

DOI: 10.19663/j.issn2095-9869.20180130001

http://www.yykxjz.cn/

顾冰宁, 刘欣, 沈和定, 王冬凤, 杨铁柱, 朱敏, 史艳梅, 李柏航. 瘤背石磺肌球蛋白重链(*MyHC*)基因的克隆与表达分析. 渔业科学进展, 2018, 39(4): 126-138

Gu BN, Liu X, Shen HD, Wang DF, Yang TZ, Zhu M, Shi YM, Li BH. Molecular cloning and analysis of the *MyHC* gene in *Onchidium struma*. Progress in Fishery Sciences, 2018, 39(4): 126-138

瘤背石磺肌球蛋白重链(*MyHC*)基因的 克隆与表达分析*

顾冰宁[#] 刘欣[#] 沈和定^① 王冬凤 杨铁柱
朱敏 史艳梅 李柏航

(上海海洋大学水产种质资源发掘与利用教育部重点实验室 水产科学国家级实验教学示范中心
海洋动物系统分类与进化上海高校重点实验室 上海 201306)

摘要 肌球蛋白是动物体内重要的功能性马达蛋白, 调控机体的信号传导、肌肉收缩和细胞器运动, 为探究肌肉在瘤背石磺(*Onchidium struma*)由海洋向陆地进化过程中发挥作用的分子机制, 实验以石磺科 4 种贝类转录组数据为基础, 采用 RACE 方法从瘤背石磺肌肉中首次克隆到肌球蛋白重链(Myosin heavy chain, *MyHC*)基因 cDNA 的全长并进行组织表达分析。研究结果显示, 瘤背石磺 *MyHC* 基因 cDNA 全长为 7566 bp, 包括 5895 bp 的开放阅读框, 228 bp 的 5'端非翻译区, 1443 bp 的 3'端非翻译区, 共编码 1964 个氨基酸。预测该基因编码的蛋白质由 31713 个原子组成, 分子式为 C₉₇₆₅H₁₅₈₉₇N₂₈₄₉O₃₁₅₀S₅₂, 分子量约为 225.28 kDa, 理论等电点为 5.56, N 端信号肽具有 29 个氨基酸长度。瘤背石磺 *MyHC* 具有 2 个保守结构域 MYSc_class_II 和 Myosin_tail_1, 且亲水性氨基酸集中在 Myosin_tail_1 区。系统进化树分析显示, 瘤背石磺 *MyHC* 与光滑双脐螺(*Biomphalaria glabrata*) *MyHC* 的亲缘关系最近。RT-PCR 结果显示, *MyHC* 基因在各个组织中均有表达, 腹足和背部皮肤表达量最高, 腹部皮肤、口球、肺囊中高表达, 肝胰腺、蛋白腺、两性腺中微量表达($P < 0.05$)。 *MyHC* 基因在瘤背石磺的主要运动器官腹足中高表达, 说明肌肉运动对瘤背石磺湿地环境的两栖适应性具有至关重要的作用。实验结果为今后进一步研究瘤背石磺肌球蛋白重链基因进行原核表达及多克隆抗体的制备奠定了良好基础, 更为深入探讨海洋无脊椎动物从海洋向陆地进化的研究提供有参考意义的分子依据。

关键词 瘤背石磺; *MyHC*; 基因克隆; 组织表达

中图分类号 S917.4 文献标识码 A 文章编号 2095-9869(2018)04-0126-13

瘤背石磺(*Onchidium struma*)是生活在潮间带高潮区滩涂的一种肺螺类, 隶属于软体动物门(Mollusca)、腹足纲(Gastropoda)、肺螺亚纲(Pulmonata)、

石磺科(Onchidioidea), 是一种具有重要营养价值和药用价值的两栖性贝类, 广泛分布于赤道两侧的热带和亚热带海区(吴旭峰等, 2010), 国内主要分布在江苏

* 国家自然科学基金项目(41276157)和上海高校水产学一流学科建设项目共同资助 [This work was supported by National Natural Science Foundation of China (41276157), and Shanghai University Fisheries First-Rate Discipline Construction Projects]. #: 共同第一作者。顾冰宁, E-mail: 604923111@qq.com; 刘欣, E-mail: dxliu2090@163.com

① 通讯作者: 沈和定, 教授, E-mail: hdshen@shou.edu.cn

收稿日期: 2018-01-30, 收修改稿日期: 2018-03-28

以南沿海地区。石磺科贝类呈现从海洋向陆地扩展的梯度分布状态(Dayrat, 2009), 深入研究其辐射过程会更好地了解海洋无脊椎动物由海栖向陆栖的宏观进化模式(Bouchet *et al*, 2005)。近年来, 其线粒体基因组(Sun *et al*, 2014; 王冬凤等, 2015)、群体多样性分析(姚理想等, 2015)以及分子标记开发(王成暖等, 2014)的研究较多, 但与性状相关功能基因的研究报道较少。石磺的真皮含有大量纵横排列的肌纤维, 能显著增强皮肤的弹性和韧性(Liu *et al*, 2014)。作为两栖性最为明显(不能长时间淹没在水中, 也不能长期在干燥环境中生活)的瘤背石磺, 腹足中含有的大量肌纤维, 在其及时快速躲避海水淹没等运动过程中起着非常重要的作用。

肌球蛋白(Myosin)是骨骼肌纤维粗肌丝的基本组成单位, 约占肌原纤维蛋白的二分之一以上, 具有结构蛋白和功能蛋白的双重性质, 是动物体内重要的功能性马达蛋白(Knight *et al*, 2000), 在机体的信号传导、肌肉收缩、细胞器运动等方面起着重要的调控作用(Hooper *et al*, 2005)。肌球蛋白分子为“Y”形不对称六聚体结构, 由2条重链亚基和4条轻链亚基组成, 重链亚基由位于氨基端球状结构的肌球蛋白头部和位于羧基端具有 α -螺旋结构的轻酶解肌球蛋白构成, 分子量约为200 kDa, 肌球蛋白头部含有1个ATP和肌动蛋白(actin)结合位点(Craig *et al*, 2006), 轻链亚基具有使肌球蛋白在生理条件下形成丝状的能力(Pett *et al*, 2000), 分子量约为16~20 kDa。单个肌球蛋白分子没有功能, 不能参与细胞生理过程, 必须聚合成排列有序的粗肌纤维才能完成一个完整生命活动。

近年来, 随着生命科学技术的发展, 生物学界对各类物种的相关结构基因, 特别是肌球蛋白重链(MyHC)基因的研究也越来越多。目前, 在对肌球蛋白基因的研究中, 大多数的研究对象是脊椎动物, 而有关无脊椎动物的研究报道甚少。目前, 已经对多个物种的肌球蛋白重链基因进行了克隆及表达, 包括人(*Homo sapiens*) (Jaenicke *et al*, 1990)、大鼠(*Rattus norvegicus*) (Strehler *et al*, 1986)、鸡(*Gallus gallus*) (Molina *et al*, 1987)、鹅(*Anser anser*) (李艳艳等, 2015)以及鱼类中的大麻哈鱼(*Oncorhynchus keta*) (Iwami *et al*, 2002)、鲤鱼(*Cyprinus carpio*) (Muramatsu *et al*, 2005)、翘嘴鲈(*Siniperca chuatsi*) (陈之航等, 2017)等, 与这些物种 MyHC 基因相对应的核酸分子杂交探针也陆续制备出来, 已达到对 MyHC 进一步深入细致研究的效果(Lied *et al*, 1985)。

本研究是在实验室团队对石磺肌肉理化性质研究的基础上, 选择在我国分布广泛、产量大、两栖性

最为明显的瘤背石磺作为实验对象, 以石磺科贝类转录组数据库为基础, 差异分析后筛选、克隆了瘤背石磺 MyHC 基因的 cDNA 全长, 对其序列进行同源性分析, 并测定分析其在不同组织中的表达状态, 为研制 MyHC 基因核酸杂交探针提供了基础资料, 也为进一步分析石磺科贝类的肌肉生长规律、探究无脊椎动物从海生到陆生的适应性进化过程及分子机制作有益尝试。

1 材料与方法

1.1 材料

本实验所用瘤背石磺采自上海市崇明岛, 采回样品放置于北向实验室暂养箱中进行暂养, 定时向暂养箱中投放玉米粉, 清理排泄物, 增添清水, 并及时清除死亡个体。实验进行前体重为(10.68±0.11) g, 肌肉舒展状态下体长为(5.82±0.12) cm。

1.2 方法

1.2.1 RNA 提取和 cDNA 合成 取瘤背石磺的背部皮肤、腹部皮肤、口球、腹足、肝胰腺、蛋白腺、两性腺、肺囊部位, 采用 Trizol(TaKaRa, 大连)试剂提取总 RNA, 提取方法参照使用说明书, 用 NanoDrop 2000 (Thermo Scientific, 美国)检测提取的 RNA 浓度, 并用 1%琼脂糖凝胶电泳检测 RNA 的完整性。以瘤背石磺腹足总 RNA 为模板, 用 SMARTer™ RACE cDNA Amplification Kit(Clontech, 日本)按照操作指南合成 RACE cDNA 第一条链。

1.2.2 MyHC 基因克隆 根据已有的瘤背石磺转录组数据, 通过基因差异表达分析、GO 功能显著性富集分析和 KEGG pathway 显著性分析, 从转录组数据库中筛选出一条与肌肉生长相关的关键序列(S_Unigene106_c0_seq1), 在 NCBI 进行 BLAST 比对, 结果为 MyHC 基因。利用 Primer Premier 5.0 软件设计引物 MyHC-F1、MyHC-R1(表 1), 扩增目的基因片段, 25 μ l 体系: 10 \times PCR Buffer 2.5 μ l, dNTP Mixture (各 2.5 mmol/L) 0.5 μ l, 上下游引物(10 μ mol/L)各 1 μ l, Taq (5 U/ μ l) 0.3 μ l, ddH₂O 18.7 μ l, cDNA 1 μ l。PCR 反应程序: 94 $^{\circ}$ C 预变性 4 min; 94 $^{\circ}$ C 30 s, 55 $^{\circ}$ C 30 s, 72 $^{\circ}$ C 30 s, 35 个循环; 72 $^{\circ}$ C 延伸 10 min。扩增产物经 1.5%琼脂糖凝胶电泳检测后, 用胶回收试剂盒(天根, 北京)割胶回收目的条带, 并与 pGEM-T Easy (Promega, 美国)载体连接构建重组质粒, 热转化感受态 DH5 α 大肠杆菌后涂布于含氨苄的选择性培养基, 37 $^{\circ}$ C 过夜培养, 经蓝白斑筛选和菌液 PCR 检测后将

表 1 本研究用到的引物
Tab.1 Primers used in this study

引物 Primer	序列 Sequence (5'~3')	用途 Usage
MyHC-F1	TGCGTGGCTATCAACCCC	片段验证 Fragment verification
MyHC-R1	GCCCTCAAGCACACCGTT	片段验证 Fragment verification
MyHC-F2	AGACTGTGTCCCACTTGC	片段验证 Fragment verification
MyHC-R2	TGAGCGGACGGATGAGAT	片段验证 Fragment verification
MyHC-F3	GTCAAGAAATACCAGCAG	片段验证 Fragment verification
MyHC-R3	TAGTGATGATGATGGTGG	片段验证 Fragment verification
MyHC-3' RACE Outer	GCGGCACACCAAGTTTGACCACAT	3' RACE
MyHC-3' RACE Inner	AACGAGGGTGAATCCGGACTATA	3' RACE
MyHC-5' RACE Outer	GTAGGCATTGTCAGAGAT	5' RACE
MyHC-5' RACE Inner	AGGGGTTGATAGCCACGC	5' RACE
MyHC-RT-F	GCCTCCTCATTTGTTCTCCA	目的基因定量 Target gene quantification
MyHC-RT-R	ATCTTCTTCTCGGCTCCCTC	目的基因定量 Target gene quantification
18s-RT-F	TGTCTCCTGCCCTACCTGTT	内参基因定量 Reference gene quantification
18s-RT-R	ATTCCATGCGCAATTATTC	内参基因定量 Reference gene quantification

阳性克隆送生工生物工程(上海)股份有限公司测序。

根据获得的 *MyHC* 基因序列, 分别设计 3' 和 5'RACE 特异性引物(表 1), 按照 Clontech RACE 说明书分别进行 3' RACE 和 5' RACE 扩增, PCR 反应产物经电泳、胶回收、连接和转化后, 通过菌落 PCR, 挑选阳性克隆样品送生工生物工程(上海)股份有限公司测序, 将所获序列和已知序列进行拼接, 获得 *MyHC* 基因的全长, 并通过重测序消除不确定碱基。

1.2.3 序列分析 利用在线工具 ProtParam、NetPhos 2.0 Server、ORF Finder、Signal 4.1 等分别对蛋白质理化性质进行预测分析; 用 Phyre2 程序预测蛋白质的二级结构和三级结构; 以日本三角涡虫 (*Dugesia japonica*) 为外群采用 MrBayes-3.2.6 和 MEGA 6.0 软件用邻位法(Neighbor-Joining)构建系统进化树, 1000 次 bootstraps, 其他参数均使用默认值。

1.2.4 实时荧光定量 PCR 实验前随机挑选 9 个目测健康、大小均一的瘤背石磺, 分别取其背部皮肤、腹部皮肤、口球、腹足、肝胰腺、蛋白腺、两性腺和肺囊 8 个组织, 将每 3 个个体的相同器官或组织等量混合作为一个样品, 共制成 3 个样品组, 提取总 RNA 并将其反转录为 cDNA, -20℃ 保存备用。

根据已获得的 *MyHC* 基因序列, 设计荧光定量引物 MyHC-RT-F 和 MyHC-RT-R(表 1), 以 18S rRNA 为内参。实时荧光定量 PCR 采用 SYBR premix Ex Taq™ II 试剂盒(TaKaRa, 大连)在 CFX-96(Bio-Rad, 美国)实时荧光定量 PCR 仪上进行。采用相对定量 $2^{-\Delta\Delta Ct}$ 法(Livak *et al*, 2001)计算目的基因的相对表达量。基因的表达量用平均值±标准误(Mean±SE)表示,

使用 SPSS 19.0 进行单因素方差分析(One-way ANOVA), $P < 0.05$ 表示有显著性差异。

2 结果与分析

2.1 *MyHC* cDNA 的全长序列

瘤背石磺 *MyHC* 的 cDNA(GenBank 登录号: KU550708)全长为 7566 bp, 包括 5895 bp 的开放阅读框, 228 bp 的 5'端非翻译区(UTR), 1443 bp 的 3'端非翻译区(UTR)。3'UTR 区含有典型的多聚核苷酸加尾信号序列 AATAAA 和 Poly(A)尾巴(图 1)。NCBI 进行 Blast 比对, 显示该序列与光滑双脐螺(*Biomphalaria glabrata*)相似性为 92%, 与加州海兔(*Aplysia californica*)相似性为 89%, 与二者的相似性均较高。

2.2 *MyHC* 的结构特征

推导该序列共编码 1964 个氨基酸, 分子量约为 225.28 kDa, 理论等电点为 5.56。谷氨酸(Glu)、亮氨酸(Leu)和赖氨酸(Lys)含量最高, 分别为 13.2%、10.5%和 10.1%; 色氨酸(Trp)、半胱氨酸(Cys)和组氨酸(His)含量最少, 分别为 0.5%、0.6%和 1.4%(表 2)。带负电荷氨基酸残基(Asp+Glu)366 个, 带正电荷氨基酸残基(Arg+Lys)324 个。原子总量 31713, 分子式为 $C_{9765}H_{15897}N_{2849}O_{3150}S_{52}$ 。不稳定指数为 47.48。

信号肽的预测结果显示, 信号肽切割位点存在于第 29 和 30 个氨基酸残基之间, 表明 N-端具有一个由 29 个氨基酸组成的信号肽。磷酸化位点预测结果显示有 74 个 Ser、41 个 Thr、9 个 Tyr, 可能是磷酸

```

1 aaaccttccccccgtcaccacccctgataggtaaaacagagcgattgctctgtgcatgtg
61 gatctatTTTTtagcctctgaggggtgagcagttgtgactttggctgccaagagctccagtc
121 cgcaaggctgttctgaagtaggtcctgaacttgcctagacacagacttccctggcgagc
181 gtctctttgcttttgaaccacgtttgcgaatctctagagactttgcctgaagggcgagca
1 M S G L
241 gatattgtttccgctcccagcgctccccccacaaaactaaccacATGTCAGGACTC
5 D V N D P D F Q Y L V V D R K K M M K E
301 GACGTTAACGATCCCAGACTTCCAGTACTTGGTCGTGGATCGCAAGAAGATGATGAAGGAG
25 Q T Q S F D G K K A C W V P D P K E G F
361 CAGACCCAGAGCTTCGATGGCAAAAAGGCTTGCCTGGGTCCCCGACCCAAAGAGGGATTC
45 L A A E I Q S T K G E E I T V K I D K N
421 TTGGCTGCTGAGATCCAGAGACCAAGGGTGAAGAAATCACCGTCAAATGATAAAGAAC
65 N E T R T L K K D D I Q Q M N P P K F E
481 AACGAGACTAGAACCCTCAAGAAAGATGACATCCAACAGATGAACCCACCCAAGTTCGAG
85 K I D D M A N M T Y L N E A S V L H N S L
541 AAAATCGACGACATGGCCAACATGACCTACCTGAACGAGGCCCTCTGTCTTGACAACTTG
105 K S R Y S S G L I Y T Y S G L F C V A I
601 AAGAGCCGTTACAGCAGTGGCTTGATCTACACCTACTCCGGCCTGTTTGGCTGGCTATC
125 N P Y R R L P I Y T Q K I I N C Y R G K
661 AACCCCTACCGTCGCCTGCCATCTACACCCAGAAGATCATCAACTGCTACAGGGGCAAG
145 R R A E M P P H L F S I S D N A Y Q N M
721 AGGAGGGCTGAGATGCCTCCTCATTGTCTCCATCTCTGACAATGCCTACAGAACAATG
165 L Q D R E N Q S M L I T G E S G A G K T
781 TTGCAAGATCGTGAGAACCAGTCCATGTTGATCACTGGTGGTGGTCCGGAGCCGGAAAGCC
185 E N T K K V I M Y F A N V A A G Q Q K K
841 GAGAACACGAAGAAAGTAATCATGTACTTTGCCAATGTGGCCCGGACAGCAGAAAAA
205 E E A A E G A E K K I G N L E D Q I I Q
901 GAGGAGGCCGAGAGGGAGCCGAGAAGAAGATTGGCAACCTTGAGGACCAAATATCCAG
225 A N P V L E A F G N A K T I R N N N S S
961 GCCAATCCAGTACTGGAGGCCCTTGGCAACGCCAAGACCATTGTAATAACAACCTTCTCT
245 R F G K F I R I H F G T Q G K I A G A D
1021 CGATTCGGTAAATTCATCCGTATTCCTTCCGTTACCCAAGGCAAGATCGCTGGTCCGAT
265 I E T Y L L E K S R V T F Q Q P A E R N
1081 ATTGAGACATACCTGTGGAGAAGTCCCGTGTGACCTTCCAACAGCCCGCTGAGAGAAAC
285 Y H I F Y Q L L S P A L P F L H E K L L
1141 TACCACATTTTCTACAGCTCTCTCACCGGCCCTGCCCTTTTGGCAGGAAACTCCTG
305 V S P D P A L Y A F I N Q G S L V V D G
1201 GTCAGCCCCGACCCCGCCCTTATGCCTTCATCAACCAGGGCTCCTTGGTGGTTCGATGGA
325 I D D T E E M K I T D E A F D V L G F S
1261 ATTGACGACACTGAGGAAATGAAAATCACAGATGAAGCCTTCGATGTGCTGGGCTTCTCT
345 A E E K T S L Y R C T A A I M H F G E M
1321 GCTGAAGAAAAGACCTCCCTGTACAGATGTACCGCCCATCATGCACTTCGGTGAATG
365 K F K Q R P R E E Q A E A D G T A E A E
1381 AAATTCAGCAAAGGCCCGCGAGGAGCAGGCCGAGGCTGACGGCACAGCTGAGGCTGAG
385 K V A F L L G I N A G D L L K G L L K P
1441 AAGGTCGCCTTCTTCTGGGCATCAACGCTGGTACCTCCTGAAGGGTCTCCTGAACCC
405 K I K V G N E F V T Q G R N L N Q V L Y
1501 AAGATCAAGGTCGGCAACGAGTTCGTCACCCAAGGACGTACCTGAACAGGTGCTGTAC
425 S V A A L A K S L Y N R M F D W L V K R
1561 TCCGTGGCTGCTCTGGCCAAGTCTCTGTACAACCGTATGTTGACTGGCTGGTCAAGCGT
445 V N K T L D T K N K R Q F F I G V L D I
1621 GTCAAACAAGACCCTGGACACCAAAAACAAGAGGCGAGTTCCTCATTGGTGTACTGGATATT
465 A G F E I F D F N S F E Q L C I N Y T N
1681 GCCGGCTTTGAGATCTTCGACTTCAACAGCTTCGAGCAGCTGTGTATCAACTACACCAAC
485 E R L Q Q F F N H H M F I L E Q E E Y K
1741 GAGCGCTTGCAGCAGTCTTCAACCATCACATGTTTCACTTGGAGCAGGAGGAATCAAG
505 K E G I D W E F I D F G M D L Q A A I D
1801 AAAGAGGGCATTGACTGGGAGTTCATTGACTTCGGTATGGACTTGCAGGCCGCCATTGAT
525 L I E K P M G I L S I L E E E C M F P K
1861 CTGATTGAAAAGCCCATGGGCATCCTGTCCATCCTTGAGGAAGAGTGCATGTTCCCCAAG
545 A S D K T F L D K L N G N H L G K S K N
1921 GCCAGCGACAAGACCTTCTTGACAAGCTGAACGGCAACCACTTGGGCAAGTCCAAGAAC
565 Y G K A G K P K K A G Q V E A H F E L H
1981 TACGGCAAGGCCGCAAGCCCAAGAAAGCCGACAAGTGGAGGCTCACTTCGAGCTGCAT
585 H Y A G S V P Y N I T G W L E K N K D P
2041 CACTACGCTGGCAGTGTGCCCTACAACATCACAGGCTGGCTGGAGAAGAACAAGGATCCA
605 L N E T V V E L L S H S K E T L V A Y L
2101 CTGAACGAGACCGTTGTTGAGCTGCTTAGCCACTCCAAGGAAACCCCTGGTGGCCTACCTG
625 F S P P E G A E E T S A K T A Q R K K G
2161 TTCTCCCCACCTGAGGGCGTGAAGAAACGTCAGCCAAGACCCGCCAGAGGAAGAAAGGA
645 G S F Q T V S H L H K E S L N K L M K N
2221 GGAAGTTTCCAGACTGTGTCCCACTTGACAAAGGAGTCCCTGAACAAGCTGATGAAGAAC
665 L Y S T H P H F V R C I I P N E F K Q P
2281 TTGTACAGCACTCATCCTCACTTCGTGCGCTGCATCATCCCCAACGAGTTCAAGCAGCCA
685 G E I D S Y L V L H Q L Q C N G V L E G
2341 GGTGAGATCGACTCCTACTTGGTGTGACCAGCTGCAATGTAACGGTGTGCTTGAGGGC
705 I R I C R K G F P N R I I Y S E F K Q R
2401 ATCCGTATTTGCCGTAAGGGATTCCCAACAGAATCATCTACTCTGAGTTCAAGCAGCGA
725 Y S I L A P N A I P Q G F V D G K K V T

```

2461 TACTCCATCCTGGCCCCAACGCCATCCCCAGGGATTGCTTGTATGGCAAGAAAGTCACT
 745 E N I L L A L Q L D T A E Y R L G T T K
 2521 GAGAACATCCTGCTGGCCCTGCAACTGGACACCGCTGAGTACCGTCTGGGTACCACCAAG
 765 V F F K A G V L G T L E D M R D E R L S
 2581 GTGTTCTTCAAGGCCGGTGTGCTGGGTACCCTTGGAGACATGCGTGACGAGCGTCTGTCT
 785 K I I S M F Q A H I R G Y L M R K S Y K
 2641 AAGATCATCTCCATGTTCCAGGCTCACATCAGAGGTTACCTCATGCGCAAATCCTACAAG
 805 K L Q D Q R V G L S V I Q R N I R K W L
 2701 AAAGTGCAGGACCAGAGAGTGGGCTGAGCGTCATCCAGCGTAACATCCGTAAATGGCTG
 825 G L K N W L W W R L Y V K V K P L L N I
 2761 GGAAGTGAAGAACTGGCTGTGGTGGAGACTGTACGTCAAGTCAAGCCCTCCTCAACATC
 845 A R A E D D M K K E E E L A K T K T E
 2821 GCCCGTGTGAGGATGACATGAAGAAGAAGGAGAGGAGTTGGCCAAGACCAAGACCGAG
 865 L E K T E K L R K E L E E Q N V G L L Q
 2881 CTGGAGAAGACAGAGAACTGCGCAAGGAGCTGGAGGAACAGAACGTTCGGTCTGCTCCAG
 885 A K N D L Y I Q L Q A E Q D N L I D A E
 2941 GCCAAGAAGGACTTGTACATCCAACCTGCAAGCCGAGCAAGACAACCTGATCGATGCCGAG
 905 E K I E K L I S Q K A E Y E S Q I K E M
 3001 GAGAAGATCGAGAAGCTCATCTCCAGAAAGGCGAGTACGAGTCCAGATCAAGGAGATG
 925 E E R L L D E E D A A A E L E E K K K K
 3061 GAAGAGAGACTGTGGATGAGGAGGACGCCGCTGCCGAGCTGGAGGAGAAGAAGAAGAA
 945 M E A E N K A L K E D I E D L E N S L Q
 3121 ATGGAAGCCGAGAACAAGCCCTCAAGGAAGACATCGAGGATCTGGAGAACTCTCTGGCC
 965 K A E Q E K Q T K D N Q I K T L Q D E M
 3181 AAGGCCGAACAAGAGAACAACCAAGACAACCAGATCAAGACCCTGCAGGACGAGATG
 985 A A Q D E Q L A K V N K E K K N L E E Q
 3241 GCCGCCAGGACGAACAGCTGGCCAAGGTCAACAAAGAGAAGAAGAACCTGGAAGAGCAA
 1005 Q K K T L E D L Q A E E D K V N H L N K
 3301 CAGAAGAAGACCTGGAAGACCTCCAGGCCGAGGAGCAAAAGTCAACCACCTCAACAA
 1025 L K T K L E Q T L D E L E D N L E R E K
 3361 CTCAAGACCAAGTGGAAACAGACTCTGACGAGTTGGAAGACAACCTGGAGAGAGAGAA
 1045 K V R G D V E K A K R K L E Q D L K S T
 3421 AAGGTGCGTGGAGATGTTGAAAAGGCAAAAGAGGAAGTTGGAGCAGGACCTGAAGAGCAC
 1065 Q E T V E D L E R S K R E L E E N V R R
 3481 CAAGAGACAGTGGAAAGATCTGGAGAGATCCAAGAGGGAGCTGGAAAGAGAATGTCCGAGG
 1085 K D L E I N N L N S K L E D E Q S L V A
 3541 AAGGACTTGGAAATCAACAACCTCAACTCCAAACTCGAAGACGAACAGAGCTTAGTTGCT
 1105 Q L Q R K I K E L Q A R I E E L E E E L
 3601 CAGCTCCAGAGAAGATCAAGAACCTCCAGGCTCGATTGAGGAGCTCGAGGAGGAATG
 1125 E A E R Q S R A K V E K Q R S E L S R E
 3661 GAAGCTGAACGCCAGTCCAGAGCTAAGTTGAGAAACAGAGAAGCAACTGTCTCGTGAG
 1145 L E E L S E R L D E A G G A T A A Q V E
 3721 TTGGAAAGCTGAGCGAGCGCTTGGACGAGGCGCGTGGCGCCACCGCCGCTCAGGTCGAG
 1165 L N K K R E Q E L L K L R R D L E E S A
 3781 CTCAACAAGAAGCGGAGCAGGAACCTGCTCAAACTGCGCCGTGACCTGGAGGAGAGCGCC
 1185 L Q H E A Q I S A I R K K Q Q D A A N E
 3841 CTGCAGCAGGAGGCCAGATCTGCCATCAGGAAGAAGCAGCAGGATGCCGCAACGAG
 1205 M A D Q I D Q L Q K V K N K V E K E K A
 3901 ATGGCTGACCAGATCGACCAGCTCCAGAAGGTCAAGAACAAGGTTGAGAAAGAAAAGGCC
 1225 Q L K S E V G D L Q A Q I E H A G K N K
 3961 CAACTGAAATCTGAGGTTGGAGATCTCCAGGCCAGATCGAGCAGCCGGAAGAACAAG
 1245 G M S E K M T K Q V E A Q I S E L N A K
 4021 GGTATGTCCGAGAAGATGACCAAAACAGGTGGAGGCTCAGATCTCCGAGCTGAACGCCAAG
 1265 L D A A N R Q I Q E L Q G Q K A R S A Q
 4081 TTGGACGCCGCCAACCCGAGATCCAGGAGCTGCAAGGCCAGAAGGCTCGCTCTGCCAG
 1285 E N A D L T R Q L E E A E H R V G A L T
 4141 GAGAACGCTGACCTGACCCGCCAAGTGGAGGAGGCGGACCGCCTGGGAGCTCTGACC
 1305 K E K N A L A A A L E E A K R A L E D E
 4201 AAGGAGAAGAAGCCCTGGCCGCCCTGGAGGAGGCCAAGAGAGCCCTGGAGGATGAG
 1325 T R A R Q K L A S E V R N L T S D L D A
 4261 ACCCGCGCCGAGAAAGTTGGCTCTGAAGTCCGCAACCTGACCTCTGACCTGGATGCT
 1345 L R E Q L E E E Q E G R S D L Q R Q L S
 4321 CTGCGCAGCAGCTGGAGGAGGAACAGGAGGACGCTCTGACCTCCAGAGACAGCTCAGC
 1365 K A N S E V Q V W R S K F E S E G T Q R
 4381 AAGGCTAACTCTGAGGTTCCAGGTATGGCGATCCAAATTCGAATCCGAGGGCACCCAGAGA
 1385 A E E L E E G R R K L Q A K L S D A E Q
 4441 GCTGAGGAGCTCGAGGAGGCGAGGCTAAGCTGCAGGCCAAACTGAGTGACGCCGAGCAG
 1405 A A E A A N L K V Q A L E K A K S R L Q
 4501 GCCGCCGAGGCTGCCAACCTCAAAGTCCAGGCCCTGGAGAAGGCTAAGAGCAGACTGCAA
 1425 G E L E D L M I D V E R S N A N A S N L
 4561 GGTGAACCTGGAAGATCTTATGATTGACGTTGAGCGCTCCAACGCCAATGCCAGCAACTG
 1445 E K K Q R A F D K T I A E W Q A K C N D
 4621 GAGAAGAAGCAGCGCTTTTCGACAAGACCATGCTGAATGGCAGGCCAAGTGAATGAC
 1465 L Q A E L E N A Q K E A R G Y S A E L F
 4681 CTCCAGGCTGAACCTGGAGAAGCCCAAGGAGCCAGAGGATATTCAGCAGAACCTCTTC
 1485 R V K A Q I E E A N D S V E A L R R E N
 4741 CGCGTCAAGGCTCAGATTGAGGAAGCAACGACAGCGTGGAGGCTCTTCGCTGAGAAC
 1505 K N L A D E I H D L T D Q L G E G R S
 4801 AAGAACTTGGCTGATGAAATCCACGATCTGACTGACCAGTTGGGAGAGGGCCGCGAG

```

1525 T H E L E K A R K R L E M E K E E L Q A
4861 ACACACGAGCTGGAGAAGGCTCGCAAGAGGCTGGAGATGGAGAAGGAGGAACCTCAAGCC
1545 A L E E A E S A L E Q E E A K V M R A Q
4921 GCTTGGAAGAGGCCGAGTCTGCCCTGGAACAGGAGGAGGCCAAGGTCATGCCGCTCAG
1565 L E I S S V R S E I D R R I Q D K E E E
4981 CTGGAGATCTCATCCGCTCCGCTCAGAGATCGACCGCCGCATCCAGGATAAGGAAGAGGAG
1585 F E N T R R N H Q R A M D S M Q A S L E
5041 TTTGAGAACACACGCCGCAACCACCAGAGAGCCATGGACTCCATGCAGGCCAGTCTGGAG
1605 A E A K G K A E A L R I K K K L E Q D I
5101 GCTGAAGCCAAGGCCAAGGCCGAGGCCCTGCGCATCAAGAAGAACTGGAACAGGACATC
1625 N E L E V A L D Q A N R G K A E F E K N
5161 AACGAGCTGGAGGTGGCTCTGGACCAGGCAACAGAGGCCAAGGCCGAGTTTGAGAAGAAC
1645 V K K Y Q Q T I K E M Q T Q I E D E Q R
5221 GTCAAGAAATACCAGCAGACAATCAAGGAAATGCAGACACAGATCGAGGATGAGCAGAGA
1665 Q R E E A R E S Y N M A E R R C T I L Q
5281 CAACGAGAGGAGGCCCGTGAATCATAACAATGGCCGAGCGCCGCTGCACAATCCTGCAG
1685 A E I D E L R A A L E Q A E R A R K G A
5341 GCTGAGATTGATGAGCTTAGAGCTGCTCTTGAACAGGCTGAGAGAGCCCCGAAGGAGCT
1705 E N E L N D A N D R V N E L S A Q N S S
5401 GAGAACGAGCTGAACGATGCCAATGACCGCGTCAACGAACTGTCCGCCCAGAACAGCTCC
1725 I Q G Q K R K L E G D V Q A M Q T D L D
5461 ATTCAAGGACAGAAGAGGAAGTTGGAGGGTGTGTACAGCCATGCAGACTGACCTGGAC
1745 E M S N E L R A S E E R S K K A M A D A
5521 GAGATGAGCAACGAACTGAGAGCCTCCGAGGAGAGATCCAAGAAGGCTATGGCTGATGCC
1765 A R L A D E L R Q E Q E H S L Q V E K L
5581 GCCCGTGGCTGATGAACTGCGCCAGGAACAGGAACACAGCTTGCAGGTAGAGAAGTTG
1785 R K G L E G A L K D T Q V R L E E A E A
5641 AGGAAGGACTGGAAGGTGCCCTCAAGGACACCCAGGTCCGCTCGGAGGAGGCTGAGGCC
1805 Q A L K G G K K M I Q K L E Q R V R L Q
5701 CAGGCCTCAAGGGAGGCAAGAAGATGATCCAGAACTGGAACAGAGAGTCAGGGAAGT
1825 E T E L D N E Q R R H A E T Q K N M R K
5761 GAGACAGAAGTGGACAACGAGCAGCGCCGCTGCGGAGACCCAGAAGAACATGCGCAAG
1845 A D R R L K E L A F Q A D E D R K N Q E
5821 GCTGACCGCAGACTCAAGGAGTTGGCCTTCCAGGCTGACGAGGACCGCAAGAACCAGGAA
1865 R L Q D L V D K L Q N K I K T Y K R Q V
5881 CTTCCAGGATCTGGTGGCAAAATGCAAAACAAGATCAAGACCTACAAGAGACAAAGTT
1885 E E A E E I A A I N L A K Y R K V Q N E
5941 GAAGAGGCCGAGGAGATTGCCGCATCAATCTGGCCAAGTACCGCAAGGTGCAGAAGCAG
1905 L E D A E E R A D T A E G S L S K L R A
6001 CTGGAAGATGCCGAGGAGCCGCTGACACCGCCGAGGGCTCTCTCTCAAGCTGAGAGCC
1925 K N R S S V S V S R T S A S P S F P G P
6061 AAGAACAGGAGCTCCGTGTCCGTGTACGCACTAGTGTTCACCCAGCTTCCCAGGCC
1945 S V S A S R A S V D R S A R A P S Q S
6121 AGTTCGGTGTCTGCCAGCCGAGCCAGCGTGGACAGATCAGCACGCGCCCCGTCCCAATCA
1965 *
6181 TAAactctgatgccacttttttagttctatgtcaatgatttgcgagtcttcagctttcttc
6241 aatgtacgcgccacaccaagtttgaccacatccttccgcttcggccactccgctcaatgag
6301 gcatcgtaccgggtccgcaggagttgaagtgaccaccggcggggtagatgacgtcactc
6361 catcgtaatctaccaggccgcagacctcaactcggcctcctgcaggcccaggtcgaccact
6421 gcaaccaccaccgctacaaccaccactaccactactaccaccaccatcatcactagt
6481 ggcattacctagaattctgacaacgaggggtggaatccggactatacttattattagact
6541 aactttttttgtgtgttccaaactgagatgtgaccggttacggaatgaagaaaacaa
6601 gtacaacaacgacaacaacggcaacagcaacagcaagatcaacaacaacaacaacaaca
6661 caacaacaacgacacatctctctagaaaacttgatttgctcgtgtttactgactcgtgt
6721 agtgtgctcgggttttttcatgttgtaaaaataatacatgtccactggctctatagccc
6781 atccgagatgtgttccatataaggtcagttaatgttggtttgtttttttgtttttat
6841 aaagtctgttagacatcttcaaccatacgtgttcacctctccatgataagtaactagtgt
6901 cacacaaaagctgtaccgctgttcttcttaagcctgcgcgcgcccacaaatgggtacacac
6961 tcgacagtgtagtgactcggggtactatcacgacaacacgcccagagtttctccatcctct
7021 gcatgcacacgagaccgatcagaacttacagaaacaccagactgttaggtcattatcata
7081 agatcaagttagcttaatgaataaaaacttaactcggacggtgaaagatgtgctcaagactt
7141 agcaaacatcattgattgccaacctgaagacgacgaagagtaggctggctctagaacgac
7201 aacaagaacaataatgataatgataataataataatcaacaggggaagtaaccaagtt
7261 actgcaaggtgactgattgtgtgggaggtgtgcttgctaattgtacgcccggcgacag
7321 agaattcatcctttgtacaaaagattagaagtattttaatttaattttttttctgtgtt
7381 attttctttgttctgtggaatcccgtttatgagcaaaatagatttatcacctttttat
7441 taacaaggtttttctgtattcgttatatttgaataaaacacatgaatgcatcttgccca
7501 ggcttttgggacttgatgtgtttgtctgtatgttcggtcogttcgttctgcacgaaaaa
7561 aaaaaa

```

图 1 瘤背石磺 MyHC cDNA 序列和推测的氨基酸序列

Fig.1 Nucleotide and deduced amino acid sequences of *O. struma* MyHC

基因序列编码区用大写字母表示, 非编码区用小写字母表示。下划线表示 Poly(A)尾巴, 双下划线表示 Myosin_tail_1 结构域, 阴影部分为 AATAAA 加尾信号

The UTR and ORF regions are shown in lowercase and uppercase letters, respectively; underline is the polyadenylation signal; Myosin_tail_1 domain is double underlined; the shadow is the putative eukaryotic polyadenylation signal (AATAAA)

表 2 瘤背石磺 MyHC 多肽链的氨基酸组分
Tab.2 Amino acid composition of MyHC from *O. struma*

Table with 6 columns: Amino acid, Amount, Percentage(%), Amino acid, Amount, Percentage(%). Rows list amino acids like Ala (A), Arg (R), Asn (N), Asp (D), Cys (C), Gln (Q), Glu (E), Gly (G), His (H), Ile (I), Leu (L), Lys (K), Met (M), Phe (F), Pro (P), Ser (S), Thr (T), Trp (W), Tyr (Y), Val (V) with their respective counts and percentages.

化位点, 糖基化位点预测结果显示有 4 个 N 糖基化位点。利用 Blastp 将蛋白序列与 GenBank 数据库所收录的全部蛋白序列进行比对, 结果显示, 瘤背石磺肌球蛋白重链按其结构有 4 个 ATP 结合位点, 1 个 P-loop、1 个 SH1 helix、1 个 switch II region、1 个 switch I region、2 个 converter subdomain、1 个 purine-binding loop、1 个 relay loop, 同时包括 2 个保守结构域

MYSc_class_II 和 Myosin_tail_1 结构域, MyHC Myosin_tail_1 结构域与其他物种 MyHC 序列进行比对见图 2, 肌球蛋白重链三级预测结构如图 3。

2.3 同源性分析

为分析瘤背石磺 MyHC 的进化地位, 根据 MyHC 氨基酸序列, 使用 MrBayes-3.2.6 和 MEGA6.0(前者

Sequence alignment table showing amino acid sequences for O. struma, B. glabrata, A. californica, O. bimaculoides, H. bleekeri, and S. esculenta across various positions (1, 26, 86, 121, 146, 181, 204, 239, 240, 239, 239, 239, 241). Includes a Consensus line for each section.

<i>O. struma</i>	264	ELEELSERLDEAGGATAAQVELNKKREQELLKLRRLDEESALQHEAQISATLRKKKQDAAN
<i>B. glabrata</i>	299	ELEELGERLDEAGGATAAQVELNKKREQELLKLRRLDEESALQHEAQISATLRKKKQDAAN
<i>A. californica</i>	300	ELEDLGERLDEAGGATAAQVELNKKREQELLKLRRLDEESALQHEAQISSLRKKKHQDAAN
<i>O. bimaculoides</i>	299	ELEEVGERLDEAGGATAAQVELNKKREQELLRLRRDLDEESSMQHEAQIATLRKKNTBATN
<i>H. bleekeri</i>	299	ELEELGERLDEAGGATAAQVELNKKREQELLRLRRDLDEESNMQHEBQIATLRKKKNQBATN
<i>S. esculenta</i>	299	ELEELGERLDEAGGATAAQVELNKKREQELLRLRRDLDEEGTMQHEBQIATLRKKKNQBATN
Consensus	301	***. . . *****. *****. *****. . . ***. ** . . ***. . . *
<i>O. struma</i>	324	EMADQIDQLQKVKNKVEKEKAQLKSEVGDQLQAQLEHAGKNKGMSEKMTKQVEAQISELNA
<i>B. glabrata</i>	359	ELADQIDQLQKVKTKEKEKSQLKSEADDLRAQLEHGSKNKGMSEKMTKQVEAQISELNS
<i>A. californica</i>	360	EMADQVDQLQKVKSRLDKEKNQLKGEVDDLQAQLEHSSKNKGMSEKMSKQVEAQISELNA
<i>O. bimaculoides</i>	359	ELSDQIDQLQKVKSRLKEKESQLRGEVDDLQQLLEHVGNKRGCSDEKMSKQVEAQISELNG
<i>H. bleekeri</i>	359	ELGDQIDQLQKVKSRLKEKEKIQLRAPLDDVQSQLEHAGKNKRGCSDEKMSKQVEAQISELNA
<i>S. esculenta</i>	359	ELGDQIDQLQKVKSRLKEKESQLRGEVDDLQSQLEHAGKNKRGCSDEKMSKQVEAQISELNG
Consensus	361	*. *. *****. . . ***. *. . . *. . . *. . . *. . . *. . . *. . . *. . . *
<i>O. struma</i>	384	KLDAANRQIQELGGQKARSAQENADLTRQLEEAHVRVGLTKKKNALAAALEEAKR
<i>B. glabrata</i>	419	KLDAANRQIQELGGQKNRAAQENADLTRQLEEAHVKVGLTKERTALATQLEETKR
<i>A. californica</i>	420	KLDAARBITQIQELGGQKNRGSQESADLTRQLEEAHVRVGLTKKKTSLASQLEEAKR
<i>O. bimaculoides</i>	419	KIDEQARSVSELTSQKSRLQGENADLTRQLEEAHNVGQLTKLKNLNHQLLEEAKR
<i>H. bleekeri</i>	419	KIDDDQRSVSELTSQKSRLQTEAADLTRQLEEAHNVGQLTKLKS SSMGANLEDAKR
<i>S. esculenta</i>	419	KIDDDQRSVSELSSQKSRLQTEAADLTRQLEEAHNVGQLSKLKS SSSLNLEDAKR
Consensus	421	*. *. *. . . *. . . *****. *. . * *. . . *

图 2 瘤背石磺 MyHC 和其他物种 MyHC Myosin_tail_1 结构域氨基酸序列的多重比对

Fig.2 Alignment of the deduced amino acid sequences of *O.struma* and other reported MyHC Myosin_tail_1 domain sequences

“*”表示保守的氨基酸残基,“.”表示相似性氨基酸

“*” indicates highly conserved amino acid residue; “.” represents similar amino acid

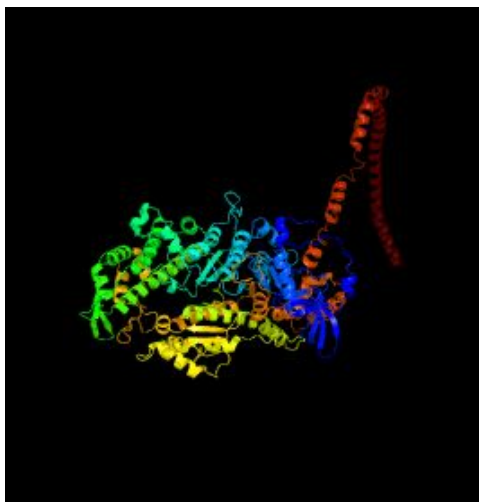


图 3 预测的瘤背石磺 MyHC 三级结构

Fig.3 The predicted tertiary structures of *O.struma* MyHC

为 MrBayes 分析结果,后者为 MEGA 分析结果)构建了包括瘤背石磺在内的 19 个物种 MyHC 序列(表 3)的系统进化树。系统进化树结果显示,无脊椎动物和脊椎动物分为两支,瘤背石磺和光滑双脐螺的 MyHC 聚类成一个分支,然后与加州海兔聚在一起,海湾扇贝(*Argopecten irradians*)、欧洲大扇贝(*Pecten maximus*)和长牡蛎(*Crassostrea gigas*)聚在一起,日本矛鱿鱼(*Heterololigo bleekeri*)、金乌贼(*Sepia esculenta*)和章鱼(*Octopus bimaculoides*)聚在一起,脊椎动物方面按照形态学分类地位聚类(图 4)。

表 3 其他物种 MyHC 的 NCBI 登录号

Tab.3 Accession numbers of MyHC from other species

蛋白 Protein	登录号 Accession No.
人 <i>Homo sapiens</i>	NM_005963
黑猩猩 <i>Pan troglodyte</i>	XM_003953128.2
猕猴 <i>Macaca mulatta</i>	NM_001195293
猪 <i>Sus scrofa</i>	NM_001104951.1
小鼠 <i>Mus musculus</i>	NM_030679
大鼠 <i>Rattus norvegicus</i>	NM_001135158.1
非洲爪蟾 <i>Xenopus laevis</i>	NM_001127883
翘嘴鲈 <i>Siniperca chuatsi</i>	HQ829291
斑马鱼 <i>Danio rerio</i>	P02564
日本矛鱿鱼 <i>Heterololigo bleekeri</i>	ACD68201.1
金乌贼 <i>Sepia esculenta</i>	ACD68202.1
章鱼 <i>Octopus bimaculoides</i>	CDG41623.1
光滑双脐螺 <i>Biomphalaria glabrata</i>	XP_013068756.1
加州海兔 <i>Aplysia californica</i>	XP_012941300.1
长牡蛎 <i>Crassostrea gigas</i>	XP_011442517.1
欧洲大扇贝 <i>Pecten maximus</i>	AAD5284.1
海湾扇贝 <i>Argopecten irradians</i>	AAC46490.1
日本三角涡虫 <i>Dugesia japonica</i>	CAB64662.1

2.4 不同组织的表达差异分析

以 18S rRNA 为内参基因,荧光定量 PCR 检测瘤背石磺 MyHC 基因在背部皮肤、腹部皮肤、口球、腹足、肺囊、肝胰腺、蛋白腺和两性腺 8 个组织中的表达,结果显示,MyHC 基因在各个组织中均有表达;

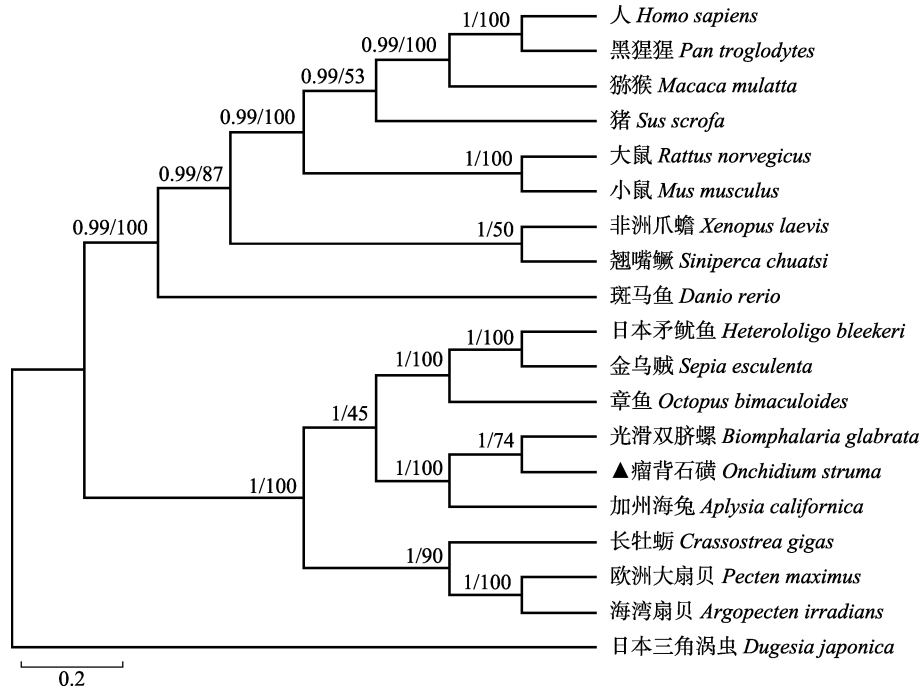


图 4 基于 MyHC 氨基酸序列构建的系统树

Fig.4 Phylogenetic analysis of *O.struma* MyHC and other reported MyHC sequences

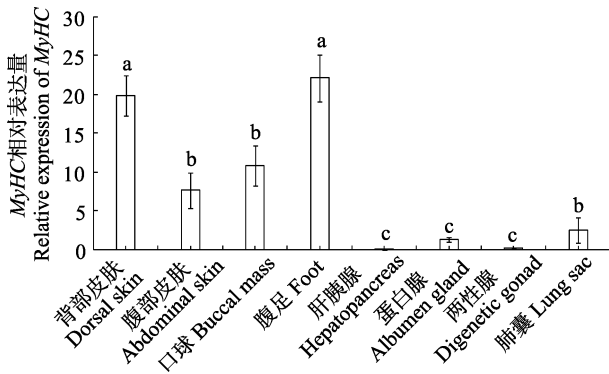


图 5 MyHC 在瘤背石磺不同组织的表达

Fig.5 qRT-PCR analysis of the expression profiles of MyHC in different tissues of *O.struma*

相同字母表示差异不显著($P>0.05$),
不同字母表示差异显著($P<0.05$)

Same letters above the bars indicate no significant difference ($P>0.05$), statistical significance is presented by different letters ($P<0.05$)

腹足和背部皮肤表达量最高,腹部皮肤、口球和肺囊中高表达,肝胰腺、蛋白腺和两性腺中微量表达且表达差异显著($P<0.05$) (图 5)。

3 讨论

3.1 瘤背石磺 MyHC 基因的存在形式

系统进化树显示,瘤背石磺的MyHC与光滑双脐

螺的MyHC聚成一个分支,其次是加州海兔,这与它们的形态学分类结果一致,而且从系统进化树也能看出各物种从低等到高等、从水生到陆生的进化规律,表明肌球蛋白重链是研究生物进化良好的备选基因。

同时有研究显示,脊椎动物中大部分都含有多种肌球蛋白重链基因和亚型,在禽类中至少有 9 种 MyHC 基因被鉴定(Kropp *et al*, 1987; Bandman *et al*, 2000),在哺乳动物中至少也有 13 种,其中,8 种来自于横纹肌,5 种来自于平滑肌和非肌肉组织(Weiss *et al*, 1999; Flück *et al*, 2003)。在鱼类肌球蛋白重链多基因家族研究中,分别从虹鳟(*Oncorhynchus mykiss*)的红肌、白肌和胸肌中获得 3 种不同的 MyHC cDNA 片段(Weaver *et al*, 2001);从翘嘴鲈中克隆得到红肌肌球蛋白重链 cDNA 片段(易潭等, 2014)。Nakaya 等(1995)发现肌球蛋白分子 α -螺旋结构会因基因亚型的差异出现显著的不同,Imai 等(1997)和 Hurliaux 等(1991)发现了在不同温度下的不同类型的 MyHC cDNA。斑马鱼(*Danio rerio*)的骨骼肌纤维也存在 3 种不同的类型,应存在不同的 MyHC 基因的亚型,而不同的 MyHC 基因的亚型在不同的肌纤维中的表达也存在较大的差异。瘤背石磺中是否存在 MyHC 基因亚型,以及存在哪几种 MyHC 基因亚型还有待进一步研究判断。

3.2 瘤背石磺 MyHC 基因蛋白结构分析

本研究克隆获得瘤背石磺 MyHC 具有典型的保

守结构组件, 4个ATP结合位点, 1个SH1螺旋, 3个Loop环, 2个转换区域, 2个转换子区域, 这些氨基酸残基组成了瘤背石磺 MyHC 两个保守结构域 MYSc_class_II 和 Myosin_tail_1 的特性。其中, MYSc_class_II 位于第33~256位氨基酸, Myosin_tail_1 位于第881~1925位氨基酸, 位于蛋白尾部, 由卷曲的肌球蛋白重链尾部区域构成, 为粗肌丝提供了结构支撑。已知 MyHC 家族蛋白的典型结构主要是由 N 端球状盘绕的头部和 C 端 α -螺旋型的尾部组成, 头部有 ATP 结合位点、Loop1 和 Loop2 环区, 是 MyHC 家族中的保守区, 同时也是肌球蛋白动力产生的主要区域, Loop1 表面环状结构决定肌纤维收缩速率以及持续的动力, 而 Loop2 表面环状结构则影响 MyHC 上游肌动蛋白激活的 ATP 酶水解活性(Holmes *et al.*, 2002)。另外, 在 MyHC 中还发现了 SH1 螺旋。SH1 修饰肌球蛋白亚片段能大大提高其热稳定性, SH1 可以大大减少整个肌球蛋白亚片段的改变(Lubov *et al.*, 2007), 其对肌纤维运动等功能有待进一步研究验证。本研究获得的瘤背石磺 MyHC 的蛋白理化性质的预测结果发现, 瘤背石磺 MyHC 蛋白 MYSc_class_II 结构域疏水性氨基酸所占的比例较大。目前, 刘希良等(2012)根据已经克隆的鳊鱼 MyHC 序列制备出原核表达载体和多克隆抗体。瘤背石磺肌球蛋白重链 Myosin_tail_1 保守区域中亲水性氨基酸较多, 属于亲水性氨基酸密集区域, 该实验结果对深入研究石磺肌球蛋白重链基因的原位杂交实验、Western blot 检测奠定了基础。

在对目前哺乳动物骨骼肌中 MyHC 基因进行研究时, 已经发现有多种途径可以调控该基因的同型异构体表达, 其中有两途径研究较多, 分别是通过肌源调节因子(Myogenic regulatory factor, MRF)和活化 T 细胞核因子(Nuclear factor of activated T cells, NFAT)(Devoto *et al.*, 1996)。前者由 MyoD、myogenin、myf5 和 MRF4 这4个蛋白组成, 后者调控途径的主要方式是通过调节 Ca^{2+} 浓度的变化。研究表明, 当肌肉收缩时, 释放 Ca^{2+} 数量增多, 相应的细胞浆内 Ca^{2+} 浓度就会上升, 而升高的 Ca^{2+} 浓度会通过一系列的化学反应来激活下游相应的蛋白, 由此使得细胞质中某些蛋白质进行脱磷酸化, 随后进入细胞核中, 与一个保守的 DNA 结合位点相结合, 以此来调节相应转录水平的变化(Tao *et al.*, 2004)。Kubis 等(1997)在对兔培养的肌细胞进行研究时, 将 Ca^{2+} 载体加入培养基中, 使肌浆中 Ca^{2+} 浓度缓慢升高, 发现此时肌细胞中慢型肌球蛋白重链和轻链的比例较高。在对石磺

MyHC 基因构成的蛋白质进行结构分析时, 发现钙调素结合区段以及许多磷酸化位点, 例如蛋白激酶 C 的活性必须依赖 Ca^{2+} 的存在。虽然现在没有确切的证据表明石磺内有相似的调控机制来控制 MyHC 基因的特异性表达, 但可以为今后探究石磺的肌球蛋白的调控机制提供基础依据。

3.3 MyHC 基因的表达与环境适应性分析

研究证实, MyHC 基因的表达量会显示出对于周围环境变化的适应和调节, Altringham 等(1985)发现鲤鱼会通过 MyHC 基因的表达量的不同来调节 ATP 酶的活性以适应温度的变化。Nihei 等(2006)通过对鲤鱼胚胎和肌肉组织中的 MyHC 基因表达量的研究发现, 鲤鱼 MyHC 基因不仅在不同发育阶段、不同肌肉类型表现出表达量的差异, 而且在不同温度下也会存在这种差异。此外, 在其他鱼类品种中, Weaver 等(2001)利用 RT-PCR 检测幼鱼阶段虹鳟的平滑肌、骨骼肌和心肌中 MyHC 基因表达, 显示平滑肌表达 MyHC 基因的表达略高于骨骼肌和心肌, 在进一步对鱼苗期和幼鱼期的平滑肌中 MyHC 基因表达量进行比较时发现, 幼鱼期 MyHC 基因表达量高于鱼苗期。本研究中, MyHC 在腹足部位高表达, 同时在其他各个组织中都进行表达, 充分证明肌球蛋白既是一种不可缺少的结构蛋白, 同时又能作为功能蛋白发挥作用; 腹足作为瘤背石磺的运动器官, MyHC 在该器官中表达量最高, 增强了瘤背石磺的运动能力, 为瘤背石磺逐渐适应陆生环境提供了有力的条件。

4 结论

本研究从瘤背石磺转录组差异表达基因中筛选到 MyHC Unigenes, 进行了 cDNA 序列克隆、结构分析、进化分析以及 qRT-PCR 表达水平分析。结果表明, 瘤背石磺 MyHC 基因的结构特征主要有: 存在多个 N-糖基化位点, MYSc_class_II 和 Myosin_tail_1 两个保守结构域, N 末端具有多个 ATP 结合位点等特异性功能区域。系统进化树显示, 瘤背石磺 MyHC 基因与光滑双脐螺聚成一支, 其符合形态学分类且物种等级相近; 表达水平分析显示, 瘤背石磺 MyHC 基因具有不同的组织分布, 在组织运动最发达的腹足有高水平的表达。通过对瘤背石磺 MyHC 基因结构、功能预测、组织表达以及分子进化的分析, 为进一步探究石磺肌球蛋白的调控机制提供基础依据; 为探究肌肉运动对瘤背石磺湿地环境两栖适应性的相关联系; 对研究瘤背石磺肌球蛋白重链基因进行原核表达及

多克隆抗体的制备提供基础资料,更为深入探讨海洋无脊椎动物从海洋向陆地进化的研究提供有参考意义的分子依据。

参 考 文 献

- Altringham JD, Johnston IA. Changes in tension generation and ATPase activity in skinned muscle fibres of the carp following temperature acclimation. *Pflügers Archiv- European Journal of Physiology*, 1985, 403(4): 449–451
- Bandman E, Rosser BWC. Evolutionary significance of myosin heavy chain heterogeneity in birds. *Microscopy Research and Technique*, 2000, 50(6): 473–491
- Bouchet P, Rocroi JP. Classification and nomenclator of gastropod families. *Malacologia: International Journal of Malacology* (Hackenheim, Germany: ConchBooks), 2005, 47(1): 1–397
- Chen ZH, Dong JJ, Sun CF, *et al.* cDNA cloning and analyses of two myosin heavy chain isoforms of mandarin fish (*Siniperca chuatsi*) based on transcriptome sequencing. *Progress in Fishery Sciences*, 2017, 38(3): 51–61 [陈之航, 董浚键, 孙成飞, 等. 基于转录组测序对翘嘴鲌(*Siniperca chuatsi*)2种肌球蛋白重链基因的克隆与分析. *渔业科学进展*, 2017, 38(3): 51–61]
- Craig R, Woodhead JL. Structure and function of myosin filaments. *Current Opinion in Structural Biology*, 2006, 16(2): 204–212
- Dayrat B. Review of the current knowledge of the systematics of Onchidiidae (Mollusca: Gastropoda: Pulmonata) with a checklist of no minal species. *Zootaxa*, 2009, 2068(2068): 1–26
- Devoto SH, Melancon E, Eisen JS, *et al.* Identification of separate slow and fast muscle precursor cells in vivo, prior to somite formation. *Canadian Metallurgical Quarterly*, 1996, 122(11): 3371–3380
- Flück M, Hoppeler H. Molecular basis of skeletal muscle plasticity—from gene to form and function. *Reviews of Physiology, Biochemistry and Pharmacology*, 2003, 146(1): 159–216
- Holmes JM, Whiteley NM, Magnay JL, *et al.* Comparison of the variable loop regions of myosin heavy chain genes from Antarctic and temperate isopods. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, 2002, 131(3): 349–359
- Hooper SL, Thuma JB. Invertebrate muscles: Muscle specific genes and proteins. *Physiological Reviews*, 2005, 85(3): 1001–1060
- Huriaux F, Vandcalle P, Focant B. Myosin heavy chain isoforms in white, red and ventricle muscles of barbel (*Barbus barbus* L). *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, 1991, 100(2): 309–312
- Imai J, Hirayama Y, Kikuchi K, *et al.* cDNA cloning of myosin heavy chain isoforms from carp fast skeletal muscle and their gene expression associated with temperature acclimation. *Journal of Experimental Biology*, 1997, 200(1): 27–34
- Iwami Y, Ojima T, Jnoue A, *et al.* Primary structure of myosin heavy chain from fast skeletal muscle of Chum salmon *Oncorhynchus keta*. *Comparative Biochemistry & Physiology Part B Biochemistry & Molecular Biology*, 2002, 133(2): 257–267
- Jaenicke T, Diederich KW, Haas W, *et al.* The complete sequence of the human beta-myosin heavy chain gene and a comparative analysis of its product. *Genomics*, 1990, 8(2): 194–206
- Knight AE, Molloy JE. Muscle, myosin and single molecules. *Essays in Biochemistry*, 2000, 35(35): 43–59
- Kropp KE, Gulick J, Robbins J. Structural and transcriptional analysis of a chicken myosin heavy chain gene subset. *Journal of Biological Chemistry*, 1987, 262(34): 16536–16545
- Kubis HP, Haller EA, Wetzel P, *et al.* Adult fast myosin pattern and Ca²⁺ induced slow myosin pattern in primary skeletal muscle culture. *Proceedings of the National Academy of Sciences of the United States of America*, 1997, 94(8): 4205–4210
- Li YY, Shao F, Wu XJ, *et al.* Molecular cloning, sequence and expression characterization of myosin heavy chain 1 gene (*MyHCl*) in goose (*Anser anser*) at embryonic period. *Journal of Agricultural Biotechnology*, 2015, 23(3): 359–368 [李艳艳, 邵芳, 吴小娟, 等. 鹅肌球蛋白重链1基因 *MyHCl* 全长 cDNA 克隆、序列及胚胎期表达特征分析. *农业生物技术学报*, 2015, 23(3): 359–368]
- Lied E, von der Decken A. Purification of fish muscle myosin heavy chain and quantification of the specific polyribosome-bound polypeptide. *Biochemical Journal*, 1985, 232(2): 467–470
- Liu C, Wei LL, Shen HD, *et al.* Complete mitochondrial genome of *Peronia verruculata* (Gastropoda: Pulmonata: Systellomatophora: Onchidiidae). *Mitochondrial DNA*, 2014, 27(2): 918–920
- Liu XL, Wang KZ, Cheng J, *et al.* Prokaryotic expression and preparation of polyclonal antibody of myosin heavy chain genes in *Siniperca chuatsi*. *Journal of Natural Science of Hunan Normal University*, 2012, 35(6): 67–73 [刘希良, 王开卓, 成嘉, 等. 鳊鱼肌球蛋白重链基因的原核表达及多克隆抗体的制备. *湖南师范大学自然科学学报*, 2012, 35(6): 67–73]
- Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2^{- $\Delta\Delta C_t$} method. *Methods*, 2001, 25(4): 402–408
- Lubov S, Valeria M, Elena S, *et al.* Nucleotide-induced and actin-induced structural changes in SH1-SH2-modified myosin subfragment 1. *Journal of Muscle Research and Cell*

- Motility, 2007, 28(1): 67–78
- Molina MI, Kropp KE, Gulick J, *et al.* The sequence of an embryonic heavy chain gene and isolation of its corresponding cDNA. *Journal of Biological Chemistry*, 1987, 262(14): 6478–6488
- Muramatsu UM, Kikuchi KH, Ikeda D, *et al.* The complete genomic sequence of the carp fast skeletal myosin heavy chain gene. *Gene*, 2005, 349(2): 143–151
- Nakaya M, Watabe S, Ooi T. Differences in the thermal stability of acclimation temperature-associated types of carp myosin and its rod on differential scanning calorimetry. *Biochemistry*, 1995, 34(9): 3114–3120
- Nihei Y, Kobiyama A, Ikeda D, *et al.* Molecular cloning and mRNA expression analysis of carp embryonic, slow and cardiac myosin heavy chain isoforms. *Journal of Experimental Biology*, 2006, 209(1): 188–198
- Pette D, Staron RS. Myosin isoform, muscle fiber types and transitions. *Microscopy Research & Technique*, 2000, 50(6): 500–509
- Strehler EE, Strehler-Page MA, Perriard JC, *et al.* Complete nucleotide and encoded amino acid sequence of a mammalian myosin heavy chain gene. Evidence against intron-dependent evolution of the rod. *Journal of Molecular Biology*, 1986, 190(3): 291–317
- Sun BN, Chen C, Shen HD, *et al.* Species diversity of Onchidiidae (Eupulmonata: Heterobranchia) on the mainland of China based on molecular data. *Molluscan Research*, 2014, 34(1): 62–70
- Tao Y, Kobayashi M, Liang CS, *et al.* Temperature-dependent expression patterns of grass carp fast skeletal myosin heavy chain genes. *Comparative Biochemistry & Physiology Part B: Biochemistry and Molecular Biology*, 2004, 139(4): 649–656
- Wang CN, Shen HD, Zheng P. Evaluation of genetic diversity in the *Onchidium struma* from different geographical populations in China by inter simple sequence repeat (ISSR). *Biotechnology Bulletin*, 2014(9): 208–212 [王成暖, 沈和定, 郑培. 中国沿海瘤背石磺不同地理群体遗传多样性的 ISSR 标记研究. *生物技术通报*, 2014(9): 208–212]
- Wang DF, Shen HD, Wu X. Comparative histology of the skin from three species of Onchidiidae. *Chinese Journal of Zoology*, 2015, 50(3): 437–444 [王冬凤, 沈和定, 吴欣. 石磺科 3 种贝类皮肤显微结构比较. *动物学杂志*, 2015, 50(3): 437–444]
- Weaver FE, Stauffer KA, Coughlin DJ. Myosin heavy chain expression in the red, white, and ventricular muscle of juvenile stages of rainbow trout. *Journal of Experimental Zoology*, 2001, 290(7): 751–758
- Weiss A, McDonough D, Wertman B, *et al.* Organization of human and mouse skeletal myosin heavy chain gene clusters is highly conserved. *Proceedings of the National Academy of Sciences of the United States of America*, 1999, 96(6): 2958–2963
- Wu XF, Shen HD, Wu WJ, *et al.* Comparison on morphology of Onchidiidae in eastern coast of China. *Chinese Journal of Zoology*, 2010, 45(6): 92–100 [吴旭峰, 沈和定, 吴文健, 等. 我国华东沿海 4 种石磺形态学比较. *动物学杂志*, 2010, 45(6): 92–100]
- Yao LX, Zhou N, Shen HD, *et al.* Genetic diversity and differentiation of *Platevindex mortoni* from the coastal area of China based on mitochondrial COI gene. *Journal of Fishery Sciences of China*, 2015, 22(4): 729–739 [姚理想, 周娜, 沈和定, 等. 中国沿海平疣桑椹石磺 COI 基因的遗传多样性与遗传分化. *中国水产科学*, 2015, 22(4): 729–739]
- Yi T, Liu XL, Bin SY, *et al.* Molecular cloning and expression analysis of slow myosin heavy chain 1(*sMyHC1*) gene cDNA in mandarin fish (*Siniperca chuatsi*). *Journal of Guangxi Normal University*, 2014, 32(2): 175–180 [易潭, 刘希良, 宾石玉, 等. 鳊鱼 *Siniperca chuatsi* 红肌 *sMyHC1* 基因 cDNA 的克隆及其表达分析. *广西师范大学学报*, 2014, 32(2): 175–180]

(编辑 冯小花)

Molecular Cloning and Analysis of the *MyHC* Gene in *Onchidium struma*

GU Bingning, LIU Xin, SHEN Heding^①, WANG Dongfeng, YANG Tiezhu,
ZHU Min, SHI Yanmei, LI Bohang

(Key Laboratory of Exploration and Utilization of Aquatic Genetic Resources, Ministry of Education;
National Demonstration Center for Experimental Fisheries Science Education; Shanghai Universities Key
Laboratory of Marine Animal Taxonomy and Evolution, Shanghai Ocean University, Shanghai 201306)

Abstract *Onchidium struma* belongs to Mollusca, Pulmonata, and was regarded as the transitional taxa from water to land because of its amphibious characteristics. Having similar characters to amphibians, *O. struma* is a good subject for the study of marine invertebrates extending landward. As the basic component of muscle units, myosin directly influences the muscle growth and meat quality *via* its molecular diversity and composition. It is a ubiquitous eukaryotic motor protein that interacts with actin to generate the force for cellular movements, ranging from cytokinesis to muscle contraction. In this study, we used rapid amplification cDNA ends (RACE) methods to obtain the full-length cDNA of the myosin heavy chain (*MyHC*) gene in *O. struma*. We performed bioinformatic and expression pattern analysis of *MyHC* mRNA in different tissues detected by real-time fluorescent quantitative PCR (qRT-PCR). The full length of the *MyHC* cDNA sequence consists of 7566 base pairs (bp) comprising a 228 bp 5' untranslated region (UTR), a 1443 bp 3' UTR, and a 5895 bp open reading frame (ORF) which encodes 1964 amino acids. *MyHC* gene was expressed in various tissues; its highest expression was found in the foot and its lowest in the hepatopancreas ($P < 0.05$). Moreover, the *MSTN* protein was predicted to be composed of 31,713 atoms and its formula is $C_{9765}H_{15897}N_{2849}O_{3150}S_{52}$, with a calculated relative molecular weight of 225.28 kDa and a *pI* of 5.56. The result of signal peptide prediction shows that the N-terminal has a signal peptide of 29 amino acids in length. Additionally, common features were found in the *MyHC* of *O. struma*, including MYSc class II and myosin tail I domain. Molecular phylogenetic analysis shows that *O. struma* is closely related to *Biomphalaria glabrata*. This study provided a novel myosin heavy chain gene sequence in *O. struma* and the results indicate that the *MyHC* gene is important for the growth and development of this animal, as well as its muscle characterization. Furthermore, the results revealed that *MyHC* is not only an essential structural protein, but also a functional protein in *O. struma*. Conclusively, *MyHC* is an excellent candidate gene for studying biological evolution.

Key words *Onchidium struma*; *MyHC*; Gene cloning; Tissue expression

^① Corresponding author: SHEN Heding, E-mail: hdshen@shou.edu.cn