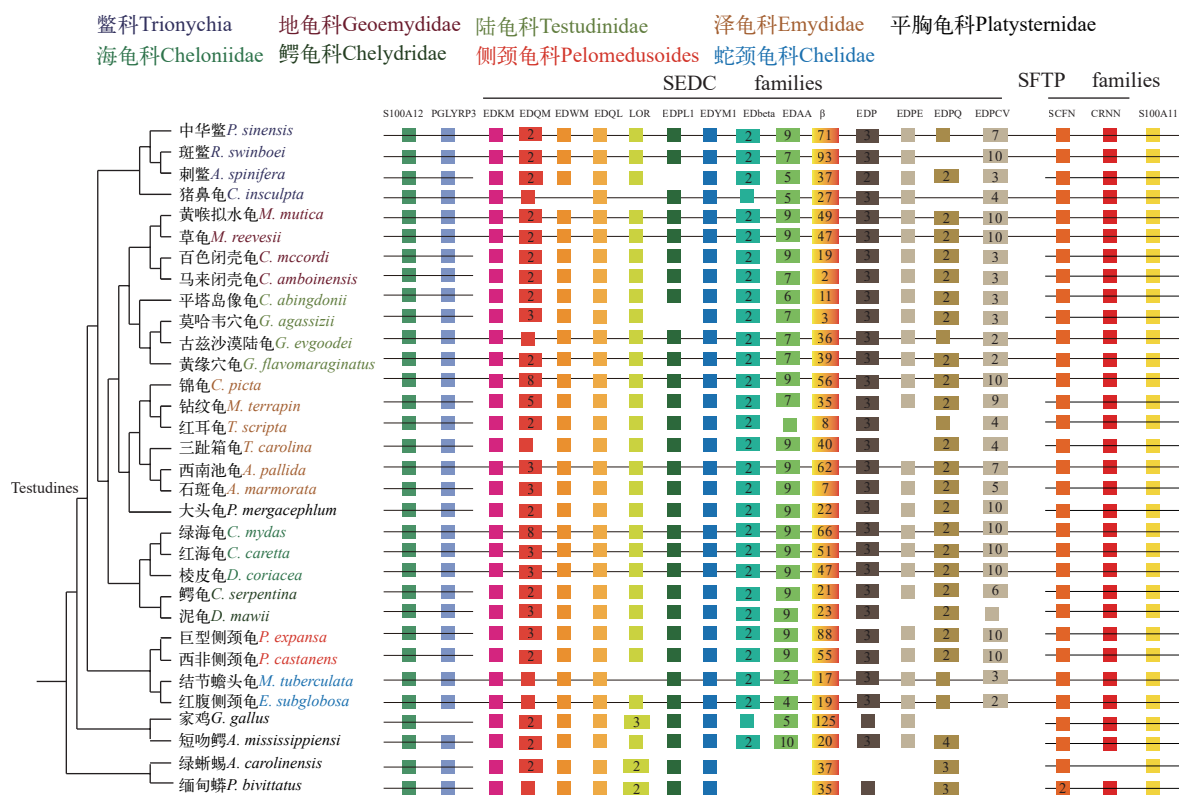


图1 龟鳖与其他爬行动物*EDC*基因的基因组分布

Fig. 1 Genomic distribution of *EDC* genes in turtles and other reptiles

左图为龟鳖和其他爬行动物的物种树; 右图显示了表皮分化复合体(*EDC*)基因的组成; 着色的矩形代表 *EDC* 基因的不同成员; 连接矩形的实线代表基因间的物理基因组连锁; 矩形中的数字表示基因数量

The left side illustrates a species tree of turtles and other reptiles, while the right side displays the distribution of the Epidermal Differentiation Complex (*EDC*) genes. Colored squares represent the identified *EDC* genes, and solid lines connecting the gene boxes indicate physical genomic linkage. The number in the rectangle indicates the gene numbers

对其四类基因家族成员的氨基酸含量进行计算和比较。龟鳖的*S100A*、*PGLYRP3*和*SFTP*家族成员间的氨基酸组成保守, *SEDC*家族的氨基酸组成高度多样化。龟鳖*SEDC*蛋白氨基酸组成中富含甘氨酸(G)和丝氨酸(S), 或半胱氨酸(C)和脯氨酸(P), 或赖氨酸(K)和谷氨酰胺残基(Q)。其中EDPQ和EDPCV的半胱氨酸(C)和脯氨酸(P)的总含量超过了氨基酸残基总数的50%, LOR的甘氨酸(G)和丝氨酸(S)的总含量超过了氨基酸残基总数的50% (图 2A)。富含甘氨酸/丝氨酸(G/S)的基因分布在EDC的左半区域, 而编码富含半胱氨酸/脯氨酸(C/P)的基因则集中在EDC的右半区域(图 2B), 这表明EDC基因是通过串联重复事件产生的。EDQM、LOR和SFTP(SCFN、CRNN)蛋白序列中部的氨基酸序列以重复序列为主, 其中鳖科LOR比龟科具有更长的重复序列(图 2C), 这可能来源于染色体交叉过程中的不等互换。

龟鳖甲壳存在硬壳(例如地龟科、陆龟科和泽龟科)软壳(鳖科)和棱皮(海龟科为代表)三种类型, 这提示其表皮的遗传基础可能存在类群差异。相对地龟科、泽龟科和海龟科, 鳖科的EDQL蛋白序列显著富含精氨酸(R)。相对地龟科、陆龟科, 鳖科的EDYM1蛋白序列显著富含甘氨酸(G)。相对地龟科、陆龟科和泽龟科, 鳖科的EDAA蛋白序列显著富含半胱氨酸(C)、丝氨酸(S), 鳖科的CRNN蛋白序列显著富含谷氨酸(E)、组氨酸(H)和精氨酸(R), 但谷氨酰胺残基(Q)含量较低(图 2D)。相对泽龟科, 鳖科的EDPE蛋白序列显著富含苏氨酸(T); 相对地龟科和陆龟科, 鳖科的SCFN蛋白精氨酸(R)含量较低; 地龟科的EDPL1比鳖科显著富含赖氨酸(K)。这些结果表明鳖科作为软壳类群在SEDC(EDQL、EDYM1、EDAA)和SFTP(CRNN、SCFN)蛋白氨基酸含量不同于硬壳、棱皮类群。

相对地龟科、陆龟科和泽龟科, 海龟科的EDKM蛋白序列显著富含异亮氨酸(I)和精氨酸(R); 地龟科、陆龟科和泽龟科的EDKM相对海龟科显著富含精氨酸(R)。相对陆龟科、海龟科的EDKM和EDQL蛋白序列分别显著富含亮氨酸(L)和苏氨酸(T); 相对泽龟科, 海龟科EDPE和EDPQ蛋白序列分别显著富含谷氨酸(E)、缬氨酸(V)和赖氨酸(K)。由此可见, 海龟壳的SEDC(EDKM、EDQL、EDPE、EDPQ)蛋白氨基酸含量不同于硬壳类群。此外, 相对陆龟科, 地龟科的EDPL1显著富含谷氨酸(E); 泽龟科的EDPL1分别比海龟科、陆龟科分别显著富含谷氨酸(E)和苏氨酸(T)。

综上, 龟鳖SEDC蛋白氨基酸组成高度多样化,

SEDC和SFTP蛋白氨基酸含量存在类群(硬壳、软壳、棱皮)特异性, 这支持了龟鳖甲壳表皮结构的多样性。

2.3 龟鳖SFTP蛋白结构域保守

为了探明龟鳖*S100A*、*PGLYRP3*和*SFTP*功能域和蛋白空间结构的保守性, 选取龟鳖每个科中基因质量最好的序列进行分析比较。龟鳖*S100A11*和*SFTP*家族成员(SCFN和CRNN)功能域高度保守, 都具有相同*S100*结构域的 α 螺旋结构(图 3A), 支持了EDC基因起源于*S100A*基因的假说。值得一提的是, 所有SEDC家族成员并不存在*S100*结构域的 α 螺旋结构。此外, 基序(motif)分析表明龟鳖*S100A*和*SFTP*家族成员具有高度保守的motif序列, 仅鳖科的*PGLYRP3*缺失了motif3(图 3B)。综上, 龟鳖*S100A*和*SFTP*功能域和蛋白空间结构高度保守, 仅*PGLYRP3*的motif3存在鳖科特异性丢失。

2.4 龟鳖特有的EDPCV起源于EDP1, 串联重复实现扩张

为了揭示龟鳖SEDC基因家族的演化历史, 我们选取家鸡*S100A11*定根并对所研究的28个龟鳖SEDC基因进行了系统发育分析。龟鳖SEDC家族成员基因都形成了独立的单系群(Monophyletic group, 包含一个共同祖先及其所有后代), 这说明龟鳖SEDC家族基因成员具有相同的祖先起源并保守(图 4A)。其中LOR、EDQL、EDWM和EDQM系统发育分支构成单系群(图 4A), 并且在染色体上依次分布(图 1A), 其共同祖先起源并通过串联复制实现。值得一提的是, 龟鳖特有的EDPCV基因和EDP1基因构成姐妹分支, EDPCV很可能起源于EDP1基因。多序列结果表明EDPCV和EDP1具有最高的序列相似性, EDP1的多个脯氨酸(P)突变为半胱氨酸(C)和缬氨酸(V)。综上, 本文推断龟鳖特有的EDPCV基因起源于EDP1基因的复制, 在新功能化(Neofunctionalization)作用下发生氨基酸突变(富含半胱氨酸和缬氨酸, 前者形成更多的二硫键, 增加硬度和强度), 并通过基因的串联复制实现EDPCV基因的扩张(图 4B), 进而实现龟鳖表型创新(甲壳、表皮角质化等)。

3 讨论

3.1 龟鳖甲壳形态演化与EDC基因簇扩张的关系

本研究结果表明龟鳖独特形态(甲壳)的演化涉及表皮分化基因的特定适应, 这些基因位于羊膜动物特有EDC基因簇。在龟鳖动物演化过程中, EDQM、EDAA、EDP、 β 角蛋白基因和龟鳖特有的EDPCV通过串联复制实现扩张(图 1)。这些发生

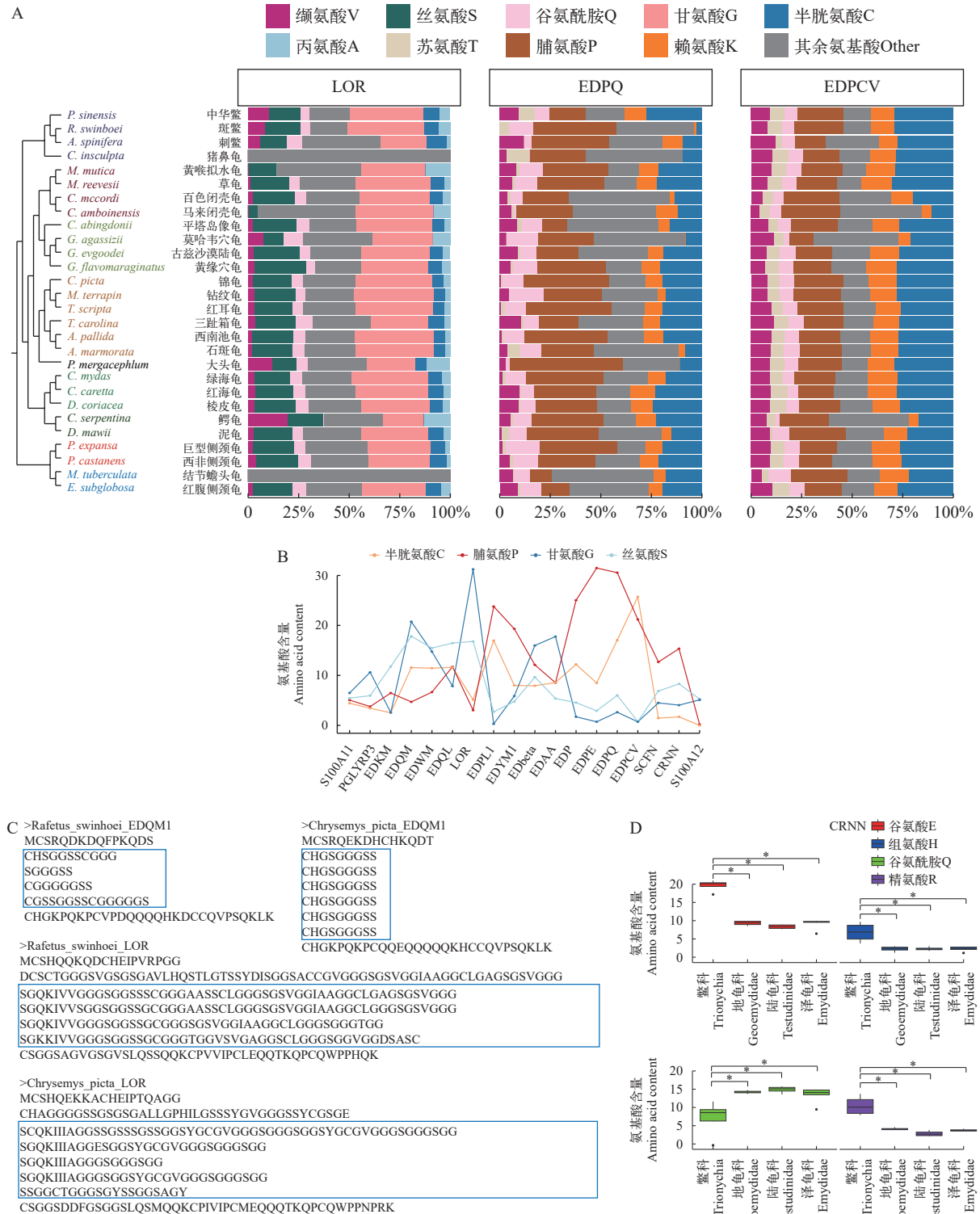


图 2 龟鳖EDC蛋白氨基酸组成

Fig. 2 Amino acid contents of turtle EDC proteins

A. 龟鳖不同物种 EDC 蛋白的氨基酸组成; B. 龟鳖EDC基因氨基酸含量均值百分比趋势图; 蛋白展示顺序和图 1 中EDC基因分布顺序一致; C. 鳖、龟SEDC示例(EDQM1和LOR)蛋白氨基酸序列, 重复序列用蓝色矩形标记; D. 不同龟鳖科间 CRNN 蛋白氨基酸组成的比较; 统计显著性为: $0.01 < *P < 0.05$

A. The diagram shows the amino acid contents of EDC proteins across turtle species; B. The average distributions of amino acid content in turtles. The protein data are shown in the order of the corresponding genes in the EDC (Fig. 1); C. Amino acid sequences of exemplary SEDC proteins (EDQM1 and LOR) from softshell and hardshell turtles are shown, with sequence repeats highlighted in blue rectangles; D. The distribution of amino acid content of CRNN protein sequences across various turtle families; The level of significance: $0.01 < *P < 0.05$

扩张的基因在不同发育时期的甲壳(背甲和腹甲)组织中存在差异表达(数据未展示)。本文推断这些在龟鳖扩张的基因可能通过新功能化作用实现组织、时空特异性表达,从而促进了甲壳(例如盾片)的产生。此外,在龟鳖发育过程中,龟鳖特有的EDPCV和EDC其他家族成员的表达时序,调控关系对于解析甲壳的发生、发育至关重要,也可成为新基因—新表型研究的范式。

3.2 龟鳖EDC蛋白的氨基酸组成特征与表皮结构的多样性

EDC编码表皮角质形成细胞的结构蛋白^[1, 5, 33],其中SEDC编码的蛋白质作为皮肤表面机械弹性结构的交联成分发挥其功能^[23, 33]。EDC家族成员编码蛋白的相对丰度和分子相互作用类型可能调节角质化产物的物理、化学性质,例如柔韧的角质层和坚硬的背甲盾片等^[11]。本研究结果表明发生扩张的SEDC家族成员显著富含不同的氨基酸成分,例如含多聚谷氨酰胺(Q)的EDQM、富含脯氨酸(P)和半胱氨酸(C)的EDPCV、富含芳香族氨基酸的EDAA。这些蛋白家族特异的序列特征可能促进与角质形成细胞结构蛋白间不同类型的相互作用^[53]。谷氨酰胺和半胱氨酸(在EDQM和EDPCV中)是实现EDC蛋白分子间交联的主要部位,分别通过转谷氨酰胺化和二硫键作用^[7, 33, 54]。富含半胱氨酸的蛋白质通过形成更多的二硫键,从而提高羽毛、毛发和爪子的硬度和强度^[20, 21]。LOR和EDQM序列中位于转谷氨酰胺化位点间的甘氨酸可能利于蛋白序列长度的变化,从而实现角质形成细胞角质化过程中细胞蛋白包膜的压缩^[33]。本研究发现相比地龟科、陆龟科和泽龟科,鳖科的EDAA蛋白序列具有显著更高的半胱氨酸含量,并且EDQM和LOR的氨基酸重复序列存在龟、鳖差

异(图2C)。此外,隶属于SEDC家族的EDQL、EDYM1、EDKM、EDPE、EDPQ和EDPL1也存在类群(软壳、硬壳和棱皮)特异的氨基酸富集情况。上述结果支持了龟鳖甲壳表皮结构的多样性,虽然缺乏直接的实验证据。

SFTP基因簇编码的蛋白在表皮对环境的屏障中起着至关重要的作用,存在于哺乳动物、爬行动物、鸟类和现存两栖动物的共同祖先中^[6, 11]。SFTP基因在其他类群中经历了不同程度的扩张和丢失。在鲸豚类中海豚保留了LCE和FLG基因^[9],而鲸鱼的LCE和FLG发生了丢失^[8]。在爬行动物中,绿蜥蜴丢失了CRNN^[11],而蛇类SCFN出现了两个拷贝^[13]。龟鳖SFTP不存在种类和拷贝的差异(图1),这表明在演化历程中的保守性。SCFN基因主要在鸡胚胎皮肤的表皮层中表达,也在指甲中表达^[55]。SCFN蛋白在鸡胚胎的喙、趾、皮肤和舌头中表达^[55, 56]。CRNN蛋白主要在表皮的颗粒层、毛囊的内根鞘以及食管上皮组织中表达^[55, 57]。此外,不含角质形成细胞的组织缺乏CRNN和SCFN表达^[55, 57]。CRNN的中心结构域含有高含量的谷氨酰胺,成为其在角质形成细胞角化过程中的转谷氨酰胺靶标^[55]。同时,在表皮分化过程中,精氨酸脱亚胺酶会催化蛋白质的翻译后修饰,从而生成游离的氨基酸,以维持角质层细胞的稳定性和屏障功能^[58]。本研究发现鳖科SCFN的精氨酸(R)含量与地龟科、陆龟科存在显著差异。鳖科CRNN的谷氨酰胺含量与地龟科、陆龟科和泽龟科存在显著差异(图2D)。SFTP蛋白序列在鳖科和龟科之间也存在重复序列差异。这些结果提示SFTP蛋白氨基酸含量的差异可能导致了龟、鳖表皮结构的不同。

3.3 EDC基因的演化机制

以往假说认为,在原始EDC中相邻S100A和

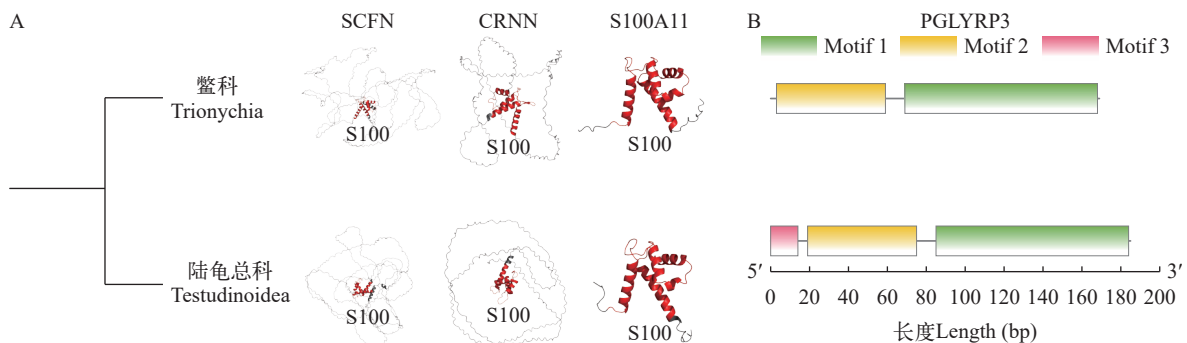


图3 龟鳖目SFTP的空间结构和PGLYRP3的motif分析

Fig. 3 Spatial structure of SFTP and motif analysis of PGLYRP3 in turtles

A. 鳖科和陆龟总科SFTP和S100A11蛋白的空间结构,红色 α 螺旋标记为S100结构域; B. 矩形标记为PGLYRP3蛋白motif区域

A. The spatial structures of the SFTP and S100A11 proteins in the families of Trionychia and Testudinoidea, with the red α -helix labeled as the S100 domain; B. The rectangle marks the protein motif regions of the PGLYRP3

PGLYRP基因经历了一系列融合、复制、外显子缺失、内含子缺失及编码序列末端的变化, 最终形成了SEDC、SFTP和EDKM基因。随后SEDC和SFTP基因经历了广泛的基因复制和序列修改, 生成了序列高度多样化的蛋白质, 其中只有少数序列元件在部分蛋白中保留^[11]。本研究发现起源于

EDPCV是龟鳖特有的, 并且起源于龟鳖共同祖先。系统发育分析和多序列比对证实EDPCV起源于EDP1, 脯氨酸(P)突变为半胱氨酸(C)和缬氨酸(V), 最终通过串联复制实现基因扩张。本研究印证了EDC起源的假说——“复制、突变、串联复制”, 也为龟鳖的表型创新研究提供了靶标。

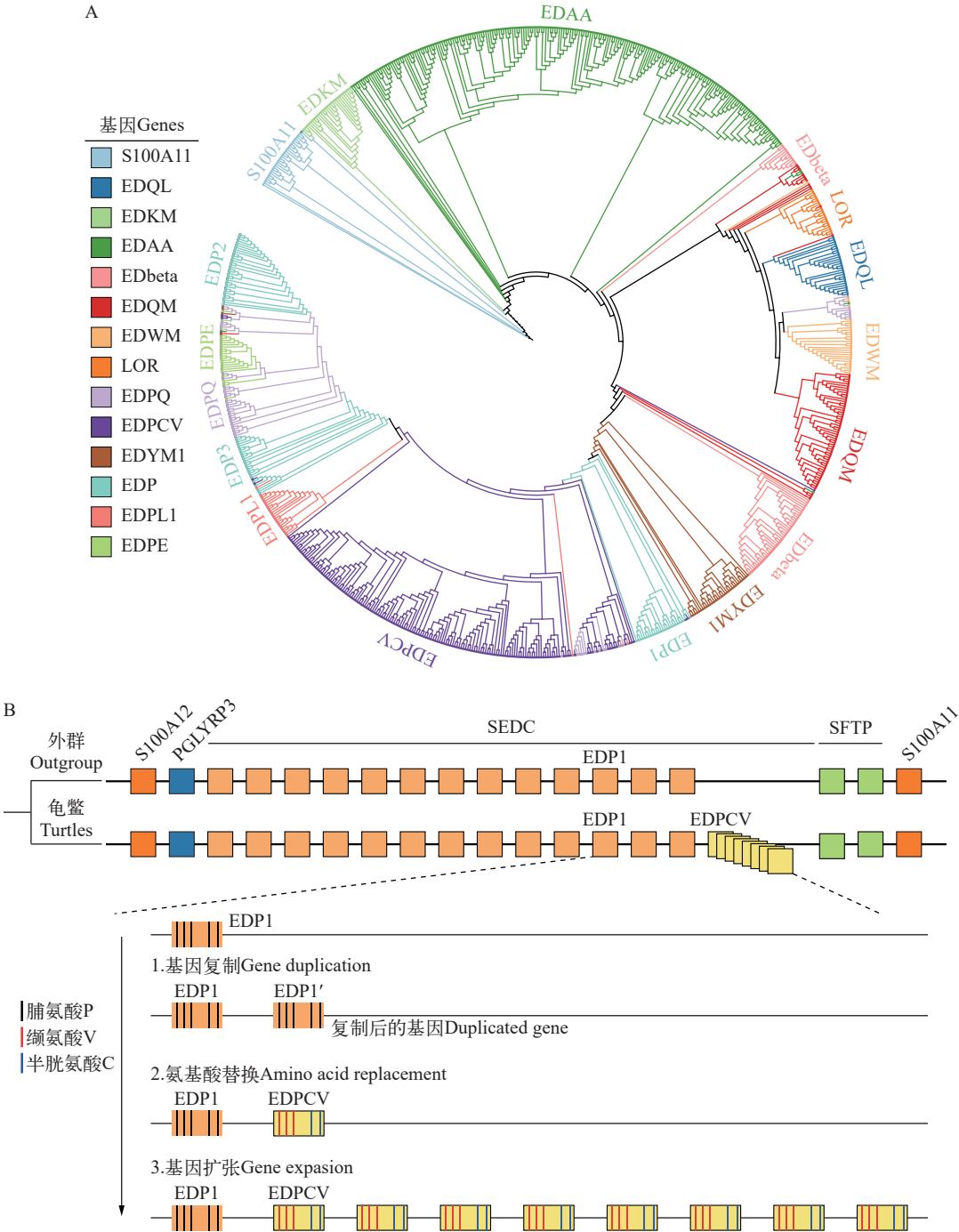


图 4 龟鳖目SEDC基因系统发育分析及演化

Fig. 4 Phylogenetic analysis and the evolution of SEDC in turtles

A. 龟鳖SEDC基因系统发育树; 颜色标注不同家族成员; B. 龟鳖EDPCV基因起源及多样化模型

A. Phylogenetic tree of SEDC genes in turtles; Colors indicate SEDC subfamilies; B. A scenario for the origin and diversification of EDPCV in turtles

综上,本研究通过比较基因组学揭示了龟鳖表皮的演化与EDC的串联复制扩张,氨基酸富集密切相关。EDC类群特异的基因组成变化促进了龟鳖表皮对不同环境和生活方式的适应性进化。龟鳖特异EDPCV基因的起源和演化历史的重塑将为甲壳表皮结构的遗传分子相关研究提供理论基础和参考。

(作者声明本文符合出版伦理要求)

参考文献:

- [1] Henry J, Toulza E, Hsu C Y, *et al.* Update on the epidermal differentiation complex [J]. *Frontiers in Bioscience* (Landmark edition), 2012, **17**(4): 1517-1532.
- [2] Sobiak B, Graczyk-Jarzynka A, Leśniak W. Comparison of DNA methylation and expression pattern of S100 and other epidermal differentiation complex genes in differentiating keratinocytes [J]. *Journal of Cellular Biochemistry*, 2016, **117**(5): 1092-1098.
- [3] Mischke D, Korge B P, Marenholz I, *et al.* Genes encoding structural proteins of epidermal cornification and S100 calcium-binding proteins form a gene complex ("epidermal differentiation complex") on human chromosome 1q21 [J]. *Journal of Investigative Dermatology*, 1996, **106**(5): 989-992.
- [4] Holthaus K B, Eckhart L. Development-associated genes of the epidermal differentiation complex (EDC) [J]. *Journal of Developmental Biology*, 2024, **12**(1): 4.
- [5] Kyriiotou M, Huber M, Hohl D. The human epidermal differentiation complex: cornified envelope precursors, S100 proteins and the 'fused genes' family [J]. *Experimental Dermatology*, 2012, **21**(9): 643-649.
- [6] Mlitz V, Hussain T, Tschachler E, *et al.* Filaggrin has evolved from an "S100 fused-type protein" (SFTP) gene present in a common ancestor of amphibians and mammals [J]. *Experimental Dermatology*, 2017, **26**(10): 955-957.
- [7] Kalinin A E, Kajava A V, Steinert P M. Epithelial barrier function: assembly and structural features of the cornified cell envelope [J]. *BioEssays*, 2002, **24**(9): 789-800.
- [8] Holthaus K B, Lachner J, Ebner B, *et al.* Gene duplications and gene loss in the epidermal differentiation complex during the evolutionary land-to-water transition of cetaceans [J]. *Scientific Reports*, 2021, **11**(1): 12334.
- [9] Strasser B, Mlitz V, Fischer H, *et al.* Comparative genomics reveals conservation of filaggrin and loss of caspase-14 in dolphins [J]. *Experimental Dermatology*, 2015, **24**(5): 365-369.
- [10] Davis A, Greenwold M J. Evolution of an epidermal differentiation complex (EDC) gene family in birds [J]. *Genes*, 2021, **12**(5): 767.
- [11] Strasser B, Mlitz V, Hermann M, *et al.* Evolutionary origin and diversification of epidermal barrier proteins in amniotes [J]. *Molecular Biology and Evolution*, 2014, **31**(12): 3194-3205.
- [12] Holthaus K B, Strasser B, Sipos W, *et al.* Comparative genomics identifies epidermal proteins associated with the evolution of the turtle shell [J]. *Molecular Biology and Evolution*, 2016, **33**(3): 726-737.
- [13] Holthaus K B, Mlitz V, Strasser B, *et al.* Identification and comparative analysis of the epidermal differentiation complex in snakes [J]. *Scientific Reports*, 2017(7): 45338.
- [14] Holthaus K B, Strasser B, Lachner J, *et al.* Comparative analysis of epidermal differentiation genes of crocodilians suggests new models for the evolutionary origin of avian feather proteins [J]. *Genome Biology and Evolution*, 2018, **10**(2): 694-704.
- [15] Holthaus K B, Alibardi L, Tschachler E, *et al.* Identification of epidermal differentiation genes of the tuatara provides insights into the early evolution of lepidosaurian skin [J]. *Scientific Reports*, 2020, **10**(1): 12844.
- [16] Sachslehner A P, Eckhart L. Evolutionary diversification of epidermal barrier genes in amphibians [J]. *Scientific Reports*, 2022, **12**(1): 13634.
- [17] Moltrasio C, Romagnuolo M, Marzano A V. Epigenetic mechanisms of epidermal differentiation [J]. *International Journal of Molecular Sciences*, 2022, **23**(9): 4874.
- [18] Simpson C L, Patel D M, Green K J. Deconstructing the skin: cytoarchitectural determinants of epidermal morphogenesis [J]. *Nature Reviews Molecular Cell Biology*, 2011, **12**(9): 565-580.
- [19] Alibardi L. Corneous beta proteins of the epidermal differentiation complex (EDC) form large part of the corneous material of claws and rhamphothecae in turtles [J]. *Protoplasma*, 2020, **257**(4): 1123-1138.
- [20] Yu J, Yu D W, Checkla D M, *et al.* Human hair keratins [J]. *Journal of Investigative Dermatology*, 1993, **101**(1): S56-S59.
- [21] Shelley W B. Physiology, biochemistry, and molecular biology of the skin [J]. *JAMA*, 1992, **268**(4): 545.
- [22] Scott I R, Harding C R, Barrett J G. Histidine-rich protein of the keratohyalin granules Source of the free amino acids, urocanic acid and pyrrolidone carboxylic acid in the stratum corneum [J]. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1982, **719**(1): 110-117.
- [23] Candi E, Schmidt R, Melino G. The cornified envelope: a model of cell death in the skin [J]. *Nature Reviews Molecular Cell Biology*, 2005, **6**(4): 328-340.
- [24] Bragulla H H, Homberger D G. Structure and functions of keratin proteins in simple, stratified, keratinized and cornified epithelia [J]. *Journal of Anatomy*, 2009, **214**(4): 516-559.
- [25] Li Y I, Kong L, Ponting C P, *et al.* Rapid evolution of Beta-keratin genes contribute to phenotypic differences that distinguish turtles and birds from other reptiles [J].

- Genome Biology and Evolution*, 2013, **5**(5): 923-933.
- [26] Moustakas-Verho J E, Zimm R, Cebra-Thomas J, *et al.* The origin and loss of periodic patterning in the turtle shell [J]. *Development*, 2014, **141**(15): 3033-3039.
- [27] Moustakas-Verho J E, Cherepanov G O. The integumental appendages of the turtle shell: an evo-Devo perspective [J]. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 2015, **324**(3): 221-229.
- [28] Dalla Valle L, Nardi A, Toni M, *et al.* Beta-keratins of turtle shell are Glycine-proline-tyrosine rich proteins similar to those of crocodilians and birds [J]. *Journal of Anatomy*, 2009, **214**(2): 284-300.
- [29] Huang B Y, Lou L Y, Fan J W, *et al.* Evolutionary analyses and sequence feature of *Hox* genes in turtles [J]. *Acta Hydrobiologica Sinica*, 2023, **47**(10): 1617-1627. [黄保友, 楼灵媛, 樊嘉伟, 等. 基于全基因组的龟鳖目 *Hox* 基因序列特征及进化分析 [J]. *水生生物学报*, 2023, **47**(10): 1617-1627.]
- [30] Ogushi K. Anatomische Studien an der japanischen dreikralligen Lippenschildkröte (*Trionyx japonicus*). I. Mitteilung [J]. *Morphologisches Jahrbuch*, 1911(43): 1-106.
- [31] Nagashima H, Shibata M, Taniguchi M, *et al.* Comparative study of the shell development of hard- and soft-shelled turtles [J]. *Journal of Anatomy*, 2014, **225**(1): 60-70.
- [32] Gilbert S F, Cebra-Thomas J A, Burke A C, *et al.* How the Turtle Gets Its Shell [M]. *Biology of Turtles: Structures to Strategies*. CRC Press, Boca Raton, 2007.
- [33] Eckhart L, Lippens S, Tschachler E, *et al.* Cell death by cornification [J]. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 2013, **1833**(12): 3471-3480.
- [34] Yenmiş M, Ayaz D. The story of the finest armor: developmental aspects of reptile skin [J]. *Journal of Developmental Biology*, 2023, **11**(1): 5.
- [35] Alibardi L. Adaptation to the land: The skin of reptiles in comparison to that of amphibians and endotherm amniotes [J]. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 2003, **298B**(1): 12-41.
- [36] Altschul S F, Gish W, Miller W, *et al.* Basic local alignment search tool [J]. *Journal of Molecular Biology*, 1990, **215**(3): 403-410.
- [37] 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.
- [38] Birney E, Clamp M, Durbin R. GeneWise and genome-wise [J]. *Genome Research*, 2004, **14**(5): 988-995.
- [39] Wilkins D. gggenes: draw gene arrow maps in 'ggplot2' [M]. R package version 0.4. 0. 2019.
- [40] Kolde R, Kolde M R. Package 'pheatmap' [J]. *R package*, 2015, **1**(7): 790.
- [41] 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.
- [42] Wickham H. Reshaping data with the reshapePackage [J]. *Journal of Statistical Software*, 2007, **21**(12): 1-20.
- [43] Wickham H. ggplot2 [J]. *Wiley Interdisciplinary Reviews*, 2011, **3**(2): 180-185.
- [44] Cramer P. AlphaFold2 and the future of structural biology [J]. *Nature Structural & Molecular Biology*, 2021, **28**(9): 704-705.
- [45] Mooers B H M. Shortcuts for faster image creation in PyMOL [J]. *Protein Science*, 2020, **29**(1): 268-276.
- [46] Bailey T L, Boden M, Busch F A, *et al.* MEME SUITE: tools for motif discovery and searching [J]. *Nucleic Acids Research*, 2009, 37(Web Server issue): W202-W208.
- [47] Yang M, Derbyshire M K, Yamashita R A, *et al.* NCBI's conserved domain database and tools for protein domain analysis [J]. *Current Protocols in Bioinformatics*, 2020, **69**(1): e90.
- [48] Chen C, Chen H, Zhang Y, *et al.* TBtools: an integrative toolkit developed for interactive analyses of big biological data [J]. *Molecular Plant*, 2020, **13**(8): 1194-1202.
- [49] Katoh K, Toh H. Recent developments in the MAFFT multiple sequence alignment program [J]. *Briefings in Bioinformatics*, 2008, **9**(4): 286-298.
- [50] Waterhouse A M, Procter J B, Martin D M A, *et al.* Jalview Version 2-a multiple sequence alignment editor and analysis workbench [J]. *Bioinformatics*, 2009, **25**(9): 1189-1191.
- [51] Minh B Q, Trifinopoulos J, Schrempf D, *et al.* IQ-TREE version 2.0: tutorials and Manual Phylogenomic software by maximum likelihood [J]. *Nucleic Acids Research*, 2022, **44**(W1): W232-W235.
- [52] Letunic I, Bork P. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation [J]. *Nucleic Acids Research*, 2021, **49**(W1): W293-W296.
- [53] Niessen C M. Tight junctions/adherens junctions: basic structure and function [J]. *Journal of Investigative Dermatology*, 2007, **127**(11): 2525-2532.
- [54] Rice R H, Winters B R, Durbin-Johnson B P, *et al.* Chicken corneocyte cross-linked proteome [J]. *Journal of Proteome Research*, 2013, **12**(2): 771-776.
- [55] Mlitz V, Strasser B, Jaeger K, *et al.* Trichohyalin-like proteins have evolutionarily conserved roles in the morphogenesis of skin appendages [J]. *The Journal of Investigative Dermatology*, 2014, **134**(11): 2685-2692.
- [56] Mlitz V, Hermann M, Buchberger M, *et al.* The trichohyalin-like protein scaffoldin is expressed in the multilayered periderm during development of avian beak and egg tooth [J]. *Genes*, 2021, **12**(2): 248.
- [57] Contzler R, Favre B, Huber M, *et al.* Cornulin, a new

member of the “fused gene” family, is expressed during epidermal differentiation [J]. *Journal of investigative dermatology*, 2005, **124**(5): 990-997.

[58] Méchin M C, Enji M, Nachat R, *et al.* The peptidylargi-

nine deiminases expressed in human epidermis differ in their substrate specificities and subcellular locations [J]. *Cellular and Molecular Life Sciences*, 2005, **62**(17): 1984-1995.

COMPARATIVE GENOMIC IDENTIFICATION AND ANALYSES OF EPIDERMAL DIFFERENTIATION COMPLEX GENES IN TURTLES

CHEN Meng-Rao^{1,2}, LI Fang^{1,2}, CHEN Qi-Ran^{1,2}, HUANG Bao-You^{1,2}, SUN Wei^{1,2}, QIAN Guo-Ying^{1,2},
GE Chu-Tian^{1,2} and WANG Zong-Ji^{1,2}

(1. Institute of Animal Sex and Development, Zhejiang Wanli University, Ningbo 315100, China; 2. College of Biological and Environmental Sciences, Zhejiang Wanli University, Ningbo 315100, China)

Abstract: The epidermal differentiation complex (*EDC*) is a cluster of genes highly associated with keratinocytes differentiation and epidermal keratinization. This gene cluster plays an important role in the differentiation of epidermal tissues in amniotes, the formation of various skin appendages, and the keratinization process. Therefore, exploring the sequence characteristics and evolution of *EDC* genes in turtles can help elucidate the genetic basis and evolutionary history of their shell epidermis traits. Currently, there is a lack of systematic comparative analysis of the sequence characteristics and evolutionary levels of *EDC* genes across different turtle groups. This study performed comparative genomics identification of *EDC* in 28 turtles, including chromosomal localization, amino acid composition, domain analysis, protein spatial structure analysis, and the reconstruction of evolutionary history. The *EDC* genes in turtles exhibit the same four gene families found in other reptiles, with members of the *SEDC* gene family expanding through multiple tandem repeat events. The amino acid composition of turtle *SEDC* proteins is highly diverse, with *SEDC* and *SFTP* protein amino acid content showing group-specific characteristics (hard-shelled, soft-shelled, leatherback), which supports the diversity of shell epidermal structures in turtles. The functional domains and protein spatial structures of *S100A* and *SFTP* are highly conserved across turtle species, with motif 3 of *PGLYRP3* showing a loss specific to the family Trionychidae. The turtle-specific *EDPCV* originates from *EDPI*, and mutations have enriched cysteine and valine, facilitating gene expansion through tandem duplication. This study provides a data foundation and theoretical basis for further research on the role of *EDPCV* in turtles.

Key words: *EDC*; Amino acid content; Gene expansion; *EDPCV* gene; Turtles

附录 S1 本项目研究龟鳖目物种的基因组信息

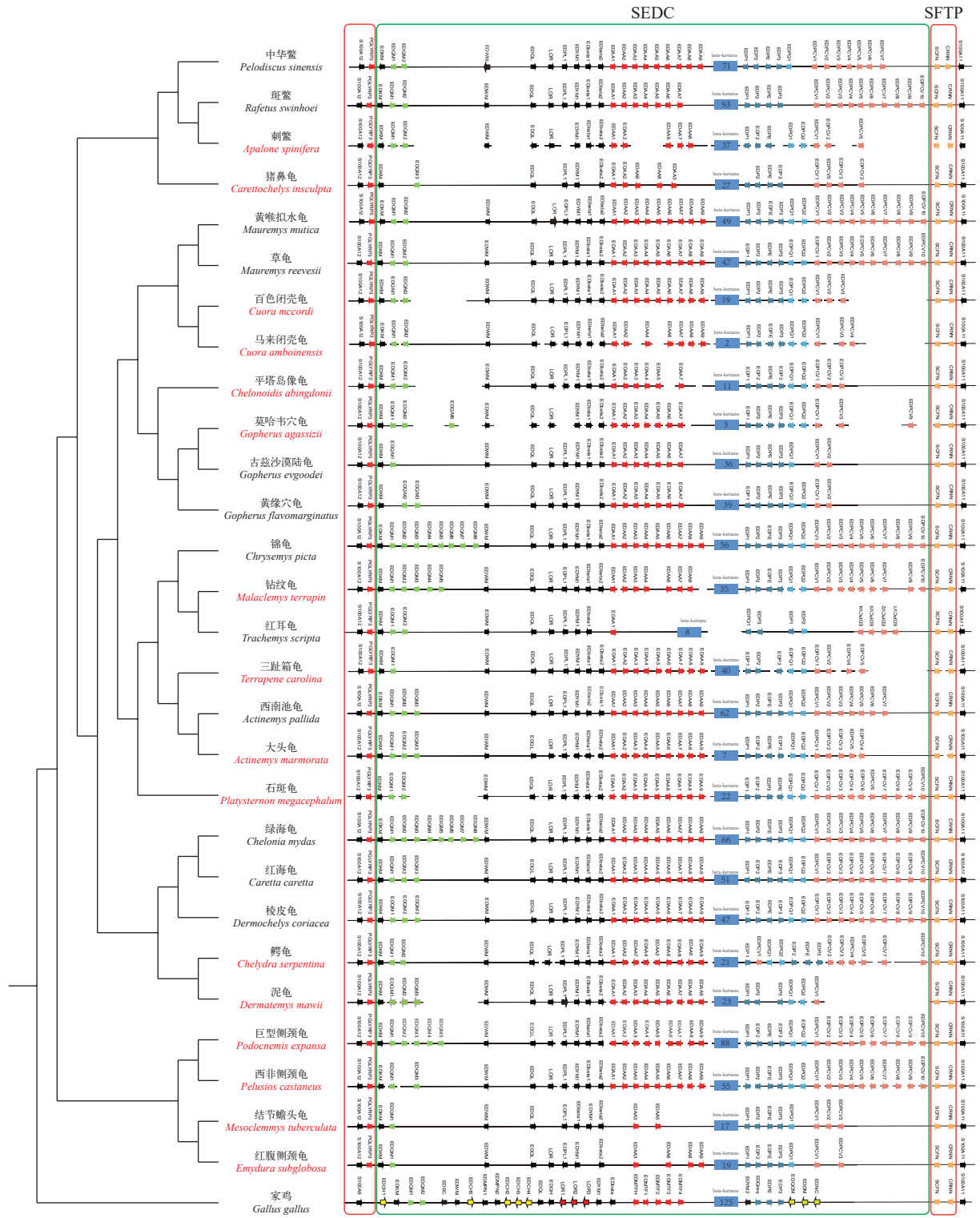
Appendix S1 Information of turtle genomes studied in this study

编号	物种名	拉丁名	基因组长度 (Mb)	Scaffold N50 (bp)	基因组组装水平	查询号
1	石斑龟	<i>Actinemys marmorata</i>	2388	13640393	Scaffold	GCA_009430475.1
2	西南池龟	<i>Actinemys pallida</i>	2326	140349533	Chromosome	GCA_023634205.1
3	刺鳖	<i>Apalone spinifera</i>	1931	2306994	Scaffold	GCA_000385615.1
4	红海龟	<i>Caretta caretta</i>	2134	130956235	Chromosome	GCF_023653815.1
5	猪鼻龟	<i>Carettochelys insculpta</i>	2356	45881824	Scaffold	GCA_007922185.1
6	绿海龟	<i>Chelonia mydas</i>	2134	134428053	Chromosome	GCF_015237465.1
7	平塔岛象龟	<i>Chelonoidis abingdonii</i>	2300	1277207	Scaffold	GCF_003597395.1
8	鳄龟	<i>Chelydra serpentina</i>	2401	21135443	Scaffold	GCA_007922165.1
9	锦龟	<i>Chrysemys picta bellii</i>	2481	20675201	Chromosome	GCA_000241765.5
10	马来闭壳龟	<i>Cuora amboinensis</i>	2214	247606	Scaffold	GCA_004028625.2
11	百色闭壳龟	<i>Cuora mccordi</i>	2390	32628679	Scaffold	GCA_003846335.1
12	泥龟	<i>Dermatemys mawii</i>	1866	34365009	Scaffold	GCA_007922305.1
13	棱皮龟	<i>Dermochelys coriacea</i>	2164	137568771	Chromosome	GCF_009764565.2
14	红腹侧颈龟	<i>Emydura subglobosa</i>	1986	44759016	Scaffold	GCA_007922225.1
15	莫哈韦穴龟	<i>Gopherus agassizii</i>	2184	228109	Scaffold	GCA_002896415.1
16	古兹沙漠陆龟	<i>Gopherus evgoodei</i>	2298	147425149	Chromosome	GCF_007399415.2
17	黄缘穴龟	<i>Gopherus flavomarginatus</i>	2460	162938067	Chromosome	GCF_025201925.1
18	钻纹龟	<i>Malaclemys terrapin</i>	2439	437266	Contig	GCA_001728815.2
19	黄喉拟水龟	<i>Mauremys mutica</i>	2484	135065811	Chromosome	GCF_020497125.1
20	草龟	<i>Mauremys reevesii</i>	2367	130473029	Chromosome	GCF_016161935.1
21	结节蟾头龟	<i>Mesoclemmys tuberculata</i>	2027	46415247	Scaffold	GCA_007922155.1
22	西非侧颈龟	<i>Pelusios castaneus</i>	2044	14055032	Scaffold	GCA_007922175.1
23	大头龟	<i>Platysternon megacephalum</i>	2319	7221511	Scaffold	GCA_003942145.1
24	巨型侧颈龟	<i>Podocnemis expansa</i>	2447	37101370	Scaffold	GCA_007922195.1
25	斑鳖	<i>Rafetus swinhoei</i>	2237	131976892	Chromosome	GCA_019425775.1
26	三趾箱龟	<i>Terrapene carolina triunguis</i>	2571	24249581	Scaffold	GCF_002925995.2
27	红耳龟	<i>Trachemys scripta elegans</i>	2126	140411086	Chromosome	GCF_013100865.1
28	中华鳖	<i>Pelodiscus sinensis</i>	2202	3400000	Scaffold	GCA_000230535.1

附录 S2 本项目研究龟鳖EDC基因缩写及全名

Appendix S2 Tentative abbreviations and full names of EDC genes in turtle

基因名缩写Gene name abbreviation	EDC基因全称Full names of EDC genes
S100A12	S100 calcium binding protein A12
PGLYRP3	Chrysemys picta Pglyrp3 Peptidoglycan recognition protein 3
EDKM	Epidermal Differentiation protein containing a KKLIQQ Motif
EDQM1	Epidermal Differentiation protein containing a glutamine (Q) Motif 1
EDQM2	Epidermal Differentiation protein containing a glutamine (Q) Motif 2
EDQM3	Epidermal Differentiation protein containing a glutamine (Q) Motif 3
EDWM	Epidermal Differentiation protein containing WYDP Motif
EDQL	Epidermal Differentiation protein rich in glutamine (Q), close to Loricrin
LOR	Loricrin
EDPL1	Epidermal Differentiation Proline-rich protein, close to Loricrin, 1
EDYM1	Epidermal Differentiation protein containing Y Motif 1
EDbeta	Epidermal Differentiation protein beta-keratin
EDAA1	Epidermal Differentiation protein rich in Aromatic Amino acids 1
EDAA2	Epidermal Differentiation protein rich in Aromatic Amino acids 2
EDAA3	Epidermal Differentiation protein rich in Aromatic Amino acids3
EDAA4	Epidermal Differentiation protein rich in Aromatic Amino acids4
EDAA5	Epidermal Differentiation protein rich in Aromatic Amino acids5
EDAA6	Epidermal Differentiation protein rich in Aromatic Amino acids6
EDAA7	Epidermal Differentiation protein rich in Aromatic Amino acids7
EDAA8	Epidermal Differentiation protein rich in Aromatic Amino acids8
EDAA9	Epidermal Differentiation protein rich in Aromatic Amino acids9
EDAA10	Epidermal Differentiation protein rich in Aromatic Amino acids10
EDP1	Epidermal Differentiation protein rich in Proline 1
EDP2	Epidermal Differentiation protein rich in Proline 2
EDP3	Epidermal Differentiation protein rich in Proline 3
EDPE	Epidermal Differentiation protein rich in Proline and glutamic acid (E)
EDPQ1	Epidermal Differentiation protein rich in Proline and glutamine (Q) 1
EDPQ2	Epidermal Differentiation protein rich in Proline and glutamine (Q) 2
EDPCV1	Epidermal Differentiation protein rich in Proline, Cysteine and Valine 1
EDPCV2	Epidermal Differentiation protein rich in Proline, Cysteine and Valine 2
EDPCV3	Epidermal Differentiation protein rich in Proline, Cysteine and Valine 3
EDPCV4	Epidermal Differentiation protein rich in Proline, Cysteine and Valine 4
EDPCV5	Epidermal Differentiation protein rich in Proline, Cysteine and Valine 5
EDPCV6	Epidermal Differentiation protein rich in Proline, Cysteine and Valine 6
EDPCV7	Epidermal Differentiation protein rich in Proline, Cysteine and Valine 7
EDPCV8	Epidermal Differentiation protein rich in Proline, Cysteine and Valine 8
EDPCV9	Epidermal Differentiation protein rich in Proline, Cysteine and Valine 9
EDPCV10	Epidermal Differentiation protein rich in Proline, Cysteine and Valine 10
SCFN	Scaffoldin
CRNN	Cornulin
S100A11	S100 calcium binding protein A11

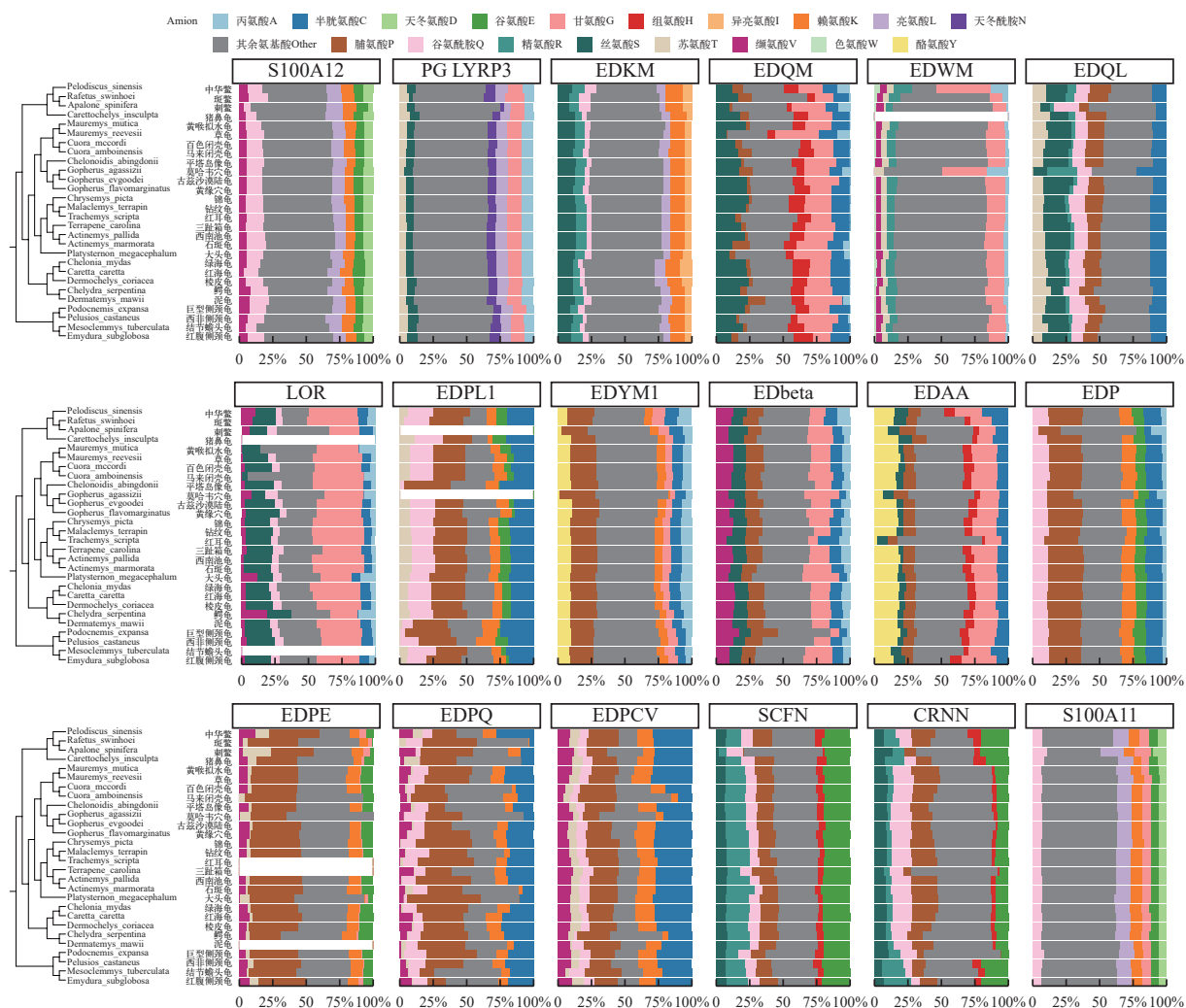


附图 S1 龟鳖目物种EDC基因鉴定结果与染色体线性分析

Appendix Fig. S1 *EDC* gene identification results and chromosome linearity analysis of turtles species

左边红色物种表示基因组是非染色体水平；黑色物种是染色体水平；绿色方框表示*SEDC*基因，红色方框表示*SFTP*基因；箭头方向表示正负链，不同颜色箭头表示不同类型基因

The red species on the left indicates that the genome is at the non-chromosomal level, while the black species represents the chromosomal level; Green boxes denote *SEDC* genes, whereas red boxes signify *SFTP* genes; The direction of the arrows indicates the positive and negative strands, with arrows of different colors representing various types of genes



附图 S2 28 种龟鳖物种 EDC 各个蛋白氨基酸含量分析

Appendix Fig. S2 Protein amino acids content analysis of EDC protein in 28 turtle species

左图描绘了龟鳖物种的系统发育树，右图展示了 18 种 EDC 蛋白中含量最高的 6 种氨基酸的百分比组成。术语“其他”是指所有剩余氨基酸的累积百分比

The left image depicts the phylogenetic tree of turtle species, while the right image illustrates the percentage composition of the six amino acids with the highest content in 18 EDC proteins; The term ‘Other’ refers to the cumulative percentage of all remaining amino acids. The protein data are shown in the order of the corresponding genes in the EDC (Fig. 1)

A >Chrysemys_picta_CRNN
MTQLLSNKGIIINAFYVFAKKDGCPTLSKGLRQLIHQEFADVTVPQGLQTIDKLLQLLDTSDGRDLDFNGFLVLIFQVAKACYGEVSQG
QRPCHGGSSASQGEANCERTKEPP

TPERDPSRQAPEPQ
TPERDSIPCQAPEPQ
IPEQDSSPCQAPELQ
IPEQDSNPCQAPEAQ
IPERDPSPCQAPELQ
TPEODSSPRQAPEPQ
IPEQDSSPRQAPEPQ
IPEQDSIPCQAPEAQ
IPERDPSPHQAPEPQ
TPERDPSPHQGLEPQ
TPERDPSCHQAPEPQ
IPEQDPSLHQAPEPQ
TPEODSSSRHITPEPQ
IPEQEPSPCRGPEPQ
TPEQDLSPSQAPEPQ

TPEQDLSHSETQLPPTQQRNQGADPTEPAAQASKSSQCLYSWHSQKPRPFPHWWPP

>Rafetus_swinhoei_CRNN
MPQLLDSISTISVFYTHGEKAEGCSTISRREMKRFLRREFADITLTPYDPPTIEAVLQLLDHDEDGAVDFNEFLLLVFRVAKVCYWYLPQK
QRLSPRAEKGPNKGQGAPEGGRAEGSREQP

UEERSYQQPLDPE
PREEEWSLHQPHEPE
PRGEERSHCQPHELE
PRKEERRHQPREFE
LREDEQSRHQPREPE
PKGEERSLHQPHEPE
PREEERSYQQPLEPEP
RGEEQSHHQPLDPE
PRKEEQSLHQPPEPK
PREEEQSRHQPREPE
PREGERSHQPSKPE
PRQEESHHQPREQE
PRKDTGSHHQPREPE
PREEEKSHCQPREPE
PREEQRSCHQPEPE
PREEEKS

B >Chrysemys_picta_SCFN
MPQLLDSISTISVFYKPGKKDEDCSTISKREVKRFIQGFADITVNPYDAHAIEAVLQL
LDHDDGGVEFNEFLLLVFRVAKVCYWYLPQKQRLPQRTETGLSGERRQEPAGRAEGSRDQ

PKPLGTEKGYETWEPETRETERSRRQPRE
PEPRGDERANYETYE
PEKREEERRRRQPHG
PEPRRDERSWYGTQRSEHRENEERSHHRSC
PEPRGDERSRYERRDLETREVDRRRHPHE
RERREDERSHQPYEPPEQSRERRCLEPESRQDEKSGYWPRESEERSYETCE
PEAREEERSHRQPREPEPRRDERSRPHQPEPQGDERSHYRAHA
PETREERSRCPQQQPEPRRDERSRYEIYDHTREEERSRCHPHE
PEHEEERSHRQPREPEPEHEEERSADSHPRK
PESREDEGSRRQPREPEPREEERSRRQPREPEPREDEGSRRQPRE
PEPRDEGSRRQPREPEPREDEERRRRQPHPEPREEDNRSRRQPHG
PELRGDCWSRHQSCPEAREGERSRCEPCEPEQREEKSRHQPRE
PQWREDEGSRRQPREPEPREDEGAADSHDERSRRQPREPEPREDEGSRRQPRE
PEPREDEGSRRQPRE
PEPREGERCSQSQTPEPHHEGNLPPSSEPVENEGTLREPHDPESTDSSRRCPATHEPPPTGDEGSQLPRESAAPGDEGNPQKTSESEPREDDGSQ
POLCEPEQKGGVLRQTPESQPLEEVESQEPQPRGPEPRTEGSRHQPPQGGGEASHLQPDIEPQ
GDGSRHRPRDPQEERERSHQAREPLPRLGEEPEGEWSQSHPPEP
ANAEETPGETDDVYKASLPCNPLYVYLLEQTVEKQLYLAPPHQDHP

>Rafetus_swinhoei_SCFN
MPQLLDSISTISVFYTHGEKAEGCSTISRREMKRFLRREFADITLTPYDPPTIEAVLQL
LDHDEDGAVDFNEFLLLVFRVAKVCYWYLPQKQRLSPRAEKGPNKGQGAPEGGRAEGSREQ

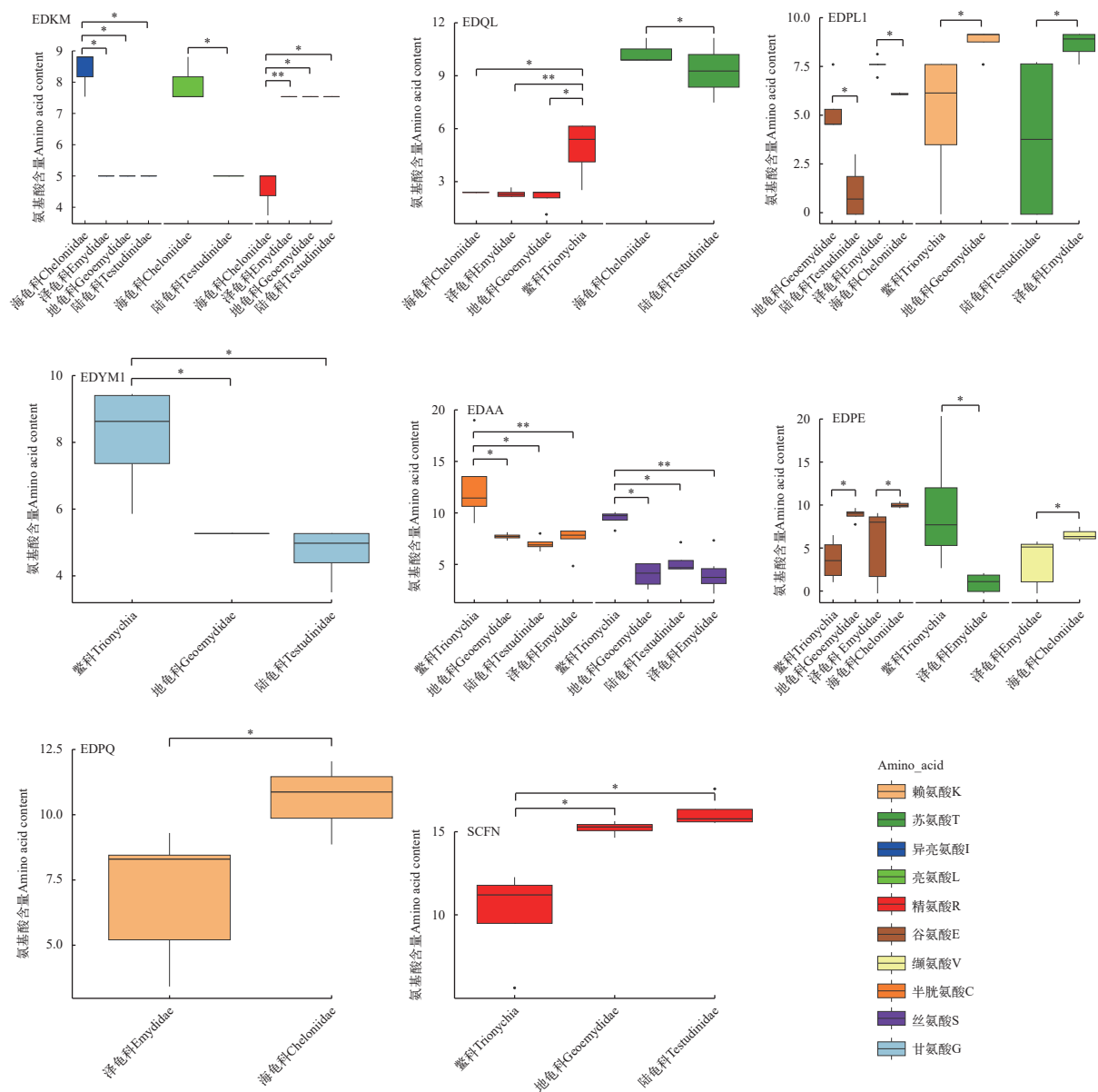
PCELEPRGAEGGKGYEMPEPEARETEQSCQRSHRQPRE
PEPRGEERSLHQPPE
PEPKGEERSHRQPRE
PEPRGEERSLHQPWPEPKGEERSLHQTRE
PEPKGEERTRYQPHPETIEKERSLHKPHE
PELRRREERSYRQPLDPEPREEEWSLHRPHELEPRKEERRHQPREFEPRKDEQSLHQPPE
PEPKGEERSHRQPREPEPREEERSLCKPHEPELKEEEQSYRQPLE
PEPRGEERSHRQPHLESRKEERSHQPRELEPREEERTRHQPRE
PETIEKERSLHKPHEPELREERSYQQPLDPEPREEEWSLHQPHE
PEPRGEERSHCQPHELEPRKEERRHQPREFELREDEQSRHQPRE
PEPKGEERSLHQPHEPEPREEERSYQQPLEPEPRGEEQSHHQPLD
PEPRKEEQSLHQPPEPKPREEEQSRHQPREPEPREGERSHHQPSK
PEPRQEESHHQPREPEPRKDTGSHHQPREPEPREEEKSHCQPRE
PEPREEQRSCHQPPPEPREEEKSCROPPPE
PEAGVEEVYQQPGGEPYQEEGHLPPWREPEAREDAQTORDADAPGDHTRCRAMPKPSPTGEVGIPLQPWDSLAGGEGSRQAATDSERREDSQ
POLHEPEHREGDGLRETPESEPPQAAAESQHQPCLDEEQREGSLQPPPPGGEGSCLPLDTEPPQ
ADRSLLPCKEPQEDASKRSSPQARELQPWLGAPQEGEWSQSHPEP
SSQEETPGETAEDTKACLCNPLYVYLLEQTVEKQLYLAPPHQEWPP

附图 S3 龟鳖SFTP蛋白包含重复序列

Appendix Fig. S3 Turtles SFTP protein contains repeated sequences

班鳖和锦龟两个代表性物种，展示了SFTP蛋白包含的重复序列，A代表CRNN蛋白序列，B表示SCFN蛋白序列；重复序列用蓝色矩形框标注

Two representative species, the *Rafetus swinhoei* and *Chrysemys picta*, show the repetitive sequences contained in the SFTP protein, (A) represents the CRNN protein sequence, and (B) represents the SCFN protein sequence; Repeated sequences are marked with a blue marker

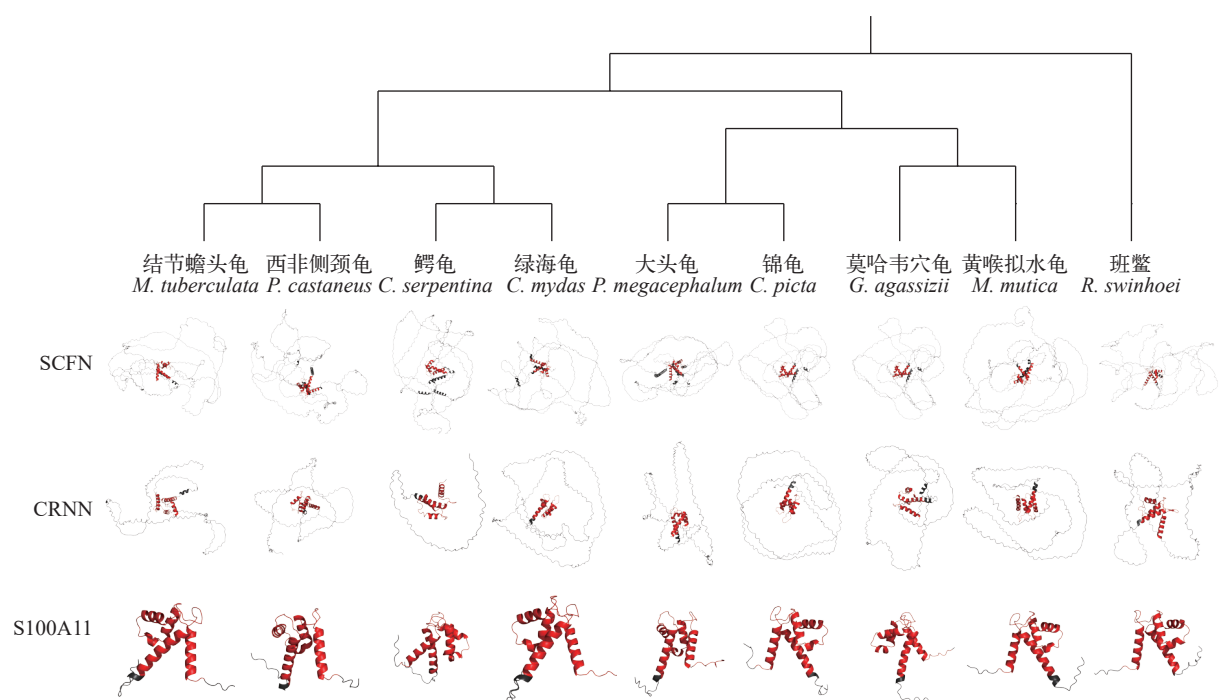


附图 S4 龟鳖类群EDC蛋白序列氨基酸显著差异分析

Appendix Fig. S4 Analysis of significant amino acid differences in EDC protein sequences of turtle Family

箱线图说明了不同家族中 EDC 蛋白氨基酸序列的差异，突出了统计显著性。一个星号(*)表示两个家族之间的氨基酸含量差异在 $0.01 < P < 0.05$ ，两个星号(**)表示两个家族之间的氨基酸含量差异为 $0.001 < P < 0.01$

The boxplot illustrates the differences in amino acid sequences of the EDC protein among various turtle families, highlighting statistical significance; An asterisk (*) indicates that the difference in amino acid content between the two families falls within the range of $0.01 < P < 0.05$, while two asterisks (**) signify that the difference in amino acid content between the two families is $0.001 < P < 0.01$

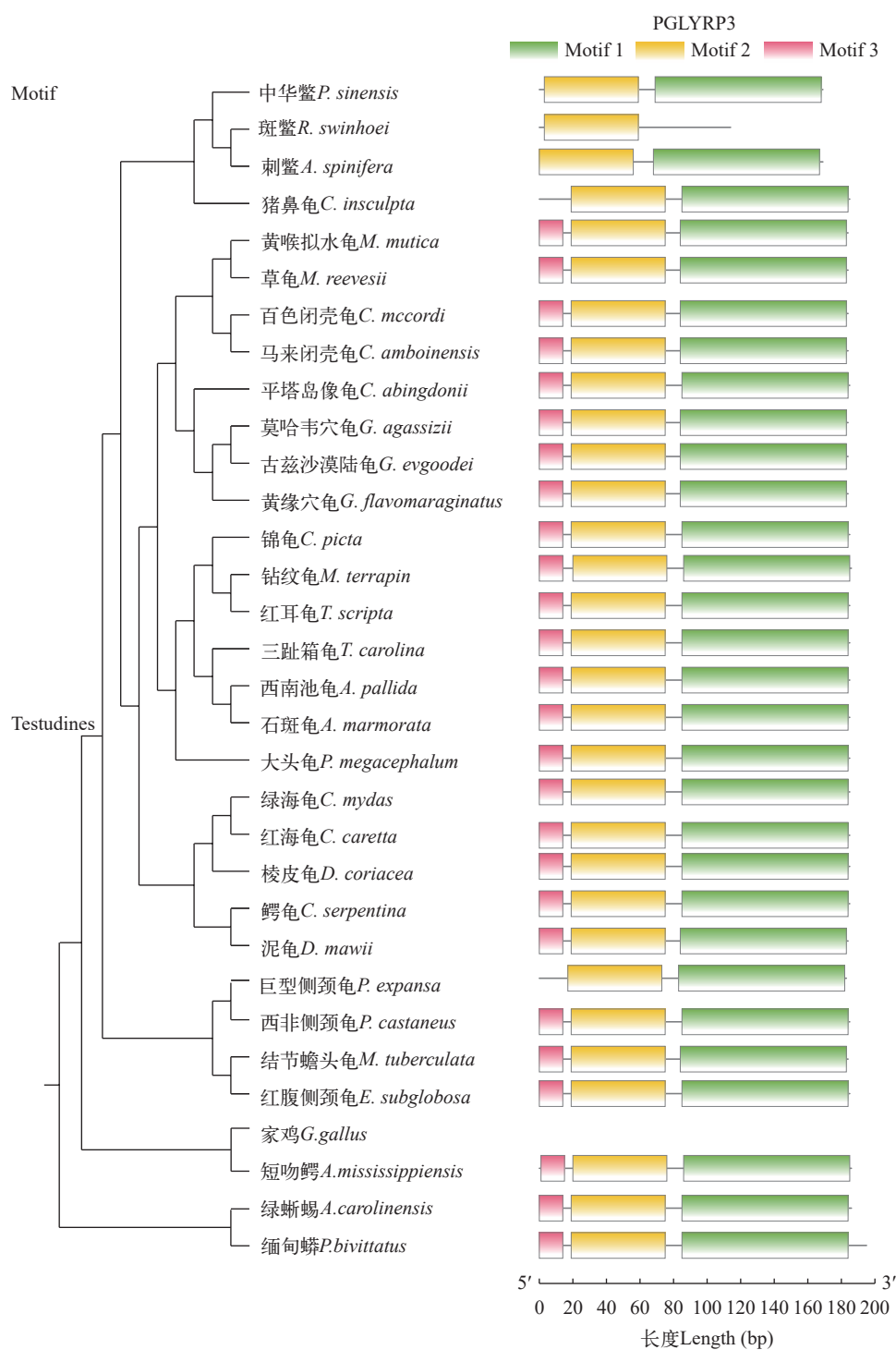


附图 S5 SFTP和S100A11蛋白结构空间模型和结构域

Appendix Fig. S5 Structural spatial models and domains of SFTP protein and S100A11 protein

上半部分展示了九个龟鳖类的代表物种树，下半部分描述了SFTP蛋白和S100A11蛋白的空间结构; 红色 α 螺旋突出显示了 S100 结构域

The upper section illustrates the representative species tree of nine turtle species, while the lower section depicts the spatial structures of the SFTP protein and the S100A11 protein; The red α helix highlights the S100 domain



附图 S6 龟鳖目和其它爬行动物PGLYRP3蛋白的motif预测

Appendix Fig. S6 Motif prediction of PGLYRP3 gene in Turtles and other reptiles

左图显示了本研究中分析的龟鳖和其它爬行动物的物种树; 右边展示了PGLYRP3蛋白motif结果

Left shows a species tree of turtles and other reptiles analyzed in this study; Right side shows the PGLYRP3 protein motif results

