

秀丽白虾 4 种 C 型凝集素基因的克隆、组织表达与进化分析

王 敏, 吴 萍*, 叶元土, 石桃丹, 魏育红

(苏州大学基础医学与生物科学学院, 江苏 苏州 215123)

摘要: C 型凝集素是一类 Ca^{2+} 依赖活性的糖蛋白, 在虾蟹类的免疫防御中发挥着重要作用。采用 RT-PCR 及 RACE 技术首次克隆了秀丽白虾 C 型凝集素家族 4 个基因 (*EmCTL-1*、*EmCTL-2*、*EmCTL-3*、*EmCTL-4*) 的 cDNA 全长。这 4 个 C 型凝集素基因的 cDNA 全长分别为 1 390、1 119、1 364 和 1 299 bp, 分别含 996、969、1 026 和 1 044 bp 的开放阅读框, 各编码 332、322、341 和 347 个氨基酸的蛋白。其氨基酸序列与罗氏沼虾相应的 C 型凝集素的相似性分别为 40%、37%、69% 和 73%, 与其他海水虾类的相似性在 30%~41% 之间。阳性选择模型分析显示虾类的 C 型凝集素基因在进化过程中经历了阳性选择。进化树分析表明, 秀丽白虾的 C 型凝集素基因与罗氏沼虾的亲缘关系最近, 并且除 *EmCTL-4* 属于十足类 C 型凝集素亚群 A 外, 其他均属于亚群 C。组织表达分析表明, *EmCTL-1* 主要在血液中表达, *EmCTL-2* 主要在性腺、血液和胸神经节中表达, 而 *EmCTL-3* 和 *EmCTL-4* 则主要在精巢中表达。研究表明, 秀丽白虾的 C 型凝集素基因具有较快的进化特点, 同时又具备凝集素行使免疫功能时必需的保守氨基酸残基及钙离子结合位点, 精巢和血液是秀丽白虾 C 型凝集素基因主要的表达器官。

关键词: 秀丽白虾; C 型凝集素; 克隆; 组织分布; 阳性选择

中图分类号: Q 785; S 966. 12

文献标志码: A

凝集素 (lectin) 是指可选择性地识别糖及其缀合物并与其非共价可逆结合的一类非酶、非免疫来源的蛋白质^[1]。自 1988 年首次在蓖麻种子中被发现以来, 已证明它广泛存在于病毒、细菌、酵母、真菌和所有植物及动物细胞的表面^[2]。其中, 动物凝集素可分为 C 型、S 型、L 型、P 型、R 型、I 型等多个家族^[3-4]。C 型凝集素 (C-type lectins, CTLs) 是一类 Ca^{2+} 依赖活性的糖蛋白, 它含有至少一个保守的糖基识别结构域 (CRD), 可专一地与异物表面特定的糖基结合并吸附、凝集异物, 从而促进机体清除异物^[1]。近年来, 随着虾蟹类分子免疫学研究的深入, 已经证明 C 型凝集素在虾蟹类的免疫防御中发挥了重要作用, 包括凝集、吞噬、抗病毒、诱导酚氧化酶系统的激活和激发呼吸爆发等方面^[5], 已在中国明对虾 (*Fenneropenaeus chinensis*)^[6]、凡纳滨对虾 (*Litopenaeus vannamei*)^[7-8]、斑节对虾

(*Penaeus monodon*)^[9]、日本囊对虾 (*Marsupenaeus japonicus*)^[10]、墨吉对虾 (*F. merguensis*)^[11]、罗氏沼虾 (*Macrobrachium rosenbergii*)^[12]、中华绒螯蟹 (*Eriocheir sinensis*) 等经济虾蟹中成功克隆^[13]。

秀丽白虾 (*Exopalaemon modestus*) 隶属于长臂虾科 (Palaemonidae) 白虾属 (*Exopalaemon*), 在我国南北各地的大中型湖泊中分布广泛, 其中以生活在江苏太湖水域的秀丽白虾最为有名, 其色白壳薄、通体透明、味道鲜美, 是闻名遐迩的“太湖三白”之一。近年来, 秀丽白虾的人工繁殖及养殖技术已获突破并逐年推广, 其繁殖生物学研究、种群遗传多样性研究已取得了一定的进展, 但其遗传背景、免疫学、分子生物学等基础研究则相当薄弱。迄今为止, 已报道的秀丽白虾的核酸序列仅有 16SrRNA (GenBank 序列号 EU493144、DQ194971、DQ647672)、28SrRNA (GenBank 序列

收稿日期: 2013-07-12 修回日期: 2013-09-13

资助项目: 国家自然科学基金项目 (31101880/C190101); 苏州市应用基础研究计划 (SYN201103); 2011 苏州大学公派出国访学资助项目

通信作者: 吴 萍, E-mail: suzhou_wuping@163.com

号 EU008718)和细胞色素氧化酶亚基 I (CO I) 基因(GenBank 序列号 AB235307)3 种,而编码秀丽白虾功能蛋白的基因研究几为空白。进行秀丽白虾 C 型凝集素基因的研究可丰富其基因资源,为进一步研究其先天免疫反应机制提供基础资料,也可研究其他经济虾类,特别是长臂虾科种类的抗病机制提供借鉴。

本研究从本实验室秀丽白虾转录组测序的结果中鉴定了秀丽白虾 C 型凝集素基因的部分序列,利用 PCR 和 RACE (rapid amplification of cDNA ends)技术克隆了秀丽白虾 C 型凝集素家族 4 个基因 (*EmCTL-1*、*EmCTL-2*、*EmCTL-3*、*EmCTL-4*)的 cDNA 全长,对其蛋白质结构和可能的功能进行了预测分析,研究了这 4 个基因的组织表达特征和遗传进化,以期秀丽白虾的遗传种质研究及免疫学研究提供依据。

1 材料与方 法

1.1 实验材料

实验虾 秀丽白虾于 2012 年 5 月购自苏州市太湖东山镇,体质量 1.915 ~ 3.723 g,体长 4.3 ~ 6.8 cm。在室内经曝气的自来水中暂养 3 d 后,挑取健康活泼的个体 36 尾,活体解剖取出性腺、鳃、肠、胸神经节、肝胰腺、心脏、肌肉等组织。二乙基焦碳酸酯 (DEPC)处理水配制的磷酸缓冲液 (pH 7.4) 中漂洗,经过液氮冷冻后,放于 -80 °C 保存,用于总 RNA 的提取。血液的采集通过使用已加入灭菌抗凝剂的注射器从已经冰水冷冻麻醉后个体的第二步足基部抽取,血液与抗凝剂最终体积比为 1:1,4 °C 10 000 r/min 离心 10 min 后收集沉淀,进行总 RNA 的提取。

主要试剂 DH5 α 大肠杆菌感受态细胞、琼脂糖购自 Invitrogen 公司;DEPC 购自上海生工生物公司;PrimeScriptTM RTase 逆转录酶、RNAiso Plus 试剂、DNA 聚合酶 I、T₄ DNA 连接酶、RNA 酶抑制剂、dNTP pMD19-T 载体购自 TaKaRa 公司;DNA Marker 购自北京全式金生物技术公司。SMARTTM RACE cDNA 扩增试剂盒购自 Clontech 公司。

1.2 实验方法

总 RNA 的提取和 cDNA 的合成 用 RNAiso Plus 提取各组织的总 RNA,用 DNA 酶 DNase I (RNase Free) 消化后,溶于 20 μ L RNase Free 的水中,经紫外分光光度计检测 RNA 的浓

度、1% 的琼脂糖凝胶电泳检测 RNA 的完整性后,取 3 μ g 总 RNA 按照 PrimeScriptTM 1st strand cDNA Synthesis Kit 说明书合成 cDNA 第一链,-20 °C 保存备用。

秀丽白虾 lectin 基因、 β -actin 基因核心序列克隆 根据本实验室秀丽白虾转录组测序的结果中所鉴定的 lectin 基因、 β -actin 基因 cDNA 序列设计引物 EmCTL-1-F、EmCTL-1-R、EmCTL-2-F、EmCTL-2-R、EmCTL-3-F、EmCTL-3-R、EmCTL-4-F、EmCTL-4-R、Em-ActinF 和 Em-ActinR (表 1)。以秀丽白虾卵巢 cDNA 为模板,25 μ L 反应体系含如下成分:10 \times PCR buffer 2.5 μ L、25 mmol/L 的 MgCl₂ 2 μ L、10 mmol/L 的 dNTP 0.5 μ L、上下游引物各 1 μ L、rTaq 酶 0.3 μ L、无菌水 17.7 μ L。PCR 扩增条件为 94 °C 4 min;94 °C 1 min、58 °C 1 min、72 °C 1 min,共 30 个循环,72 °C 延伸 7 min。PCR 产物经 1.0% 琼脂糖凝胶电泳、割胶、回收纯化后克隆至 pMD19-T 载体,转化到感受态 DH5 α ,进行蓝白斑筛选,菌液 PCR 检测阳性克隆送上海英俊生物技术有限公司进行测序,所得的 cDNA 片段与 GenBank 数据库中的已知序列进行 BLAST 同源性分析。

lectin 基因全长 cDNA 序列的获得 采用 SMARTTM RACE cDNA 扩增试剂盒分别扩增秀丽白虾 lectin 基因的 3'末端和 5'末端。按说明书在 0.2 mL 离心管中依次加入总 RNA 1 μ g、5'-CDS primer 1 μ L、SMART II A oligo 1 μ L、5 \times First-Strand buffer 2 μ L、20 mmol/L DTT 1 μ L、10 mmol/L dNTP Mix 1 μ L、MMLV Reverse Transcriptase 1 μ L 用于合成 5' -RACE-Ready cDNA;加入总 RNA 1 μ g、3'-CDS primer 1 μ L、sterile H₂O 1 μ L、5 \times First-Strand buffer 2 μ L、20 mmol/L DTT 1 μ L、10 mmol/L dNTP Mix 1 μ L、MMLV Reverse Transcriptase 1 μ L 用于合成 3' -RACE-Ready cDNA。根据上述克隆测序已获得的秀丽白虾 *EmCTL-1*、*EmCTL-2*、*EmCTL-3*、*EmCTL-4* 基因的 cDNA 片段序列,分别设计 3'和 5'RACE 引物(表 1)。分别以 5'-RACE-Ready cDNA 和 3' -RACE-Ready cDNA 为模板,采用巢式 PCR 进行 2 轮扩增,以获得目的产物。首轮 PCR 反应体系组成为:RACE-Ready cDNA 2.5 μ L、10 \times Advantage 2 PCR buffer 5 μ L、10 mol/L dNTP Mix 1 μ L、10 \times UPM 5 μ L、10 μ mol/L 末端特异引物 1 μ L、50 \times

Advantage 2 Polymerase Mix 1 μL 和 PCR-Grade Water 34.5 μL 。PCR 反应条件为 94 $^{\circ}\text{C}$ 4 min; 94 $^{\circ}\text{C}$ 1 min、68 $^{\circ}\text{C}$ 30 s、72 $^{\circ}\text{C}$ 3 min 共 25 个循环; 72 $^{\circ}\text{C}$ 延伸 10 min。用 1.2% 琼脂糖凝胶电泳检测 PCR 产物, 将 PCR 产物稀释 50 倍后进行巢式 PCR 反应, 反应体系组成为: 稀释后的首轮 PCR 产物 5 μL 、10 \times Advantage 2 PCR buffer 5 μL 、10 mmol/L

dNTP Mix 1 μL 、10 $\mu\text{mol/L}$ NUP 1 μL 、10 $\mu\text{mol/L}$ 末端特异引物 1 μL 、50 \times Advantage 2 Polymerase Mix 1 μL 、PCR-Grade Water 36 μL 。巢式 PCR 反应条件为 94 $^{\circ}\text{C}$ 4 min; 94 $^{\circ}\text{C}$ 1 min、68 $^{\circ}\text{C}$ 30 s、72 $^{\circ}\text{C}$ 3 min 共 20 个循环; 72 $^{\circ}\text{C}$ 延伸 10 min。RACE 产物纯化、克隆和测序方法同上。

表 1 引物名称及序列

Tab. 1 Parameters of primer pairs used in this study

引物名称 primer	引物序列(5'-3') primer sequence(5'-3')
EmCTL-1-F	GCATAGTTGCTGCTGGGACA
EmCTL-1-R	CTCGGTTGGTAAATCTTAGTAGTG
EmCTL-2-F	ACTTGGGACAGCGTTTCGTC
EmCTL-2-R	GTGGAAATAGAAGAAATCAACCG
EmCTL-3-F	GTGCGGCGTCAGTTACTC
EmCTL-3-R	CTACAACATCGTATTAGATAGATTA
EmCTL-4-F	AAGAACCCACGCAGAACG
EmCTL-4-R	CTGGAGGATTTAGCATAGTTTA
Em-ActinF	TGGTGGTCATGGGTGAGAGG
Em-ActinR	TCGCTAGGCGGTGCTCGT
EmCTL-1-3outer	AAGGGAGGGAAGTTGGGTTTGGA
EmCTL-1-3inner	TCTTCAGGAGGGCACTTGGCTTTG
EmCTL-1-5outer	CCAAAGCCAAGTGCCCTCCTGAA
EmCTL-1-5inner	CATCTGGTACGGGACTGGGCATT
EmCTL-2-3outer	ATGGATGGGCGGCTCTGACCTTA
EmCTL-2-3inner	CTGAAGCCCAGCCAGAAGAGGAA
EmCTL-2-5outer	TCAGAGCCGCCATCCATAGGTC
EmCTL-2-5inner	TCGGTTCCTGGTGCCCTGATTGT
EmCTL-3-3outer	GACGACCACCAAGAAGGGCAATG
EmCTL-3-3inner	CTGGCACGACGGAAGCAGGGTAA
EmCTL-3-5outer	CATCCTTACCCTGCTCCGTCGTG
EmCTL-3-5inner	CATTGCCCTTCTTGGTGGTCGTC
EmCTL-4-3outer	AGTTGCCCCCATTTTCGTTGA
EmCTL-4-3inner	CACCGTTTGGGGGAACAGTATCA
EmCTL-4-5outer	CGAGAAATGCTGTTCGCAGTCCA
EmCTL-4-5inner	CGCCCTCAAGAGAATGGCACACC
SMART II A Oligonucleotide	AAGCAGTGGTATCAACGCAGAGTACGCGGGXXXXX
3'-RACE CDS Primer A	AAGCAGTGGTATCAACGCAGAGTAC(T)30VN(N = A, C, G, or T; V = A, G, or C)
5'-RACE CDS Primer	(T) ₂₅ VN(N = A, C, G, or T; V = A, G, or C)
UPM(Long)	CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT
UPM(Short)	CTAATACGACTCACTATAGGGC
NUP(Nested Universal Primer A)	AAGCAGTGGTATCAACGCAGAGT

注: X 代表 SMART 引物序列中特有的不明碱基

Notes: X means undisclosed base in the proprietary SMARTer oligo sequence

lectin 基因的同源性比对、全长 cDNA 序列特征分析及进化树构建 通过 Vector NTI Suite 软件对获得的 cDNA 序列进行拼接。利用 NCBI 的 ORF Finder 程序对所得序列作开放阅读框分析 (<http://www.ncbi.nlm.nih.gov/projects/gorf/orfig.cgi>), 预测编码氨基酸序列。信号肽和结构

域分析分别采用 SignalP 3.0 程序 (<http://www.cbs.dtu.dk/services/SignalP/>) 和 Motif-Scan 程序 (<http://hits.isb-sib.ch/cgi-bin/PFSCAN>)。利用 SWISS MODEL 蛋白质结构同源建模工具 (<http://www.expasy.ch/swissmod/SWISS-MODEL.html>), 预测所得序列的蛋白三维结构, 并用

PyMOL 软件观察分析。秀丽白虾 C 型 Lectin 与其他物种氨基酸序列的比较采用 Clustal W 多序列比较软件 (<http://www.ebi.ac.uk/Tools/clustalw2/index.html>) 和 BioEdit 进行分析;用 MEGA 5.0 软件以邻接法 (neighbour-joining, NJ) 构建系统进化树,用自展法 (bootstrap) 进行 1 000 次重复检验。

分子进化分析 选取含有 2 个 CRD 结构域的 lectin 基因,包括中国明对虾 (EU834290.1)、凡纳滨对虾 (DQ871245.1)、墨吉对虾 (FJ751773.1)、斑节对虾 (DQ871244.1)、罗氏沼虾 (JQ349148 和 JQ349150) 以及本研究获得的秀丽白虾的 4 个 lectin 基因序列,根据序列对位排列结果,以密码子置换模式 M0、M3、M1a、M2a、M7 和 M8 来计算每个位点上的 ω 值 ($\omega = dN/dS$)^[14]。采用 M0、M3、M1a、M2a、和 M7、M8 模型进行成对比较,用 PAML 4.0 软件包中的 CODEML 程序对 Lectin 蛋白编码序列检测阳性选择位点^[15]。用似然法检验 (LRT) 检测模型间差异是否有统计学意义^[16], 所得结果用 χ^2 进行显著性检验。

不同组织中 lectin 基因的 mRNA 表达 通过上述方法获得的不同组织的 cDNA 模板,以 EmCTL-1-F/R、EmCTL-2-F/R、EmCTL-3-F/R、EmCTL-4-F/R 为 RT-PCR 特异引物,以本研究获

得的秀丽白虾的 β -actin 基因 (GenBank 注册号: KC568294) 序列设计内参引物 Em-Actin-RTF (5'-CTGGCACCACTATCTACATCG-3') 和 Em-Actin-RTR (5'-TGATGTGGGTCATCATTTTCGC-3')。PCR 反应体系同上,PCR 扩增条件为 94 °C 4 min; 94 °C 1 min, 58 °C 1 min, 72 °C 1 min, 共 25 个循环, 72 °C 延伸 7 min, 反应重复 3 次。

2 结果

2.1 秀丽白虾 4 种 C 型凝集素基因的克隆

以秀丽白虾卵巢 cDNA 为模板,分别以 ECTL-1-F、EmCTL-1-R; EmCTL-2-F、EmCTL-2-R; EmCTL-3-F、EmCTL-3-R 和 EmCTL-4-F、EmCTL-4-R 为引物,在 1 116 bp、768 bp、602 bp 及 1 022 bp 处各获得一特异性条带,克隆测序后,结果经 BlastX 分析确定为秀丽白虾 C 型凝集素基因 lectin1、lectin2、lectin3 和 lectin4 目的片段,分别命名为 Em-CTL1、Em-CTL2、Em-CTL3 和 Em-CTL4。3' 端和 5' 端分别经巢式 PCR 扩增得到特异性条带,测序结果经分析拼接后,获得此 4 个基因的全长 cDNA 序列 (图 1)。Em-CTL1、Em-CTL2、Em-CTL3 和 Em-CTL4 的 cDNA 全长分别为 1 390、1 119、1 364 和 1 299 bp。

```

A 1   acatgggatagaaacggtgaaacgccagaggatgggagcatctcttctttgtacaggtagacctgtttcactccaccatcaca
88  atgattttcaaatccaatttcataacttccttcactattcataggtatgtctgtgtggacaatcgacacccaactaaggaaacctat
    M I F K S N F I L P S L F I G I V A A G T I D T N T K E T Y 30
178  gaatgtocatctccttaccagcctgttggccagacaaattactgcttactcatcgacgatogtatcgaggccacatgggatgacatgaga
    E C P S P Y Q P V G P D N Y C L L I D D R I G G T W D D M R 60
268  aaattctgcaaggaaatggaaatggcgaacttggccttcttccccgacgcaaacctcctttacgagggtgagtgagcatcaacaagaacgga
    K F C K E W N G D L A F F P D A N L L Y E V V Q Y I N K N G 90
358  ctgaacgactctagccttctggattggagcagaggatattgaaagggagggaaagttgggtttggaccaataaccacaagacgtccgaatg
    L N D S S F W I G A E D I E R E G S W V W T N N H K D V R M 120
448  cacactccccctctggcctccttcccgatatgaacagggacccaacagaggactcagtcagaactgtgcatccctcaacaagataaatat
    H T P L W A P S R Y E Q E P T G G L S Q N C A S L N K D K Y 150
538  ttcttctgtagatgatggctactgtcaagaacogcttcggagttatttgcgatccactaattcaacgcaagaaaaatcaagtgaataacct
    F F V D D G Y C Q D R F G V I C D P L I H A K K K S S E I P 180
628  gacgaggtcgtcgaagaggagattccagtagcaggagaatgccagtcocogtaccagatgatcggtgacggctgcttcttattgatgtc
    D E V V E E E I P V A G E C P V P Y Q M I G D G C F L I D V 210
718  gtcaagaccaataacctggtatgactcgaataatctatgcaagctcttggaggaaactggctaaaaatgatgacccaatattattggga
    V K T N T W Y D S K N L C E A L G G K L A K I D D A N L L G 240
808  gatctttatgaatacttctcaaataggagatccatgtttctttatgtagcgtggttccagatgatcttcaggagggcacttggctttgg
    D L Y E Y F S N R S I H V S L W I G G S D D L Q E G T W L W 270
898  gaagacatgtcactcgtcaaaatgggaacacccttctgggcttcaacatgtacgaggaacaagaacggatggccataacaggagaac
    E D M S L V K M G T P F W G F N M Y E E Q [E P D] G H N R E N 300
988  tgcctttcgtcatagcacaattctttcttcaatgatcttcgtgtaacaacctagctggtgttgcatgctcagtatgatatgagg
    C L S L H S D N Y F F [F N D] L P C N N L A G V A C Q Y D M R 330
1078 ctatgatccctcgtttttatatacagtgataaagatcttgtaaatgaagaacatcgctcttatttagttatgacctgcttcttctcgtagt
    L * 331
1168 atcaattattcttatgagcttccaaagcatactgcagttcaatcaaatatcatatcactactaagatttaaccaaccgagacaggatttat
1258 tttatttcatacaaaatttattatataaggcttcacaatgttggtcttgttatactggttatacctgtaatacattataatatttaata
1348 caaacatgcccattaaaaaaaaaaaaaaaaaaaaaaaaa

```

B 1 acatggggacattatcctcaagcatatagtcaagtgctgctgctgacagcgttaacaatt
 63 atggagctgtccagggttacgttgctctttcctgacctctcctctgggttggccgttgacagaaaatggatagcaatattgtttgc
M E L S R V T L L L F L T S P L G L A V A E N V D S N I V C 30
 153 gatccaccttccagatattggacgtcattgcttggtgattgacacgtccatcatggacttggacagcgttctgctcagttctccgc
D P P F R D I G R H C L V I D T S I I G T W D S V R Q F C R 60
 243 aacctggaggagaattagtgaaatgcaacatgcaacatcctccatgacattttcagcttcatcaaggaatcctattctagactcc
N L G G E L V N M H D A N I L H D I F S F I K E N P I L D S 90
 333 gagagctactggattggacatacgacatactagatgaggcaaatgggttggctttacaacaacacctggttogaatgaacctct
E S Y W I G A Y D I L D E G K W V W S Y N N T H V R M N T P 120
 423 ctatggggcgtacaatcagggaaccaggaaccgagggcgtctacgaaactgtccatgcttaacaagattatcatattacgtc
L W G V Q S G H Q E P T G G V Y E N C A M L N K D Y H Y Y V 150
 513 tttgatgacaactgtgatcaacaaatgcccgttttttgccttcgacaaaaaatggctgaagcccagccagaggaagagaaagat
F D D N C D Q P N A V I C A S T K K M A E A Q P E E E E K D 180
 603 aatccccaaagacagagtgcccttcgacctctcatcatgacgtatgtatgtcttggctgacgcactgaaagaaaccccttggtct
N T P R R E C P S P F I M I G N V C L L V D A L K E N P W S 210
 693 gaacaacagatgatctgacgtcccttggaggtcacatggcaagaatagaggatgccaacgacttggcgatctctacgactctcata
E Q Q M I C S A L G G H M A R I E D A N D L G D L Y D Y L I 240
 783 gcaaacgacgtcagaatgacctatggatggcgctctgacctagaatggaagcattgggttgggaaacaacgcaaggtcccc
A N D V R M D L W M G G S D L R M E D D W V W E N N A K V P 270
 873 atgggcacgccattttgggtcaagaactcgcgcatgcaggagcctgcaggagcctctatgagaactccctctccctctctcggtt
M G T P F W G Q E T R G M Q E P A G G L Y E N C L S L S S V 300
 963 gatttctctatttccacgatgacagtgctcagcgaatgaaagcgttggttgocagatgacatttaattctatgattctatcat
D F F Y F H D D E C Q R M K G V V C Q Y D I * 322
 1053 aataatagataaattgttaccataaataaaggaggaaattaaactgaaaaaaaaaaaaaaaaaaaaa

C 1 acatgggggtcacaacaatta
 22 ggtacttggaggaaactacagattcaactttctagcttttaacaggcaccctctttaatttggcacttggattttgaatcaacgaaa
M K S I V V L F C L F L G I C T G T W A K A A S K S D D Y P 30
 202 tataccggctattgggattcgaccacaacacctggatcaagatgcacaggcgtctcgaagaggctggaagccattgtattttctagac
Y T C Y W D S T T T P G S R C T G V F E E V G S H C I F V D 60
 292 aaagatctctctggacctggcaggacatgaggatgtctgccagctccttggggagatctagtaaacctacaggatgcaaatttttg
K D L S G T W Q D M R D V C H V L G G D L V N L Q D A N F L 90
 382 gcagatataaagattacttacatatttatggccataccgctgacttactgggtgggaagcaacgaccagtgacggaggagctctgg
A D I K D Y L H I Y G H T G V T Y W V G S N D Q V T E G V W 120
 472 cagtgattaaacaccattcgcaagtgaagtaggaacgcccttggggcgaacctccccgacaggagacaacggccagacaggaggtct
Q W I N D H S Q V K V G T P L W G D L P D R R Q R P T G G S 150
 562 gcgaaaactgcgtatactttattatgaagatcaactattctatgacgatactcagtgccgctcagttactcattatttgcgaaca
A E N C G I L Y Y E D H Y F M H D T Q C G V S Y S L I C E Q 180
 652 tacaattccaaggaatagaagttacgagctaacgccagtaggacggatattgtgcaagaagatttccaataactgccatccccg
Y N F Q G I E V T R A N A S R T D I V Q E D S P N N C P S P 210
 742 tacgtatctgtagctggaagatgcatcttcatcaacaacctggctaaaaattcatggcagctcacaagacctctgagagcctcgga
Y V I V A G R C I F I N T L A K N S W Q S H K T S C E S L G 240
 832 ggttacatgggtcaagctggatgacttaactctcggcgtattttgactatctgacacagtcaggcataagccatgagttctggatc
G Y M V K L D D A N F L G A I F D Y L T Q S G I S H E F W I 270
 922 ggocctctgacgaccacaagaaggcaatggtctggcagcaggaagcagggttaagatggcactccctctggggagacaacggc
G A S D D H Q E G Q W F W H D G S R V R M G T P F W G D N G 300
 1012 aatatcactcagaatcctgacggaggatcacgaaaattgggtatgatcaactacagatcattactctctgatgatgaggaatgc
N I T Q N P D G G V T E N C G M I N Y D D H Y F F D D E E C 330
 1102 agtccagtcattacgtctttgcgaaacacctagtagccagtgccagagaagctaatatcacaggataaaaagctctgtaaatgttt
S A S H Y V I C E T T * 341
 1192 gtttaactatcctaactagatgtttagagacttctctatagattctgtattattcatcggtatataaagatctgtttgtaagatt
 1282 caagaaaattggatttttcaatagttgttaataaattattgaaatataaaaaaaaaaaaaaaaaaaaaaaaaa

D 1 acatggggcattaccattgtataactttcaagaacagcaaaatcctgaccggccaacagca
65 atggggatcaaatcctgtaccttctactgacaatgacgggggacatcaaggcttccctcagtcattatgaagaaattatcaca
M G Y Q I L Y L L L T I A V V T S K A S P Q S Y E E I I S Q 30
155 attatgaaggagctgcagaagaagaacccacgacagcgaacacgacctgtcctccgatgttcagttgatagaggcaaatgtctttac
I M K E S A E E E P T Q N G T A C P P M F S L I G G K C L Y 60
245 ttgcctcctttgtcaacaaacaaagttgcctgcaaggcaggtgtccatctcttgaggcgcaatggcagccatccgtactgcgact
F A S F V Q Q N Q V A C R Q V C H S L E G E L A A I R T A T 90
335 gaactccaggaataatacagagatataaaggagagactcgtcgaatggcctttggatagacggaaacctttgatggcaagaacat
E L Q E I I R E I K R R G L V E M G F W I D G T F D G K E H 120
425 tggaaacttctcaaatacagacctagttocactgggaacgaccttctggtagcagccgaagataataacccaagtcogacaatgatccc
W N F S N T D L V P L G T P F W Y A A E D N N P S P T N D P 150
515 ctacttaaatgtcaagcatataatagctccgatcttaataaagacatggactcgaacagcatttctgcogctttgtgaacat
L L K C A S I N N E S G F L M K D M D C E Q H F S P L C E H 180
605 gaccggctagtcttgacgaaagcgaagaagatcaagaagaggaaacggacatctggaggctgagaacattagttgccccatttttc
E P A S L D E S E E D Q E E E P D H L E A E N I S C P P F F 210
695 gttgacatcgggcaattgtatagcattcataacttggctgaaatccctggattgcagcaagcagcgtgcaatgccatcactagt
V D I S G N C I A F I T W S E I P W I A A K Q T C N A I T S 240
785 ggcgcagagcttctaatgactaaagacatagactactgcgacatctttacaactattttcagaggaaaggactagatggacataatttc
G A E L L M T K D I E L L R H L Y N Y F H E E G L D G H N F 270
875 tggataggtgccacagatgtgtcgaagaaggcgtatggacatacaatggataaaaagtagatagaggaccgttttgggggaacagt
W I G A T D V A E E G V W T Y N G I K V D I E A P F W G N S 300
965 atcactctaggtgtcctgagcctgagcaatgatgctgaaatgtctgctgctgacatccgaggccaacactacttcagggaacaaa
I T L G V M E P D G N D A E N C L A L T S A G Q H Y F R D K 330
1055 agttgcgatgaaccttacaacccaatgatgcacagcaaacggctgtctttgtagaacaataaacggagtcoccttctgttcattgtttctg
S C D E P Y N P I C T A K P L S L * 347
1145 gtacgtggtagctgtacacgttttatcagagttatctacctgttaataataaactatgctaaatcctccagcattgtacagtcaggga
1235 aacctctcatttgataataacaaataataaaacaaaaaataaaacaaaaa

图 1 秀丽白虾 4 种 C 型凝集素基因 cDNA 及推导的氨基酸序列

CRD 结构域用下划线表示,信号肽序列用双下划线表示,加尾信号用阴影表示,Ca²⁺结合域 2 用方框表示。A: EmCTL-1; B: EmCTL-2; C: EmCTL-3; D: EmCTL-4

Fig. 1 Nucleotides and deduced amino acids sequence of 4 C-lectin full-length cDNA in *E. modestus*

Single underlines indicate the CRD domains, the cleavage site of the putative signal peptide are double underlined, shadows indicate the poly(A) signals, the motif of Ca²⁺ binding sites 2 are boxed. A: EmCTL-1; B: EmCTL-2; C: EmCTL-3; D: EmCTL-4

表 2 秀丽白虾 C 型 lectin 基因的核苷酸序列特征

Tab. 2 Characteristics of nucleotide sequences of *Em-CTL1*, *Em-CTL2*, *Em-CTL3* and *Em-CTL4* in *E. modestus*

	<i>Em-CTL1</i>	<i>Em-CTL2</i>	<i>Em-CTL3</i>	<i>Em-CTL4</i>
GenBank 注册号 accession number	KC568290	KC568291	KC568292	KC568293
cDNA 全长/bp cDNA full length	1 390	1 119	1 364	1 299
开放阅读框/bp ORF	996	969	1 026	1 044
编码氨基酸数 number of amino acids	332	322	341	347
预测蛋白分子量/ku molecular weight	37.78	36.68	38.31	38.82
预测蛋白等电点 isoelectric point	4.23	4.21	4.38	4.20
5'非编码区/bp 5'UTR	87	62	111	64
3'非编码区/bp 3'UTR	307	88	227	191
poly(A)长度/bp poly(A) length	28	22	31	29

2.2 秀丽白虾 4 种 C 型凝集素基因的全长序列特征分析

秀丽白虾 4 种 C 型凝集素基因 *Em-CTL1*、*Em-CTL2*、*Em-CTL3* 和 *Em-CTL4* 的核苷酸序列特征如表 2 所示。分别编码 332、322、341 和 347 个氨基酸的蛋白,在 N 端分别含有 19、21、20 和 17 个氨基酸的信号肽序列。氨基酸序列用 SMART 在线软件分析,结果发现 *Em-CTL1*、*Em-CTL2*、

Em-CTL3 和 *Em-CTL4* 均含有 2 个 CRD 模块(图 3-a),并且 2 个 CRD 都有 6 个保守的半胱氨酸残基(图 3-b)。*Em-CTL2* 和 *Em-CTL3* 的 CRD1 由 128 个氨基酸残基组成,*Em-CTL1* 和 *Em-CTL4* 的 CRD1 氨基酸数目分别为 127 和 126;*Em-CTL2* 和 *Em-CTL3* 的 CRD1 模式为 C-X10-C-X17-C-X79-C-X15-C-X7-C,*Em-CTL1* 和 *Em-CTL4* 则分别为 C-X12-C-X17-C-X78-C-X15-C-X7-C 和 C-X10-C-

X17-C-X77-C-X15-C-X7-C。*Em-CTL1* 和 *Em-CTL3* 的 CRD2 由 127 个氨基酸残基组成, *Em-CTL2* 和 *Em-CTL4* 的 CRD2 氨基酸数目分别为 129 和 130; 同时 *Em-CTL1*、*Em-CTL2* 和 *Em-CTL3* 的 CRD2 模式均为 C-X10-C-X17-C-X77-C-X15-C-X7-C; 而 *Em-CTL4* 则为 C-X10-C-X17-C-X80-

C-X15-C-X7-C。

由图 2 可见, 秀丽白虾 4 种 C 型凝集素的 Ca^{2+} 结合域 2 各不相同, *Em-CTL1* 的 Ca^{2+} 结合域 2 为氨基酸残基 EPD 和 FND, 而 *Em-CTL2*、*Em-CTL3* 和 *Em-CTL4* 的此结构则分别为 EPA 和 FHD、WHD 和 FDD、EPD 和 FRD(图 2)。

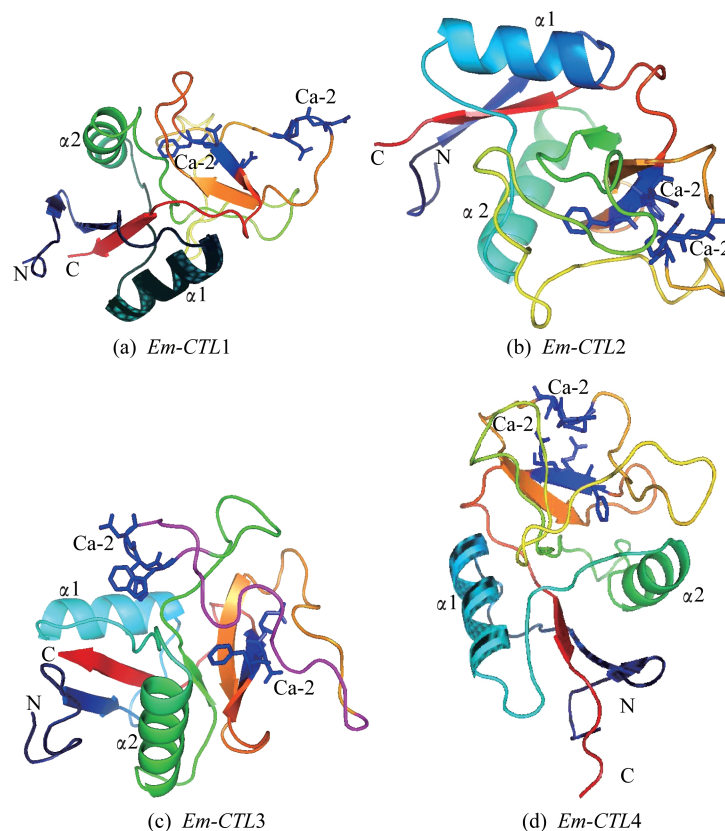


图 2 秀丽白虾 4 种 C-型凝集素的三维结构模型

α 指示 α 螺旋, Ca-2 指示 Ca^{2+} 结合域 2

Fig. 2 3D structure models of C-Lectin in *E. modestus*

α indicates alpha helix and Ca-2 indicates Ca^{2+} binding sites 2

以秀丽白虾 *Em-CTL1*、*Em-CTL2*、*Em-CTL3* 和 *Em-CTL4* 基因编码的氨基酸序列与 GenBank 数据库中的已知序列作 BlastP 分析, 结果显示 *Em-CTL1* 与罗氏沼虾的 Lectin1 的相似性为 40%, 与斑节对虾、凡纳滨对虾、墨吉对虾和中国明对虾的 Lectin 的相似性分别为 31%、31%、32% 和 30%; *Em-CTL2* 与罗氏沼虾的 Lectin2 的相似性为 73%, 与中国明对虾、墨吉对虾、凡纳滨对虾和日本囊对虾的相似性分别为 33%、30%、32% 和 38%; *Em-CTL3* 与罗氏沼虾的 Lectin3 的相似性为 69%, 与墨吉对虾、印度明对虾 (*Fenneropenaeus indicus*)、日本囊对虾和

中国明对虾的相似性分别为 30%、41%、35% 和 31%; *Em-CTL4* 与罗氏沼虾的 Lectin4 的相似性为 37%, 与凡纳滨对虾、中国明对虾、短沟对虾 (*Penaeus semisulcatus*)、斑节对虾和墨吉对虾的相似性分别为 39%、37%、37%、36% 和 36%。与罗氏沼虾、凡纳滨对虾、中国明对虾、斑节对虾等 C-Lectin 的多重序列比对发现, Lectin 蛋白在秀丽白虾和其他虾类间在 CRD 结构域上具有一定的保守性, 含有高度保守的半胱氨酸(C)、色氨酸(W)、天冬氨酸(D)等残基(图 3)。

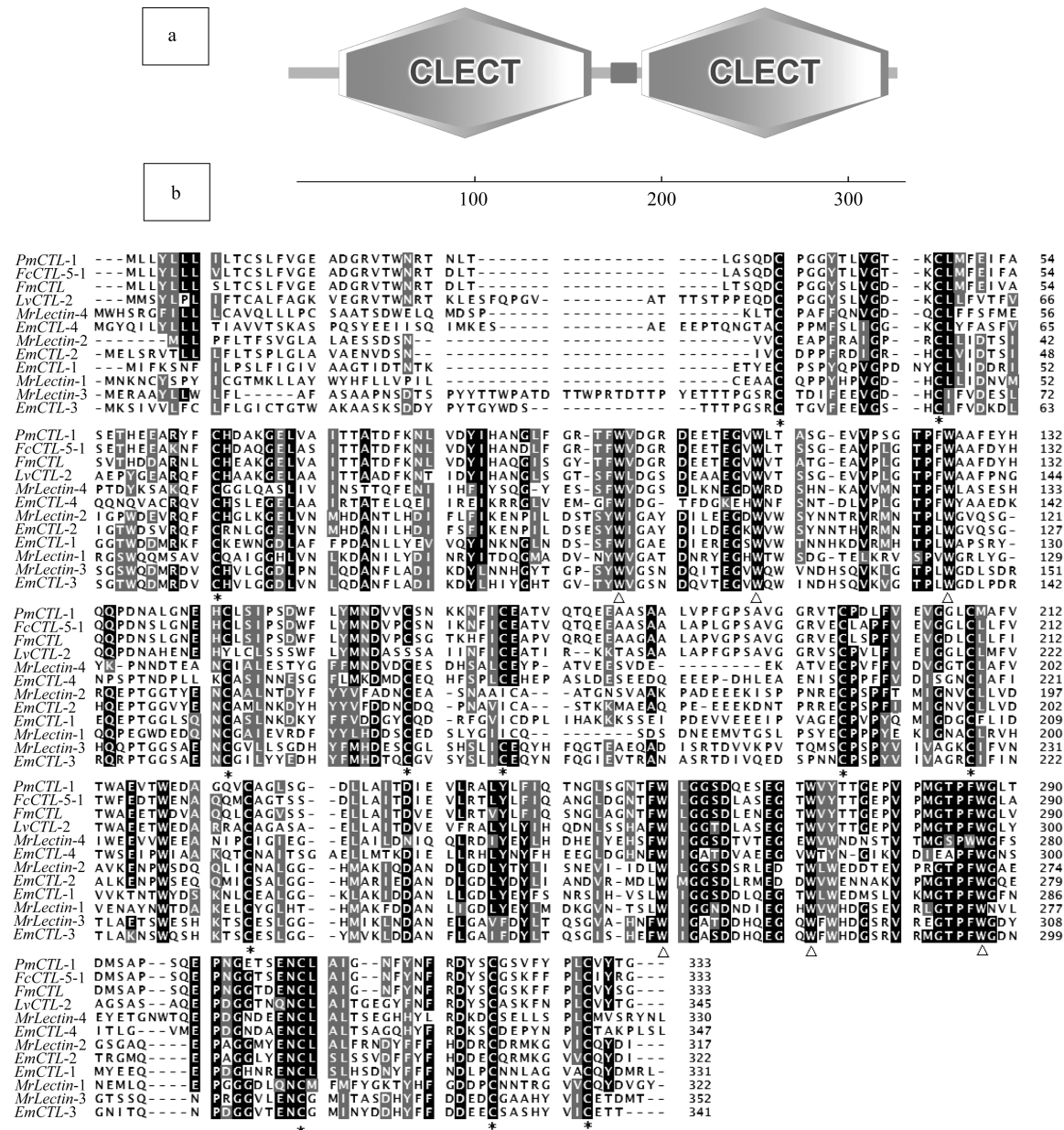


图 3 秀丽白虾 C-lectin 基因编码序列的结构及其氨基酸序列比对

a. C 型凝集素基因的 2 个 CRD 模块; b. C 型凝集素氨基酸序列的多重比对

阴影表示相同和相似的氨基酸,星号表示 2 个 CRD 结构域中保守的半胱氨酸残基,△表示 2 个 CRD 结构域中保守的色氨酸残基

Fig. 3 The structure and alignment of the deduced amino acid sequences of C-Lectin in E. modestus

a. CRD1 and CRD2 domains of C-Lectin in E. modestus; b. Alignment of the deduced amino acid sequence of C-Lectin

Shaded letters represent conserved amino acid in these proteins. Conserved cysteine residues in CRD1 and CRD2 were indicated with asterisk. Conserved tryptophane residues in CRD1 and CRD2 were indicated with triangle. MrLectin-1, MrLectin-2, MrLectin-3, MrLectin-4 (Macrobrachium rosenbergii, AFN20597. 1; AFN20598. 1; AFN20599; AFN20600. 1); FcCTL-5-1 (Fenneropenaeus chinensis, EU834289. 1); FmCTL (F. merguensis, FJ751773. 1); PmCTL-1 (Penaeus monodon, DQ871244. 1); LvCTL-2 (Litopenaeus vannamei, DQ871245. 1)

2.3 秀丽白虾 C 型凝集素基因的进化分析

阳性选择分析 经过序列比对,密码子校正后,用 PAML4.0 分析了含 2 个 CRD 结构域的虾类 lectin 基因的阳性选择情况,结果显示,结构中具有 2 个 CRD 结构域的虾类 lectin 基因在进化过程中

经历了阳性选择。经过 LRT 检验,M2a、M3 和 M8 模型比 M1a、M0 和 M7 模型更适合 lectin 基因。在模型 M2a 中,有 3.8% (ω = 15.35) 的位点属于阳性选择位点;在模型 M8 中,有 5.6% (ω = 6.32) 的位点属于阳性选择位点。

系统进化分析 分别选择了不同十足类 C-Lectin 的典型蛋白以 NJ 法构建了系统进化树,结果如图 4 所示。秀丽白虾的 *EmCTL-2*、*EmCTL-3*、*EmCTL-4* 首先分别与罗氏沼虾的 *Lectin-2*、*Lectin-3*、*Lectin-4* 聚类,再依次与其他虾类聚类。不同十足类的 C-Lectin 可分成 4 个亚群(A、B、C、D),在已克隆的秀丽白虾 4 种 C 型凝集素中,*EmCTL-1*、*EmCTL-2*、*EmCTL-3* 都属于亚群 C,*EmCTL-4* 则属于亚群 A。由图 4 可见,亚群 A 中,除斑节对虾的凝集素 1 (*PmCTL-1*) 仅含有一个 CRD 结构域外,其他虾类都含有 2 个 CRD 结构域;亚群 B 中,除中国明对虾 (*FcCTL*) 外,其他

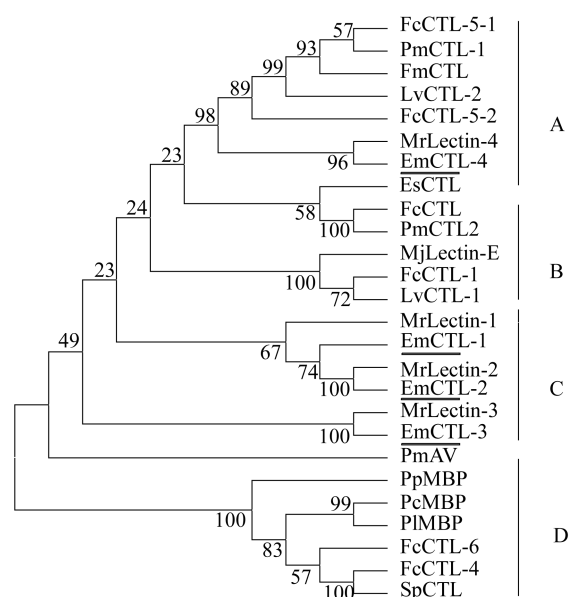


图 4 不同十足类 C-Lectin 的 NJ 系统进化树
秀丽白虾的 4 种 C-Lectin 用下划线标出

Fig. 4 Phylogenetic tree constructed from the amino acid sequence of shrimp C-Lectin by neighbour-joining

Lectins from *E. modestus* are marked in underline *FcCTL-5-1*, *FcCTL-5-2*, *FcCTL-1*, *FcCTL*, *FcCTL-4*, *FcCTL-6* (*Fenneropenaeus chinensis*, EU834289. 1; EU834290. 1; DQ167572. 1; AY871270. 1; EU834293. 1; EU834291. 1), *PmCTL-1*, *PmAV*, *PmCTL-2* (*Penaeus monodon*, DQ871244. 1, AY302750. 1, DQ078266. 1), *MrLectin-1*, *MrLectin-2*, *MrLectin-3*, *MrLectin-4* (*Macrobrachium rosenbergii*, AFN20597. 1; AFN20598. 1; AFN20599; AFN20600. 1), *FmCTL* (*F. merguensis*, FJ751773. 1), *LvCTL-2*, *LvCTL-1* (*Litopenaeus vannamei*, DQ871245. 1; EF583939. 1), *MjLectin-E* (*Marsupenaeus japonicus*), *PcMBP* (*Procambarus clarkii*, FJ410911. 1), *PIMBP* (*Pacifastacus leniusculus*, AY861653. 1), *EsCTL* (*Eriocheir sinensis*, AGC31447. 1), *SpCTL* (*Scylla paramamosain*, AGC96521. 1), *PpMBP* (*Portunus pelagicus*, ABM65756. 1)

虾蟹类都仅含有 1 个 CRD 结构域;亚群 C 中所有种类都具有 2 个 CRD 结构域,而亚群 D 中的所有种类,其凝集素蛋白都仅具有一个 CRD 结构域。

2.4 秀丽白虾 4 种 C 型凝集素基因的组织分布

以 β -actin 为内参,半定量 RT-PCR 结果显示, *EmCTL-1* 主要在秀丽白虾的血液、精巢和卵巢中表达,其中血液中的表达量最高,在胸神经节及肝胰腺中表达量相对较低,在心脏中微弱表达,在鳃和肌肉中不表达。 *EmCTL-2* 主要在卵巢、精巢、血液、胸神经节和心脏中表达,在肝胰腺中表达量相对较低,在肌肉和鳃中不表达。 *EmCTL-3* 主要在精巢、卵巢、肝胰腺、血液、心脏和胸神经节中表达,其中在精巢中的表达量最高。 *EmCTL-4* 主要在精巢、卵巢、心脏、胸神经节和肝胰腺中表达,在血液中表达量相对较低,在精巢中表达量最高(图 5)。

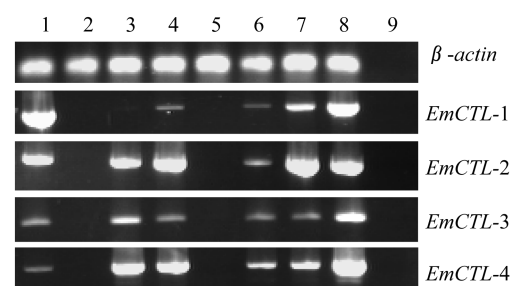


图 5 4 种 C-lectin mRNA 在秀丽白虾不同组织中的差异性表达

1. 血液; 2. 肌肉; 3. 心脏; 4. 胸神经节; 5. 鳃; 6. 肝胰腺; 7. 卵巢; 8. 精巢; 9. 阴性对照

Fig. 5 The significance expression of 4 C-lectin mRNA in different tissues of *E. modestus* by RT-PCR

1. hemolymph; 2. muscle; 3. heart; 4. thoracic ganglion; 5. gill; 6. hepatopancreas; 7. ovary; 8. testis; 9. a negative control

3 讨论

3.1 秀丽白虾 C 型凝集素基因的结构特征

C 型凝集素代表一个依赖于钙离子参与才能识别碳水化合物配体的糖原结合蛋白家族,在脊椎动物中高度保守,而在无脊椎动物中则表现出一定的多样性^[5]。仅虾类而言,在日本囊对虾已报道的 *C-lectin* 中,均只包含一个 CRD 结构域,罗氏沼虾已报道的 4 个 *C-lectin* 基因,都具有 2 个 CRD 结构域^[12],而中国明对虾和凡纳滨对虾,则存在 *C-lectin* 基因既有包含 1 个 CRD 的,又有含 2 个 CRD 的情况^[17-19]。本研究克隆获得的秀丽

白虾的 4 个 *C-lectin* 基因,都含有 2 个典型的 CRD 结构域,这一点与罗氏沼虾高度相似。并且,在这 4 个 C 型凝集素蛋白中,都含有 2 个 Ca^{2+} 结合域 2,或称为糖基结合位点,其功能与碳水化合物化合物的识别有关。CRD 结构域中,主要由 2 个基序与钙离子共同作用完成对糖基分子的结合,第一个典型的基序为 EPN 或 QPD,第二个典型的基序为 WND。EPN 和 EPD 可对甘露糖进行识别,QPD 则可识别半乳糖^[20-21]。本研究显示,秀丽白虾 EmCTL-1、EmCTL-2、EmCTL-3 和 EmCTL-4 的糖基结合位点分别为 EPD/FND、EPA/FHD、WHD/FDD 和 EPD/FRD,表明 EmCTL-1、EmCTL-2 和 EmCTL-4 可能对甘露糖进行结合。在虾类中,这 2 个基序存在一定的多态性,有 EPD、EPN、WTD、QPD、EPS、LND、FRD、VND、QPT、FRD 等多种形式^[5]。同样,秀丽白虾 4 种 *C-lectin* 基因的糖基结合位点也发生了一定程度的变异,其中,EmCTL-2、EmCTL-3 的糖基结合位点分别与罗氏沼虾的 Lectin2、Lectin3 一致^[12],而 EmCTL-1、EmCTL-4 则与凡纳滨对虾的一致^[22],说明在虾类中,糖基结合位点虽然多变,但在某种程度上还是存在一定的保守性。

对秀丽白虾 4 种 C 型凝集素基因的信号肽分析表明,4 个基因都存在 17~21 个氨基酸残基的信号肽,说明这 4 种 C 型凝集素全为分泌型蛋白。在昆虫中,经对氨基酸的化学修饰研究证明,序列中存在的保守的半胱氨酸、色氨酸残基,对凝集素的凝血活性起着重要作用^[23]。本研究中经多序列比对发现,秀丽白虾的 4 个 C 型凝集素基因中都存在着保守的半胱氨酸和色氨酸残基(图 3),提示着其具有一定的凝血功能。已经证实,C 型凝集素中保守的天冬氨酸残基的存在,对与 Ca^{2+} 的结合至关重要^[12],本研究所克隆的 4 个秀丽白虾 C 型凝集素基因均具有保守的天冬氨酸残基,且都具有信号肽,因此可以从结构上推测这 4 个 C 型凝集素基因可以在血淋巴中粘合入侵的微生物,具有一定的免疫功能,当然其具体的功能尚需经进一步的感染实验来证实。

3.2 秀丽白虾 C 型凝集素基因的进化分析

多序列比对表明,除了 EmCTL-2 与罗氏沼虾的 Lectin2 有 73% 的相似性外,秀丽白虾的 C-Lectin 与其他虾类的相似性为 31%~38%,相似程度总体上偏低。在有关的研究中,凡纳滨对

虾 C 型凝集素 2 与其他的物种相似性为 31%~34%,罗氏沼虾的 4 个 C 型凝集素相似性为 30%~38%^[12],可见 C 型凝集素是一个进化较快的基因,这与无脊椎动物免疫相关基因具有较高变异性的特点相吻合^[24]。同时,基于 PAML 程序的分析结果表明,包含秀丽白虾 4 个 C 型凝集素基因在内的具有 2 个 CRD 结构域的 *lectin* 基因经历了阳性选择。阳性选择是免疫相关基因的一个明显特征^[24],虾类具有 2 个 CRD 结构域的凝集素中存在阳性选择,这一点暗示着 C 型凝集素在免疫防御中的作用,同时阳性选择后造成的凝集素氨基酸的多样性更进一步说明了凝集素对入侵微生物的强大的防御能力^[12]。本研究的结果从另一个角度证明了秀丽白虾这 4 个 *lectin* 基因与免疫功能相关的确定性。

Wang 等^[5]基于 *C-lectin* 基因氨基酸序列构建的进化树显示,十足类的 *C-lectin* 基因可分成 4 组,含有 2 个 CRD 结构域的种类分别构成了 2 个亚群(A 和 C),含 1 个 CRD 结构域的种类则构成了另一亚群(D),本实验的结果和此基本一致,可见 CRD 结构域的数目对其系统进化起了重要作用。秀丽白虾与罗氏沼虾同属于长臂虾亚科,因此系统树也揭示了其进化关系,EmCTL-2、EmCTL-3、EmCTL-4 分别与罗氏沼虾的 Lectin2、Lectin3、Lectin4 显示了最近的亲缘关系。此外,系统进化树还显示,对虾类的 C 型凝集素可含有 1 个或 2 个 CRD 结构域,真虾类则以含 2 个 CRD 结构域占优势,而蟹类及螯虾类,则主要以含 1 个 CRD 结构域为主(图 4,D 组)。

3.3 秀丽白虾 C 型凝集素基因的组织表达特征

在目前所研究的十足类的 C 型凝集素基因中,部分存在仅在特定组织中表达的特征,例如中国明对虾的 *Fclectin* 基因仅在血淋巴中特异表达^[25],*Fc-hsL*, *Fc-Lec2*, *Fc-lec5* 和墨吉对虾的 *FmLC* 则在肝胰腺中特异表达^[11,17,26-27];而有的则在不同的组织中都能表达,如罗氏沼虾的 *MrLec3* 可在肝胰腺、鳃、胃和肠中表达^[12],凡纳滨对虾的 2 种 C 型凝集素基因 *LvLectin-1* 和 *LvLectin-2* 则在肝胰腺、肌肉、生殖腺、血液和鳃中都能表达^[28]。本研究以 β -actin 为内参的半定量 RT-PCR 结果表明,秀丽白虾 4 种 C 型凝集素不具有组织特异表达的特点,而是在其体内具有不同的组织表达模式。其中 *EmCTL-1* 在血液中表

达量最高, *EmCTL-2* 在卵巢中表达量最高, 在血液中的表达量相对较高, 而 *EmCTL-3* 和 *EmCTL-4* 都是在精巢中的表达量最高, 同时, *EmCTL-2* 和 *EmCTL-4* 在胸神经节中都具有较高的表达量。这和中华绒螯蟹、凡纳滨对虾的 C 型凝集素的研究结果一致, 中华绒螯蟹的 *EsLecA* 主要在精巢和胸神经节中表达, 而 *EsLecG* 则主要在胸神经节中表达^[13]; 凡纳滨对虾的 2 种 C 型凝集素基因 *LcCTLD* 和 *LvLec* 在神经节中的表达量最高^[7-8]。秀丽白虾 4 种 C 型凝集素基因的共同特点是在精巢中都有极高的表达; 虽然在肝胰腺中都有表达, 但肝胰腺并不是其表达量最高的组织; 在肌肉和鳃组织中都不表达。本研究结果显示, 秀丽白虾的精巢是 C 型凝集素基因的一个重要的表达器官, 这个结果是否提示了精巢除了执行主要的生殖功能外, 还有一部分免疫功能尚可发掘, 或者预示着精巢中高丰度 C 型凝集素基因的存在, 其本身对协助生殖细胞的正常产生起了一定的屏障作用, 这一点尚需更深入的研究来证实。免疫相关的基因同时又兼具一定的生殖功能, 这在哺乳类已有相关报道, 如大鼠附睾中特异表达的 *Bin 1b* 基因是一个与生育有关的天然抗菌肽基因^[29], 在生育期的表达量最高, 年老后下降; 另外的研究也证实, Toll-like receptors (TLRs) 在生殖系统中参与多种生理与病理过程, 不仅在雄性的生殖系统中具有明显的抗病毒效应^[30], 而且在雌性的卵巢及子宫中均行使重要功能, 在卵巢癌、盆腔炎、睾丸类固醇疾病、少精症、前列腺炎及前列腺癌等生殖系统疾病的形成中具有重要作用^[31]。C 型凝集素或者其家族中的某一成员本身是否是一类具有生殖与免疫的双功能基因, 尚需依赖后续的功能实验研究 C 型凝集素在秀丽白虾生殖过程中抗微生物感染的作用来确定, 预期可为研究虾类生殖系统的自身防御开辟一定的通路, 也可作为虾类的遗传育种提供可能的分子标记, 并为虾类遗传育种中加快种苗选育工作的进程提供有效的基因源。

参考文献:

- [1] 谢建辉, 顾建新. C 型凝集素[J]. 生命科学, 2011, 23(6): 555-562.
- [2] Sharon N, Lis H. History of lectins: from hemagglutinins to biological recognition molecules [J]. *Glycobiology*, 2004, 14(11): 53R-62R.
- [3] Drickamer K, Taylor M E. Biology of animal lectins [J]. *Annual Review of Cell Biology*, 1993, 9(2): 237-264.
- [4] Varki A, Cummings R D, Esko J D, et al. *Essentials of Glycobiology* (2nd edition) [M]. 2nd ed. New York: Cold Spring Harbor Laboratory Press, 2009: 100-180.
- [5] Wang X W, Wang J X. Diversity and multiple functions of lectins in shrimp immunity [J]. *Developmental & Comparative Immunology*, 2013, 39(1-2): 27-38.
- [6] Sun J, Wang L, Wang B, et al. Purification and characterization of a natural lectin from the plasma of the shrimp *Fenneropenaeus chinensis* [J]. *Fish & Shellfish Immunology*. 2008, 25(3): 290-297.
- [7] Junkunlo K, Prachumwat A, Tangprasittipap A, et al. A novel lectin domain-containing protein (LvCTLD) associated with response of the whiteleg shrimp *Penaeus (Litopenaeus) vannamei* to yellow head virus (YHV) [J]. *Developmental and Comparative Immunology*, 2012, 37(3-4): 334-341.
- [8] Zhang Y, Qiu L, Song L, et al. Cloning and characterization of a novel C-type lectin gene from shrimp *Litopenaeus vannamei* [J]. *Fish & Shellfish Immunology*, 2009, 26(1): 183-192.
- [9] Ma T H, Benzie J A, He J G, et al. PmLT, a C-type lectin specific to hepatopancreas is involved in the innate defense of the shrimp *Penaeus monodon* [J]. *Journal of Invertebrate Pathology*, 2008, 99(3): 332-341.
- [10] Yang H, Luo T, Li F, et al. Purification and characterisation of a calcium-independent lectin (PjLec) from the haemolymph of the shrimp *Penaeus japonicus* [J]. *Fish & Shellfish Immunology*, 2007, 22(1-2): 88-97.
- [11] Rattanaporn O, Utarabhand P. Molecular cloning of a C-type lectin with two CRD domains from the banana shrimp *Fenneropenaeus merguensis*: Early gene up-regulation after *Vibrio harveyi* infection [J]. *Journal of Invertebrate Pathology*, 2011, 106(2): 196-204.
- [12] Ren Q, Li M, Du J, et al. Immune response of four dual-CRD C-type lectins to microbial challenges in giant freshwater prawn *Macrobrachium rosenbergii* [J]. *Fish & Shellfish Immunology*, 2012, 33(2): 155-167.
- [13] Jin X K, Li W 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.
- [14] Yang Z, Nielsen R, Goldman N, *et al.* Codon-substitution models for heterogeneous selection pressure at amino acid sites [J]. *Genetics*, 2000, 155 (1): 431 – 449.
- [15] Yang Z. PAML 4: phylogenetic analysis by maximum likelihood [J]. *Molecular Biology and Evolution*, 2007, 24(8):1586 – 1591.
- [16] Anisimova M, Bielawski J P, Yang Z. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution [J]. *Molecular Biology and Evolution*, 2001, 18(8):1585 – 1592.
- [17] Sun Y D, Fu L D, Jia Y P, *et al.* A hepatopancreas-specific C-type lectin from the Chinese shrimp *Fenneropenaeus chinensis* exhibits antimicrobial activity [J]. *Molecular Immunology*, 2008, 45 (2): 348 – 361.
- [18] Liu Y C, Li F H, Dong B, *et al.* Molecular cloning, characterization and expression analysis of a putative C-type lectin (Fclectin) gene in Chinese shrimp *Fenneropenaeus chinensis* [J]. *Molecular Immunology*, 2007, 44(4):598 – 607.
- [19] Wei X, Liu X, Yang J, *et al.* Two C-type lectins from shrimp *Litopenaeus vannamei* that might be involved in immune response against bacteria and virus [J]. *Fish & Shellfish Immunology*, 2012, 32 (1): 132 – 140.
- [20] Drickamer K. Engineering galactose-binding activity into a C-type mannose-binding protein [J]. *Nature*, 1992, 360(6400):183 – 186.
- [21] Kolatkar A R, Weis W I. Structural basis of galactose recognition by C-type animal lectins [J]. *The Journal of Biological Chemistry*, 1996, 271(12): 6679 – 6685.
- [22] Ma T H, Tiu S H, He J G, *et al.* Molecular cloning of a C-type lectin (LvLT) from the shrimp *Litopenaeus vannamei*: early gene down-regulation after WSSV infection [J]. *Fish & Shellfish Immunology*, 2007, 23(2):430 – 437.
- [23] 徐小超. 蜘蛛抱蛋 (*Aspidistra elatior* Bulme) 凝集素的分离、纯化及性质研究 [D]. 四川: 四川大学生命科学学院, 2007.
- [24] Hayes M L, Eytan R I, Hellberg M E. High amino acid diversity and positive selection at a putative coral immunity gene (*tachylectin-2*) [J]. *BMC Evolutionary Biology*, 2010, 10(5):150 – 164.
- [25] Liu Y C, Li F H, Dong B, *et al.* Molecular cloning, characterization and expression analysis of a putative C-type lectin (Fclectin) gene in Chinese shrimp *Fenneropenaeus chinensis* [J]. *Molecular Immunology*, 2007, 44(4):598 – 607.
- [26] Zhang X W, Xu W T, Wang X W, *et al.* A novel C-type lectin with two CRD domains from Chinese shrimp *Fenneropenaeus chinensis* functions as a pattern recognition protein [J]. *Molecular Immunology*, 2009, 46(8 – 9):1626 – 1637.
- [27] Xu W T, Wang X W, Zhang X W, *et al.* A new C-type lectin (FcLec5) from the Chinese white shrimp *Fenneropenaeus chinensis* [J]. *Amino Acids*, 2010, 39(5):1227 – 1239.
- [28] Wei X, Liu X, Yang J, *et al.* Two C-type lectins from shrimp *Litopenaeus vannamei* that might be involved in immune response against bacteria and virus [J]. *Fish & Shellfish Immunology*, 2012, 32 (1): 132 – 140.
- [29] Li P, Chan H C, He B, *et al.* An antimicrobial peptide gene found in the male reproductive system of rats [J]. *Science*, 2001, 291(5509):1783 – 1785.
- [30] Wang T, Zhang X, Chen Q, *et al.* Toll-like receptor 3-initiated antiviral responses in mouse male germ cells *in vitro* [J]. *Biology of Reproduction*, 2012, 86 (4): 1 – 10.
- [31] Girling J E, Hedger M P. Toll-like receptors in the gonads and reproductive tract: emerging roles in reproductive physiology and pathology [J]. *Immunology and Cell Biology*, 2007, 85 (6): 481 – 489.

Molecular cloning, expression characterization and phylogenetic analysis of four novel C-type lectin genes in white shrimp (*Exopalaemon modestus*)

WANG Min, WU Ping*, YE Yuantu, SHI Taodan, WEI Yuhong
(School of Biology and Basic Medical Sciences, Soochow University, Suzhou 215123, China)

Abstract: *Exopalaemon modestus* is one of the main shrimps which has high economic value. But its genetic information and functional gene background are rarely known so far. In this study, the cDNAs of four novel C-type lectins (designated as *EmCTL-1*, *EmCTL-2*, *EmCTL-3* and *EmCTL-4*) were cloned from the ovary of white shrimp, *E. modestus*, by means of PCR and 5' and 3' rapid amplification of cDNA ends (RACE). The full-length cDNAs consist of 1 390, 1 119, 1 364 and 1 299 bp with one open reading frame consisting of 996, 969, 1 026 and 1 044 bp, encoding 332, 322, 341 and 347 amino acids respectively. Their deduced amino acid sequences contained a putative signal peptide of 19, 21, 20 and 17 amino acid residues respectively. The four C-type lectin genes all contained two carbohydrate recognition domains, CRD1 and CRD2. The CRD1 was composed of 128 residues in *EmCTL-2* and *EmCTL-3*, 127 residues in *EmCTL-1* and 126 residues in *EmCTL-4*. And the CRD2 was composed of 127 residues in *EmCTL-1* and *EmCTL-3*, 129 residues in *EmCTL-2* and 130 residues in *EmCTL-4*. Both the CRD1 and CRD2 had 6 conserved cysteine residues and 3 conserved tryptophane residues. The motif of Ca^{2+} binding sites 2 were EPA and FND in *EmCTL-1*, EPA and FHD in *EmCTL-2*, WHD and FDD in *EmCTL-3* and EPD and FRD in *EmCTL-4*. So *EmCTL-1*, *EmCTL-2* and *EmCTL-4* can be deduced binding with mannose in immune reactions. And the four C-type lectin genes shared 40%, 37%, 69% and 73% identity with the parallel C-type lectin genes in *Macrobrachium rosenbergii*, existing in Palaemonidae. The CODEML program implemented in the PAML package was used to test for variations in selective pressures of shrimp and prawns C-type lectin genes with two CRDs, included the four genes acquired in this study. The result showed that the shrimp and prawns lectin genes with two CRDs had undergone positive selection. According to the LTR test, 3.8% of the sites were driven by positive selection in the M2a model, while 5.6% of the sites were driven by positive selection in the M8 model. Phylogenetic analysis showed that the lectins from the decapods could be divided into four clear clusters. Cluster C included the dual-CRD lectins while cluster D included the single-CRD lectins, and cluster A included the dual-CRD lectins except the PmCTL-1 in tiger shrimp, cluster B included the single-CRD lectins except the FcCTL in Chinese shrimp. It indicated that the numbers of CRD played a key role in phyletic evolution. The results showed that *EmCTL-1*, *EmCTL-2*, *EmCTL-3* belonged to cluster C and *EmCTL-4* belonged to cluster A. Four novel C-type lectin genes in *E. modestus* had a close evolutionary relationship to the C-type lectins of giant freshwater prawn. Tissue distribution analysis using semi-quantitative RT-PCR methods showed that *EmCTL-1* was mainly expressed in hemolymph, testis and ovary. *EmCTL-3* and *EmCTL-4* could be detected mainly in testis, ovary, hepatopancreas, hemolymph, heart and thoracic ganglion while *EmCTL-2* was mainly expressed in ovary, testis, hemolymph, thoracic ganglion and heart. The results showed that the hemolymph was the main expression tissue of four C-type lectin genes in *E. modestus*. Interestingly, and the testis was the organ with highest expression level of C-type lectin transcripts in this study. It can be speculated that the C-type lectins in *E. modestus* may agglutinate microorganisms in the hemolymph for possessing signal peptides, conserved cysteine residues, conserved tryptophane residues, high expression level in hemolymph and observed positive selection.

Key words: *Exopalaemon modestus*; C-type lectin; cloning; tissue expression pattern; positive selection

Corresponding author: WU Ping. E-mail: suzhou_wuping@163.com