

# 中华绒螯蟹翻译起始因子 4G 基因在幼体发育阶段的 mRNA 表达模式与适应性进化分析

吴海霞<sup>1</sup>, 时昕晔<sup>1,2</sup>, 周开亚<sup>1</sup>, 严洁<sup>1</sup>, 李鹏<sup>1</sup>

(1. 南京师范大学生命科学学院, 江苏 南京 210023)

(2. 南京师范大学海洋科学与工程学院, 江苏 南京 210023)

**[摘要]** 真核生物翻译起始因子 4G(eIF4G) 充当衔接 eIF4F 复合体的骨架蛋白, 在帽依赖性翻译起始中起关键作用。为探究中华绒螯蟹(*Eriocheir sinensis*) *eIF4G* 基因在幼体发育阶段的作用, 本研究利用 RACE 技术克隆了中华绒螯蟹 *eIF4G* 基因(命名为 *EseIF4G*) 的全长 cDNA 序列, 利用实时荧光定量 PCR 技术检测该基因在幼蟹不同发育阶段的 mRNA 表达模式, 并探讨其在节肢动物进化过程中是否发生了适应性进化。生物信息学分析显示, *EseIF4G* 基因的 cDNA 全长为 3 769 bp, 包含一个 2 379 bp 的开放阅读框, 编码 792 个氨基酸; 该编码蛋白具有保守的 MIF4G、MA3 和 W2 结构域, 无跨膜区、疏水性区域和信号肽, 是不稳定且非分泌型蛋白。荧光定量 PCR 检测显示, *EseIF4G* 基因在中华绒螯蟹受精卵、溞状幼体 I ~ V 期、大眼幼体期和仔蟹 I ~ III 期共 10 个幼体发育阶段广泛表达, *EseIF4G* 基因 mRNA 表达量在卵至溞状幼体 IV 期相对较高, 而在溞状幼体 V 期至仔蟹 II 期相对较低, 在溞状幼体 V 期降至最低, 与在仔蟹 II 期的表达量没有显著差异 ( $P > 0.05$ )。选择压力分析显示, PAML 的位点模型和分支位点模型没有鉴定出正选择位点, 分支模型在中华绒螯蟹末端分支没有检测到正选择信号; 利用 Datamonkey 在 *EseIF4G* 的氨基酸水平鉴定出两个正选择位点(448G 和 655N)。现有数据表明, *EseIF4G* 基因可能参与中华绒螯蟹幼蟹发育过程, 特别是短尾化过程, 调控 mRNA 翻译和蛋白质相互作用; 同时, 其在节肢动物进化过程中进化相对保守。

**[关键词]** 中华绒螯蟹, 翻译起始因子基因, cDNA 克隆, mRNA 表达, 适应性进化

**[中图分类号]** Q951+3 **[文献标志码]** A **[文章编号]** 1001-4616(2022)03-0096-12

## The Identification, Adaptive Evolutionary Analyses and mRNA Expression Pattern of *eIF4G-Like* Gene in Larval Development Stages of *Eriocheir sinensis*

Wu Haixia<sup>1</sup>, Shi Xinye<sup>1,2</sup>, Zhou Kaiya<sup>1</sup>, Yan Jie<sup>1</sup>, Li Peng<sup>1</sup>

(1. School of Life Sciences, Nanjing Normal University, Nanjing 210023, China)

(2. School of Ocean Science and Engineering, Nanjing Normal University, Nanjing 210023, China)

**Abstract:** Eukaryotic translation initiation factor 4 gamma(eIF4G) plays a critical role in the initiation of cap-dependent translation by acting as a scaffolding protein for the assembly of eIF4F complex. To investigate the role of *eIF4G* gene from the Chinese mitten crab during larval development stages, the full-length cDNA of *eIF4G-like* gene(denoted as *EseIF4G*) was cloned and identified by RACE(rapid amplification of cDNA ends) PCR approach, and the expression patterns were detected in larval developmental stages. Furthermore, the present study explored whether adaptive evolution of *eIF4G* gene occurred during arthropods evolution by site model, branch model and branch-site model implemented in CODEML program of the PAML package, and SLAC(fixes effects likelihood), FEL(fixes effects likelihood) and FUBAR(fast unconstrained bayesian approximation) methods implemented in Datamonkey website. Bioinformatics analyses revealed that the full-length cDNA of *EseIF4G* is composed of 3 769 nucleotides with an open-reading frame of 2 379 base pairs, which encoding 792 amino acids and containing MIF4G, MA3 and W2/eIF5C domain. *EseIF4G* is an unstable, no-secretory and omini- $\alpha$

收稿日期: 2022-01-06.

基金项目: 国家自然科学基金项目(31000954)、江苏省高等学校自然科学研究重大项目(19KJA330001)、江苏省研究生科研与实践创新计划项目(KYCX20\_1241).

通讯作者: 李鹏, 博士, 副教授, 研究方向: 动物生态与进化生物学. E-mail: lipeng@njnu.edu.cn

protein, and has no transmembrane, hydrophobic and signal region. Real-time PCR analysis indicated that the *EseIF4G* gene was widely expressed in the embryo (O) and nine different developmental stages of larvae (Z1: the first zoeal stage, Z2: the second zoeal stage, Z3: the third zoeal stage, Z4: the fourth zoeal stage, Z5: the fifth zoeal stage, M: megalopa stage, J1: the first juvenile crab stage, J2: the second juvenile crab stage, J3: the third juvenile crab stage). The expression levels of *EseIF4G* mRNA were relatively higher from the embryo stage to stage Z4, and were relatively lower at stage Z5, M, J1 and J2. The expression levels of *EseIF4G* mRNA sharply declined to the lowest at stage Z5, which had no significant difference with stage J2 ( $P > 0.05$ ), and then increased significantly at stage M and J1 ( $P < 0.05$ ). Segregating sites analysis indicated that site 97 and site 457 were the segregating sites of EseIF4G protein, which were located in MIF4G domain and MA3 domain, respectively. Selective pressure analysis showed that overall, there was evidence for strong purifying selection under one-ratio model, with a  $\omega$  value of 0.0355 (significantly less than 1). No positively selected sites were found by site model in PAML, but two shared amino acid sites (448G and 655N) were found by FEL and FUBAR method in Datamonkey. When the terminal branch of *E. sinensis* was set as foreground, we did not detected any positively selected sites in PAML. Present data suggested that *EseIF4G* gene involving in the early developmental process of *E. sinensis*, was possible to serve as a regulator of mRNA translation process and proteins interaction, and was relatively conserved in the evolution of arthropod. These results provide new insights into the mechanisms of brachyurization metamorphosis in decapod crustaceans.

**Key words:** *Eriocheir sinensis*, eukaryotic translation initiation factors, cDNA cloning, mRNA expression, adaptive evolution

真核翻译起始因子 (eukaryotic translation initiation factors, eIFs) 是一类调控起始密码子周围功能核糖体复合物组装的蛋白质, 在真核细胞蛋白质翻译起始过程中发挥重要作用. 真核生物的基因表达受基因转录和翻译水平的调节; 在翻译级联的四个阶段中, 起始是关键的速率限制步骤, 因此 eIFs 是蛋白质生物合成的关键调节剂<sup>[1]</sup>. 但随着研究的深入, 发现 *eIFs* 的表达异常会导致一些基因无法正常翻译, 致使细胞恶变、肿瘤、神经退行性疾病的发生<sup>[2]</sup>. 例如, eIFs 家族与恶性肿瘤的发生、发展和预后关系密切, 主要体现在基因表达水平的变化上, 已有研究在子宫内膜癌中观察到 *eIF4G*、*peIF2 $\alpha$* 、*eIF3H* 的显著过表达, *eIF3C* 表达显著降低<sup>[3]</sup>.

目前 eIFs 家族成员共有 20 余种, 通常以复合物的形式实现其生物学活性<sup>[4]</sup>; 其中对 *eIF2*、*eIF3*、*eIF4*、*eIF5* 和 *eIF6* 的研究较多. eIF4 家族由 eIF4A、eIF4B、eIF4E、eIF4G 等亚基组成 (称作 eIF4F), 真核细胞蛋白质合成通常从该复合物与 mRNA 5' 端的 7-甲基化鸟嘌呤核苷帽状结构 ( $m^7GpppN$ ) 特异性结合开始<sup>[4]</sup>. eIF4G 作为组装 eIF4F 复合物的骨架蛋白, 与 RNA 解旋酶 eIF4A 有两个结合位点, 与 mRNA 帽结合蛋白 eIF4E、核糖体结合复合物 eIF3、促分裂原激活的蛋白激酶 1 (mitogen-activated protein kinase interacting kinase, MNK1) 和聚腺苷酸结合蛋白 (poly-A-binding protein, PABP) 各有一个结合位点<sup>[5-6]</sup>. eIF4G 与 PABP 相互作用以使 mRNA 有效环化, 从而增强 eIF4F 复合物的帽结合活性以及 60S 与 40S 核糖体亚基的连接<sup>[7]</sup>. eIF4G 与 eIF4E 结合, 增加 eIF4E 介导的帽识别和 eIF4F 复合物的帽结合活性; eIF4G 与 eIF3 的结合有助于将 40S 核糖体亚基带入 eIF4F 复合体<sup>[8]</sup>. 此外, eIF4G 将 43S 预启动复合物 (PIC, 包括 40S 核糖体亚基、eIF3 复合物和相关因子) 募集到 mRNA 5' 非翻译区 (terminal untranslated region, UTR), 随后启动扫描寻找起始密码子<sup>[9]</sup>.

从酵母、植物到哺乳动物, 迄今已在许多物种中鉴定出 eIF4G, 其一般具有 3 个典型结构域: MIF4G、MA3 和 W2/eIF5C 结构域, 均为由  $\alpha$  螺旋组成的 HEAT 重复片段结构. 其中哺乳动物细胞具有 3 种同工型, 即 eIF4GI、eIF4GII 和 p97/NAT1/DAP-5<sup>[10]</sup>. eIF4GI 和 eIF4GII 都具有 eIF4A、eIF4E、eIF3 和 PABP 的结合位点, 但 eIF4GI 与 eIF4GII 的浓度在不同细胞中有所不同, 通常 *eIF4GI* 的表达浓度高于 *eIF4GII*<sup>[11]</sup>. 有研究发现 *eIF4G* 基因表达水平的上升与细胞恶性转化、抑制凋亡有关, 而表达水平下降与营养缺乏、热休克、缺氧等因素有关<sup>[12]</sup>. Das 等报道 eIF4G 可以通过充当细胞 mRNA 代谢的衔接分子, 总体上调整 mRNA 的功能状态<sup>[13]</sup>. Orellana 等研究表明, 当缺乏合成代谢刺激时, 脂多糖 LPS 诱导的蛋白质合成减少与核糖体效率和 eIF4E·eIF4G 组装体降低有关, 因此发育过程涉及核糖体数目和翻译因子丰度降低, 以及 eIF2 $\alpha$  磷酸化增加<sup>[14]</sup>. 红细胞特异性 RNA 结合蛋白 RBM38 与 eIF4G 相互作用, 促进 mRNA 的翻译, 从而动态控制哺乳动物红细胞的生成过程<sup>[15]</sup>. Ghosh 和 Lasko 研究发现, 果蝇 *eIF4GI* 基因敲低并不影响雄性生育能力, 而 *eIF4GII* 在精子发生过程中的作用却很明显; 这些蛋白质在双重敲除后, 发现其在精子发生早期功

能冗余<sup>[16]</sup>. 以上所有发现均证明 eIF4G 家族在蛋白质合成中起关键调节作用,从而调控机体的代谢、生长、发育和繁殖过程.

中华绒螯蟹(*Eriocheir sinensis*)属于真蟹类方蟹科,其形态在发育早期经历从似虾型向蟹型进化的“短尾化(Brachyurization)”过程<sup>[17]</sup>. 短尾化期间,蟹的外部形态发生显著变化,包括腹部退化、折叠在头胸甲下方并插入到步足之间或特殊腔内<sup>[18]</sup>. 目前,对短尾化变态的分子机制研究甚少,主要是对相关基因表达量变化的探究<sup>[19-22]</sup>,比较功能基因在幼体不同发育阶段或不同部位的表达差异;以及中华绒螯蟹幼体发育阶段的比较转录组学分析<sup>[23]</sup>,筛选了 1016 个差异表达基因参与变态、感官知觉和免疫过程;但尚未研究 *eIF4G* 基因在甲壳动物变态发育过程的作用,且未从进化生物学方面来探讨相关基因是否在甲壳动物进化过程中发生适应性进化. 本研究利用 RACE(Rapid amplification of cDNA ends)技术克隆中华绒螯蟹 *eIF4G* 基因全长 cDNA 序列(命名为 *EseIF4G*),运用生物信息学方法分析 *eIF4G* 基因的理化性质、功能结构,并通过 qRT-PCR(Quantitative real-time PCR)方法探究该基因在中华绒螯蟹幼蟹 10 个发育时期的 mRNA 表达差异情况,以及从进化生物学角度对中华绒螯蟹 *EseIF4G* 进行选择压力分析,为进一步探究该基因在中华绒螯蟹早期发育过程的分子进化机制以及十足目甲壳动物的短尾化变态发育研究提供参考.

## 1 材料与方法

### 1.1 实验样品采集

实验材料取自江苏泰州兴化中堡镇中华绒螯蟹养殖场,取仔蟹 I 期腹部组织约 30 mg,加入 RNA 保护试剂,使用无 RNase 的手术剪剪碎混匀后保存于 -80 °C 备用. 另取中华绒螯蟹受精卵(oosperm, O)、蚤状幼体 I 期(the first zoea stage, Z1)、蚤状幼体 II 期(the second zoea stage, Z2)、蚤状幼体 III 期(the third zoea stage, Z3)、蚤状幼体 IV 期(the fourth zoea stage, Z4)、蚤状幼体 V 期(the fifth zoea stage, Z5)、大眼幼体期(megalopa stage, M)、仔蟹 I 期(the first crab stage, J1)、仔蟹 II 期(the second crab stage, J2)、仔蟹 III 期(the third crab stage, J3)整体组织样品各约 30 mg,加入 RNA 保护试剂,剪碎匀浆后于 -80 °C 保存备用.

### 1.2 *EseIF4G* 基因 cDNA 全长扩增

按 RNeasy Mini Kit(Qiagen, Germany)说明书提取总 RNA,使用 BioPhotometer 核酸和蛋白质定量仪(Eppendorf, Germany)检测 RNA 的浓度和纯度( $OD_{260}/OD_{280}$ ). 采用 Super SMART™ PCR cDNA Synthesis Kit(Clontech, USA)合成 cDNA,用于后续 RACE 实验克隆 *EseIF4G* 基因,以及实时荧光定量 PCR(qRT-PCR)检测.

根据课题组前期构建的中华绒螯蟹抑制性消减杂交文库<sup>[20]</sup>中获得的一个 492 bp 的表达序列标签(Expressed sequence tags, EST)序列(GenBank 登录号:GE329190)设计 RACE 反应特异引物(见表 1). 以仔蟹 I 期腹部组织 cDNA 为模板,用 *EseIF4G* 基因特异性引物和通用引物 10×Universal Primer A Mix(UPM)进行 RACE PCR 扩增. 5' RACE 和 3' RACE 产物进行切胶纯化和载体连接,连接产物转入 *Escherichia coli* 细胞 DH5 $\alpha$ (TaKaRa, 中国大连)并经蓝白斑筛选,选取 3 个阳性克隆送往生工生物工程(上海)股份有限公司测序.

### 1.3 *EseIF4G* 基因序列分析

将拼接获得的 *EseIF4G* 基因全长 cDNA 序列用 NCBI(<http://www.ncbi.nlm.nih.gov/>)进行 cDNA 全长完整性验证、开放阅读框查找(open reading frame, ORF). 使用 ExPASy 服务器(<https://web.expasy.org/>)<sup>[24]</sup>的 ProtParam 和 ProtScale 工具分析 *EseIF4G* 蛋白的理化性质和疏水性. 使用 SignalP 程序(<http://www.cbs.dtu.dk/services/SignalP>)<sup>[25]</sup>预测信号肽, MotifScan 程序([http://myhits.isb-sib.ch/cgi-bin/motif\\_scan](http://myhits.isb-sib.ch/cgi-bin/motif_scan))查找 motif, SMART 软件(<http://smart.embl-heidelberg.de/>)<sup>[26]</sup>分析结构功能域, PSORT II 软件(<http://psort.hgc.jp/form2.html>)进行亚细胞定位, TMHMM v.2.0 软件(<http://www.cbs.dtu.dk/services/TMHMM-2.0/>)<sup>[27]</sup>预测跨膜区. 使用在线软件 PSIPRED(<http://bioinf.cs.ucl.ac.uk/psipred/>)预测二级结构, SWISS-MODEL(<http://www.expasy.org/swissmod/SWISS-MODEL.html>)预测三级结构<sup>[24]</sup>. 使用 MEGA X 软件构建邻接树(Neighbor-joining tree, NJ tree)来验证 *EseIF4G* 蛋白在进化过程中与其他同源蛋白的关系<sup>[28]</sup>.

表 1 本研究所用的主要引物  
Table 1 Primers used in this study

引物名称	引物序列(5'→3')	应用
10×Universal Primer A Mix(UPM)	Long-CTAATACGACTCACTATAGGGCAAGCAGCTGGTATCAACGCAGAGT Short-CTAATACGACTCACTATAGGGC	5'&3'RACE PCR
GSP1-eIF5'	CAGTTTCTCTGGAATCGCTGTTGG	5'RACE
GSP2-eIF5'	ATGCCAAAGTGGTTTCTGTGTTA	
GSP3-eIF5'	CCATTTTGCCTTTGCTGTTGC	
GSP1-eIF3'	ACTATCCTGGCAAAGGAAAGCAC	3'RACE
GSP2-eIF3'	GAAGAACGGGCAACAGCAAAG	
GSP3-eIF3'	TCCATTCTTCATCGTTGCATCC	
eIF-ORFf	TTCTATGTACGCACAGCTGTG	ORF
eIF-ORFr	CTTGGCATTATCATTTGAGCATCCTC	
qβ-actin-F	GCATCCACGAGACCACTTACA	Real-time PCR
qβ-actin-R	CTCCTGCTTGCTGTATCCACATC	
qeIF-F	CGATTCAAGCCCAATCGCTG	
qeIF-R	TGGATAGGGAACGCACCTTCG	

1.4 实时荧光定量 PCR (qRT-PCR) 检测

用 Prime Primer 5 软件设计引物 qβ-actin-F 和 qβ-actin-R(表 1)用于 qRT-PCR 检测,选择 *β-actin* 基因(GenBank 登录号: HM053699.1)作为内参基因,其引物 qβ-actin-F 和 qβ-actin-R 引自文献<sup>[29]</sup>. 使用 SYBR® Premix Ex-Taq™ II 试剂盒(TaKaRa, 中国大连)在 ABI Step One Plus™ 仪器上检测 *EseIF4G* 基因在中华绒螯蟹幼体 10 个不同发育阶段的 mRNA 表达情况. PCR 反应和实验样本均设计 3 次重复,采用  $2^{-\Delta\Delta C_t}$  方法计算相对表达量<sup>[30]</sup>. 用 SPSS 25.0 软件进行单因素方差分析和 Tukey's 法多重比较,以(平均值±标准差)表示,显著性差异水平以  $\alpha=0.05$  表示,图中用不同的小写字母标注.

1.5 适应性进化分析

1.5.1 序列获取

为研究中华绒螯蟹 *EseIF4G* 基因的适应性进化,选择了其他 13 种代表性节肢动物物种的 *eIF4G* 基因序列加入分析,涉及甲壳亚门、螯肢亚门和六足亚门,并选择黑头软口鲦(*Pimephales promelas*)作为外群. 部分物种的 *eIF4G* 基因序列从 NCBI 数据库直接下载,或从 NCBI 下载的基因组中用近缘物种的序列执行本地 blast 获取,本研究所用的直系同源基因在 NCBI 上的登录号和本地获取所用物种基因组信息见表 2.

表 2 本研究所用物种基因或基因组信息  
Table 2 Species genes or genomes information used in this study

亚门 Subphylum	纲 Phylum	物种名 Scientific name	登录号 Accession
甲壳亚门 Crustacea	软甲纲 Malacostraca	中华绒螯蟹 <i>Eriocheir sinensis</i>	KF199900
	软甲纲 Malacostraca	灰眼雪蟹 <i>Chionoecetes opilio</i>	GCA_016584305.1
	软甲纲 Malacostraca	三疣梭子蟹 <i>Portunus trituberculatus</i>	GCA_008373055.1
	软甲纲 Malacostraca	斑节对虾 <i>Penaeus monodon</i>	XM_037921888.1
	软甲纲 Malacostraca	南美白对虾 <i>Penaeus vannamei</i>	XM_027378269.1
	软甲纲 Malacostraca	普通卷甲虫 <i>Armadillidium vulgare</i>	GCA_004104545.1
	软甲纲 Malacostraca	方鼻卷甲虫 <i>Armadillidium nasatum</i>	GCA_009176605.1
	鳃足纲 Branchiopoda	大型溞 <i>Daphnia magna</i>	XM_032929506.1
	颚足纲 Maxillopoda	鹅颈藤壶 <i>Pollicipes pollicipes</i>	XM_037234747.1
螯肢亚门 Chelicerata	肢口纲 Merostomata	美洲鲎 <i>Limulus polyphemus</i>	XM_022399930.1
	蛛形纲 Arachnida	温室拟肥腹蛛 <i>Parasteatoda tepidariorum</i>	XM_016047656.1
	蛛形纲 Arachnida	穹蛛 <i>Stegodyphus dumicola</i>	XM_035364169.1
六足亚门 Hexapoda	昆虫纲 Insecta	小蜜蜂 <i>Apis florea</i>	XM_012486943.2
	昆虫纲 Insecta	第二堆砂白蚁 <i>Cryptotermes secundus</i>	XM_023866510.1
外群 Outgroup		黑头软口鲦 <i>Pimephales promelas</i>	XM_039692321.1

1.5.2 构建系统发育树

使用 PRANK 软件对上述获取的 *eIF4G* 基因序列进行密码子水平的比对<sup>[31]</sup>. 使用 Gblocks 软件过滤比对后序列,删除一些低质量或高变异度的序列区域,仅留进化保守的区域用于后续分析<sup>[32]</sup>. 使用 IQ-TREE 自动测试并选择最佳替代模型,构建最大似然树(maximum likelihood tree, ML tree)<sup>[33]</sup>. 该系统发育树同时也被用于构建后续选择压力分析的树文件. 从 TimeTree(<http://www.timetree.org/>) 获取该 15 个

物种的系统发育关系<sup>[34]</sup>.

### 1.5.3 检测特异性位点

选择其中6种甲壳动物,将其编码区序列翻译为氨基酸序列,PRANK 比对后使用 FasParser 软件鉴定中华绒螯蟹 *EseIF4G* 基因的特异性位点 (Segregating sites)<sup>[35]</sup>.

### 1.5.4 选择压力分析

非同义替代率 (Non-synonymous substitution rate, dN) 与同义替代率 (Synonymous substitution rate, dS) 的比值  $\omega$  (dN/dS), 通常作为检测选择压力的标准, 其中  $\omega > 1$ ,  $\omega = 1$  和  $\omega < 1$  分别代表基因受到正选择 (Positive selection), 中性选择 (Neutral selection) 和纯化选择 (Purify selection). 本研究使用 PAML 4.9 软件包的 CODEML 程序, 选择位点模型 (Site model)、分支模型 (Branch model) 和分支位点模型 (Branch-site model) 进行进化分析<sup>[36]</sup>.

使用位点模型检测 *eIF4G* 基因在节肢动物数据集中受正选择的位点, 比较 M8a 和 M8 模型, 利用似然率检测方法 (Likelihood ratio tests, LTRs), 计算  $2\Delta\ln L$  值与自由度之间的卡方分布, 若  $P < 0.05$ , 则两个模型之间有显著差异<sup>[37]</sup>. 在 M8 模型中, 利用经典贝叶斯方法 BEB (Bayes empirical bays) 计算位点的后验概率 (Posterior probability, PP), 若  $PP > 0.95$ , 则被认为是潜在的正选择位点<sup>[38]</sup>. 在分支模型中, 首先比较 one-ratio 和 free-ratio 模型以确定支系是否具有不同的进化速率; 其次, 将中华绒螯蟹末端分支设为前景支, 比较 one-ratio 模型与 two-ratio 模型来判断前景分支是否受正选择. 使用分支位点模型检测某一进化谱系受正选择作用影响的位点, 本研究中同样将中华绒螯蟹的末端分支设为前景支进行分析, 若 Ma0 和 Ma 模型差异显著且  $\omega > 1$  时, 接受正选择模型 Ma, 同时所检测的正选择位点  $PP > 0.95$ , 则被认为是正选择位点.

此外, 使用 Datamonkey (<http://www.datamonkey.org/>) 中的固定效应似然法 (fixes effects likelihood, FEL)、单一似然祖先计数法 (single likelihood ancestor, SLAC) 和 FUBRA (Fast unconstrained bayesian approximation) 检测正选择位点, 设置 FEL 和 SLAC 显著水平小于 0.2, FUBAR 后验概率值大于 0.8<sup>[39]</sup>. 最后, 整合 PAML 和 Datamonkey 筛选的正选择位点, 将两种方法以上检测到的位点视为显著的正选择位点, 并标注在模拟的中华绒螯蟹 *eIF4G* 蛋白三维结构中.

## 2 结果与讨论

### 2.1 中华绒螯蟹 *EseIF4G* 基因 cDNA 全长序列的克隆与分析

通过 RACE PCR 扩增与测序获得 *EseIF4G* 基因序列片段, 使用 DNASTar 7.1 软件拼接获得的各序列, 获得中华绒螯蟹 *EseIF4G* 基因的 cDNA 全长序列为 3 769 bp (图 1), GenBank 登录号为 KF199900. 其包括 568 bp 的 5'-UTR、822 bp 的 3'-UTR 和 2 379 bp 的 ORF. 其编码的蛋白质具有 792 个氨基酸, 分子量为 89.6 kDa, 理论等电点 pI 为 6.07. *EseIF4G* 蛋白是不稳定蛋白和非分泌型蛋白, 没有明显的疏水性区域, 没有跨膜区且不具有信号肽, 定位于细胞核可能性高达 91.3%.

推导的氨基酸序列进行同源比对分析后发现, 本实验克隆的 *EseIF4G* 蛋白与其他物种显示出较高的同源性, 例如与克氏原螯虾 (*Procambarus clarkii*, GenBank 登录号: ASW35070.1) *eIF4G* 蛋白的同源性高达 84%, 与南美白对虾 (*P. vannamei*, GenBank 登录号: XP\_027234070.1) *eIF4G2*-like 蛋白的同源性为 86%, 与灰眼雪蟹 (*C. opilio*, GenBank 登录号: KAG0719495.1) *eIF4G2* 蛋白的同源性为 83%, 与普通卷甲虫 (*A. vulgare*, GenBank 登录号: RXG61596.1) *eIF4G2* 蛋白的同源性为 60% 等.

用 SMART 分析 *EseIF4G* 蛋白的结构域, 结果表明 *EseIF4G* 蛋白含有 MIF4G 结构域、MA3 结构域和 W2/*eIF5C* 结构域, 前两个结构域是 *eIF4G* 蛋白家族共有的典型结构域. 用 Motif Scan 搜索 *EseIF4G* 蛋白的 motif, 结果表明 *EseIF4G* 蛋白可能还含有如下功能基序: 6 个 N 端糖基化 (Asn 糖基化) 位点、2 个环腺苷酸 (cyclic adenosine monophosphate, cAMP) 磷酸化位点、11 个酪蛋白激酶 II (casein kinase II, CK2) 磷酸化位点、7 个 N-酰基化 (N-myristoylation) 位点、8 个蛋白激酶 C (protein kinase C, PKC) 磷酸化位点、1 个 NLS\_BP 核定位信号和 1 个富含 Gln 区等. 使用 PSIPRED 预测 *EseIF4G* 蛋白的二级结构, 结果表明 *EseIF4G* 蛋白包含 30 个  $\alpha$  螺旋, 无  $\beta$ -折叠和 31 个无规则卷曲. SWISS-MODEL 对 *EseIF4G* 蛋白进行同源建模, 显示该蛋白与模板 3l6a.1.A 蛋白的匹配度最高, 构建的蛋白三维结构图如图 2 所示, 与 PSIPRED 预测结果一致, 该蛋白为全  $\alpha$  蛋白.



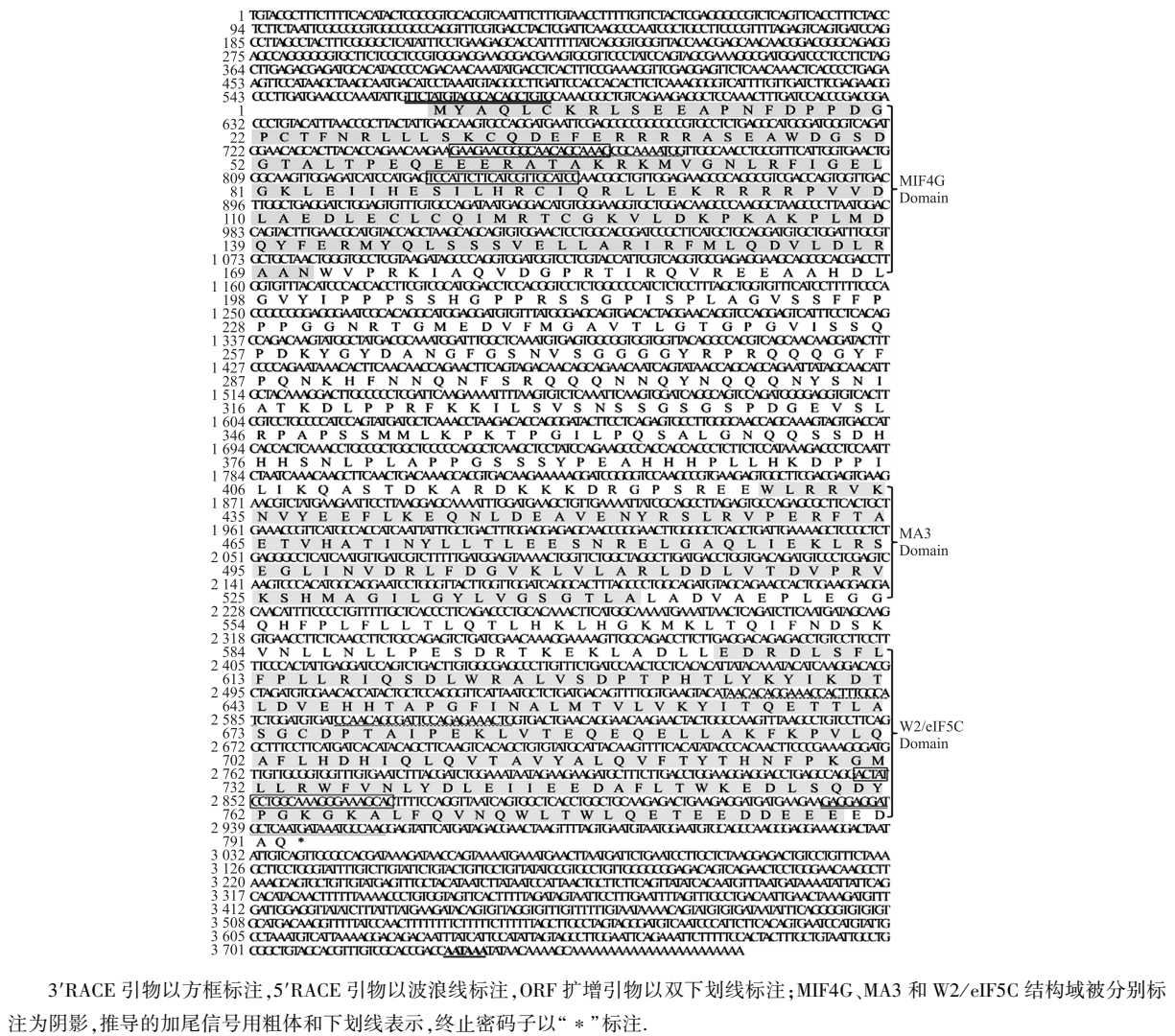


图 1 中华绒螯蟹 *EseIF4G* 基因的 cDNA 全长序列和推导的氨基酸序列结构

Fig. 1 The full-length cDNA and predicted amino acid sequence of *EseIF4G* gene from *Eriocheir sinensis*

## 2.2 多序列比对和系统发育分析

为了更好地研究 *EseIF4G* 与其他同源蛋白的关系,从 NCBI 上下载了包括脊椎动物和无脊椎动物在内的 22 个不同物种的 eIF4G 蛋白序列,使用 MEGA X 软件基于邻接法构建了 NJ 系统发育树(图 3)。系统发育分析显示,*EseIF4G* 与无脊椎动物分支(Crustacea+Insecta+Merostomata,甲壳纲+昆虫纲+肢口纲)聚在了一起;同时,中华绒螯蟹与南美白对虾(*P. vannamei*)、普通卷甲虫(*A. vulgare*)形成了甲壳纲分支,该分支与昆虫纲分支和肢口纲分支完全分离。

## 2.3 中华绒螯蟹 *EseIF4G* 基因在其不同幼体发育阶段的 mRNA 表达模式

利用 qRT-PCR 技术对 *EseIF4G* 基因在中华绒螯蟹其不同幼体发育阶段的 mRNA 表达进行检测,结果如图 4 所示。以受精卵期的基因表达量作为参照,*EseIF4G* 基因在中华绒螯蟹 10 个不同的幼体发育阶段广泛表达。总体上,*EseIF4G* mRNA 表达量在受精卵、溞状幼体 I、II、IV 期相对较高,相反在溞状幼体 V 期、大眼幼体期、仔蟹 I、II 期相对

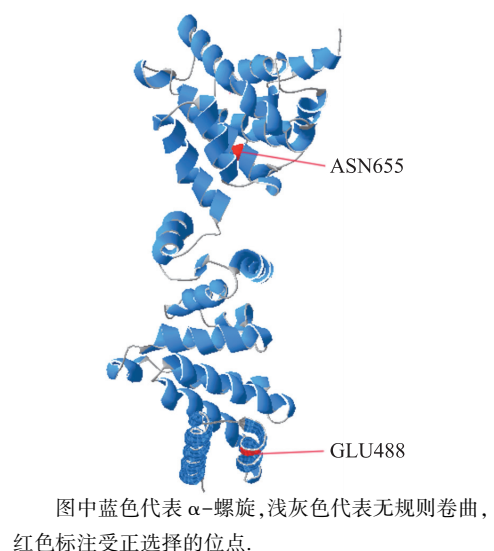
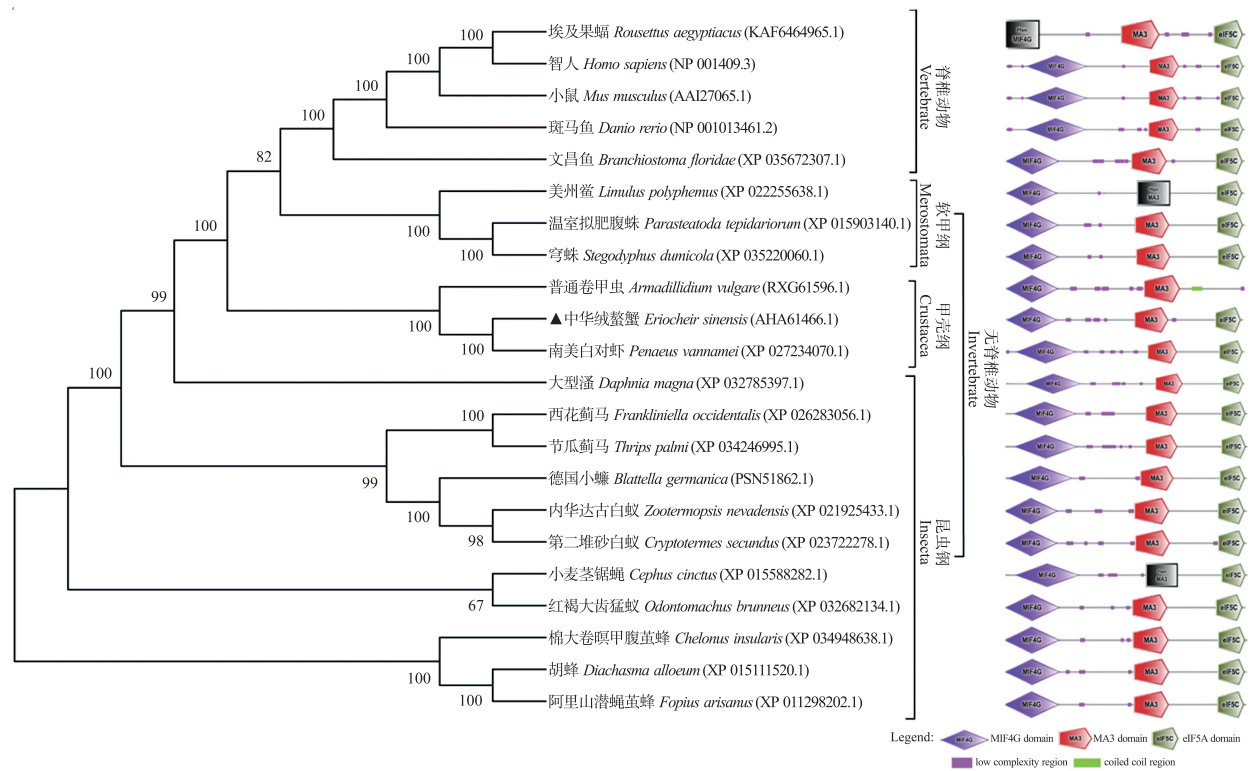


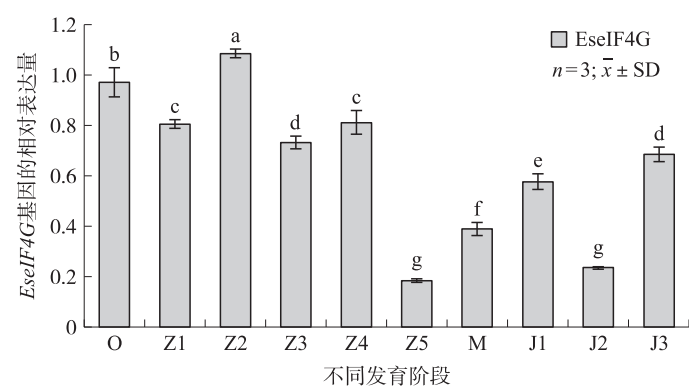
图 2 中华绒螯蟹 *EseIF4G* 蛋白的三维结构  
Fig. 2 Predicted three-dimensional structure of *EseIF4G* from *Eriocheir sinensis*

较低。 *EseIF4G* 基因的表达量在短尾化期间显著波动。 *EseIF4G* 基因在溞状幼体 II 期表达量最高,而在溞状幼体 V 期和仔蟹 II 期表达量最低 ( $P>0.05$ )。 值得注意的是,在 Z5 期急剧下降至最低后, *EseIF4G* 基因表达量在大眼幼体期和仔蟹 I 期逐步增加 ( $P<0.05$ )。



每一枝上的数值表示 Neighbor-Joining bootstrap 值;氨基酸序列号在物种名之后,不同分支被不同颜色覆盖;  
各物种 eIF4G 蛋白的结构域在对应分支后方显示。

图 3 基于邻接法构建的 *EseIF4G* 及其同源蛋白的聚类树状图  
Fig. 3 Phylogenetic tree of *EseIF4G* and its homologs bases on neighbor-joining method



O:受精卵,Z1:溞状幼体 I 期,Z2:溞状幼体 II 期,Z3:溞状幼体 III 期,Z4:溞状幼体 IV 期,Z5:溞状幼体 V 期,M:大眼幼体期,J1:仔蟹 I 期,J2:仔蟹 II 期,J3:仔蟹 III 期。以卵期基因的表达量为参照,采用(平均值±标准差)表示,不同的小写字母表示不同发育阶段表达量的差异水平显著 ( $P<0.05$ ),内参基因  $\beta$ -actin,  $n=3$ 。

图 4 中华绒螯蟹 *EseIF4G* 基因在其不同幼体发育阶段的相对表达量  
Fig. 4 The relative expression level of *EseIF4G* gene in different larval development stages of *Eriocheir sinensis*

2.4 中华绒螯蟹 *EseIF4G* 基因的适应性进化分析

2.4.1 系统发育关系

利用包括中华绒螯蟹在内的 14 种节肢动物以及外群黑头软口鲗的 *eIF4G* 基因序列,基于最大似然法构建了系统发育树。结果如图 5 显示, *eIF4G* 基因树分为三个主要的进化枝:甲壳亚门进化枝 (Crustacea clade)、六足亚门进化枝 (Hexapoda clade) 和螯肢亚门进化枝 (Hexapoda clade)。比较 *eIF4G* 基因树

(图 5-a) 和物种树(图 5-b)后,发现甲壳动物 *eIF4G* 基因树与物种树的拓扑结构略有差异. 在物种树中,鳃足纲的大型蚤(*D. magna*)和颚足纲的鹅颈藤壶(*P. pollicipes*),与其余甲壳动物聚成一个大的进化枝;而在基因树中,大型蚤和鹅颈藤壶,与六足亚门昆虫纲的两个物种聚为一个进化枝;其余的进化枝(例如软甲纲、螯肢亚门)的拓扑结构完全一致.

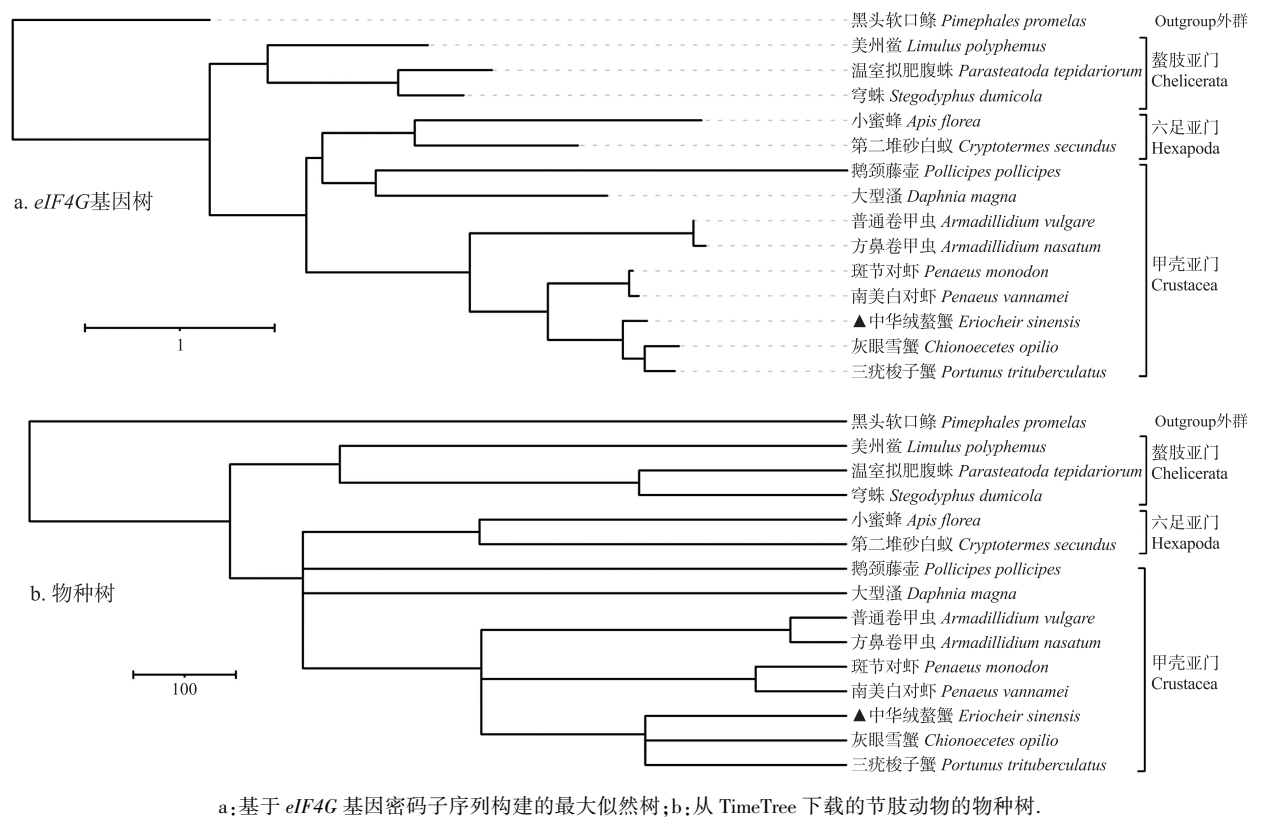


图 5 节肢动物和外群 *eIF4G* 基因的系统发育树

Fig. 5 Phylogenetic tree of *eIF4G* genes from arthropods and outgroup species

2.4.2 特异性位点检测

对中华绒螯蟹 *EseIF4G* 蛋白的特异性位点检测结果如图 6 所示,位点 97、457 为特异性位点,分别位于 MIF4G 结构域和 MA3 结构域. 位点 97 处的氨基酸由谷氨酰胺 (Gln, Q) 变为精氨酸 (Arg, R), 位点 457 处的氨基酸由赖氨酸 (Lys, K) 变为精氨酸 (Arg, R).

		97	457	
<i>Eriocheir sinensis</i> ▲	R	C I Q <b>R</b> L L E	R S L <b>R</b> V P E	
<i>Chionoecetes opilio</i>	R	C I Q Q L L E	R C L K V P E	
<i>Penaeus monodon</i>	R	C I Q Q L L E	R S L K L P E	
<i>Penaeus vannamei</i>	R	C I Q Q L L E	R S L K L P E	
<i>Armadillidium vulgare</i>	N	C I Q Q L L E	R G I K I P D	
<i>Daphnia magna</i>	S	C I Q Q L L A	R D C K V P D	
<i>Pollicipes pollicipes</i>	M	C I Q Q L L E	K E I K S I E	

图中特异性位点用红色标记.

图 6 中华绒螯蟹 *EseIF4G* 基因的特异性位点

Fig. 6 Segregating sites of *EseIF4G* gene in *Eriocheir sinensis*

2.4.3 选择压力分析

为研究中华绒螯蟹 *EseIF4G* 基因是否受到正选择作用,首先假设系统发生树所有分支有同一个  $\omega$  值, one-ratio 模型结果显示  $\omega$  值为 0.035 5, 显著小于 1, 表明该基因整体上受纯化选择作用(表 3). Free-ratio 模型假定系统发生树上每一支和节点都具有独立的  $\omega$  值, 发现 free-ratio 模型显著优于 one-ratio 模型, 提示不同支系具有不同的进化速率. 在分支模型的 two-ratio 中, 将中华绒螯蟹的末端分支设置为前景支, 其余分支设为背景支, two-ratio 模型拟合效果并没有明显优于 one-ratio 模型 ( $P=0.440$ ), 表明中华绒螯蟹 *EseIF4G* 基因与其他近缘甲壳动物有相似的进化速率(表 3). 其次, 使用位点模型对 *EseIF4G* 基因检测,



M8 模型明显优于 M8a 模型 ( $P<0.01$ ),但是没有检测出后验概率大于 0.95 的正选择位点(表 3).最后,使用分支位点模型,将中华绒螯蟹设置为前景支,发现 Ma 模型和 Ma0 模型之间没有显著差异 ( $P>0.05$ ).

表 3 基于分支模型、位点模型和分支位点模型对 *EseIF4G* 基因的选择压力分析

models	-lnL	2ΔlnL	df	<i>P</i> values	$\omega$ values	positively selected sites( $PP>0.95$ )
one-ratio( M0)	17 234.722				0.035 5	—
free-ratio vs M0	4 185.715	26 098.014	24	nearly 0	variable	—
two-ratio vs M0 (branch <i>E. sinensis</i> )	17 234.424	4.795	1	0.440	$\omega_{\text{foreground}} = 0.346$ $\omega_{\text{background}} = 0.054\ 1$	—
M8a	4 147.220	25 557.419	1	nearly 0	1	Not allowed
M8	16 925.930				66.598	
Ma0	17 155.614	0	1	1.000	$\omega_0 = 0.031\ 0, \omega_1 = 1, \omega_2 = 1$	Not allowed
Ma	17 155.614				$\omega_0 = 0.031\ 0, \omega_1 = 1, \omega_2 = 1$	

在 Datamonkey 网站上选用 SLAC、FEL 和 FUBAR 3 种方法检测,设置 SLAC 和 FEL 显著水平为 0.2, FUBAR 后验概率为 0.8. 分析结果显示, FEL 检测出 3 个正选择位点(448、512 和 655), FUBAR 检测出 2 个正选择位点(448 和 655),而 SLAC 没有检测出正选择位点(表 4). 将 PAML 和 Datamonkey 2 种方法检测的结果整合起来, *EseIF4G* 基因检测出 2 个正选择位点(448 和 655),标注在 eIF4G 蛋白三维结构中,如图 2 所示.

表 4 PAML 和 Datamonkey 鉴定 *EseIF4G* 基因的正选择位点

Table 4 Positively selected sites of *EseIF4G* gene by PAML and Datamonkey

Site	PAML		Datamokey		
	branch-site( Ma)	site( M8)	FEL( $P<0.2$ )	FUBAR( $PP>0.8$ )	SLAC( $P<0.2$ )
448			0.030	0.868	
512			0.140		
655			0.023	0.824	

eIF4G 是 eIF4F 复合体的重要组成部分,在翻译过程中可作为组装 mRNA、核糖体亚基和帽结合蛋白的骨架蛋白. 本研究克隆了中华绒螯蟹 *EseIF4G* 基因的全长 cDNA 序列. 系统发育分析发现, *EseIF4G* 蛋白序列具有高度保守的结构域,其在无脊椎动物和脊椎动物的整个物种中高度保守.

中华绒螯蟹胚胎与幼体发育包括受精卵、5 个蚤状幼体期、1 个大眼幼体期和 3 个仔蟹期,此后幼蟹经过多次蜕皮发育为成蟹. 在蚤状幼体 V 期至仔蟹 I 期,中华绒螯蟹幼体形态发生显著变化,即“短尾化”过程. 短尾化之前,蟹类幼体呈类似虾的体型;短尾化过程中,形态结构变化过程分为两个截然相反的过程:分化和整合. 一些部位高度分化(尤其是附肢)或融合在一起(尤其是腹神经节);某些部位的功能得到加强,例如胸部;某些部位的功能减弱了,例如腹部<sup>[18]</sup>. 短尾化过程完成后,蟹的显著形态变化减少,但体型随蜕壳而不断变大. 本实验中,在中华绒螯蟹幼体发育的 10 个阶段均检测到 *EseIF4G* 基因表达,表明 *EseIF4G* 基因可能在中华绒螯蟹的幼体发育过程中起调节作用. 其次,随着幼蟹的形态变化, *EseIF4G* 基因的表达量也随之波动,在蚤状幼体 V 期(Z5)突然下降,然后在大眼幼体期(M)显著升高,这个显著变化暗示 *EseIF4G* 基因可能在中华绒螯蟹的短尾化过程中起调节作用. 此外,中华绒螯蟹幼体经历生活环境由咸淡水转为淡水的过程:从受精卵到大眼幼体期前期生活在江河口或沿海水域等咸淡水中,而后在大眼幼体后期则迁移到淡水中<sup>[21,40]</sup>. 中华绒螯蟹 *EseIF4G* 基因表达变化的趋势与其生活史对应,说明 *EseIF4G* 基因在中华绒螯蟹适应生活环境的渗透压调节方面有一定作用. 已有研究表明,中华绒螯蟹 *EjsHSP90*、*EjsANK* 和 *EjsPCNA* 基因在仔蟹 I 期腹部组织的表达量均高于在大眼幼体期腹部组织<sup>[19-20]</sup>, *EsScr* 和 *EsAntp* 基因在蚤状幼体 V 期的表达量最低且显著高于在大眼幼体期<sup>[21]</sup>. 上述基因的表达模式的变化与 *EseIF4G* 基因高度相似,验证了 *EseIF4G* 基因随着幼蟹的形态变化和生活环境的迁移,可能参与中华绒螯蟹幼蟹的早期发育,特别是短尾化过程.

筛查中华绒螯蟹 *EseIF4G* 蛋白的特异性位点,发现位点 97、457 为特异性位点,分别位于 MIF4G 结构域和 MA3 结构域. eIF4G 的中间结构域称为 MIF4G 结构域,已被证明介导多种蛋白质-蛋白质相互作用,例如与翻译起始因子 eIF4A 和 eIF3、RNA 的结合. 同时,除了 eIF4F 复合体外,其他含 MIF4G 结构域的蛋

白质还参与了 mRNA 成熟、监视、翻译和降解等信号通路<sup>[41]</sup>。eIF4G 的 C 端结构域也被称为 MA3 结构域,具有调控功能,可与 ATP 依赖性 RNA 解旋酶 eIF4A 结合以促进翻译起始;程序性细胞死亡 4 蛋白(PDCD4)也具有两个 MA3 结构域,与 eIF4G-MA3 竞争结合 eIF4A,从而抑制蛋白质的合成<sup>[42]</sup>。因此推测这两个特异位点可能与中华绒螯蟹的幼体发育阶段的 mRNA 组装、翻译、降解相关,但需要进一步的实验验证。

基于 one-ratio 模型的 PAML 选择压力分析显示  $\omega$  值为 0.035 5,显著小于 1,表明该基因整体上受纯化选择作用;free-ratio 与 one-ratio 比较显示不同的节肢动物谱系具有不同的进化速率。进一步,将中华绒螯蟹的末端分支设为前景支的分支模型和分支位点模型,以及整体水平上的位点模型都没有检测到正选择信号,一定程度上说明中华绒螯蟹 *EseIF4G* 基因进化保守,纯化选择在该基因进化过程中起关键作用,与其他节肢动物具有相同或类似的功能。但在选定的节肢动物中,使用 Datamonkey 的 FEL 和 FUBAR 方法,共同鉴定到两个正选择位点(448 和 655),分别位于 MA3 和 W2 结构域。W2 结构域是 eIF4G 蛋白羧基末端小却保守的结构域,因包含两个保守的色氨酸得名;充当潜在的翻译调节剂参与到 eIF2B、eIF5 与 eIF2 亚基,eIF4G 与蛋白激酶 MNK1 的相互作用<sup>[43]</sup>。结果表明,这些位点在节肢动物 eIF4G 蛋白进化中起重要作用,调控 mRNA 翻译过程和蛋白质相互作用,可能参与各种节肢动物外部形态、生理活动等的差异调控。未来可通过定向诱变等实验探明 *EseIF4G* 基因在中华绒螯蟹幼体发育阶段发挥的作用,为深入理解中华绒螯蟹短尾化的遗传学基础提供证据。

### 3 结论

本研究首次克隆了中华绒螯蟹 *EseIF4G* 基因全长 cDNA 序列,其编码蛋白具有 MIF4G、MA3 和 W2/eIF5C 3 个保守结构域,在进化上与甲壳动物最近,并且在无脊椎动物长期进化过程中高度保守。*EseIF4G* 基因在中华绒螯蟹幼体发育各个阶段广泛表达,在溞状幼体 V 期表达量最低,随后在大眼幼体期表达量显著升高,可能参与调控中华绒螯蟹的短尾化和渗透压适应过程。同时,检测到 eIF4G 蛋白两个中华绒螯蟹特异性位点和两个节肢动物潜在正选择位点,这些可能是调控 mRNA 翻译和蛋白质互作的位点,参与节肢动物生长、发育等过程。本研究为进一步探索甲壳动物幼体短尾化调控机制提供了基础资料。

### [参考文献]

- [1] CHEN Q, YANG B, NASS N, et al. Impact of eukaryotic translation initiation factors on breast cancer: still much to investigate[J]. *Cancers*, 2020, 12(7): 1–19.
- [2] VALÁSEK L S. ‘Ribozoomin’-translation initiation from the perspective of the ribosome-bound eukaryotic initiation factors (eIFs) [J]. *Current protein & peptide science*, 2012, 13(4): 305–330.
- [3] SMOLLE M A, CZAPIEWSKI P, LAPIŃSKA-SZUMCZYK S, et al. The prognostic significance of eukaryotic translation initiation factors (eIFs) in endometrial cancer[J]. *International journal of molecular sciences*, 2019, 20(24): 6169.
- [4] 吴继华, 张建中. 真核翻译起始因子家族与恶性肿瘤相关性的研究进展[J]. *现代肿瘤医学*, 2015, 23(23): 3525–3528.
- [5] PRÉVT D, DARLIX J L, OHLMANN T. Conducting the initiation of protein synthesis: the role of eIF4G[J]. *Biology of the cell*, 2003, 95(3–4): 141–156.
- [6] VILLA N, DO A, HERSHEY J W, et al. Human eukaryotic initiation factor 4G (eIF4G) protein binds to eIF3c, -d, and -e to promote mRNA recruitment to the ribosome[J]. *Journal of biological chemistry*, 2013, 288(46): 32932–32940.
- [7] HOWARD A, ROGERS A N. Role of translation initiation factor 4G in lifespan regulation and age-related health[J]. *Ageing research reviews*, 2014, 13(1): 115–124.
- [8] SONENBERG N, HINNEBUSCH A G. Regulation of translation initiation in eukaryotes: mechanisms and biological targets[J]. *Cell*, 2009, 136(4): 731–745.
- [9] JACKSON R J, HELLEN C U, PESTOVA T V. The mechanism of eukaryotic translation initiation and principles of its regulation[J]. *Nature reviews molecular cell biology*, 2010, 11(2): 113–127.
- [10] GRADI A, IMATAKA H, SVITKIN Y V, et al. A novel functional human eukaryotic translation initiation factor 4G[J]. *Molecular cell biology*, 1998, 18(1): 334–342.
- [11] COLDWELL M, SACK U, COWAN J, et al. Multiple isoforms of the translation initiation factor eIF4GII are generated via use

- of alternative promoters, splice sites and a non-canonical initiation codon[J]. Biochemical journal, 2012, 448(1): 1–11.
- [12] KEIPER B D, GAN W, RHOADS R E. Protein synthesis initiation factor 4G[J]. International journal of biochemistry & cell biology, 1999, 31(1): 37–41.
- [13] DAS S, DAS B. eIF4G3—an integrator of mRNA metabolism? [J]. Fems yeast research, 2016, 16(7): 1–9.
- [14] ORELLANA R A, WILSON F A, GAZZANELO M C, et al. Sepsis and development impede muscle protein synthesis in neonatal pigs by different ribosomal mechanisms[J]. Pediatric research, 2011, 69(6): 473–478.
- [15] ALVAREZ-DOMINGUEZ J R, ZHANG X, HU W. Widespread and dynamic translational control of red blood cell development[J]. Blood, 2017, 129(5): 619–629.
- [16] GHOSH S, LASKO P. Loss-of-function analysis reveals distinct requirements of the translation initiation factors eIF4E, eIF4E-3, eIF4G and eIF4G2 in *Drosophila* spermatogenesis[J]. PLoS one, 2015, 10(4): e0122519.
- [17] TANG B, ZHOU K, SONG D, et al. Molecular systematics of the Asian mitten crabs, genus *Eriocheir* (Crustacea: Brachyura) [J]. Molecular phylogenetics and evolution, 2003, 29(2): 309–316.
- [18] GUINOT D, BOUCHARD J M. Evolution of the abdominal holding systems of brachyuran crabs (Crustacea, Decapoda, Brachyura) [J]. Zoosystema, 1998, 20(4): 613–694.
- [19] LI P, ZHA J, HUANG H, et al. Identification, mRNA expression and characterization of a novel *ANK-like* gene from Chinese mitten crab *Eriocheir japonica sinensis* [J]. Comparative biochemistry and physiology part b: biochemistry & molecular biology, 2009, 153(4): 332–339.
- [20] LI P, ZHA J, KONG Y, et al. Identification, mRNA expression and characterization of proliferating cell nuclear antigen gene from Chinese mitten crab *Eriocheir japonica sinensis* [J]. Comparative biochemistry and physiology part A: molecular & integrative physiology, 2010, 157(2): 170–176.
- [21] 江贤峰. 中华绒螯蟹 *EsScr* 和 *EsAntp* 基因的鉴定及其在幼体发育过程中的表达分析[D]. 南京: 南京师范大学, 2015.
- [22] CHEN M, YUE Y, HE J, et al. Screening and identification of microRNAs during larval metamorphic development of Chinese mitten crab *Eriocheir sinensis* [J]. Aquaculture research, 2020(51): 2322–2355.
- [23] SONG C, CUI Z, HUI M, et al. Comparative transcriptomic analysis provides insights into the molecular basis of brachyurization and adaptation to benthic lifestyle in *Eriocheir sinensis* [J]. Gene, 2015, 558(1): 88–98.
- [24] GASTEIGER E, HOOGLAND C, GATTIKER A, et al. The proteomics protocols handbook: protein identification and analysis tools on the ExPASy server[M]. Totowa NJ: humana press, 2005.
- [25] ALMAGRO-ARMENTEROS J J, TSIRIGOS K D, SØNDERBY C K, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks[J]. Nature biotechnology, 2019, 37(4): 420–423.
- [26] IVICA L, PEER B. 20 years of the SMART protein domain annotation resource[J]. Nucleic acids research, 2018, 46(D1): D493–D496.
- [27] KROGH A, LARSSON B, HEIJNE G V, et al. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes[J]. Journal of molecular biology, 2001, 305(3): 567–580.
- [28] KUMAR S, STECHER G, LI M, et al. MEGA X: molecular evolutionary genetics analysis across computing platforms[J]. Molecular biology and evolution, 2018, 35(6): 1547–1549.
- [29] JIN X, LI W, CHENG L, et al. Two novel short C-type lectin from Chinese mitten crab, *Eriocheir sinensis*, are induced in response to LPS challenged[J]. Fish shellfish immunology, 2012, 33(5): 1149–1158.
- [30] LIVAK K J, SCHMITTGEN T D. Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta C_T}$  method[J]. Methods, 2001, 25(4): 402–408.
- [31] LÖYTYNOJA A. Phylogeny-aware alignment with PRANK [C]//Russell D. Methods in molecular biology. Totowa NJ: humana press, 2014.
- [32] GERARD T, JOSE C. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments[J]. Systematic biology, 2007, 56(4): 564–577.
- [33] LAM-TUNG N, SCHMIDT H A, ARNDT V H, et al. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies[J]. Molecular biology and evolution, 2015(1): 268–274.
- [34] KUMAR S, STECHER G, SULESKI M, et al. TimeTree: a resource for timelines, timetrees, and divergence times [J]. Molecular biology and evolution, 2017, 34(7): 1812–1819.
- [35] SUN Y. FasParser: a package for manipulating sequence data[J]. Zoological research, 2017, 38(2): 110–112.
- [36] YANG Z. PAML 4: phylogenetic analysis by maximum likelihood [J]. Molecular biology and evolution, 2007, 24(8):

- 1586–1591.
- [ 37 ] BIELAWSKI J P, YANG Z. Maximum likelihood methods for detecting adaptive evolution after gene duplication[ J ]. Journal of structural and functional genomics, 2003, 3( 1–4 ) : 201–212.
  - [ 38 ] YANG Z, WONG W S W, NIELSEN R. Bayes empirical bayes inference of amino acid sites under positive selection[ J ]. Molecular biology and evolution, 2005, 22( 4 ) : 1107–1118.
  - [ 39 ] WEAVER S, SHANK S D, SPIELMAN S J, et al. Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes[ J ]. Molecular biology and evolution, 2018, 35( 3 ) : 773–777.
  - [ 40 ] WARNER G F. The biology of crabs[ M ]. New York: Van Nostrand Reinhold Company, 1977.
  - [ 41 ] PONTING C P. Novel eIF4G domain homologues linking mRNA translation with nonsense-mediated mRNA decay. [ J ]. Trends in biochemical sciences, 2000, 25( 9 ) : 423–426.
  - [ 42 ] SUZUKI C, GARCES R G, EDMONDS K A, et al. PDCD4 inhibits translation initiation by binding to eIF4A using both its MA3 domains[ J ]. Proceedings of the national academy of sciences of the United States of America, 2008, 105( 9 ) : 3274–3279.
  - [ 43 ] ARAVIND L, KOONIN E V. Eukaryote-specific domains in translation initiation factors; implications for translation regulation and evolution of the translation system[ J ]. Genome research, 2000, 10: 1172–1184.

[ 责任编辑:黄 敏 ]