

可口革囊星虫 cDNA 文库的构建及蚯蚓血红蛋白基因序列的分析

王孟前¹, 李 晔¹, 苏秀榕¹, 李太武^{1,2}, 贺静静¹, 王中华¹, 周 君¹

(1. 宁波大学生命科学与生物工程学院, 浙江 宁波 315211; 2. 宁波城市职业技术学院, 浙江 宁波 315110)

摘要:提取可口革囊星虫总 RNA, 反转录成 cDNA, 将大于 500bp 的 cDNA 片段连接入载体 pBluescript II SK (+), 电转化至 DH10B 宿主菌中, 得到原始文库. 对 cDNA 文库的滴度、重组率和插入片段的大小进行检测, 并挑取 500 个单菌落进行测序. 结果表明, cDNA 文库的库容量为 3.06×10^5 cfu, 重组率达 97.2%. 经测序共得到 437 条序列, 分析得到 211 个单基因簇 (unigene), 其中 115 条序列有相关同源性. 测序得到蚯蚓血红蛋白 cDNA 全长序列为 823bp, ORF 编码 120 个氨基酸, 蛋白分子量为 13.63kD, 等电点为 5.78. NCBI 上同源性比对结果则表明, 与 *Siphonosoma cumanense* 的蚯蚓血红蛋白同源性最高为 67%. 可口革囊星虫 cDNA 文库的构建为今后克隆和研究可口革囊星虫的相关功能基因奠定了坚实的基础.

关键词:海洋生物学; 可口革囊星虫; cDNA 文库; 序列分析; 蚯蚓血红蛋白

DOI: 10.3969/J. ISSN. 1000-8160. 2010. 01. 004

中图分类号: Q 178.53

文献标识码: A

文章编号: 1000-8160(2010)01-0020-07

可口革囊星虫 (*Phascolosoma esculenta*) 属星虫动物门 (Sipuncula), 革囊星虫纲 (Phascolosomatidea), 革囊星虫目 (Phascolosomatiformes), 革囊星虫科 (Phascolosomatidae), 革囊星虫属 (*Phascolosoma*) 动物. 民间又称其为海蛆、沙虫、海丁、土笋等. 其体长最长可达 10 cm, 为我国的特有种, 广泛分布于浙江、福建、广东、广西、海南和台湾等省区^[1]. 它肉质脆嫩、味道鲜美, 含有丰富的营养物质, 有“海洋冬虫夏草”之称, 我国沿海居民常食用^[2-3]. 可口革囊星虫含有多种活性物质, 能够调节机体多种机能, 含有丰富的蛋白质、多种氨基酸、微量元素等, 具有多种药理活性^[4-5]. 由于寒武纪大爆发时期没有星虫的化石记录, 其体内的一些活性物质就显得尤为重要, 蚯蚓血红蛋白 (hemerythrin) 就是其中的 1 种. 蚯蚓血红蛋白是 1 种含铁的金属蛋白, 是天然的非血红素载氧体. 它存在于多毛类 (Polychaeta)、曳鳃类 (Priapulida)、星虫类 (Sipunculia) 和腕足类 (Brachiopoda) 等为数不多的无脊椎动物体内, 并在这些动物体内担负着运送和贮存氧的重任^[6-7].

目前, 国内外对可口革囊星虫还没有开展功能基因组学研究, 没有构建可口革囊星虫文库的相关报道. 国内外对可口革囊星虫的研究主要集中在分类、养殖、组织学、生理生态、繁殖和生物活性等方面^[8]. 目前登录到 GenBank 中的可口革囊星虫的基因数为 7 条, 相应的蛋白质序列数据为 6 条, 且大多是部分序列, 全长基因很少. 本研究工作构建了可口革囊星虫的 cDNA 文库, 并对测序的结果进行了分析, 为今后对其功能基因组的深入研究奠定了基础.

1 材料和方法

1.1 材料

研究用的可口革囊星虫于 2008 年 2 月取自浙江省宁波奉化市开苗海淡水养殖场. RNA 提取试剂盒、cDNA 文库构建所需试剂、载体 pBluescript II SK (+)、宿主菌 DH10B 和 DNA 分子量 Marker 均为 TaKaRa 产品. 电脉冲仪为 BIORAD 公司产品.

收稿日期: 2009-03-13

基金项目: 国家自然科学基金资助项目 (40776075); 浙江省重大科技攻关项目 (2006C13089)

作者简介: 王孟前 (1985 ~), 女, 硕士研究生; E-mail: qiezi030931@yahoo.com.cn

通讯作者: 苏秀榕, 教授; E-mail: suxiurong@nbu.edu.cn

1.2 方法

1.2.1 总 RNA 的提取 取可口革囊星虫血液按照 TaKaRa RNAiso 试剂盒的说明书提取总 RNA,电泳检测抽提结果。

1.2.2 mRNA 的分离与纯化 利用 Oligotex-dT30 < Super > mRNA 纯化试剂盒分离纯化 mRNA,紫外测定 mRNA 的浓度和纯度并电泳检测 mRNA 质量。 A_{260}/A_{280} 的比值大于 1.8 的 Poly(A) + RNA,用于 cDNA 文库的构建。

1.2.3 cDNA 的合成及文库的构建 取纯化后的 mRNA 约 1.50 μg ,以 Oligo(dT)18 Anchor 引物[5'-(GA)10ACTAGTCTCGAG(T)18V-3'(V:A or C or G)],参照 cDNA 文库构建试剂盒说明书合成双连 cDNA,电泳检测 cDNA 的质量。补平 cDNA 末端后连接 *EcoR* I 接头并末端磷酸化,再用 *Xho* I 酶切。切胶回收去除 500bp 以下的小片段。利用 T4 DNA 连接酶把加工好的双链 cDNA 连接到 pBluescript II SK (+) 载体的 *EcoR* I 和 *Xho* I 位点上;连接产物电转化至 DH10B(电转化条件为 1.5kV,200 Ω ,25 μF)得到初始文库。取一部分转化产物涂板,37 $^{\circ}\text{C}$ 培养过夜后用质粒抽提试剂盒抽提质粒,-80 $^{\circ}\text{C}$ 保存。

1.2.4 文库的库容及滴度检测 将初始文库的菌液稀释 10、 10^2 、 10^3 倍数的菌液涂板,37 $^{\circ}\text{C}$ 培养过夜。计菌落数,计算文库的库容和滴度。

1.2.5 插入片段大小及重组率检测 在得到转化细菌中随机挑取 36 个克隆于无菌水中,沸水浴 10min 变性后,用试剂盒中的 T7、T3 引物进行 PCR 扩增,检测插入 DNA 的片段长度和重组率。PCR 体系:TaKaRa E $\times$ TaqHS(0.5 U/ mm^3) 0.1 mm^3 ,10 $\times$ Buffer (Mg^{2+}) 1 mm^3 ,dNTP Mixture (各 2.5 mmol/dm^3) 1 mm^3 ,T7 引物 (5 pmol/mm^3) 0.5 mm^3 ,T3 引物 (5 pmol/mm^3) 0.5 mm^3 , dH_2O 10 mm^3 。PCR 反应程序:95 $^{\circ}\text{C}$ 5min,94 $^{\circ}\text{C}$ 30s,55 $^{\circ}\text{C}$ 30s,72 $^{\circ}\text{C}$ 3min,最后 72 $^{\circ}\text{C}$ 10min,共 35 个循环。电泳检测扩增产物。

1.2.6 序列测定及分析 挑取 500 个单菌落,由上海英骏生物技术有限公司进行序列测定,同时测序的单克隆——对应于 96 孔板,-80 $^{\circ}\text{C}$ 保存,以备后用。测序结果用软件进行分析比对。

2 结果

2.1 总 RNA 和 mRNA 质量检测

可口革囊星虫总 RNA 和 mRNA 电泳结果见图 1。其结果显示:总 RNA 的 3 条主带为 28S、18S 和 5S RNA,18S RNA 的量大于 28S RNA;分离纯化后的 mRNA 含量较高,呈弥散分布。可见总 RNA 和 mRNA 比较完整,没有降解。

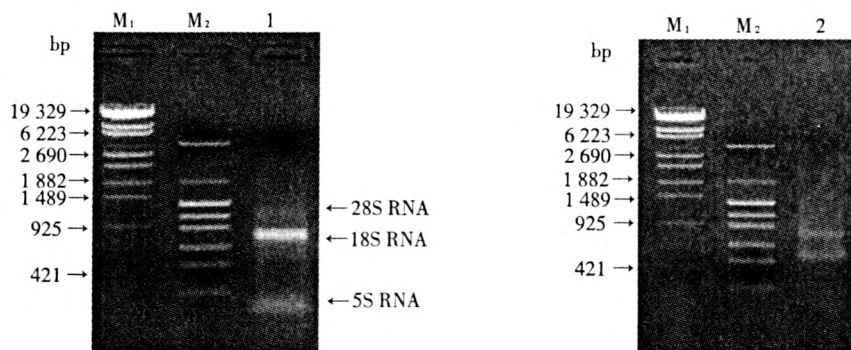


图 1 总 RNA 和 mRNA 电泳图

Fig. 1 Electrophoresis for total RNA and mRNA

M₁: λ -EcoT14 I,M₂:pHY,1:总 RNA,2: mRNA

2.2 cDNA 的检测

双链 cDNA 产物电泳结果见图 2。其结果显示:双链 cDNA 主要集中于 0.75 ~ 1.00 kb,割胶回收去除

500 bp 以下小片段(图 2)。

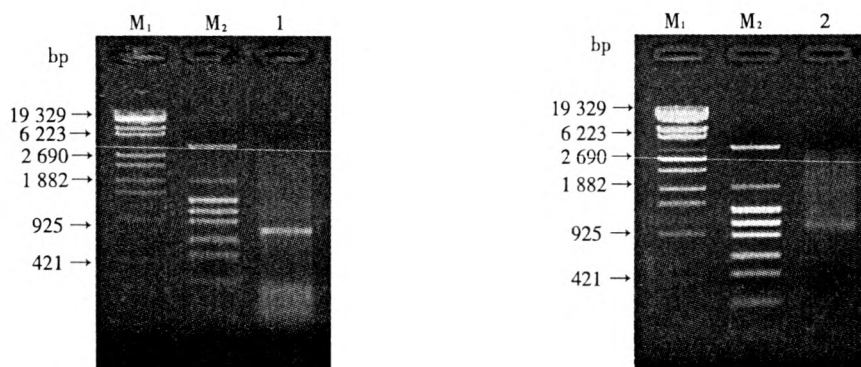


图 2 双链 cDNA 电泳图

Fig. 2 Electrophoresis for double-stranded cDNA

M₁: λ-EcoT14 I, M₂: pHY, 1: 合成的双链 cDNA, 2: 去除短片段的 cDNA

2.3 cDNA 文库质量的检测

cDNA 初始文库的滴度约为 1.7×10^5 cfu/cm³, 库容为 3.06×10^5 cfu, 重组率为 97.2%, 插入片断的长度均在 750bp 以上, 大多数片段大于 1 000bp(图 3)。

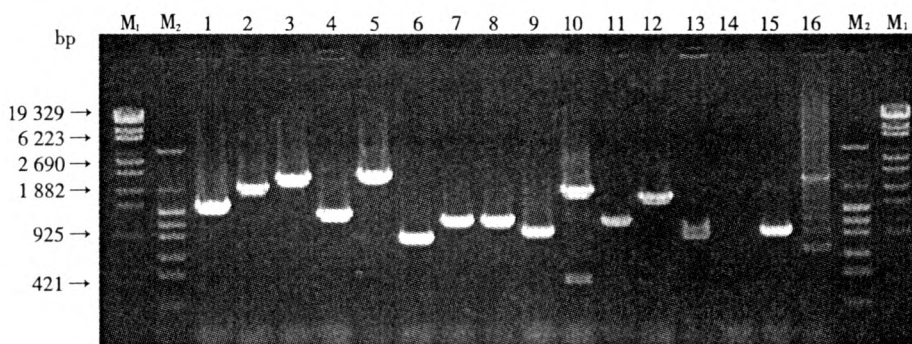


图 3 cDNA 文库的插入片断大小的鉴定结果

Fig. 3 Characterizing the inserted fragment size of cDNA library

M₁: λ-EcoT14 I, M₂: pHY, 1~16: 随机选取的 16 个克隆

2.4 测序结果初步分析

测序得到 437 个可口革囊星虫血液基因序列, 去除低质量序列和载体序列后, 长度大于 500bp 的有 425 条序列, 通过软件比对分析得到 26 个重叠群 (contigs), 185 个单拷贝 (singletons), 共计单基因簇 211 个, 其中 115 条序列有相关同源性, 其他的 96 条无同源性。绝大多数基因与营养代谢、生理学过程、酶活性、细胞过程及转录调节过程有关, 少数与信号转导及细胞分裂分化等功能有关, 如蚯蚓血红蛋白 (hemerythrin)、铁蛋白 (ferritin)、热休克蛋白 (heat-shock protein70, HSP70)、α-微管蛋白 (alpha-tubulin)、肌动蛋白 (actin)、细胞色素氧化酶 (cytochrome c oxidase subunit)、NADH 脱氢酶 (NADH dehydrogenase)、细胞分裂调控蛋白 (cell division control protein)、转录调控蛋白 (transcriptional regulatory protein) 和锌指蛋白 (zinc finger protein) 等。其中, 文库中测序得到的 ferritin、HSP70 与 NCBI 上登入的可口革囊星虫的 [ferritin (EU091352)、HSP70 (EU416330)] 两两比对, 同源性分别为 97% 和 99%, 仅仅个别的碱基序列存在着一定的差异。这说明测序结果可信。

2.5 蚯蚓血红蛋白序列的分析

2.5.1 序列分析 文库中测序得到的蚯蚓血红蛋白的全长 cDNA 序列为 823bp。该序列包含一个 360bp 的

	MGF EI PEPY VWDE SF KVFY DNL DDEHKGLFQ GI FALGNGPADA AN	
<i>Phascolosoma esculenta</i>	... PV . D.. D. . NQ. . Q. . . . L. S. M. AN. GSQ . A	45
<i>Periserrula leucophryna</i>	... D.... FK.. A.. E... KK. AI. DS. . . VC. . NN-. D.	44
<i>Golfingia vulgaris</i>	... P.. D.. . . . P.. RT.. TAI T. . D. . . H. ARD -DNKD .	44
<i>Hirudo medicinalis</i>	 E. G. . . A. . N. SKS NGA	45
<i>Perinereis aibuhitensis</i>	.. YD FK... M. QI. NA. . DVCG . NN-. N.	44
<i>Theromyzon tessulatum</i>	. V..... Q... T. E... EK. . E. K. . KD. SDS . . CSET	45
<i>Riftia pachyptila</i>	. S.. V. S.. . . . D.. R. A. . . . S.. QA. . KC. AK CAENR A	45
<i>Sipunculus nudus</i>	. P. PV . D. F... T.. Q... FE. . EK. R. I. . AL D-NNPS .	44

	L KK L VKVTAN HF AD E E GMM QSA NYAGFDSHKK AHEDF LA VLR GLS	
<i>Phascolosoma esculenta</i>	. DN AYA . LNS . . LF . . AA . . V. K. . . . YPP D.. . . EVH . K.	90
<i>Periserrula leucophryna</i>	. . . M. D.... . . S. K. DF K.... . . V. I. . .	89
<i>Golfingia vulgaris</i>	. GE. RRC . GK.. LN. QV.. . AS Q. . . . Y. E.... . DE. I HQ. DHWK	89
<i>Hirudo medicinalis</i>	. . H. S.. IDE . . SH . . D.. KK. S. SDF . N.... . V.. Q. S. G. . .	90
<i>Perinereis aibuhitensis</i>	. . MIE K. V.. K. . DF K.... . . I. . . .	89
<i>Theromyzon tessulatum</i>	. E. . . . LIED . . T. . . . E.. K. KS. EDL I. S.. VET . K. VK	90
<i>Riftia pachyptila</i>	. AD VD R.. AKT . FS. LPE . N. I. SE . V. KIK S..	90
<i>Sipunculus nudus</i>	. . L. YR.... . . E. . . F. . T. . . . G. YR... VL K.. SFK	89

	APVDDDKVN YAKD WLW NHIK GIDFKYK GKL	
<i>Phascolosoma esculenta</i>	V. L. AGF . . . T.... Q... T.....	120
<i>Periserrula leucophryna</i>	... G... LH	119
<i>Golfingia vulgaris</i>	G----- DS . W. T T.....	114
<i>Hirudo medicinalis</i>	S.. AN . . . HW.....	120
<i>Perinereis aibuhitensis</i>	... P... IK . . . E.....	119
<i>Theromyzon tessulatum</i>	... SEENIKM . . E.....	120
<i>Riftia pachyptila</i>	.. L.. AT . AF . . Q.....	120
<i>Sipunculus nudus</i>	... P. KD . YF Q... T.....	119

图5 可口革囊星虫蚯蚓血红蛋白氨基酸序列同源性比较

Fig. 5 Homology between *P. esculenta* and other species based on hemerythrin amino sequence

■: Octamerization interface

3 讨论

一个具有应用价值的 cDNA 文库,一般要求文库的容量应不小于 1.7×10^5 cfu 及插入片段的平均大小不小于 1 kb^[9],这样才能确保 cDNA 片段的完整性,同时基因组 cDNA 中的任 1 条序列至少有 1 个拷贝存在于重组文库中. 本研究所构建的 cDNA 文库的滴度约为 1.7×10^5 cfu/cm³,库容为 3.06×10^5 cfu,重组率为 97.2%,表明该文库是完整的、有效的,并且质量较高.

在获得的 211 种基因序列中,115 条序列在 NCBI 上比对存在与其有一定同源性的基因,96 条序列无同源性,在 115 条基因中有部分基因是全长序列. 115 个基因大致可分为 6 大类:营养代谢、生理学过程、细胞学过程、信号传导、转录活性和其他. 其中与营养代谢相关基因有 25 种,如铁蛋白、NADH 脱氢酶、酰基辅酶 A 氧化酶、谷氨酸酰胺转移酶、细胞色素 C 氧化酶亚基、硫酸软骨素合酶、乙醇脱氢酶、天冬氨酸甲基解氨酶、氨基转移酶、半胱氨酸合酶、NAD-谷氨酸脱氢酶等. 与生理学过程相关基因有 19 种,如蚯蚓血红蛋白、金属蛋白酶、脱水应答蛋白、DNA 错配修复蛋白、SOD、HSP70、复制结构域蛋白、核酸内切酶 I、血管内皮生长因子、焦磷酸激酶、腺苷三磷酸酶、脱氧核糖核酸外切酶等. 与细胞学过程相关基因有 11 种,如半胱氨酸蛋白酶抑制剂、 α -微管蛋白、网状细胞结合蛋白、细胞分裂调控蛋白、层粘连蛋白受体等. 与信号传导相关基因有 14 种,如绿色荧光蛋白、ATP 结合蛋白、识别蛋白、ABC 运输通透酶、传感蛋白、细胞因子信号 3 抑制剂、信号肽酶、跨膜螺旋受体等;与转录活性相关基因有 7 种,如锌指蛋白、转录调控蛋白、RNA 解螺旋酶、mRNA 帽子结合蛋白 eIF4E、转录-修复结合因子、苏氨酸-tRNA 合成酶等. 其他相关基因 39 种.

在动物界的进化过程中,出现了 3 种重要的呼吸蛋白,即蚯蚓血红蛋白(hemerythrin)、血蓝蛋白(hemocyanin)和血红蛋白(hemoglobin),其中蚯蚓血红蛋白是最古老的呼吸蛋白^[10]. 自从在古氏仿革囊星虫(*Phascolopsis gouldii*)体腔内分离得到蚯蚓血红蛋白,其结构和性质立即引起了化学工作者和生物学家的浓

厚兴趣,从而对蚯蚓血红蛋白的研究工作一度展开^[11-12]。蚯蚓血红蛋白单体由 4 条 α 螺旋组成,内含 2 个自由 Fe^{2+} ,但只有第 2 个 Fe^{2+} 能与 O_2 进行可逆性结合^[13]。蚯蚓血红蛋白无较大的别构效应和协同性,所以运载氧的效率显然不如脊椎动物的血红蛋白^[14]。但它的出现实现了以呼吸蛋白运载氧的形式替代了依靠气体渗透作用被动摄取氧的方式,这是生物界呼吸史上一次质的飞跃。

蚯蚓血红蛋白根据 Fe 的氧化态可分为含 Fe^{3+} 的高铁蚯蚓血红蛋白(methemerythrin)和氧合蚯蚓血红蛋白(oxyhemerythrin);含 Fe^{2+} 的脱氧蚯蚓血红蛋白(deoxyhemerythrin)和蚯蚓肌血红蛋白(myohemerythrin);还有一些以混合价态 Fe(2 价、3 价)形式存在的蚯蚓血红蛋白(hemerythrin)^[7]。通过氨基酸序列的同源性比较可知,蚯蚓血红蛋白在生物进化过程中相对保守,其同源性都在 40% 以上,最高为 67%。本实验从可口革囊星虫文库中得到蚯蚓血红蛋白的序列,是对蚯蚓血红蛋白研究的一个重要补充,也为独立成门的星虫动物门的进化分类提供了一定的参考,同时也为深入探索蚯蚓血红蛋白的作用途径和机理奠定了基础。

参考文献:

- [1] 李凤鲁. 中国沿海革囊星虫属(星虫动物门)的研究[J]. 青岛海洋大学学报:自然科学版,1988,19(3):78-90.
- [2] 纪其雄,丁理发,苏秀榕,等. 可口革囊星虫体液细胞组成及营养成分研究[J]. 水产科学,2007,26(7):387-389.
- [3] 周迎松,丁理发,徐继林,等. 可口革囊星虫主要营养要素的分析[J]. 营养学报,2007,29(4):413-414.
- [4] 周化斌,张永普,吴洪喜,等. 可口革囊星虫的营养成分分析与评价[J]. 海洋湖沼通报,2006(2):62-68.
- [5] 蒋定文,沈先荣,贾福星,等. 海洋星虫提取物的营养分析及免疫调节作用的初步观察[J]. 中国生化药物杂志,2004,25(2):96-97.
- [6] Andrew L F, Stephen J L. Reactions of non-heme iron centers with dioxygen in biology and chemistry [J]. Chemical Reviews, 1994, 94(3):771-803.
- [7] Ronald E S. Dioxygen and hemerythrin[J]. Chem Rev, 1994, 94:715-726.
- [8] 邓中日,黄勃,方再光. 可口革囊星虫体壁形态和组织学结构[J]. 水产科学,2007,27(6):335-339.
- [9] 萨姆布鲁克 J, 弗里齐 E F, 曼尼阿蒂斯 T. 分子克隆实验指南[M]. 金冬雁, 黎孟枫, 译. 2 版. 北京:科学出版社,1995:432-448.
- [10] 郭祖宝. 动物体内呼吸色素的种类及其分布[J]. 生物学教学,2006,31(1):64.
- [11] Laurent C, Marie C S, Jean M, et al. Cloning and expression analysis of a cDNA that encodes a leech hemerythrin [J]. Biochimica et Biophysica Acta, 2001, 1518:282-286.
- [12] Stefano V, Enrico N, Xavier B, et al. Molecular evolution and phylogeny of Sipunculan hemerythrins [J]. Journal of Molecular Evolution, 2006, 62:32-41.
- [13] Magnus K A, Hazes B, Ton-That H, et al. Crystallographic analysis of oxygenated and deoxygenated states of arthropod hemocyanin shows unusual differences[J]. Proteins: Structure, Function and Genetics, 1994, 19:302-309.
- [14] Liberatore F A, Truby M F, Klippenstein G L. The quaternary structure of *Phascolosoma agassizii* coelomic hemerythrin [J]. Arch Biochem Biophys, 1974, 160(1):223-229.

Construction and analysis of a cDNA library for *Phascolosoma esculenta*

WANG Meng-qian¹, LI Ye¹, SU Xiu-rong¹, LI Tai-wu^{1,2}, HE Jing-jing¹, WANG Zhong-hua¹, ZHOU Jun¹

(1. Faculty of Life Science and Biotechnology, Ningbo University, Ningbo 315211, China;

2. Ningbo City College of Vocational Technology, Ningbo 315110, China)

Abstract: Total RNA was extracted from the blood of healthy *Phascolosoma esculenta* with Trizol reagent. The first-strand cDNA was synthesized from purified mRNA, and subsequently the second-strand cDNA was generated via *E. coli* DNA polymerase I. The fractionated cDNA fragments longer than 0.5 kb were ligated into pBluescript II SK (+) express vector, and transformed into DH10B by electroporation. Then, the library titre, recombinant rate and length of inserted cDNAs were measured, respectively. The results indicated that this library reached 3.06×10^5 cfu in capacity and the percentage of recombination was as high as 97.2%. A total of 437 clones was sequenced, in which 211 cDNA were merged. 115 sequences were assigned putative functions based on their similarity to known genes. The full-length cDNA of hemerythrin was 823 bp, including an open reading frame (ORF) encoding a polypeptide of 120 amino acids with predicted molecular weight of 13.63 kD and theoretical isoelectric point of 5.78. The blasting in NCBI indicated that the homology between *Siphonosoma cumanense* and *Phascolosoma esculenta* was 67%. In conclusion, an excellent cDNA expression library had been constructed successfully in this report, which would lay solid foundation for further characterization of the genes of *Phascolosoma esculenta*.

Key words: marine biology; *Phascolosoma esculenta*; cDNA library; sequences analysis; hemerythrin

DOI:10.3969/J. ISSN. 1000-8160. 2010. 01. 004

(责任编辑:郭水伙)

ORF, 编码 120 个氨基酸;在 cDNA 序列上 5' 端有起始密码子 ATG, 其附近序列符合 Kozak 规则, 3' 端有终止密码子及 polyA 加尾信号 AATAAA (图 4)。

2.5.2 同源性分析 蚯蚓血红蛋白在 NCBI 上进行蛋白水平上同源性比较, 得到同源关系相对较近的 7 个物种的氨基酸序列 (图 5): 光裸方格星虫 (*Sipunculus nudus*, CAG14944)、欧医蛭 (*Hirudo medicinalis*, AAS49167)、管状蠕虫 (*Riftia pachyptila*, ABW24415)、双齿围沙蚕 (*Perinereis aibuhitensis*, ACI24391)、赫氏戈芬星虫 (*Golfingia vulgaris*, CAG14945)、*Periserrula leucophryna* (AAQ63634)、整嵌晶蛭 (*Theromyzon tessulatum*, AAG01808), 同源性分别为 62%、51%、46%、49%、47%、46%、41%, 与库岛管体星虫 (*Siphonosome cumanense*, P22766) 的同源性最高为 67%。

```

1                               TCACTAAAGGGAACAAAA
19  GCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGCTGCAGGC
79  TCGACAGCTTGCTCTTCACTTGTCTCGAGTTTCAGCGTAGCTCAGTCAAGTCAACCGTAGCC
139 ATGGGTTTCCAGTTCCAGATCCATATGTTTGGGATGAGTCCTTCAAGGTCTTCTATGAT
1   M G F P V P D P Y V W D E S F K V F Y D
199 GACCTGGACAACCAGCATAAGCAGCTTTTCCAGGACTTTTCAGCATGGGTGCTAACCCA
21   D L D N Q H K Q L F Q G L F S M G A N P
259 GGCAGCCAGGCTGCCCTTGACAACGCCTACGCTGTTCTGAACAGCCACTTCCTCTTCGAG
41   G S Q A A L D N A Y A V L N S H F L F E
319 GAGGCGCCCATGCAGGTCCGCAAGTACGCTGGCTACCCCCACACAAGAAGGCCCATGAT
61   E A A M V Q A K Y A G Y P P H K K A H D
379 GACTTCTTGGCTGAGGTCCATGAAAAGAGCGTCCCACTCGATGCCGGTTTCGTCAACTAT
81   D F L A E V H G K S V P L D A G F V N Y
439 ACAAAGGACTGGCTGGTACAACACATCAAGACGATTGACTTCAAGTACAAGGCCAAGTTG
101  T K D W L V Q H I K T I D F K Y K G K L
499 TAAACTGCCTGCAGTATTCCTCAGATTCTCCCGATCACGGAACACGACTGCTCTCATC
121 *
559 GGACTAAGCAACCTCATCTTTCATATGTTTAGTGAAATTCCCTCTGTTTCCGCTCTTGA
619 CCTATGCCTCCGTAACTCTGTTATAAGTGACTTTCCGGGACTCCTGTCGTGAGTGGGTT
679 CTACTTCCGGTTGATCACCAATGCAGTTGGAAGTTTCAGGCAACAATCTCATCTGCTAAT
739 CTATTAGGACTCTTTTTTGATTTCTTTCTTTCTTTTAAGAGGATGTG AATAAA TTGTG
799 ATATAATAAAAAAAAAAAAAAAAAA

```

图 4 可口革囊星虫蚯蚓血红蛋白编码基因 cDNA 全长及推导的氨基酸序列

Fig. 4 Nucleotide and deduced amino acid sequences of hemerythrin cDNA of *P. esculenta*

■: 氧结合位点, -: 家族签名序列, □: polyA 加尾信号

2.5.3 功能位点分析 将目标蛋白用相关的分析软件进行了分子量预测和结构功能位点的分析。PratParam 对蛋白质基本物理化学参数的分析结果表明: 蚯蚓血红蛋白的分子量约为 13.63 kD, 理论等电点为 5.78, 负电核残基 (Asp + Glu) 总数为 15 个, 正电核残基 (Arg + Lys) 总数为 11 个; 目标蛋白有 7 个氧结合位点: ²⁶His、⁵⁶His、⁶⁰Glu、⁷⁵His、⁷⁹His、¹⁰⁸His、¹¹³Asp (图 4); 16 个 Octamerization interface (图 5)。ScanProsite 对氨基酸参数的分析结果显示: HFLFEEAAMQVAKYAGYPHKKAH 为 hemerythrin 家族特征性序列。