

异齿裂腹鱼 (*Schizothorax oconnori*) CLOCK 蛋白结构和功能的生物信息学分析

周建设^{1,2}, 张 驰¹, 王万良¹, 曾本和¹, 王且鲁¹, 王金陵¹

(1. 西藏自治区农牧科学院水产科学研究所, 西藏 拉萨 850032; 2. 河南农业大学牧医工程学院, 河南 郑州 450000)

摘 要:为探讨异齿裂腹鱼 CLOCK 蛋白的结构、功能信息, 本文通过生物信息学的方法对异齿裂腹鱼 CLOCK 蛋白的理化性质、亚细胞定位、跨膜区和信号肽、二级结构、三级结构、蛋白互作网络、系统进化等进行分析。异齿裂腹鱼 CLOCK 蛋白分子式为 $C_{4389}H_{6978}N_{1278}O_{1405}S_{37}$, 相对分子量 101315.62 Da, 氨基酸含量最高的是谷氨酰胺为 146 个, 半衰期为 30 h, 不稳定系数为 68.37, 属不稳定蛋白, 理论等电点 (PI) 为 6.44, 属于酸性蛋白质, 脂肪系数为 69.96, CLOCK 属于亲水性蛋白。蛋白均定位于细胞核内, 不存在信号肽序列和跨膜区, 存在一个核定位信号序列为“KAKRVSRNKSEKKRR”。异齿裂腹鱼 CLOCK 蛋白具有疏松的结构特征, 含有 111 个磷酸化位点, CLOCK 蛋白与多个时钟相关的蛋白互作, 在物种进化上与金线鲃属鱼类 CLOCK 蛋白具有较高的同源性。

关键词:异齿裂腹鱼; CLOCK 蛋白; 生物信息学

中图分类号:S965.199 **文献标识码:**A

Bioinformatic Analysis of Structure and Function of *Schizothorax oconnori* CLOCK Protein

ZHOU Jian-she^{1,2}, ZHANG Chi¹, WANG Wan-liang¹, ZENG Ben-he¹, WANG Qie-lu¹, WANG Jin-lin¹

(1. Institute of Fisheries Science, Tibet Academy of Agricultural and Animal Husbandry Sciences, Tibet Lhasa 850032, China; 2. College of Animal Science and Veterinary Medicine, Henan Agricultural University, Henan Zhengzhou 450000, China)

Abstract: To discuss the structure and functional information of *Schizothorax oconnori* CLOCK protein, in this paper, the physicochemical property, subcellular localization, transmembrane domain, signal peptides, secondary structure, tertiary structure, protein interaction network and phylogenetic of *Schizothorax oconnori* CLOCK protein were analyzed. The molecular formula of CLOCK protein was $C_{4389}H_{6978}N_{1278}O_{1405}S_{37}$ with 101315.62Da relative molecular weight. The Gln (Q) was the highest content with 146 number in all amino acid of CLOCK protein and had 30 hours half-life period, 68.37 coefficient of instability, 6.44 isoelectric point, 69.96 coefficient of fat. The protein was acid, instability and hydrophilic and was located in nuclear, there was no signal peptide sequence and transmembrane domain, nuclear localization signal sequence was ‘KAKRVSRNKSEKKRR’. *Schizothorax oconnori* CLOCK protein with architectural feature of loose contained 111 phosphorylation sites, and interacted with multi-the clock protein, in the evolution of species, the CLOCK protein of *Sinocyclocheilus grahami* has a high homology.

Key words: *Schizothorax oconnori*; CLOCK protein; Bioinformatics

CLOCK 蛋白是可以调控动物生活特性的潜在的候选基因, 其等位基因的变异影响许多生物的繁殖有关^[1], 在雌性老鼠中, 编码 CLOCK 蛋白的 *clock* 基因突变可以导致异常的动情周期^[2], Leder 等人^[3]将 *clock* 基因的一个拷贝定位到一个数量性状

位点 (QTL) 区域, 该区域解释了雌虹鳟鱼产卵日期变化的 20 % ~ 50 % (取决于父母的贡献)。O'Malley 等人^[4]鉴定出 *clock* 基因的变异拷贝 (等位基因), 在秋季和春季产卵的大鳞大麻哈鱼之间, 频率显著不同。因此, *clock* 基因被认为是与动物繁殖等生活史特性有关季节性节律潜在的目标基因。*clock* 基因为节律振荡器编码出潜在的 CLOCK 蛋白, 在 CLOCK 蛋白的氨基酸序列中存在多个保守序列, 包括 C 端富含谷氨酰胺 (Q) 的结构域^[5], 这个谷氨酰胺富集区起着转录激活的作用, 同时 C 端谷氨酰胺重复区域的长度决定了 CLOCK 蛋白的转录

收稿日期: 2019 - 11 - 18

基金项目: 西藏自治区自然科学基金项目“clock 基因对异齿裂腹鱼繁殖行为微进化的响应研究” (XZ2019ZRG-103)

作者简介: 周建设 (1985 -), 男, 副研究员, 研究方向: 鱼类遗传育种与繁殖, E-mail: zjianshe@163.com。

激活潜能^[6],因此,它可以影响与动物昼夜节律相关的行为和生理功能^[7-8]。CLOCK 蛋白因 Poly-Q 片段的发现为研究与昼夜节律有关的动物生活史特性与基因之间的联系提供了切入点^[9-10]。有些物种的 Ploy-Q 长度随纬度梯度的变化而变化^[11],重要的是,Poly-Q 长度和纬度梯度变化与繁殖启动的时间相关,纬度越高,Poly-Q 长度相对较长。异齿裂腹鱼藏语俗称“棒棒鱼”,广布于雅鲁藏布江,在中国脊椎动物红色名录中被评估为“无危种”^[12]。跟其他裂腹鱼一样,不同栖息地种群产卵旺季不同,随雅鲁藏布江海拔的变化而变化,海拔越高,产卵时间越晚。栖息在高原的鱼类,种群内个体之间的繁殖开始时间存在明显差异。理解这种差异的遗传基础对我们理解鱼类能否通过微进化应对气候变化非常重要,CLOCK 蛋白 C 端 Poly-Q 的长度与繁殖启动时间相关,编码该蛋白的基因也是调控动物生活特性的潜在的候选基因。因此,本文通过生物信息学方法分析异齿裂腹鱼 CLOCK 蛋白的结构和功能,为进一步研究该蛋白在异齿裂腹鱼繁殖启动时间中的作用提供参考。

1 材料与方法

1.1 材料

对 6 尾异齿裂腹鱼卵巢进行转录组测序 (PRJ-NA527715、PRJNA527627),共获得 3 代测序数据 15G,2 代测序数据 6G。根据 NR、GO、KEGG、等数据库的注释结果和同源物种 *clock* 基因的比对,截取异齿裂腹鱼 *clock* 基因的全长 CDS 序列,通过 DNAMAN 软件将 *clock* 基因翻译成 CLOCK 蛋白用于后

续预测。

1.2 方法

登录在线数据库 (Expert Protein Analysis System: <https://www.expasy.org/b>),运用其中的 ProtParam 工具分析 CLOCK 的理化性质;ProtScale 工具分析异齿裂腹鱼 CLOCK 的亲疏水性性质;PORT II 分析 CLOCK 的亚细胞定位;Signal P4.0 分析 CLOCK 信号肽;NLStradamus 工具预测该蛋白的核定位信号;TMHMM2.0 分析 CLOCK 蛋白的跨膜区域;SOPMA 工具分析 CLOCK 蛋白的二级结构,SWISS-MODEL 预测蛋白的三级结构;Net Phos3.1 在线预测 CLOCK 蛋白 N-糖基化位点;通过 STRING 数据库构建 CLOCK 的蛋白质间网络图;JASPAR 预测可能的转录因子结合位点;通过 NCBI Blast 在线选择比对序列,用 MEGA6.0 构建进化树。

2 结果与分析

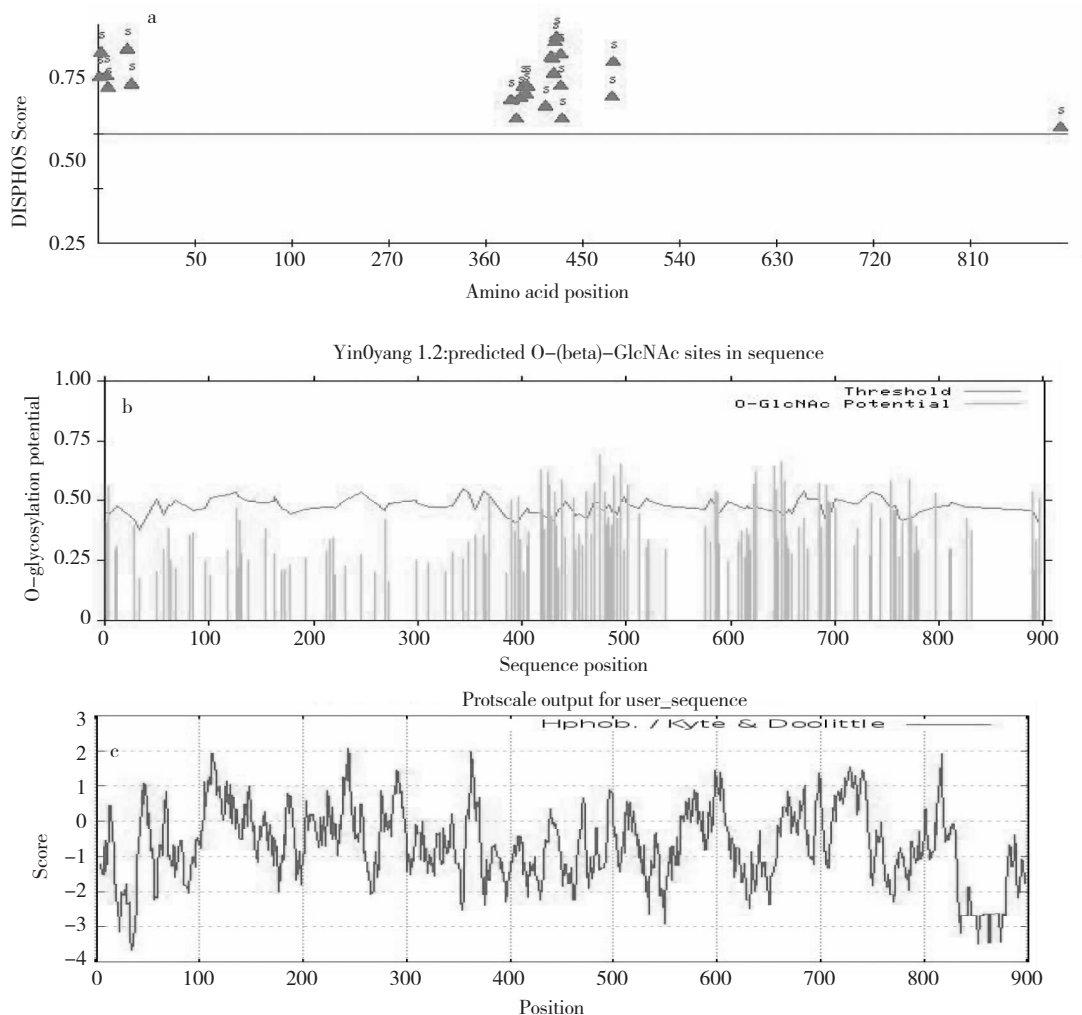
2.1 CLOCK 的理化性质分析

通过 ExPASy 数据库中 ProtParam 工具分析,CLOCK 蛋白分子式为 $C_{4389}H_{6978}N_{1278}O_{1405}S_{37}$,相对分子量 101315.62Da,其氨基酸组成见表 1,其中含有谷氨酰胺 146 个,含量最高,占到整个序列的 14.2%,没有吡咯氨基酸和硒半胱氨酸,色氨酸的含量最低,为 0.6%,CLOCK 的半衰期为 30 h,不稳定系数为 68.37,属不稳定蛋白,带负电荷的氨基酸 (Asp + Glu) 为 80 个,带正电荷的氨基酸 (Arg + Lys) 为 73 个,理论上的等电点 (PI) 为 6.44,属于酸性蛋白质,脂肪系数为 69.96,亲水性平均值为 -0.679。有 24 个丝氨酸磷酸化修饰位点 (图 1a),存在多个

表 1 异齿裂腹鱼 CLOCK 蛋白氨基酸组成

Table 1 The amino acid composition of CLOCK protein of *Schizothorax oconnori*

氨基酸 Amino acid	数目 Number	含量 (%) Content	氨基酸 Amino acid	数目 Number	含量 (%) Content
Ala (A)	57	6.3	Lys (K)	34	3.8
Arg (R)	39	4.3	Met (M)	28	3.1
Asn (N)	23	2.6	Phe (F)	29	3.2
Asp (D)	32	3.6	Pro (P)	51	5.7
Cys (C)	9	1.0	Ser (S)	84	9.3
Gln (Q)	146	16.2	Thr (T)	71	7.9
Glu (E)	48	5.3	Trp (W)	5	0.6
Gly (G)	43	4.8	Tyr (Y)	19	2.1
His (H)	26	2.9	Val (V)	39	4.3
Ile (I)	34	3.8	Pyl (O)	0	0.0
Leu (L)	84	9.3	Sec (U)	0	0.0



a 为磷酸化修饰位点预测;b 为 O 型糖基化修饰位点预测;c 为亲疏水性分析

a showed the phosphorylation modification site prediction; b showed the type O glycosylation modification site prediction; c showed the hydrophilic and hydrophobic analysis

图 1 异齿裂腹鱼 CLOCK 蛋白的理化性质分析

Fig. 1 Physical and chemical property analysis of CLOCK protein of *Schizothorax oconnori*

O 型糖基化修饰位点(图 1b)。ExPASy 数据库 Pro-rtScale 工具分析显示,CLOCK 蛋白最强亲水位点在第 34 位谷氨酸,分值 -3.667,疏水性最强位点在第 244 位丙氨酸,分值 2.067,图 1c 显示了 CLOCK 蛋白所有氨基酸的亲疏水性,整体上亲水性氨基酸多于疏水性氨基酸,CLOCK 蛋白属于亲水性蛋白。

2.2 CLOCK 蛋白的亚细胞结构定位分析

通过 PORT II 预测,CLOCK 蛋白定位于细胞外、细胞质、细胞核的可能性分别为 4.3 %、17.4 %、69.6 %,该蛋白定位于细胞核内。

2.3 CLOCK 蛋白的信号肽、核定位信号及跨膜区分析

运用 SignalP 工具在线分析异齿裂腹鱼 CLOCK 蛋白的不存在信号肽序列,运用 NLStradamus 工具预测异齿裂腹鱼 CLOCK 蛋白的核定位信号信号,结果如图 2 显示,存在一个核定位信号序列为

“KAKRVS RNKSEKKRR”。TMHMM 在线预测,异齿裂腹鱼 CLOCK 蛋白序列中无跨膜区。

2.4 CLOCK 蛋白的空间结构预测

SOPMA 预测 CLOCK 的二级结构,结果显示多肽链中 α -螺旋占 42.29 %,无规则卷曲占 39.73 %,延伸链占 13.43 %, β -转角占 4.55 %,其中无规则卷曲的占比最大(图 3a),表明异齿裂腹鱼 CLOCK 蛋白具有疏松的结构特征,为与其他分子的相互作用提供了便利。SWWI-MODEL 在线预测异齿裂腹鱼 CLOCK 蛋白的三级结构如图 3b 所示,三级结构决定了蛋白的性质。

2.5 异齿裂腹鱼 CLOCK 磷酸化位点预测

运用 NetPhos 工具预测异齿裂腹鱼 CLOCK 蛋白磷酸化位点如图 4 所示,结果表明,异齿裂腹鱼 CLOCK 蛋白含有 111 个,其中丝氨酸磷酸化位点最多,为 63 个;苏氨酸磷酸化位点 44 个,酪氨酸磷酸

化位点 4 个。

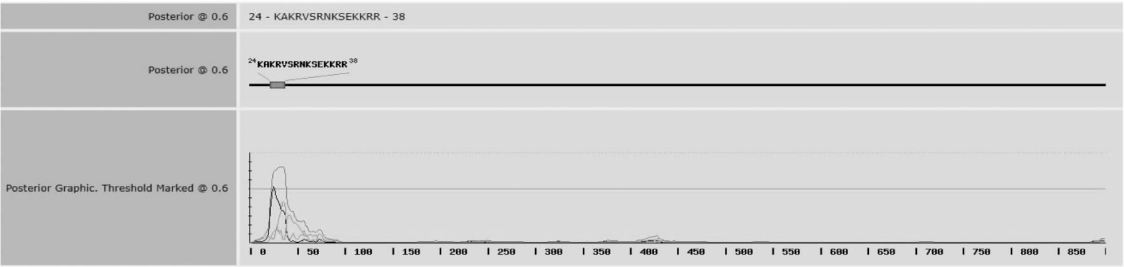
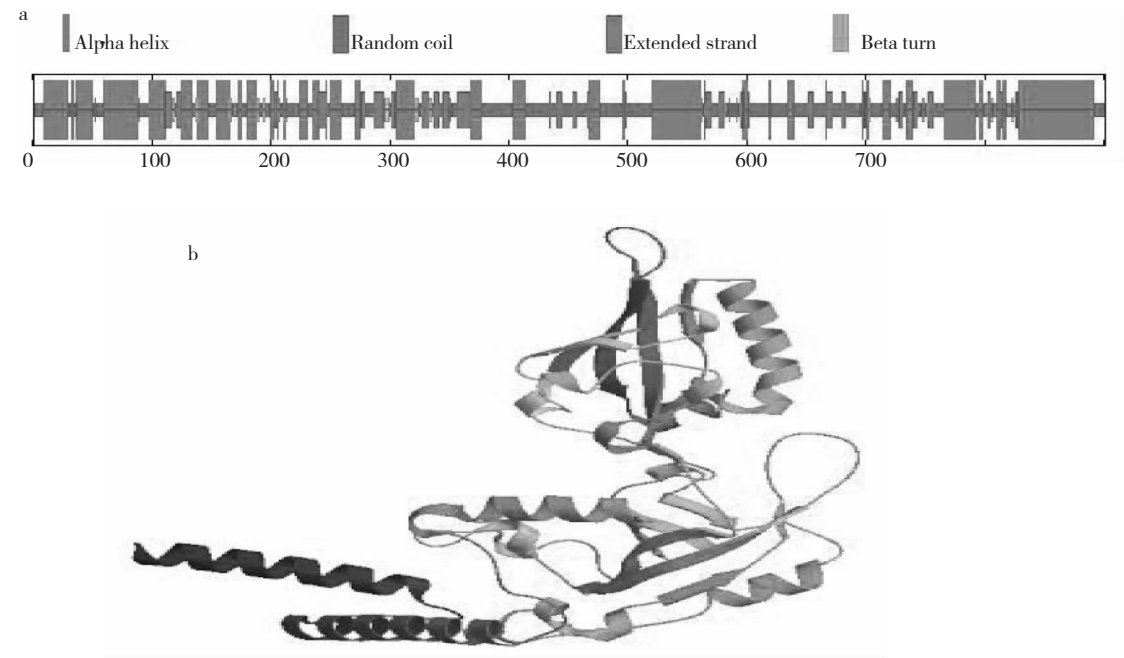


图 2 异齿裂腹鱼 CLOCK 蛋白核定位信号分析

Fig. 2 The protein localization signal analysis of CLOCK protein of *Schizothorax oconnori*



a 为二级结构, b 为三级结构
a showed secondary structure; b showed third structure

图 3 异齿裂腹鱼 CLOCK 蛋白结构分析

Fig. 3 The protein structure analysis of CLOCK protein of *Schizothorax oconnori*

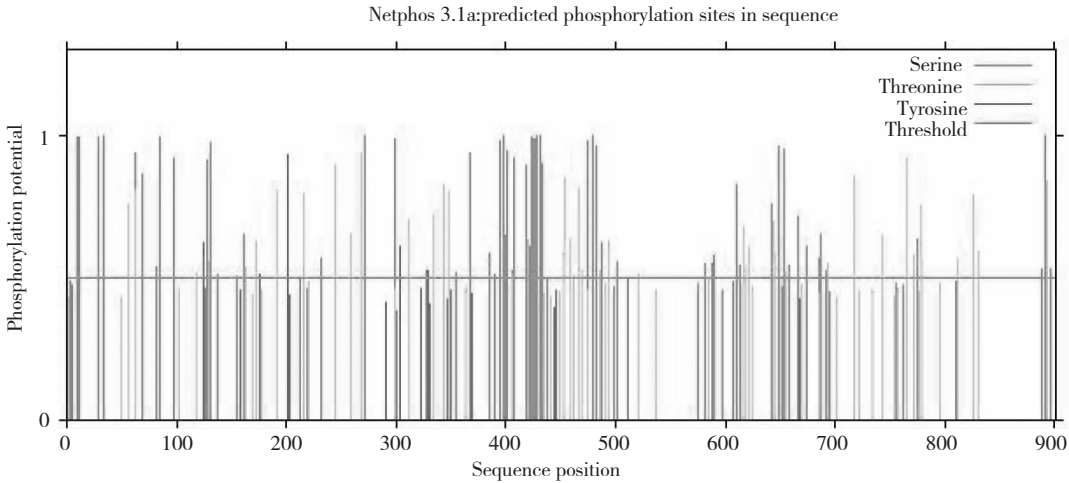


图 4 异齿裂腹鱼 CLOCK 蛋白磷酸化位点预测

Fig. 4 Phosphorylation site prediction of CLOCK protein of *Schizothorax oconnori*

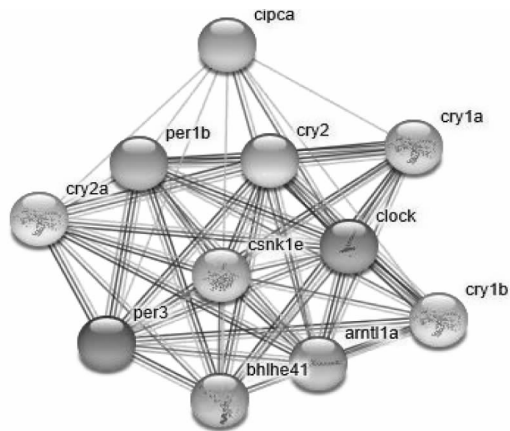


图 5 异齿裂腹鱼 CLOCK 蛋白互作网络

Fig. 5 Protein interaction network of CLOCK protein of *Schizothorax oconnori*

2.6 异齿裂腹鱼 CLOCK 蛋白互作分析

STRING 数据库预测与 CLOCK 蛋白互作的蛋白,构建互作网络,结果见图 5,包括隐花色素基因 (*cry2a*、*cry2*、*cry1a*、*cry1b*) 编码的蛋白,周期基因 (*per3*、*per1b*) 编码的蛋白,酪蛋白激酶 (*csnk1e*) 蛋白,*arntl1a*、*bhlhe41*、*cipca* 所编码的蛋白,这 10 种基因所编码的蛋白与 CLOCK 蛋白的相关系数分别为 0.966、0.979、0.992、0.875、0.850、0.903、0.942、0.919、0.848,均存在较高的蛋白互作关系。

2.7 异齿裂腹鱼 CLOCK 蛋白的系统进化分析

通过 NCBI 中 BLAST 分析,选取金线鲃属、鲫属、坑鱼属、斑马鱼、电鳗属、斑点叉尾鲷、臀点脂鲤属、丽脂鲤属、虱目鱼属、鲱鱼属、眼斑海葵鱼 CLOCK 蛋白的序列构建 NJ 树(图 6),与物种进化程度一致,异齿裂腹鱼 CLOCK 蛋白与金线鲃属鱼类 CLOCK 蛋白具有较高的同源性。

3 讨论

生物节律广泛存在于动物世界中,这种周期性是通过一系列昼夜节律基因及其蛋白的正、负转录-翻译机制来自动调节的^[13-14],例如 *clock* 基因编

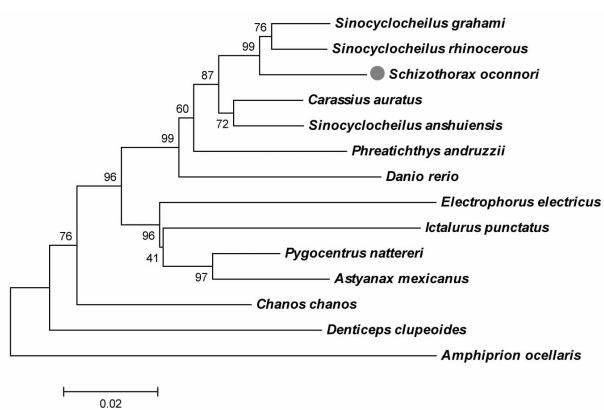


图 6 异齿裂腹鱼 CLOCK 蛋白系统进化树

Fig. 6 The phylogenetic tree of CLOCK protein of *Schizothorax oconnori*

码的 CLOCK 蛋白与另一蛋白 BMAL1 结合形成异二聚体,产生的转录激活复合体是脊椎动物昼夜节律振荡器的核心组成部分^[15-16],本研究在进行 CLOCK 蛋白互作网络分析时发现,CLOCK 蛋白与 *arntl1a* 基因编码的蛋白相互作用,*arntl1a* 基因与编码 BMAL1 蛋白的基因同物异名,同时 CLOCK/BMAL1 复合体还通过转录激活下游的时控基因来调节振荡器的输出^[17],有证据表明,CLOCK 蛋白多态性与虹鳟不同品系间的产卵时间有关^[3]。CLOCK 蛋白的 C 末端,有一个谷氨酰胺富集区,包含一个多聚谷氨酰胺 (Poly-Q) 重复^[18],本研究中异齿裂腹鱼 CLOCK 蛋白的 C 末端与其他鱼类一样也存在一个多聚谷氨酰胺重复,不同属的鱼类其 Poly-Q 长度存在明显差异(图 7)。小鼠和果蝇 *clock* 基因突变的行为和分子特征都表明了 CLOCK 蛋白 C 末端区域的重要性^[15-16]。青藏高原裂腹鱼 6~10 龄即可达性成熟,雄鱼的性成熟早于雌鱼 1~2 年,集群产卵期为每年的 3~5 月^[19],研究证明,CLOCK 蛋白 Poly-Q 的长度受海拔等高原水域生境的影响,不同栖息地裂腹鱼的产卵旺季不同^[20],即使同一栖息地同种裂腹鱼的产卵时间也有先有后,有些物种的 Ploy-Q 影响繁殖启动的时间,纬度越高,Poly-Q

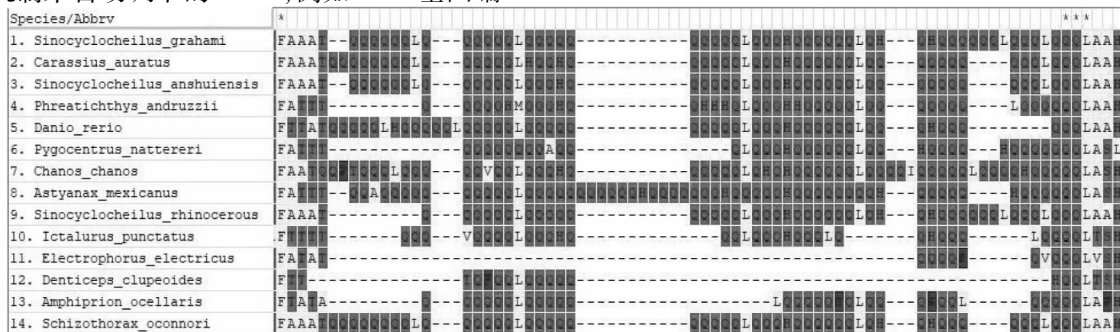


图 7 异齿裂腹鱼 CLOCK 蛋白 C 末端 Poly-Q 结构

Fig. 7 Poly-Q structure at the C-terminal of CLOCK protein of *Schizothorax oconnori*

长度相对较长^[11]。这种产卵日期的变化可能与CLOCK蛋白的多态性有关。本研究分析和预测异齿裂腹鱼CLOCK蛋白的结构和功能,其在编码区存在一个Poly-Q的富集区,推测Poly-Q富集区的变异可能影响异齿裂腹鱼繁殖的启动,这也是我们下一步需要研究的内容(图7)。

4 结 论

本研究基于异齿裂腹鱼转录组数据,获得*clock*基因全长CDS序列并进行了生物信息学分析,异齿裂腹鱼CLOCK蛋白属亲水性蛋白,不存在信号肽序列和跨膜区,存在一个核定位信号序列为“KAK-RVSRNKSEKKRR”,二级结构主要以 α -螺旋和无规则卷曲为主,其CDS区的Poly-Q富集区的变异可能影响异齿裂腹鱼繁殖的启动。

参考文献:

- [1] L. Beaver, B. Gvakharia, T. Vollintine, et al. Loss of circadian clock function decreases reproductive fitness in males of *Drosophila melanogaster*[J]. *Proceedings of the National Academy of Sciences*, 2002,99(4):2134-2139.
- [2] P. E. Chappell, R. S. White, P. L. Mellon. Circadian gene expression regulates pulsatile gonadotropin-releasing hormone (GnRH) secretory patterns in the hypothalamic GnRH-secreting GT1-7 cell line[J]. *Journal of Neuroscience*, 2003,23(35):11202-11213.
- [3] E. Leder, R. Danzmann, M. Ferguson. The candidate gene, Clock, localizes to a strong spawning time quantitative trait locus region in rainbow trout[J]. *Journal of Heredity*, 2006,97(1):74-80.
- [4] K. G. O'MALLEY, M. D. Camara, M. A. Banks. Candidate loci reveal genetic differentiation between temporally divergent migratory runs of Chinook salmon (*Oncorhynchus tshawytscha*)[J]. *Molecular Ecology*, 2007,16(23):4930-4941.
- [5] A. Johnsen, A. Fidler, S. Kuhn, et al. Avian Clock gene polymorphism: evidence for a latitudinal cline in allele frequencies[J]. *Molecular Ecology*, 2007,16(22):4867-4880.
- [6] T. K. Darlington, K. Wager-Smith, M. F. Ceriani, et al. Closing the circadian loop: CLOCK-induced transcription of its own inhibitors per and tim[J]. *Science to Breed Fish*, 1998,280(5369):1599-

- 1603.
- [7] A. Avivi, U. Albrecht, H. Oster, et al. Biological clock in total darkness: the Clock/MOP3 circadian system of the blind subterranean mole rat[J]. *Proceedings of the National Academy of Sciences*, 2001,98(24):13751-13756.
- [8] N. Hayasaka, S. I. LaRue, C. B. Green. In vivo disruption of Xenopus CLOCK in the retinal photoreceptor cells abolishes circadian melatonin rhythmicity without affecting its production levels[J]. *Journal of Neuroscience*, 2002,22(5):1600-1607.
- [9] R. Dor, C. B. Cooper, I. J. Lovette, et al. Clock gene variation in Tachycineta swallows[J]. *Ecology*, 2012,2(1):95-105.
- [10] R. Dor, I. J. Lovette, R. J. Safran, et al. Low variation in the polymorphic Clock gene poly-Q region despite population genetic structure across barn swallow (*Hirundo rustica*) populations[J]. *PLoS One*, 2011,6(12):e28843.
- [11] K. G. O'Malley and M. A. Banks. A latitudinal cline in the Chinook salmon (*Oncorhynchus tshawytscha*) Clock gene: evidence for selection on PolyQ length variants[J]. *Proceedings of the Royal Society B: Biological Sciences*, 2008,275(1653):2813-2821.
- [12] 蒋志刚, 江建平, 王跃招, 等. 中国脊椎动物红色名录[J]. *生物多样性*, 2016,24(5):500-551.
- [13] D. Bell-Pedersen, V. M. Cassone, D. J. Earnest, et al. Circadian rhythms from multiple oscillators: lessons from diverse organisms[J]. *Nature Reviews Genetics*, 2005,6(7):544-556.
- [14] P. E. Hardin. The circadian timekeeping system of *Drosophila*[J]. *Current Biology*, 2005,15(17):R714-R722.
- [15] M. W. Young and S. A. Kay. Time zones: a comparative genetics of circadian clocks[J]. *Nature Reviews Genetics*, 2001,2(9):702-715.
- [16] S. Panda, J. B. Hogenesch, S. A. Kay. Circadian rhythms from flies to human[J]. *Nature*, 2002,417(6886):329-335.
- [17] S. M. Reppert, R. Weaver. Coordination of circadian timing in mammals[J]. *Nature*, 2002,418(6901):935-941.
- [18] P. J. Mitchell and R. Tjian. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins[J]. *Science*, 1989,245(4916):371-378.
- [19] 熊飞. 青海湖裸鲤繁殖群体生物学[D]. 武汉: 华中农业大学, 2003.
- [20] 周翠萍. 宝兴裸裂尻鱼的繁殖生物学研究[D]. 雅安: 四川农业大学, 2007.