

DOI: 10.19663/j.issn2095-9869.20171219002

http://www.yykxjz.cn/

史晓丽, 张莹雪, 孟宪红, 孔杰, 栾生, 罗坤, 曹宝祥, 曹家旺, 陈宝龙. 中国对虾甘氨酸脱羧酶的基因克隆、表达及其与抗WSSV的关联分析. 渔业科学进展, 2019, 40(2): 74-82

Shi XL, Zhang YX, Meng XH, Kong J, Luan S, Luo K, Cao BX, Cao JW, Chen BL. cDNA cloning of the GLDC gene in *Fenneropenaeus chinensis* and its expression and functional analyses after WSSV infection. Progress in Fishery Sciences, 2019, 40(2): 74-82

中国对虾甘氨酸脱羧酶的基因克隆、表达及其与抗 WSSV 的关联分析*

史晓丽^{1,2} 张莹雪^{1,2} 孟宪红^{1,2①} 孔杰^{1,2} 栾生^{1,2}
罗坤^{1,2} 曹宝祥^{1,2} 曹家旺^{1,2} 陈宝龙^{1,2}

(1. 农业农村部海洋渔业可持续发展重点实验室 中国水产科学研究院黄海水产研究所 青岛 266071;
2. 青岛海洋科学与技术试点国家实验室海洋渔业科学与食物产出过程功能实验室 青岛 266071)

摘要 本研究采用 RACE 技术克隆获得中国对虾(*Fenneropenaeus chinensis*)甘氨酸脱羧酶基因(*FcGLDC*)的全长 cDNA 及 DNA 序列, 并对其进行生物信息学分析。结果显示, *FcGLDC* 基因的 cDNA 全长为 3481 bp, 其中, ORF 为 2829 bp, 5'UTR 长 17 bp, 3'UTR 长 86 bp。完整的阅读框编码 942 个氨基酸, 分子量为 104.66 kDa, 预测的理论等电点为 6.51。*FcGLDC* 基因 DNA 序列全长共 4964 bp, 包含 12 个外显子和 11 个内含子。同源性及系统进化分析显示, *FcGLDC* 基因与节肢动物的 *GLDC* 基因聚为一类; 氨基酸序列比对发现, *FcGLDC* 基因的蛋白序列与节肢动物的相似度最高, 与内华达白蚁(*Zootermopsis nevadensis*)、体虱(*Pediculus humanus corporis*)和白纹伊蚊(*Aedes albopictus*)的相似度分别为 71%、68% 和 68%。在鳃、肝胰腺和肌肉中的荧光定量 PCR 结果显示, *FcGLDC* 在肌肉中的相对表达量最高, 鳃中最低。WSSV 感染后, 该基因在鳃、肝胰腺和肌肉中呈现出了不同的时空表达特点。使用直接测序法结合质谱法, 在该基因内部发现 4 个 SNP 位点, 但是各位点与抗 WSSV 性状均不相关($P>0.05$)。本研究表明, *FcGLDC* 基因在对虾感染 WSSV 后的应答反应中或起一定作用。

关键词 中国对虾; 甘氨酸脱羧酶; 基因克隆; 表达; 关联分析

中图分类号 Q785; S966.1 **文献标识码** A **文章编号** 2095-9869(2019)02-0074-09

中国对虾(*Fenneropenaeus chinensis*)是我国重要的土著海水养殖种类之一, 近十几年来, 其养殖产业饱受病害困扰。在诸多病原当中, 白斑综合征病毒

(White spot syndrome virus, WSSV)是对对虾养殖生产毒害最大的病毒之一(Lightner, 1996), 它导致的疾病传播快、破坏性强, 且防治困难, 已给我国乃至全

* 中国水产科学研究院黄海水产研究所基本科研业务费(20603022017001)、国家自然科学基金(41676148)、泰山学者良种工程项目和现代农业产业技术体系专项资金(CARS-48)共同资助 [This work was supported by Central Public-Interest Scientific Institution Basal Research Fund, Yellow Sea Fisheries Research Institute, CAFS (20603022017001), National Natural Science Foundation of China (41676148), the Taishan Scholar Program for Seed Industry and China Agriculture Research System (CARS-48)]. 史晓丽, E-mail: shixiaoli209@163.com

① 通讯作者: 孟宪红, 研究员, E-mail: mengxianhong@ysfri.ac.cn

收稿日期: 2017-12-19, 收修改稿日期: 2018-03-10

球的对虾养殖业造成了极大危害,严重制约了对虾养殖业的健康发展。国内外学者展开了大量研究工作,如消除传染源、切断传播途径、控制水质、提高对虾抵抗力等以降低其危害,但至今效果不显著 (Yang *et al.*, 2001; He *et al.*, 2012; Yang *et al.*, 2012; 曹家旺等, 2017)。研究对虾对病原感染的免疫应答机制,可以为有效进行对虾的病害防治提供重要的理论指导。

甘氨酸脱羧酶(Glycine Decarboxylase, GLDC)是由 P-蛋白、H-蛋白、T-蛋白和 L-蛋白 4 个亚基构成的多酶复合体系(Tada *et al.*, 1987)。它作为天冬门氨酸氨基转移酶超家族(Aspartateamino transferase superfamily, AAT_1)的成员,是一种磷酸吡哆醛依赖酶。在人类中的研究表明,甘氨酸脱羧酶的突变是非酮性高甘氨酸血症(Nonketotic Hyperglycinemia, NKH)的病因,该疾病是由先天基因缺陷导致大量甘氨酸在体液中积累而引起的严重的神经功能障碍,如脑病、肌张力减退、呼吸暂停、顽固性癫痫和可能的死亡(Kanekar *et al.*, 2013)。也有学者指出, GLDC 是非小细胞肺癌(Non-small cell lung cancer, NSCLC)发生的关键酶, NSCLC 初期肿瘤的 TICs 中检测到高表达量的 LIN28B 干细胞因子和 GLDC(Zhang *et al.*, 2012)。GLDC 在糖酵解和甘氨酸/丝氨酸代谢中引起巨大变化,导致嘧啶代谢发生变化,从而影响肿瘤细胞的增殖(信波等, 2007)。迄今为止,水产动物中尚未见该基因相关的研究报道。

在中国对虾感染 WSSV 后的表达谱芯片研究中,发现甘氨酸脱羧酶基因出现显著差异表达,此外,该基因内部出现与抗病性状呈显著关联的 SNP 位点(本实验室未发表数据)。故此,本研究采用 RACE 技术克隆中国对虾的甘氨酸脱羧酶基因(以下称 *FcGLDC* 基因),并研究其在 WSSV 侵染后的时空表达特点;利用直接测序法筛选基因内部的 SNP 位点,并利用飞行时间质谱法对 SNP 位点进行分型及其与抗 WSSV 性状的关联分析。研究结果将有助于为中国对虾抗 WSSV 育种提供理论依据和实践资料。

1 材料与amp;方法

1.1 实验材料及人工 WSSV 感染实验

中国对虾实验群体取自中国水产科学研究院黄海水产研究所遗传育种中心[体重(1.01±0.21) g],实验前暂养 7 d。实验用水为砂滤海水,养殖期间水温为 24.8~25.4℃,盐度为 28,持续充氧,每天换水,并投喂配合饲料。利用单尾、等量、口饲法对 1500 尾左右的对虾进行 WSSV 感染,感染方法参见逢锦菲

等(2013),选取 WSSV 感染实验中早期死亡的 48 尾和后期死亡的 48 尾对虾分别作为敏感群体和抗性群体,将 2 个群体的肌肉组织保存于液氮中,用于后续 DNA 的提取及关联分析。

随机挑选暂养的健康中国对虾 180 尾,平均分为 2 组(WSSV 感染组和对照组),每组 3 个平行。WSSV 感染组从第五腹节处注射病毒含量 10^3 copies/ μ l 的病毒悬液 10 μ l,对照组从第五腹节处注射 PBS 10 μ l,各组分别在实验开始后的 0、6、12、24、48 和 72 h 取鳃、肝胰腺和肌肉组织,每个时间点各取 3 尾存于液氮中,用于后续 RNA 的提取及荧光定量验证。

1.2 基因组 DNA 和总 RNA 的提取及 cDNA 合成

使用 Trizol 法提取中国对虾肝胰腺组织的总 RNA,具体操作方法参照说明书(Invitrogen, 美国)。采用紫外分光光度计与琼脂糖凝胶电泳检测所提取 RNA 的质量及完整性。利用 SMART™ RACE Amplification Kit (Clontech, 美国)反转录合成 cDNA 第一条链。按照 TIANGEN 基因组 DNA 提取试剂盒的方法进行基因组 DNA 的提取。采用紫外分光光度计与琼脂糖凝胶电泳检测所提取 DNA 的质量及完整性。

1.3 中国对虾 *FcGLDC* 基因全长 cDNA 的克隆及测序

根据本实验室的中国对虾 454 转录组测序获得的甘氨酸脱羧酶基因的部分 EST 序列,使用 Primer Premier 5.0 软件设计 5'RACE 和 3'RACE 引物(表 1),利用 Advantage 2 Polymerase (AD 酶)进行扩增,扩增产物经 1.2%琼脂糖凝胶电泳检测后,利用胶回收试剂盒回收目的片段,连接到 pMD18-T simple 载体上,转化入 Top 10 感受态细胞。使用 M13 引物对克隆进行菌落 PCR 鉴定,阳性克隆经确定后送青岛擎科测序公司进行测序。

1.4 中国对虾 *FcGLDC* 基因的克隆及测序

以对虾的基因组 DNA 为模板,根据 *FcGLDC* 基因的 cDNA 序列设计引物(表 1)。利用 TaKaRa 的 *LA Taq* DNA 聚合酶进行扩增,反应体系:DNA 模板 2 μ l,正反向引物各 5 μ l, dNTP mixture (2.5 mmol/L each) 8 μ l, Buffer 5 μ l, *LA Taq* 0.5 μ l,加水补足到 50 μ l。PCR 反应程序:94℃变性 5 min; 94℃变性 30 s, 60℃退火 30 s, 72℃延伸 3 min, 35 个循环, 72℃延伸 5 min。扩增产物测序由青岛擎科测序公司进行。

1.5 基因生物信息学分析

测序结果通过 NCBI 网站上的 BLASTX 进行序列同源性比对分析,利用 DNASTAR 软件进行全长序

表1 引物序列及信息
Tab.1 Primer sequences in RACE, qRT-PCR and DNA amplification

引物 Primer	序列 Sequence(5'~3')	用途 Purpose
5' outer primer	ACGGCGAGCGTCTGCGGGTG	RACE
5' inner primer	GCTGCGTCGGTCAAGGCGTC	RACE
3' outer primer	CGAGAGTTTAAGAAGACAGC	RACE
3' inner primer	AATCCGAAGACAAAGCAG	RACE
18S F	TATACGCTAGTGGAGCTGGAA	qRT-PCR
18S R	GGGGAGGTAGTGACGAAAAAT	qRT-PCR
GLDC 1F	TCAGGGAAATCTCGGA	DNA 扩增
GLDC 1R	TGCTGGATACTGGAAAAGC	DNA 扩增
GLDC 2F	TCAGTGCCGTGCTTTTCCAG	DNA 扩增
GLDC 2R	GCAGATGTTGCTGGTGGCTT	DNA 扩增
GLDC 3F	GAAGGACTCTTACCGCCTTG	DNA 扩增
GLDC 3R	AAGGGCTATTCCAACCGTCT	DNA 扩增
GLDC 4F	GTTAATCCGACATTGGCAACT	DNA 扩增
GLDC 4R	CTCTGTGCTGAGGCAGGATT	DNA 扩增
GLDC 5F	GACTTCCTCACCCACAGCAT	DNA 扩增
GLDC 5R	TCCCTCACCCACGCGCAT	DNA 扩增
GLDC 6F	GAGGAGCTTGAACATGACC	DNA 扩增
GLDC 6R	CTGTGCTGAGGCAGGATT	DNA 扩增
GLDC 7F	GTGTGCCTGATTCCAACCTC	DNA 扩增
GLDC 7R	CATGAGGGATGCAGAAGGTC	DNA 扩增
GLDC 8F	CGAGAGTTTAAGAAGACAGC	DNA 扩增
GLDC 8R	GGGTGAATCCAGTACATAACAA	DNA 扩增

列拼接,在线生物学软件 ExPASY 对全长序列编码的氨基酸序列进行生物信息学分析,使用 MEGA 5.0 软件的邻接法(NJ)构建系统进化树。

1.6 基因的时空表达分析

以 18S rRNA 作为 Real-time PCR 反应的内参基因,采用 TaKaRa 的荧光定量 PCR 试剂盒对基因进行荧光定量 qRT-PCR 检测。所用仪器为 ABI 7500 型荧光定量 PCR 仪,各个基因引物序列见表 1。20 μ l 反应体系包含: 2 \times SYBR Green PCR Master Mix 10 μ l, 正向和反向引物各 1 μ l (2 μ mol/L), 稀释后的 cDNA 模板 1 μ l (1 : 10 稀释), 循环参数按照说明书。实验重复 3 次,以平均值进行后续数据分析。根据 $2^{-\Delta\Delta Ct}$ 方法计算每个基因的相对表达量(Livak *et al.*, 2001), 所得数据用 SPSS 18.0 进行单因素方差分析(One-way ANOVA), 比较均值的差异显著性, $P < 0.05$ 为差异显著, $P > 0.05$ 为差异不显著。

1.7 中国对虾 *FcGLDC* 基因的 SNP 位点筛查及其与抗 WSSV 性状的关联分析

取中国对虾 5 个个体的 DNA 进行混合,进行 PCR 扩增反应,扩增产物用 1.2% 琼脂糖凝胶电泳检测后将产物送青岛擎科测序公司测序。在测序结果中寻找有明显双峰的位点,将其定为 SNP 位点。之后利用飞行时间质谱法对 SNP 位点进行分型,并在敏感群体和抗性群体中对分型结果进行关联分析。

2 结果与分析

2.1 中国对虾 *FcGLDC* 基因 cDNA 全长序列的克隆

利用 Trizol 法获得中国对虾肝胰腺组织的 RNA,经 1.2% 琼脂糖凝胶电泳检测后发现, 18S 和 28S rRNA 条带清晰,说明其完整性较好,紫外分光光度计检测其 $OD_{260\text{ nm}}/OD_{280\text{ nm}}$ 为 1.86,表明其纯度较高,可用于后续实验。

中国对虾 *FcGLDC* cDNA 序列全长 3481 bp, 其中,5'UTR 长 17 bp,3'UTR 长 86 bp,ORF 为 2829 bp, 编码 942 个氨基酸,包含 94 个碱性氨基酸(K 和 R), 108 个酸性氨基酸(D 和 E), 322 个疏水氨基酸(A、I、L、F、W 和 V), 233 个亲水氨基酸(N、C、Q、S、T 和 Y), 222 个带电荷氨基酸(K、R、D 和 E), 预测蛋白分子量为 104.66 kDa, 理论等电点为 6.51(图 1)。

2.2 中国对虾 *FcGLDC* 基因的克隆

经 PCR 扩增得到中国对虾 *GLDC* 基因 DNA 序列全长,共 4964 bp, 包含 12 个外显子和 11 个内含子。所有内含子均由 GT 开头,以 AG 结尾,符合 GT-AG 法则。所有内含子都在开放阅读框内部(图 2)。

2.3 *FcGLDC* 蛋白系统进化分析

氨基酸序列分析显示, *FcGLDC* 蛋白含有天冬氨酸氨基转移酶超家族的典型保守序列 SAAILPISW(A 也可 S), 定位于第 736~744 氨基酸(图 3)。氨基酸序列比对发现, *FcGLDC* 的蛋白序列与节肢动物的相似度最高, 与内华达白蚁(*Zootermopsis nevadensis*)、体虱(*Pediculus humanus corporis*)和白纹伊蚊(*Aedes albopictus*)的相似度分别是 71%、68%和 68%, 而与哺乳动物如非洲象(*Loxodonta africana*)和草原鹿鼠(*Peromyscus maniculatus bairie*)的相似度分别为 67%和 67%。

系统进化分析表明, *GLDC* 基因在进化过程中为了脊椎动物(哺乳动物、爬行动物等)和无脊椎动物(主要为节肢动物)两个大的分支, 在无脊椎动物分支

```

1           M L E A L G Y E S L D A L T D A A M P E N I R L G
1  aggcagaccaagaaagATGCTCGAGGCGTGGGTTATGAGTCGCTGGACGCCTTGACCGACGCAGCGATGCCGAGAACATCCGACTCG
26  R D L N V Q N P M G E H A L L S R I R E I S E K N E I W R T
91  GGAGGACCTGAACGTGCAGAAATCCGATGGCGGAGCAGCGCTACTCAGCAGGATCAGGAAATCTCGGAGAAGAACAAATCTGGAGGA
56  Y I G M G Y H N C R V P H I L R N I F E N P G W T T Q Y T P
181 CCTACATCGGCATGGGCTACCACAACCTGCCCGTCCCCACATCCTCAGGAACATCTTCGAGAATCCGGGATGGACAACACAGTACACGC
86  Y Q P E V A Q G R L E S L L N Y Q T M V T D L T G L E V A N
271 CATACCAACCGGAAGTGGCCAGGGTGCCTCGAATCCCTCTGAATTACCAGACGATGGTGACTGACCTGACTGGCCTTGAAGTTGCTA
116 A S L L D E G T A A A E A V G L A Y R Q N K R R K V Y L S D
361 ATGCTTCATTGTTGGACGAAGGCACAGCAGTGCAGAAGCCGTTGGGTTGGCGTACAGGCAGAACAAACGGCGAAAAGTGTACCTGTGAC
146 K L H P Q T L A V V R T R A L P L G L E V L I G D V F D I D
451 ACAAGCTTACC CGCAGAGCTCGCCGTTGTCGAACCCCGCCCTTCCCTGGCCTTGAAGTCTCATTGGCGATGATTTGATATCG
176 F T N R D V S A V L F Q Y P D T E G T V N D F T Q V I E N A
541 ACTTACCAATCGCATGTGAGTGCCTGCTTTTCCAGTATCCAGACACTGAAGGAACCGTCAATGACTTACACAAGTTATTGAAAATG
206 Q S N G T M V V C A T D L L A L T V L V P P G D L G V D I A
631 CGCAGAGCAATGGAACAATGTTAGTGTGTCGACGACTCTTGGCTCTCACAGTCTCGTTCGCCCGGCGACTCGGCGGTGATATCG
236 V G T S Q R L G V P L G Y G G P H A G F F A C K N N L V R L
721 CCGTCGGGACTTCGACAGACTCGGGTTCCTCTGGGCTACGGAGTCTCAGCCGGCTTCTTCGCTGCAAGAAACATCTGGTGAAGC
266 M P G R M I G V T R D A N G K D S Y R L A L Q T R E Q H I R
811 TCATGCCGGGGAATGATCGGTGTCACAAGGATGCAAAATGGGAAGGACTTACC G C C T T G C T T C A A A C C G A G A C A C A C A T T C
296 R D K A T S N I C T A Q A L L A N M S A M F A V Y H G P K G
901 GTCGTGACAAAGCCACAGCAACATCTGCACAGCTCAGGCACTGCTTGC AAACATGTG CAGCAATGTTTGCAGTTTACCATGGACAAAAG
326 L K N I A N R I H N A T V I L S R G L Q S A G H I I E N D I
991 GGTAAAAAATATTGCTAATCGCATTACAACGCAACAGTCTTCTGTCAAGGGGCTT CAGT CAGTGGCCATATAATTGAGAATGATA
356 F F D T M R V N P T L A T S E I R M R A Q Q K E I N L R Y F
1081 TCTTCTTTGACACTATGAGGGTTAATCCGACATTTGGCAACTAGTGAGATCAGAATGAGAGCTCAACAGAAAGAAATTAACCTCGTTACT
386 R D E T V G I A L D E T V R K Q D L D D L F W V K D Y L S N V
1171 TCAGAGATGAGACGGTTGGAATAGCCCTTGATGAGACAGTAAGAAAGCAGGACTTGACGATTTGTTCTGGGTGTTCAACTGCAAGAACG
416 T A D K V A S N L S G A A L P K E H I S N T K F A R K S D F
1261 TTA CTG C A G A C A A G G T C G C T A G C A A T C T C A G C G G T G C T C T G C C C A A G G A A C A C A T C T C C A A T A C A A A A T T G C T C G C A A G T C G G A C T
446 L T H S I F N T Y H S E A Q L V R M K I L E N K F N C L V
1351 TCCTCACCACAGCATATTCAACACATATCATTCTGAGGCCAGTGGTGCCTACATGAAGATCCTGGAGAATAAGGATGTTTCTCTTG
476 H S M I P L G S C T M K L N S T T E M M P C S F P H F T E I
1441 TACATCCATGATCCCCTAGGCTCTTGTACGATGAAGCTAAACAGCAACAGAGATGATGCCGTGACGCTTCCACACATTCACAGAAA
506 H P F V P P E Q A L G Y R L L F E E L E H D L C E I T G Y D
1531 TCATCCATTTGTACCACCAGAGCAAGCACTTGGTTATCGGCTTCTGTTTGGAGGCTGAACATGACCTATGTGAGATCACGGGTTATG
536 K I S F Q P N T P P Q P H Y P P L P P I I S Y L D A R G E G
1621 ACAAAATTTTCATTCCAGCCAAACACTCCACCTCAACCACTTCCACCTCTCCACCATCATCTCACTTGGATGGCGGTGGGAGG
566 H R N V C L I P T S A H G T N P A S A Q M A G M K V E A V N
1711 GACACCGCAATGTGGCTGATTTCCAACCTTCAGCCATGGTACAAATCTGCCTCAGCACAGATGGCAGGAATGAAAGTAGAAGCTGTGA
596 V D K E G S I D F R H L T A K K I E K H R K N L A C L M I T Y
1801 ATGTTGACAGAAAGGGTCCATTGATTTFCAGACATCTTACTGCCAAGATTGAGAAACATCGTAAGAACTGGCATGCTGCTGATACAT
626 P S T N G V F E D T V K D V C H M V H E A G G Q V Y L D G A
1891 ATCTCTTACCAATGGTGTATTGAAAGACTGTAAAGATGTTGCCATATGGTTTATGAAGCTGGTGGTCAAGTGTACTTGGATGGAG
656 N M N A Q V G L C R P G D I G G D V S H L N L H K T F C I P
1981 CAAACATGAATCCAGGTGGGTCTGTGTGCGCCAGGAGATATTGGAGGAGATGTATCCCATCTTAATTTGCACAAGACCTTCTGCATCC
686 H G G G G P G M G P I G V K A H L I P Y L P S H P I V D P M
2071 CTCATGGTGGAGGAGCTGGCATGGACCCATTTGGAGTGAAGGCCACCTGATTCCATACTTGC AAGTCAACCCCATTTGTGACCCAA
716 A L M K K K A R S F G L V S A A P Y G S A A I L P I S W A Y
2161 TGGCCTTAATGAAGAAGAAAGCTCCGAGTTTGGCCTTGTGTGTCAGCTCCCTATGGTTCAGTGTATCCTTCCAATCTCTTGGGCT
746 V K M M G P K G L R E A T Q V A I L N A N Y M A N R L R K Y
2251 ATGTCAAGATGATGGTCCAAAGGTTTGGGAAAGTACTCAGTTGCTATCTTGAATGCTAATCACTACATGGCAATCCGCTCAGAAAAT
776 Y K V L Y T G N C G L V A H E F I L D M R E F K K T A G I E
2341 ACTACAAGTGTGTACACTGGAATTTGGATTAGTTGCTCATGAATTCATCCTTGAATGGCAGAGTTTAAAGAAGACAGCTGGAATG
806 A T D I A K R L Q D Y G F H A P T M S W P V A G S L M V E P
2431 AAGTACAGACATTGCAAGCGTCTTCAGACTATGGTTCCATGCACCACAATGTCTGGCCTGTTGCTGGTAGTTGATGGTAGAGC
836 T E S E D K A E L D R F C D M E I L I R E E I R D E Y G R
2521 CAACGGAATCCGAAGACAAGCAGAGTTGGATCGTTTCTGTGATGCAATGATCTTAATTCGAGAAGAGATCAGAGACCTGGAGGAGGCCA
866 S D P K C N P L N M S P H P L V T V T S Y N W D R P Y S R E
2611 GGAGTGATCCCAAGTGAACCCCTCAACATGTCACCCCTAGTAACTGTCACCTTCTTACAACCTGGGACCGCTGATATCCCGAG
896 Q A A F P A P F V R P E S K W W P T V G R I D D I Y G D R N
2701 AGCAAGCTGCTTCCCTGCTCATTGTTGCGGCCAGAGACAAATGGTGGCCAACTGTGGGACGATTGATGACATCTATGGCACCCTGA
926 L V C T C P P M E T S C S P F D D *
2791 ACCTGGTTTGTACCTGTCCGCCATGGAGACTTCAATGCTACCCCTTGGACGATTAACcttgcttgaaccttggttttgtaaacacctaa
2881 aattcatcatatttgcaggagaaaaagaaaaaaagaggaggattttatattttctttttctatgtataacttgatagatttggct
2971 gcaagtttatgaaccctaaatattaaaactcaaaaaattcatcccacttatgtataactactctgaaatcatcagttataaagtcag
3061 aaagttatattatagcaccagaacataaaataggcaaatgtcatcatcactttcatctatcttcatagaaagctcttattaca
3151 ggtagatttttaatacactgttaggctacactaggacaactactttgtataaactggacttctttatctctgataaagaaggctat
3241 tattaagattctttataccatgtctaaatgtagtaataaaatgaacctctttgataaataagattgacataataatttcc
3331 tgttagtaatacattgttaatactgtgttcttaacttcaaaaaaggtgtatagactattacattcattggtagtcatggattca
3421 ctacagtggtgatattttctatttttaaaaaagtagaattgaataaaaaa

```

图 1 中国对虾 *GLDC* 基因 cDNA 全长序列及其编码的氨基酸序列

Fig.1 The full-length cDNA sequence and deduced amino acid sequence of the *FcGLDC* gene
全长起始密码子 ATG 由下划线标出; 终止密码子 TAA 由*标出; 黑色方框内为保守序列
Start codon (ATG) was underlined; Stop codon (TAA) marked with asterisk; Conserved sequence marked with black box



图 2 *FcGLDC* 基因的 DNA 结构

Fig.2 The DNA structure of *FcGLDC*

红色框表示外显子, 黑色线表示内含子 Red box represents exon, black line represents intron

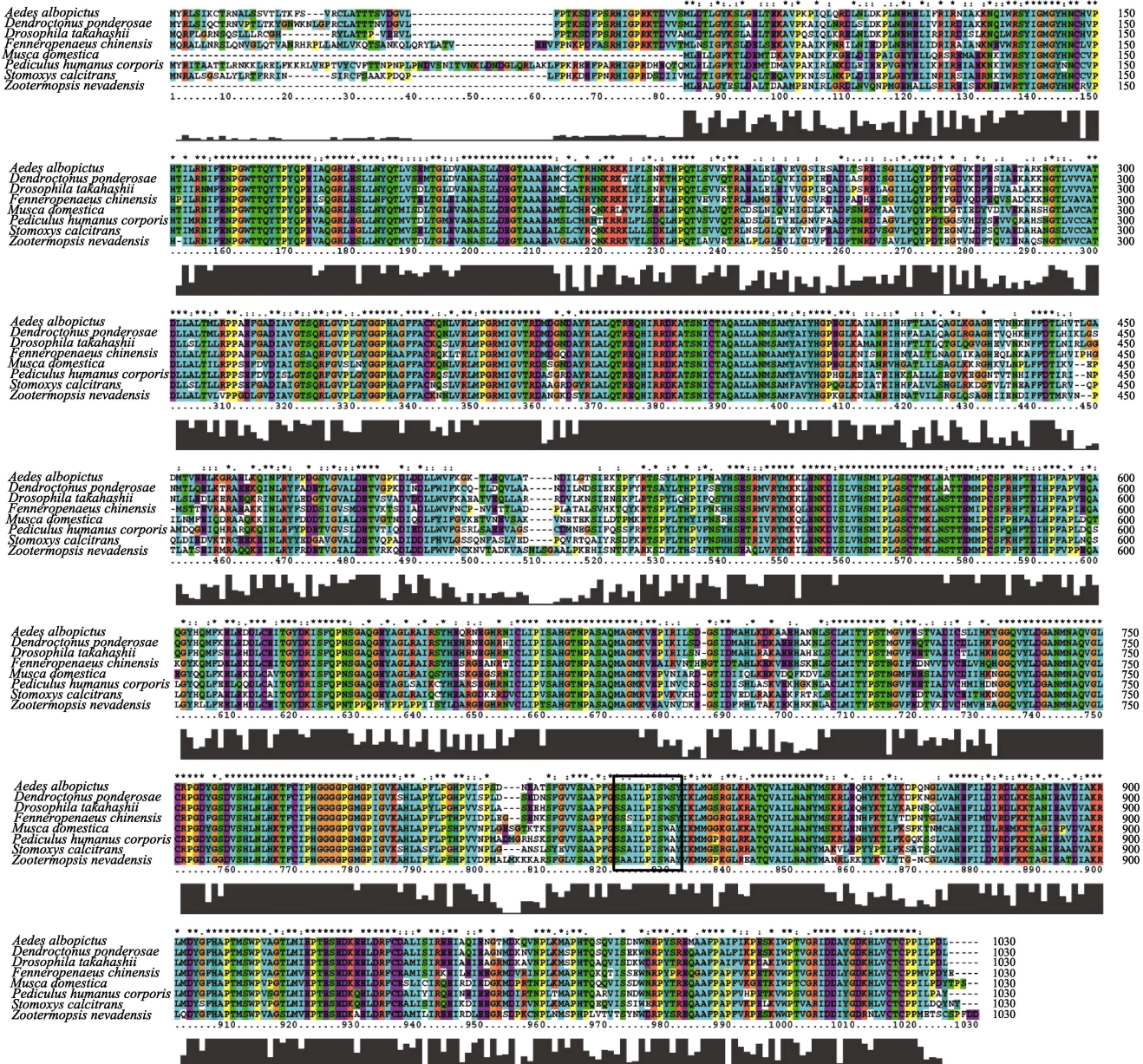


图 3 中国对虾与其他物种 GLDC 氨基酸的序列比对

Fig.3 Alignment analysis of amino acid sequence of FcGLDC from various species
黑色方框内为保守序列。Conserved sequence marked with black box

中, 中国对虾所属的甲壳亚门和家蝇 (*Musca domestica*)、埃及伊蚊 (*Aedes aegypti*) 等六足亚门的物种之间也体现了差异性 (图 4)。

2.4 中国对虾 *FcGLDC* 基因的组织表达分析

FcGLDC 基因在中国对虾各组织中的相对表达

量分析结果显示, 该基因在鳃、肝胰腺和肌肉中均有表达, 其中, 肌肉中的表达量最高, 肝胰腺中次之, 鳃中的表达量最少, 但三者间无显著差异 ($P>0.05$) (图 5A)。

FcGLDC 基因在感染 WSSV 的对虾不同组织中的相对表达量变化分析结果显示, 在鳃中, 该基因的

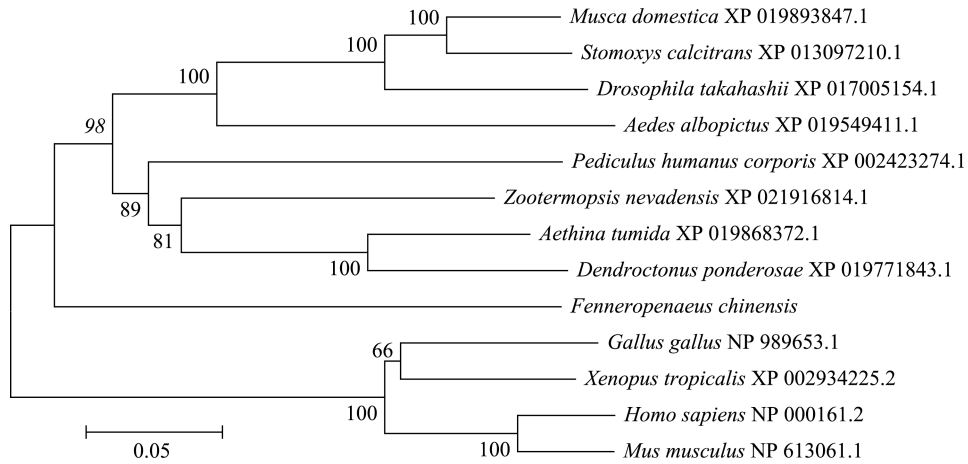


图 4 中国对虾与其他物种 GLDC 的系统进化分析

Fig.4 The phylogenetic tree analysis of FcGLDC from various species

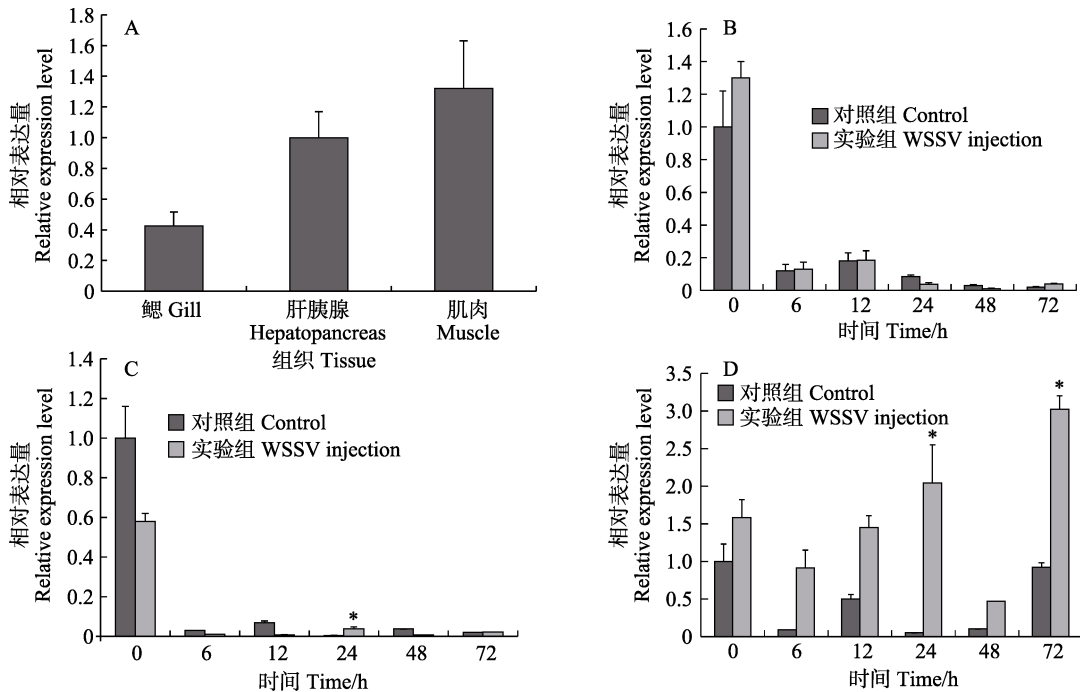


图 5 *FcGLDC* mRNA 的组织表达及其在感染后的组织表达(n=3)

Fig.5 The expression profiles of *FcGLDC* in hepatopancreas, gill, and muscle (n=3)

A: 正常组织; B: 感染后的肝胰腺; C: 感染后的鳃; D: 感染后的肌肉。*表示显著性差异
 A: Control; B: Hepatopancreas after WSSV challenge; C: Gill after WSSV challenge; D: Muscle after WSSV challenge.
 * represents significant difference

表达量在感染后的变化与对照组基本一致, 均为急剧下降, 在 24 h 时, 实验组表达量有小幅上升, 是对照组的 4.47 倍($P<0.05$)(图 5C)。在肝胰腺中, *GLDC* 基因的表达量同样在注射后明显降低, 在 12 h 有上升趋势, 而后在 48 h 时表达量达到最低。肝胰腺中, 各个时间点对照组与实验组的表达量差异均不显著 ($P>0.05$)(图 5B)。在肌肉组织中, 实验组 *GLDC* 基因的表达量在 24 h 和 72 h 时表现为 2 次峰值, 分别为对照组的 19.98 倍($P<0.05$)和 4.21 倍($P<0.05$)(图 5D)。

在 6 个时间点中, 实验组的表达量均高于对照组。

2.5 中国对虾 *FcGLDC* 基因 SNP 位点多态性分析及其关联分析

利用直接测序法, 在 *FcGLDC* 内部共发现 4 个可能的 SNP 位点, 并利用质谱法对这 4 个位点进行了成功分型。分型结果采用 PopGene 软件处理, 统计各位点的期望杂合度 H_e , 观测杂合度 H_o , 计算各位点多态信息含量 PIC, 并进行哈迪-温伯格平衡检验(表 2)。

由表 2 可知, *FcGLDC* 基因内的这 4 个 SNP 标记的 H_o 范围为 0.045~0.462, H_e 范围为 0.108~0.465。C1588-TG 和 C4406-CT 的 PIC 分别为 0.355 和 0.285, 为中度多态($0.25 < PIC < 0.5$), C4854-GA 和 C2967-AG 的 PIC 分别为 0.142 和 0.101, 属于低度多态($PIC < 0.25$)。哈温平衡分析结果显示, 只有 C1588-TG 符合哈迪-温伯格平衡检验, 且差异不显著。

表 2 SNP 位点多态性分析

Tab.2 The analysis of genetic polymorphism of SNPs

SNP 位点 SNP loci	观测杂合度 H_o	期望杂合度 H_e	PIC	哈温平衡 HW
C1588-TG	0.462	0.465	0.355	NS
C4406-CT	0.295	0.346	0.285	ND
C4854-GA	0.168	0.155	0.142	ND
C2967-AG	0.045	0.108	0.101	ND

NS: 差异不显著; ND: 差异显著

NS: No significant difference; ND: Significant difference

使用 SPSS 18.0 对基因型和抗 WSSV 性状之间的相关性进行卡方检验, 发现各位点与抗 WSSV 性状不相关($P > 0.05$)(表 3)。

表 3 SNP 位点与抗 WSSV 性状的关联分析

Tab.3 The correlation analysis of SNPs between WSSV-sensitive group and WSSV-resistant group

SNP 位点 SNP loci	位置 Location	突变类型 Mutation type	卡方值 χ^2	显著性 P
C1588-TG	内含子 Intron	无义突变 Nonsense mutation	0.513	0.52
C4406-CT	ORF	同义突变 Synonymous mutation	1.538	0.42
C4854-GA	ORF	同义突变 Synonymous mutation	0.110	0.74
C2967-AG	内含子 Intron	无义突变 Nonsense mutation	3.875	0.14

3 讨论

本研究首次报道了对虾中的甘氨酸脱羧酶, 克隆的 *FcGLDC* 基因 ORF 全长为 2829 bp, 编码 942 个氨基酸。氨基酸序列分析表明, 其含有天冬氨酸氨基转移酶超家族的典型保守序列 SAAILPISW(A 也可为 S)。蛋白的进化分析显示, 该甘氨酸脱羧酶基因与其他节肢动物的甘氨酸脱羧酶基因明显聚为一类。ORF 序列分析、蛋白保守位点分析及蛋白的系统进化分析均表明 *FcGLDC* 基因与其他物种的甘氨酸脱羧酶具有较高的相似性, 因此, 确定本研究克隆的 *FcGLDC* 基因为中国对虾的甘氨酸脱羧酶基因。

本研究分析了中国对虾感染 WSSV 后 *GLDC* 基因在鳃、肝胰腺及肌肉组织中的表达特征, 结果显示, 该基因在 3 种组织中均有表达, 且在肌肉中的表达最高。感染后, 通过分析 *GLDC* 在 3 种组织中的表达规律发现, 肌肉对该基因最敏感, 说明 WSSV 对该基因有相对比较强的刺激。考虑到对虾感染 WSSV 后出现了运动机能减退的情况(Lightner *et al*, 1998), 推测肌肉中该基因的表达与对虾的运动不协调有关。

本研究通过直接测序法得到了中国对虾 *GLDC* 基因全长 DNA 序列, 该 DNA 序列全长为 5608 bp, 包含 11 个内含子, 鉴定的外显子-内含子边界都与经典的剪接供体(GT-)和受体序列(-AG)一致。开放阅读框起于第 1 个外显子, 止于第 12 个外显子。众所周知, 真核生物的基因在每个外显子和内含子的接头区, 都有一段高度保守的共有序列, 即每个内含子的 5'端起始的 2 个核苷酸均为 GT, 3'端的 2 个核苷酸均为 AG。但也有少数基因不符合 GT-AG 法则, 如在人类 135 kb 的 *GLDC* 基因含有的 24 个内含子中, 只发现有 1 个内含子不符合 GT-AG 法则(Takayanagi *et al*, 2000)。

本研究通过直接测序法定位了中国对虾 *GLDC* 基因 DNA 全长序列中 4 个可能的 SNP 位点, 并进一步通过飞行时间质谱法对这 4 个位点进行分型检测。4 个 SNP 位点有 2 个位于内含子上, 2 个位于外显子上。位于外显子上的突变位点均在密码子的第 3 位, 属于同义突变, 未对所编码氨基酸序列造成影响。位于内含子区域的突变在转录翻译过程中被剪切, 并不参与调控过程。位点多态性分析显示, 有 3 个位点不符合哈迪-温伯格平衡, 推测这可能是杂合子数量过多或不足造成的。位点与性状的关联分析表明, 获得的 SNP 位点均与性状不相关。关联分析研究时需要考虑群体分层的问题。由于群体存在遗传背景的差异, 使得来自不同亚群的个体中等位基因频率存在差异, 即存在群体分层, 关联分析时这些位点会造成假阳性(Emahazion *et al*, 2001)。因此, 加大样本量, 并尽量使用遗传背景一致或相似的群体如采用基于家系的群体进行关联研究就显得尤为重要。本研究选用家系材料作为验证群体, 遗传背景相近, 符合关联研究所用群体的要求。水产动物中, 罗非鱼(*Oreochromis niloticus*)、大口黑鲈(*Micropterus salmoides*)、栉孔扇贝(*Chlamys farreri*)等物种的经济性状与 SNP 位点的关联分析时采用的群体多小于 200 个个体(刘福平等, 2009; 杜芳芳等, 2011; 李纪勤, 2012), 本研究采用的是感染早期死亡和后期死亡各 48 尾个体进行关联分析, 数量上基本达到了关联分析群体数量的要求。

本研究首次在中国对虾中获得了甘氨酸脱羧酶基因的全长 cDNA 序列及其 DNA 序列, 分析了 WSSV 感染后该基因在不同组织中的表达特征, 并进行了基因内部 SNP 位点多态性分析及其关联分析, 研究结果将有助于为中国对虾抗 WSSV 育种提供理论依据和实践资料。

参 考 文 献

- Cao JW, Kong J, Luo K, *et al.* Immune priming response induced by heat-inactivated WSSV on *Fenneropenaeus chinensis* at different temperature. *Progress in Fishery Sciences*, 2017, 38(2): 128–136 [曹家旺, 孔杰, 罗坤, 等. 热灭活 WSSV 诱导的中国明对虾 (*Fenneropenaeus chinensis*) 在不同温度下的免疫致敏反应. *渔业科学进展*, 2017, 38(2): 128–136]
- Du FF, Bai JJ, Li SJ, *et al.* Effects of SNPs in *POU1F1* promoter on growth traits in largemouth bass (*Micropterus salmoides*). *Journal of Fisheries of China*, 2011, 35(6): 793–800 [杜芳芳, 白俊杰, 李胜杰, 等. 大口黑鲈 *POU1F1* 启动子区域 SNPs 对生长的影响. *水产学报*, 2011, 35(6): 793–800]
- Emahazion T, Feuk L, Jobs M, *et al.* SNP association studies in Alzheimer's disease highlight problems for complex disease analysis. *Trends in Genetics*, 2001, 17(7): 407–413
- He YD, Zhang X. Comprehensive characterization of viral miRNAs involved in white spot syndrome virus (WSSV) infection. *RNA Biology*, 2012, 9(7): 1019–1029
- Kanekar S, Byler D. Characteristic MRI findings in neonatal nonketotic hyperglycinemia due to sequence changes in GLDC gene encoding the enzyme glycine decarboxylase. *Metabolic Brain Disease*, 2013, 28(4): 717–720
- Lightner DV. Epizootiology, distribution and the impact on international trade of two penaeid shrimp viruses in the Americas. *Revue Scientifique et Technique*, 1996, 15(2): 579–601
- Lightner DV, Redman RM. Shrimp diseases and current diagnostic methods. *Aquaculture*, 1998, 164(1–4): 201–220
- Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods*, 2001, 25: 402–408
- Liu FP, Bai JJ, Ye X, *et al.* Cloning of *MC4R* gene and study on the association between SNPs of *MC4R* and growth trait in tilapia (*Oreochromis niloticus*). *Journal of Fishery Science of China*, 2009, 16(6): 816–823 [刘福平, 白俊杰, 叶星, 等. 罗非鱼 *MC4R* 基因克隆及其与生长相关的 SNPs 位点. *中国水产科学*, 2009, 16(6): 816–823]
- Li JQ. Development, characterization and application of EST-SNP markers in Zhikong scallop (*Chlamys farreri*). Doctoral Dissertation of Ocean University of China, 2012 [李纪勤. 栉孔扇贝 (*Chlamys farreri*) EST-SNP 的开发及其应用. 中国海洋大学博士研究生学位论文, 2012]
- Tada K, Hayasaka K. Non-ketotic hyperglycinaemia: Clinical and biochemical aspects. *European Journal of Pediatrics*, 1987, 146(3): 221–227
- Takayanagi M, Kure S, Sakata Y, *et al.* Human glycine decarboxylase gene (GLDC) and its highly conserved processed pseudogene (psiGLDC): Their structure and expression, and the identification of a large deletion in a family with nonketotic hyperglycinemia. *Human Genetics*, 2000, 106(3): 298
- Yang F, He J, Lin X, *et al.* Complete genome sequence of the shrimp white spot bacilliform virus. *Journal of Virology*, 2001, 75(23): 11811–11820
- Yang JY, Chang CI, Liu KF, *et al.* Viral resistance and immune responses of the shrimp *Litopenaeus vannamei* vaccinated by two WSSV structural proteins. *Immunology Letters*, 2012, 148(1): 41–48
- Zhang WC, Shyhchang N, Yang H, *et al.* Glycine decarboxylase activity drives non-small cell lung cancer tumor-initiating cells and tumorigenesis. *Cell*, 2012, 148(1–2): 259
- Pang JF, Kong J, Meng XH, *et al.* Real-time PCR assay for quantifying WSSV load in *Fenneropenaeus chinensis* “Huanghai No.2”. *Oceanologia et Limnologia Sinica*, 2013, 44(3): 782–787 [逢锦菲, 孔杰, 孟宪红, 等. 中国对虾 (*Fenneropenaeus chinensis*) “黄海 2 号”人工感染 WSSV 的荧光定量分析. *海洋与湖沼*, 2013, 44(3): 782–787]
- Xin B, Liu XY, Wang W, *et al.* Coexistence of GABA and glycine in the neurons of the oral and interpolar subnuclei of the spinal trigeminal nucleus in the GAID67-GFP transgenic mouse. *Chinese Journal of Neuroanatomy*, 2007, 23(1): 45–48 [信波, 刘翔宇, 汪伟, 等. GAD67-GFP 转基因小鼠三叉神经脊束核吻侧亚核和极间亚核内 GABA 和甘氨酸共存神经元的观察. *神经解剖学杂志*, 2007, 23(1): 45–48]

(编辑 冯小花)

cDNA Cloning of the GLDC Gene in *Fenneropenaeus chinensis* and Its Expression and Functional Analyses after WSSV Infection

SHI Xiaoli^{1,2}, ZHANG Yingxue^{1,2}, MENG Xianhong^{1,2①}, KONG Jie^{1,2}, LUAN Sheng^{1,2},
LUO Kun^{1,2}, CAO Baoxiang^{1,2}, CAO Jiawang^{1,2}, CHEN Baolong^{1,2}

(1. Key Laboratory of Sustainable Development of Marine Fisheries, Ministry of Agriculture and Rural Affairs, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao 266071;

2. Laboratory for Marine Fisheries Science and Food Production Processes, Pilot National Laboratory for Marine Science and Technology (Qingdao), Qingdao 266071)

Abstract Glycine decarboxylase is a key glycolytic enzyme that is involved in both the glycolytic and gluconeogenic pathways. In this study, the glycine decarboxylase gene from the Chinese shrimp *Fenneropenaeus chinensis* (*FcGLDC*) was cloned and sequenced for the first time. The full-length cDNA sequence of *FcGLDC* was 3481 bp long and contained a 17 bp 5'-UTR, an 86 bp 3'-UTR, and a 2829-bp open reading frame that encoded a 942 amino acid peptide with a calculated molecular mass of 104.66 kDa and theoretical isoelectric point of 6.51. The full-length *FcGLDC* DNA sequence was 4964 bp long and contained 12 exons and 11 introns. Multiple sequence alignment showed its high similarity with the glycine decarboxylase genes from other arthropod species: *Zootermopsis nevadensis* (71%), *Pediculus humanus corporis* (68%), and *Aedes albopictus* (68%). Phylogenetic analysis revealed that *FcGLDC* was in the same class with those from the other arthropods. The *FcGLDC* transcript showed the highest expression in muscles and the lowest expression in the gill. Stimulation of the shrimp with white spot syndrome virus (WSSV) resulted in changes in *FcGLDC* expression profiles in the hepatopancreas, gill, and muscles. In the gill, the transcript level was approximately 4.47-fold higher than that of the control group ($P<0.05$) at 24 h post injection (hpi). In the hepatopancreas, no significant difference was observed between the WSSV-challenged group and the control group ($P>0.05$). In the muscles, the transcript levels increased significantly, by approximately 19.98-fold at 24 hpi and 4.21-fold at 72 hpi, relative to the control group levels ($P<0.05$). Four single nucleotide polymorphisms were identified, using direct sequencing and time-of-flight mass spectrometry. Association analysis indicated that there was no significant association between the genotype and WSSV resistance ($P>0.05$). These results show that *FcGLDC* is inducible and may be involved in the shrimp's immune response to pathogens.

Key words *Fenneropenaeus chinensis*; Glycine decarboxylase; Gene cloning; Expression; Association analysis

① Corresponding author: MENG Xianhong, E-mail: mengxianhong@ysfri.ac.cn