

马氏珠母贝(*Pinctada fucata*)血细胞 RNA-Seq 转录组数据中补体样组分分析*

吴 勇¹ 张 野² 许尤厚³ 梁雪茹² 王忠良²

(1. 广东海洋大学科技处 湛江 524088; 2. 广东海洋大学水产学院 湛江 524088;
3. 广西北部湾海洋生物多样性养护重点实验室(北部湾大学) 钦州 535000)

摘要 补体系统在体液免疫中发挥着重要作用,是连接先天免疫和适应性免疫的枢纽。本研究基于前期转录组 RNA-Seq 数据检索马氏珠母贝(*Pinctada fucata*)补体样组分,并对其结构域及溶藻弧菌(*Vibrio alginolyticus*)刺激后的基因表达水平进行了分析。结果表明,共获得 68 条补体样成分 unigene,分别编码 14 个含 C1q 结构域蛋白、14 个凝集素蛋白、4 个纤维蛋白原相关蛋白、12 个丝氨酸蛋白酶、10 个含硫酯键蛋白、1 个末端补体分子 C6、5 个补体受体、1 个补体因子及 7 个纤胶凝蛋白;结构域分析表明,马氏珠母贝补体样组分均含有相应的保守结构域,与其他贝类补体组分的研究结果基本一致;表达量分析表明,经溶藻弧菌刺激后,68 条补体样组分 unigene 中共有 21 条 unigene 表达上调,其中 12 条为模式识别受体 unigene。结合牡蛎等贝类补体组分的研究结果,推测贝类已形成一个含有多补体组分的原始补体系统,且可能以凝集素途径或替代途径行使补体的生物学功能。以上研究结果为深入了解马氏珠母贝及其他海洋无脊椎动物免疫防御机制、探讨补体系统的进化起源提供了新的依据。

关键词 马氏珠母贝; RNA-Seq; 补体样组分; 保守结构域
中图分类号 Q789 doi: 10.11693/hyh20190500095

与其他无脊椎动物一样,贝类缺乏适应性免疫系统,而主要依靠先天性免疫系统防御各种病原和外来物质的入侵,以维持机体的正常生命活动;非特异的先天性免疫防御被认为是宿主对抗细菌、真菌以及病毒等病原入侵的第一道防线。补体系统作为先天性免疫系统的重要组成部分,是迄今所知机体中最复杂的一个限制性蛋白水解系统(limited proteolysis system),其不仅联系着先天免疫和获得性免疫,而且广泛参与多种生理和病理过程(Beutler, 2004)。由补体系统介导的生物学功能包括吞噬作用、细胞溶解、炎症、免疫复合物溶解、凋亡细胞清除及促进体液免疫

应答等(Schmidt *et al.*, 2000)。

自 19 世纪末被发现以来(Fujita *et al.*, 2004a),补体系统在哺乳动物中得到了广泛的研究。补体系统必须被激活以启动一系列的生化反应,才能表现出各种免疫效应,并最终导致外来病原的裂解和破坏。高等动物中补体的激活过程分为前端反应和末端通路两个阶段;按激活物及激活顺序的不同,前端反应又可分为三条途径,即经典途径(classic pathway)、凝集素途径(lectin pathway)和替代途径(alternative pathway);三条途径具有共同的末端通路(terminal pathway),并形成膜攻击复合物(membrane attack

* 广东省基础与应用基础研究基金项目,2019A1515011875 号;广东海洋大学 2019 年“冲一流”省财政专项资金建设项目,231419015 号;广东省省级科技计划项目,2015A020209169 号;国家级大学生创新创业训练计划项目,CXXL2018001 号;广东海洋大学“海之帆-起航计划”大学生科技创新培育项目,qhjhzr201812 号;广西北部湾海洋生物多样性养护重点实验室(钦州学院)开放课题,2017KB03 号;国家自然科学基金项目,31202023 号。吴 勇,硕士,助理研究员,E-mail: gdouwy@126.com

通信作者: 王忠良,副教授,E-mail: leong2006@126.com

收稿日期: 2019-05-18, 收修改稿日期: 2019-07-02

complex, MAC)。此外, Huber-Lang 等(2006)发现小鼠血浆中的凝血酶(thrombin)分子可替代 C3 依赖型 C5 转化酶切割 C5 分子, 并由此提出了一条经凝血素(prothrombin)介导的补体激活途径。

补体系统是一种古老的免疫防御机制, 在无脊椎动物的免疫防御中发挥重要作用, 最近在无脊椎动物中陆续发现补体组分, 使得“补体系统在低等无脊椎动物中早已形成”的推论得到进一步确认(Nonaka *et al*, 2001; Smith *et al*, 2001; Dishaw *et al*, 2005)。如在玻璃海鞘(*Ciona intestinalis*)和文昌鱼(*Branchiostoma floridae*)的基因组中发现了数百个含补体相关结构域分子(Azumi *et al*, 2003; Huang *et al*, 2008); 在昆虫等无脊椎动物中发现模式识别分子(Pattern recognition receptor, PRR)及丝氨酸蛋白酶(serine protease, SP)、纤胶凝蛋白(ficolin)、甘露糖结合凝集素相关丝氨酸蛋白酶(MBL-associated serine protease, MASP)、C3、C2 和补体因子 B 等补体样分子(Matsushita *et al*, 2001; Nonaka *et al*, 2001; Blandin *et al*, 2004; Fujita *et al*, 2004b); 另外, 刺胞动物中也发现了 C3 样分子、B 因子/C2 以及参与形成 MAC 的蛋白分子(Nonaka *et al*, 2006; Miller *et al*, 2007)。基于海胆(*Strongylocentrotus purpuratus*)原始补体系统的研究, Smith 等(2001)提出由 C3、B 因子及 MASP 等组成的补体替代途径和凝集素途径在棘皮动物(海胆)和尾索动物(海鞘)中早就已出现。最近, Wang 等(2017)系统分析了太平洋牡蛎(*Crassostrea gigas*)全基因组数据中补体样组分的结构域和系统进化关系, 提出贝类补体系统可能通过替代途径激活。目前, 受限于少数无脊椎动物补体系统的研究结果以及全基因组信息, 无脊椎动物补体系统的组成及其激活途径的研究工作进展缓慢, 而对于贝类补体系统的研究工作更是匮乏。

近年来, 高通量测序技术的发展为研究无脊椎动物补体系统的结构特征、推断补体组分功能以及可能存在的激活途径提供了新的途径(Hibino *et al*, 2006; Huang *et al*, 2008; Zhang *et al*, 2012; Gerdol *et al*, 2015; Wang *et al*, 2017)。本文基于前期转录组 RNA-Seq 数据检索马氏珠母贝(*Pinctada fucata*)补体样组分, 并对其结构域和溶藻弧菌(*Vibrio alginolyticus*)刺激前后的基因表达水平进行了分析, 为深入了解马氏珠母贝免疫防御机制、探讨补体系统的进化起源提供新的依据。

1 材料与方法

1.1 马氏珠母贝的暂养、细菌刺激及总 RNA 提取

马氏珠母贝(*Pinctada fucata*)(平均壳长 70mm)购自广东省湛江市徐闻迈陈珍珠贝养殖场, 暂养于室内玻璃钢水槽中(80L/槽, 20 只/槽); 水温 25°C, 盐度 28, 饲养期间连续充气, 投喂螺旋藻粉; 每天 100%换水一次。

室内暂养一周后将珠母贝分为两组(每组 40 只); 第一组于闭壳肌处注射 0.1mL 浓度为 5×10^7 CFU/mL 的溶藻弧菌(*V. alginolyticus*)悬液; 第二组于闭壳肌处注射同等体积的 PBS (pH 7.4)作为对照。注射 4h 后, 闭壳肌采集血淋巴, 每只约 0.5mL; 采集的血淋巴分装至 1.5mL 离心管中, 每管 1mL; 4°C, 800g 离心 10min 收集血细胞, 并立即提取总 RNA。

1.2 RNA-Seq 文库构建、转录组测序及数据

RNA-Seq 文库构建及转录组测序的具体方法见 Wang 等(2016)。

转录组测序共获得 70407878 条 Raw reads, 经去除含有接头、重复及测序质量较低的原始读数后, 获得 56345139 条 Clean reads。使用转录组 *de novo* 组装软件 Trinity (Grabherr *et al*, 2011)对 Clean reads 进行组装, 并进行去冗余处理和进一步拼接, 共得到 74007 条 unigenes(转录组 Raw reads 已提交至 NCBI 的 SRA 数据库, 登录号为 SRP041567)。

1.3 Unigene 表达水平分析

采用 RSEM 软件(Version 1.1.21) (Li *et al*, 2011)将每个样品的 Clean reads 比对 Trinity 组装的参考序列, 得到每个样品比对的每个 unigene 上的 read 数目, 并进行 RPKM 标准化转换(Mortazavi *et al*, 2008), 进而分析 unigene 的表达水平; 采用 DEGseq 方法(version 1.12.0) (Anders *et al*, 2010)筛选差异表达 unigene, 筛选阈值为 q 值 < 0.005 且 $|\log_2 \text{Fold Change}| > 1$ 。

1.4 生物信息学分析

利用 BLAST 程序在 Nt、Nr、Swiss-Prot、KEGG、COG 及 Pfam 数据库中比对 unigene 并获得期注释信息(E 值 $< 1.0E^{-5}$); 然后, 在此注释信息中检索含 C1q 结构域蛋白(C1q、C1qDC)、凝集素(凝集素、C 型凝集素、胶原凝集素、甘露糖结合凝集素 MBL)、C3、纤维蛋白原相关蛋白(fibrinogen-related protein, FREP)、丝氨酸蛋白酶、含硫酯蛋白(thioester-containing protein, TEP)、补体受体(complement receptor, CR)、补体因子及 ficolin 等补体样组分。采用 NCBI 保守结

构域数据库 (<https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>) 及 SMART server (<http://smart.embl-heidelberg.de/>) 进行补体样组分的蛋白结构域检测。

2 结果与分析

2.1 马氏珠母贝的补体样组分

通过对马氏珠母贝 RNA-Seq 转录组注释信息的

检索, 共获得 68 条补体样成分 unigene, 分别编码 14 个含 C1q 结构域蛋白、14 个凝集素蛋白(含 2 个 MBL、2 个胶原凝集素及 10 个 C 型凝集素)、4 个 FREP、12 个丝氨酸蛋白酶(含 1 个补体 C2)、10 个含硫酯键蛋白(含 1 个补体 C4、1 个补体 C5、2 个含硫酯蛋白及 6 个 alpha 巨球蛋白)、1 个末端补体分子 C6、5 个补体受体、1 个补体因子及 7 个 ficolin(表 1)。

表 1 马氏珠母贝 RNA-Seq 数据中的补体样成分
Tab.1 Summary of complement-like components from RNA-seq dataset of *P. fucata*

分类(数量)	基因 ID	EST 长度(bp)	注释信息	注释数据库
C1q/C1qDC (14)	comp66465_c0	864	Complement C1q-like protein 2	Nr
	comp72411_c0	820	Complement C1q-like protein 2	Nr
	comp819228_c0	217	Complement C1q-like protein 3	Swiss-Prot
	comp11397_c0	1547	Complement C1q-like protein 4	Swiss-Prot
	comp15071_c0	502	Complement C1q-like protein 4	Nr
	comp314296_c0	757	Complement C1q-like protein 4	Nr
	comp37171_c0	834	Complement C1q-like protein 4	Nr
	comp51879_c0	526	Complement C1q-like protein 4	Nr
	comp55080_c0	675	Complement C1q-like protein 4	Nr
	comp70389_c0	2802	Complement C1q-like protein 4	Nr
	comp71265_c0	1461	Complement C1q-like protein 4	Nr
	comp72322_c0	2067	Complement C1q-like protein 4	Swiss-Prot
	comp754969_c0	363	Complement C1q-like protein 4	Nr
	comp487618_c0	518	C1q domain containing protein	Nr
C4/C5/C6 (3)	comp62656_c0	1854	Complement C4	Swiss-Prot
	comp377897_c0	457	Complement component C5-1	Nr
	comp1280548_c0	208	Complement component C6	Nr
含硫酯蛋白/alpha- 巨球蛋白 (8)	comp19552_c0	278	Thioester-containing protein-C	Nr
	comp72940_c0	4090	Thioester-containing protein-C	Nr
	comp74799_c0	3244	Alpha-2-macroglobin splicing variant 1	Nr
	comp74799_c1	3998	Alpha-2-macroglobulin	Nr
	comp47264_c0	425	Alpha-2-macroglobulin	Nr
	comp1426901_c0	223	Alpha-macroglobulin	KOG
	comp395606_c0	407	Alpha-macroglobulin	KOG
	comp61634_c0	1299	Alpha-macroglobulin	KOG
凝集素 (14)	comp76368_c1	3898	Lectin, mannose-binding 2	KEGG
	comp71299_c0	2296	Lectin, mannose-binding 1	KEGG
	comp737450_c0	334	Collectin-11	Nr
	comp60893_c0	577	Collectin-12	Nr
	comp193934_c0	703	C-type lectin domain family 3	Nr
	comp44859_c0	393	C-type lectin	KOG
	comp67238_c0	563	C-type lectin domain family 3	Nr
	comp72224_c0	1241	C-type lectin 2	Nr
	comp68098_c0	871	C-type lectin	Nr
	comp1135571_c0	332	C-type lectin	Nr
comp760004_c0	305	C-type lectin 5	Nr	

续表

分类(数量)	基因 ID	EST 长度(bp)	注释信息	注释数据库
凝集素 (14)	comp70764_c0	895	C-type lectin 8	Nr
	comp311441_c0	492	C-type lectin domain family 4	Nr
	comp69799_c0	1497	C-type lectin 11	Nr
	comp334865_c0	372	Fibrinogen-related protein 7-1 precursor	Nr
纤维蛋白原相关蛋白 (4)	comp63837_c0	1016	Fibrinogen-related protein	Nr
	comp379987_c0	634	Fibrinogen-related protein	Nr
	comp59038_c0	1886	Fibrinogen-related protein 1	Nr
	comp64269_c0	1856	Thymus-specific serine protease-like	Nr
丝氨酸蛋白酶 (12)	comp437030_c0	342	Testicular-specific serine protease 3	Nr
	comp121_c0	245	Serine protease 42	Nr
	comp70805_c0	2143	Serine protease	Nr
	comp72849_c0	1420	Serine protease 27-like	Nr
	comp181719_c0	844	Serine protease	Nr
	comp749730_c0	211	Serine protease mRNA	Nt
	comp74323_c0	2199	Serine protease	Swiss-Prot
	comp506917_c0	295	Serine protease 27	Swiss-Prot
	comp75471_c0	2093	Serine protease	KOG
	comp63431_c0	1296	HtrA serine peptidase 2	KEGG
	comp488509_c0	425	Complement C2	Swiss-Prot
	comp398145_c0	1235	Complement component receptor 1-like protein	Swiss-Prot
补体受体 (5)	comp68087_c0	1586	Complement receptor type 1	Swiss-Prot
	comp437543_c0	548	Complement receptor type 2	Swiss-Prot
	comp688205_c0	259	Complement receptor type 2	Swiss-Prot
补体因子 (1)	comp553342_c0	309	complement receptor type 2-like	Nr
	comp76920_c0	2275	Complement Factor H	Nr
	comp75716_c0	1432	Ficolin-2	Nr
	comp432272_c0	404	Ficolin-2-like	Nr
纤胶凝蛋白 (7)	comp74191_c0	1366	Ficolin-2	Nr
	comp50977_c0	437	Ficolin and related extracellular proteins	KOG
	comp76571_c0	2717	Ficolin and related extracellular proteins	KOG
	comp74642_c0	3228	Ficolin	KEGG
	comp1221917_c0	222	Ficolin-1	Swiss-Prot

2.2 含 C1q 结构域蛋白、甘露糖结合凝集素、纤维蛋白原相关蛋白及纤胶凝蛋白

脊椎动物中, C1q、MBL 及 ficolin 等均作为 PRR 识别外来入侵病原微生物并激活补体系统的经典途径和凝集素途径。在检索到的 14 个马氏珠母贝含 C1q 结构域蛋白(C1qDC)中, 共有 11 个 C1qDC 含有 C 末端球状结构域, 但均未能检测到 N 末端的胶原结构域。见表 2。

甘露糖结合凝集素(Mannose-binding lectin, MBL)和能结合微生物表面的多糖, 从而激活 MBL 相关丝氨酸蛋白酶和 C3。本文中, 2 个 MBL 分子中均含有 Lectin_leg 样结构域, 但未检索到 N 末端胶原结构域和 C 型凝集素样结构域; 同时, 其他 12 个 C 型凝集素基因(C-type lectin, CTL)中均发现含有 CTLD 结构域。见表 3。

表 2 马氏珠母贝 C1q/C1qDC 保守结构域分析
Tab.2 Analysis of the conserved domains in C1q/C1qDC of *P. fucata*

基因 ID	注释信息	保守结构域
comp66465_c0	Complement C1q-like protein 2	N-卷曲螺旋区, C-球状结构域
comp72411_c0	Complement C1q-like protein 2	C-球状结构域
comp819228_c0	Complement C1q-like protein 3	C-球状结构域
comp11397_c0	Complement C1q-like protein 4	信号肽, N-卷曲螺旋区, C-球状结构域
comp15071_c0	Complement C1q-like protein 4	——
comp314296_c0	Complement C1q-like protein 4	C-球状结构域
comp37171_c0	Complement C1q-like protein 4	——
comp51879_c0	Complement C1q-like protein 4	信号肽, C-球状结构域
comp55080_c0	Complement C1q-like protein 4	C-球状结构域
comp70389_c0	Complement C1q-like protein 4	——
comp71265_c0	Complement C1q-like protein 4	信号肽, C-球状结构域
comp72322_c0	Complement C1q-like protein 4	信号肽, N-卷曲螺旋区, C-球状结构域
comp754969_c0	Complement C1q-like protein 4	C-球状结构域
comp487618_c0	C1q domain containing protein	信号肽, C-球状结构域

表 3 马氏珠母贝凝集素保守结构域分析
Tab.3 Analysis of the conserved domains in lectins of *P. fucata*

基因 ID	注释信息	保守结构域
comp76368_c1	Lectin, mannose-binding 2	Lectin_leg-like 结构域, 跨膜区
comp71299_c0	Lectin, mannose-binding 1	Lectin_leg-like 结构域, 跨膜区
comp737450_c0	Collectin-11	——
comp60893_c0	Collectin-12	CTLD 结构域, 跨膜区
comp193934_c0	C-type lectin domain family 3	CTLD 结构域
comp44859_c0	C-type lectin	CTLD 结构域
comp67238_c0	C-type lectin domain family 3	CTLD 结构域
comp72224_c0	C-type lectin 2	CTLD 结构域
comp68098_c0	C-type lectin	CTLD 结构域
comp1135571_c0	C-type lectin	——
comp760004_c0	C-type lectin 5	CTLD 结构域
comp70764_c0	C-type lectin 8	CTLD 结构域, 跨膜区
comp311441_c0	C-type lectin domain family 4	CTLD 结构域
comp69799_c0	C-type lectin 11	CTLD 结构域, 跨膜区

与 MBL 结构相似的 ficolin 也是由 N 端的胶原结构域和 C 端的糖识别结构域组成。马氏珠母贝 ficolin 基因中均发现了识别糖结构的纤维蛋白原样

(Fibrinogen-like, FBG)结构域, 但同时, RNA-Seq 数据中还发现了 4 个含有 FBG 结构域的 FREP, 推测其可能具有高等动物 ficolin 相似的配体结合功能。见表 4。

表 4 马氏珠母贝纤维蛋白原相关蛋白及纤胶凝蛋白保守结构域分析
Tab.4 Analysis of the conserved domains in Fibrinogen-related proteins and ficolins of *P. fucata*

基因 ID	注释信息	保守结构域
comp334865_c0	Fibrinogen-related protein 7-1 precursor	FBG 结构域
comp63837_c0	Fibrinogen-related protein	FBG 结构域
comp379987_c0	Fibrinogen-related protein	FBG 结构域
comp59038_c0	Fibrinogen-related protein 1	卷曲螺旋区, FBG 结构域
comp75716_c0	Ficolin-2	FBG 结构域
comp432272_c0	Ficolin-2-like	FBG 结构域
comp74191_c0	Ficolin-2	FBG 结构域
comp50977_c0	Ficolin and related extracellular proteins	FBG 结构域
comp76571_c0	Ficolin and related extracellular proteins	FBG 结构域
comp74642_c0	Ficolin	FBG 结构域, 跨膜区
comp1221917_c0	Ficolin-1	FBG 结构域

2.3 丝氨酸蛋白酶

补体丝氨酸蛋白酶主要包括经典途径和凝集素途径中的 MASP、C1r 和 C1s 以及替代途径中的 B 因子和补体 C2。在马氏珠母贝 RNA-Seq 数据中, 共发现了 12 个丝氨

酸蛋白酶基因(包括 11 个丝氨酸蛋白酶和 1 个补体 C2 分子); 其中, 9 个丝氨酸蛋白酶基因中含有 SPC 结构域, 补体 C2 分子中含有一个串联 CCP 结构域, 但未发现同时含有 CCP 和 SPC 结构域的丝氨酸蛋白酶分子。见表 5。

表 5 马氏珠母贝丝氨酸蛋白酶保守结构域分析
Tab.5 Analysis of the conserved domains in serine proteases of *P. fucata*

基因 ID	注释信息	保守结构域
comp64269_c0	Thymus-specific serine protease-like	SPC 结构域
comp437030_c0	Testicular-specific serine protease 3	SPC 结构域
comp121_c0	Serine protease 42	SPC 结构域
comp70805_c0	Serine protease	—
comp72849_c0	Serine protease 27-like	SPC 结构域, 跨膜区
comp181719_c0	Serine protease	—
comp749730_c0	Serine protease mRNA	SPC 结构域
comp74323_c0	Serine protease	SPC 结构域
comp506917_c0	Serine protease 27	SPC 结构域
comp75471_c0	Serine protease	SPC 结构域
comp63431_c0	HtrA serine peptidase 2	SPC 结构域
comp488509_c0	Complement C2	串联 CCP 结构域(SCR 重复)

2.4 含硫酯键蛋白家族分子

马氏珠母贝 RNA-Seq 组装数据内共检索到 10 个含硫酯键蛋白家族分子, 包括含硫酯蛋白、补体 C4/C5 和 alpha-巨球蛋白。保守结构域分析发现, 以上含硫酯键蛋白均具有 alpha 2-巨球蛋白超家族成员

的特征保守结构域, 如硫酯键位点、受体结合区、A2M_comp 结构域、A2M_N 结构域及 A2M_N_2 结构域等。然而, 组装数据中并未检索到在补体系统活化过程中起枢纽作用且与 C4 和 C5 分子同源的补体 C3 分子。见表 6。

表 6 马氏珠母贝含硫酯键蛋白保守结构域分析
Tab.6 Analysis of the conserved domains in thioester-containing proteins of *P. fucata*

基因 ID	注释信息	保守结构域
comp19552_c0	Thioester-containing protein-C	A2M 受体结合区
comp72940_c0	Thioester-containing protein-C	A2M_N_2 结构域, 硫酯键位点, A2M 受体结合区
comp377897_c0	Complement component C5-1	细丝蛋白类型的免疫球蛋白(Ig-FLMN)结构域
comp62656_c0	Complement C4	A2M_comp 结构域, A2M 受体结合区
comp74799_c0	Alpha-2-macroglobin splicing variant 1	A2M_N_2 结构域, VWC 结构域
comp74799_c1	Alpha-2-macroglobulin	A2M_comp 结构域, A2M 受体结合区, 硫酯键位点
comp47264_c0	Alpha-2-macroglobulin	A2M_N 结构域
comp1426901_c0	Alpha-macroglobulin	A2M_N_2 结构域
comp395606_c0	Alpha-macroglobulin	A2M_comp 结构域
comp61634_c0	Alpha-macroglobulin	A2M_N 结构域

2.5 其他补体样组分

CR1、CR2、H 因子、C4 结合蛋白、衰变加速因子(decay-accelerating factor, DAF)和膜辅蛋白(membrane cofactor protein, MCP)同属于补体激活调节剂基因家族(regulator of complement activation, RCA)成员, 参与补体的活化调节。本研究中, 共发现 5 个补体受体(包含 2 个 CR1 和 3 个 CR2)及 1 个 H 因子; 经保守结构域分析

发现其蛋白质结构中均含有串联 CCP 结构域(SCR 重复), 与 RCA 家族成员一致。此外, 本研究中还检索到 1 个补体终末途径的 C6 成分, 但并未发现其含有 EGF(表皮生长因子)、SCR 等保守结构域; 经 BLASTx 比对分析发现, 其氨基酸序列与弓形虫(*Toxoplasma gondii*)含 SCR 重复结构域蛋白(SCR repeat domain-containing protein)具有 32%的同源性。见表 7。

表 7 马氏珠母贝其他补体样组分的保守结构域分析
Tab.7 Analysis of the conserved domains in other complement-like components of *P. fucata*

基因 ID	注释信息	保守结构域
comp398145_c0	Complement component receptor 1-like protein	串联 CCP 结构域(SCR 重复)
comp68087_c0	Complement receptor type 1	串联 CCP 结构域(SCR 重复)
comp437543_c0	Complement receptor type 2	串联 CCP 结构域(SCR 重复)
comp688205_c0	Complement receptor type 2	串联 CCP 结构域(SCR 重复)
comp553342_c0	Complement receptor type 2-like	串联 CCP 结构域(SCR 重复)
comp76920_c0	Complement Factor H	串联 CCP 结构域(SCR 重复)
comp1280548_c0	Complement component C6	—

2.6 补体样组分基因表达水平特征

分析溶藻弧菌刺激前后马氏珠母贝补体样组分的表达水平, 见图 1。结果表明, 68 条补体样组分 unigene 中, 共有 21 条 unigene 表达上调, 包括 C1qDC、CTL、ficolin、丝氨酸蛋白酶、C4、含硫酯键蛋白、补体受体及补体因子 H。而在表达上调的 unigene 中, 共有 12 条模式识别受体 unigene(包括 C1qDC、CTL 及 ficolin), 占表达上调 unigene 数的 57%, 这也表明模式识别受体在马氏珠母贝免疫防御过程中发挥重要作用。

3 讨论

脊椎动物补体系统是一个复杂的蛋白水解系统, 是先天性免疫系统的重要组成部分, 联系着先天性免疫和获得性免疫, 其激活后的主要效应功能表现为促进吞噬微生物病原的调理作用和补体介导的溶胞效应(Huang *et al.*, 2008)。目前, 有关无脊椎动物补体的研究大多数以脊索动物、棘皮动物等为研究对象, 如在海鞘、文昌鱼中众多补体同源分子的发现, 并由此推测其可能具有一个依赖于凝集素的原始补体系

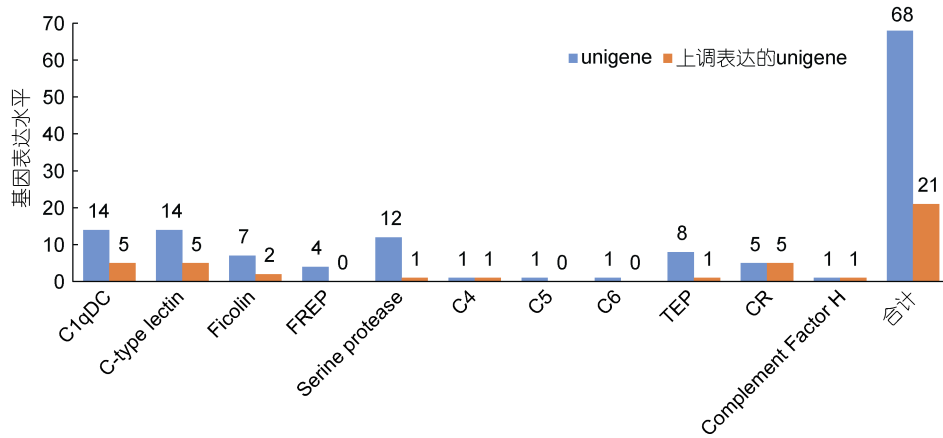


图1 溶藻弧菌刺激后马氏珠母贝补体样组分基因表达水平变化

Fig.1 Analysis of gene expression levels of complement-like components after *V. alginolyticus* challenge

统(Azumi *et al.*, 2003; Huang *et al.*, 2008)。然而有关贝类等原口动物补体同源分子及激活途径的研究目前仍处于初级阶段,尚缺乏系统性研究。本研究中,基于马氏珠母贝血细胞转录组数据共检索到 68 条补体样成分的 unigene,分别编码 C1q 结构域蛋白、凝集素蛋白、FREP、丝氨酸蛋白酶、含硫酯键蛋白、补体受体、补体因子及 ficolin 等补体组分。虽然未检索到补体系统的中心分子 C3,但这些补体同源分子的发现依然暗示它们在免疫防御中可能具有脊椎动物补体组分相似的功能,并构成了马氏珠母贝复杂的原始补体系统。

在脊椎动物中, C1q、MBL、CTL 和 ficolin 等补体组分通常作为模式识别受体在免疫应答过程中发挥病原识别功能并激活不同的补体途径(Nonaka *et al.*, 2004)。而无脊椎动物补体组分,如扇贝的 C1qDC 和 CTL 分子同样能够作为调理素参与病原相关分子模式(Pathogen-associated molecular patterns, PAMP)的识别、结合高等动物 IgG 抗体以及入侵微生物的清除(Wang *et al.*, 2012a, b; Zhang *et al.*, 2008)。此外, LPS 和革兰氏阴性菌能够诱导文昌鱼 C1q 基因(*AmphiC1q1*)的表达(Yu *et al.*, 2008); 玻璃海鞘 MBL 基因经 LPS 刺激后高度活化(Bonura *et al.*, 2009); 而 LPS、LTA 及细菌可诱导文昌鱼 ficolin 基因(*FCN1*)的上调表达(Huang *et al.*, 2011)。本研究中,部分 C1q、CTL 及 ficolin 基因在溶藻弧菌刺激后 mRNA 水平也出现了上调表达(图 1),暗示这些补体相关分子可能参与了马氏珠母贝原始补体系统的激活作用。

补体系统中的丝氨酸蛋白酶,包括 MASP 家族(MASP、C1r、C1s)和 B 因子家族(B 因子和 C2)等,在

不同补体途径激活过程中发挥重要作用。目前,海洋无脊椎动物中仅发现了文昌鱼 MASP1 和 MASP3 分子,以及海葵(*Nematostella vectensis*)和真海鞘(*Halocynthia roretzi*)的 MASP1 分子,而在其他海洋无脊椎动物中尚未有 MASP 家族分子的报道(Ji *et al.*, 1997; Kimura *et al.*, 2009)。本研究中鉴定的丝氨酸蛋白酶分子均未检测到补体控制蛋白(CCP)结构域,这与牡蛎基因组中检索的丝氨酸蛋白酶分子结构特征一致,推测双壳贝类丝氨酸蛋白酶在进化过程中丢失了 CCP 结构域(Wang *et al.*, 2017)。B 因子家族分子在文昌鱼、玻璃海鞘、海胆和欧洲沟纹蛤仔(*Ruditapes decussatus*)等无脊椎动物中均有发现,虽然其分子结构同源性较高,但仍存在一定的差异(Gross *et al.*, 2000; Yoshizaki *et al.*, 2005; He *et al.*, 2008; Prado-Alvarez *et al.*, 2009)。如海鞘的 B 因子在 N 端具有额外的 2 个 LDLR(Low-density lipoprotein receptor, 低浓度脂蛋白受体)结构域和 1 个 CCP 结构域,而本研究中马氏珠母贝 C2 分子与中国鲎(*Tachypleus tridentatus*)B 因子在 N 端均具有串联重复的 CCP 结构域(Tagawa *et al.*, 2012)。这些分子结构不同的无脊椎动物 B 因子家族成员是否具有典型的 B 因子/C2 功能仍需要进一步的验证。

C3 作为补体系统的中心组分在补体活化和生物学效应发挥过程中起着枢纽作用。因而,研究者普遍认为 C3 或其类似分子的出现是低等动物补体系统产生的主要标志,并基于 C3 或其类似分子研究补体的进化过程(Nonaka *et al.*, 2004)。马氏珠母贝 RNA-Seq 数据中并未检索到 C3 分子,仅发现了与 C3 同源的 C4、C5 分子以及同属含硫酯蛋白家族的 α_2 -巨球蛋白

和 TEP 分子, 这与昆虫中的研究结果一致(Nonaka *et al*, 2004, 2006)。在果蝇(*Drosophila melanogaster*)和按蚊(*Anopheles gambiae*)的基因组数据中同样没有发现与高等动物补体 C3 结构一致的分子, 而仅存在与 C3 同家族的 TEP 分子(如按蚊 TEP-1 分子)。功能研究发现, 病原刺激可诱导 TEP-1 基因的表达上调, 而且该分子具有与 C3 类似的、依赖分子内硫酯键的调理作用, 可以显著地促进吞噬细胞对革兰氏阴性菌的吞噬(Levashina *et al*, 2001; Blandin *et al*, 2004)。因此, 有研究者认为昆虫的 TEP 分子可替代补体 C3 并形成一个类似于脊椎动物补体系统的免疫防御系统。与此同时, 近年在太平洋螯虾(*Pacifastacus leniusculus*)、栉孔扇贝(*Chlamys farreri*)和太平洋牡蛎中也发现了多个 TEP 分子(Zhang *et al*, 2007; Wu *et al*, 2012; Wang *et al*, 2017)。但目前其功能研究还较少, 尚不能确定其是否具有脊椎动物 C3 分子的类似功能。

免疫细胞表面的特异性补体受体可与相应的补体活性分子或调节因子结合, 介导补体调理作用等生物学功能。补体受体按其发现的先后顺序依次命名为 CR1、CR2、CR3、CR4 和 CR5 等。本研究中, 检索到 2 个马氏珠母贝 CR1 分子和 3 个 CR2 分子, 且其分子结构中均含有保守的 SCR 重复结构域, 与典型的 CR1 和 CR2 分子结构相一致(Ahearn *et al*, 1989); 基因表达量分析表明, 马氏珠母贝 5 个 CR 分子经溶藻弧菌刺激后其在血细胞中的表达量均显著上调(图 1), 推测其可能位于血细胞表面并与 C3b 或类似的补体活性片段结合共同发挥补体系统的功能。同时, 结合本研究中补体调节成分(H 因子)的发现, 我们也可以推测马氏珠母贝可能具有以 C3 或类似分子为中心组分的补体系统。

4 结论

随着多种无脊椎动物全基因组序列测定的完成以及高通量测序技术的应用, 无脊椎动物补体相关成分逐渐被发现, 随之有关其功能和激活途径的研究也更加深入。本研究中, 通过对马氏珠母贝 RNA-Seq 转录组注释信息的检索和生物信息学分析, 共获得 68 条补体样成分的 unigene, 包括 C1q、C4/C5/C6、TEP 家族蛋白、MBL、CTL、SP、CR、Ficolin 以及补体因子 H 等; 虽然 RNA-Seq 数据中未发现典型的补体 C3、MASP 等关键补体成分, 但结合太平洋牡蛎、沟纹蛤仔等贝类补体成分的研究结果, 可以推测贝类具有补体的凝集素途径和替代途径;

同时, 这些激活途径在进化上经历了较大的变化, 其激活的具体方式仍需进一步研究; 此外, 马氏珠母贝 C6 样分子的发现也提示贝类补体终末途径存在的可能性。

参 考 文 献

- Ahearn J M, Fearon D T, 1989. Structure and function of the complement receptors, CR1 (CD35) and CR2 (CD21). *Advances in Immunology*, 46: 183—219
- Anders S, Huber W, 2010. Differential expression analysis for sequence count data. *Genome Biology*, 11(10): R106
- Azumi K, De Santis R, De Tomaso A *et al*, 2003. Genomic analysis of immunity in a Urochordate and the emergence of the vertebrate immune system: “waiting for Godot”. *Immunogenetics*, 55(8): 570—581
- Beutler B, 2004. Innate immunity: an overview. *Molecular Immunology*, 40(12): 845—859
- Blandin S, Levashina E A, 2004. Thioester-containing proteins and insect immunity. *Molecular Immunology*, 40(12): 903—908
- Bonura A, Vizzini A, Salerno G *et al*, 2009. Isolation and expression of a novel MBL-like collectin cDNA enhanced by LPS injection in the body wall of the ascidian *Ciona intestinalis*. *Molecular Immunology*, 46(11—12): 2389—2394
- Dishaw L J, Smith S L, Bigger C H, 2005. Characterization of a C3-like cDNA in a coral: phylogenetic implications. *Immunogenetics*, 57(7): 535—548
- Fujita T, Endo Y, Nonaka M, 2004a. Primitive complement system-recognition and activation. *Molecular Immunology*, 41(2—3): 103—111
- Fujita T, Matsushita M, Endo Y, 2004b. The lectin-complement pathway-its role in innate immunity and evolution. *Immunological Reviews*, 198(1): 185—202
- Gerdol M, Venier P, Pallavicini A, 2015. The genome of the Pacific oyster *Crassostrea gigas* brings new insights on the massive expansion of the C1q gene family in Bivalvia. *Developmental & Comparative Immunology*, 49(1): 59—71
- Grabherr M G, Haas B J, Yassour M *et al*, 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature Biotechnology*, 29(7): 644
- Gross P S, Clow L A, Smith L C, 2000. SpC3, the complement homologue from the purple sea urchin, *Strongylocentrotus purpuratus*, is expressed in two subpopulations of the phagocytic coelomocytes. *Immunogenetics*, 51(12): 1034—1044
- He Y N, Tang B, Zhang S C *et al*, 2008. Molecular and immunochemical demonstration of a novel member of Bf/C2 homolog in amphioxus *Branchiostoma belcheri*: implications for involvement of hepatic cecum in acute phase response. *Fish & Shellfish Immunology*, 24(6): 768—778
- Hibino T, Loza-Coll M, Messier C *et al*, 2006. The immune gene repertoire encoded in the purple sea urchin genome. *Developmental Biology*, 300(1): 349—365

- Huang H Q, Huang S F, Yu Y C *et al*, 2011. Functional characterization of a ficolin-mediated complement pathway in amphioxus. *Journal of Biological Chemistry*, 286(42): 36739—36748
- Huang S F, Yuan S C, Guo L *et al*, 2008. Genomic analysis of the immune gene repertoire of amphioxus reveals extraordinary innate complexity and diversity. *Genome Research*, 18(7): 1112—1126
- Huber-Lang M, Sarma J V, Zetoune F S *et al*, 2006. Generation of C5a in the absence of C3: a new complement activation pathway. *Nature Medicine*, 12(6): 682—687
- Ji X, Azumi K, Sasaki M *et al*, 1997. Ancient origin of the complement lectin pathway revealed by molecular cloning of mannan binding protein-associated serine protease from a urochordate, the Japanese ascidian, *Halocynthia roretzi*. *Proceedings of the National Academy of Sciences of the United States of America*, 94(12): 6340—6345
- Kimura A, Sakaguchi E, Nonaka M, 2009. Multi-component complement system of Cnidaria: C3, Bf, and MASP genes expressed in the endodermal tissues of a sea anemone, *Nematostella vectensis*. *Immunobiology*, 214(3): 165—178
- Levashina E A, Moita L F, Blandin S *et al*, 2001. Conserved role of a complement-like protein in phagocytosis revealed by dsRNA knockout in cultured cells of the mosquito, *Anopheles gambiae*. *Cell*, 104(5): 709—718
- Li B, Dewey C N, 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, 12: 323
- Matsushita M, Endo Y, Hamasaki N *et al*, 2001. Activation of the lectin complement pathway by ficolins. *International Immunopharmacology*, 1(3): 359—363
- Miller D J, Hemmrich G, Ball E E *et al*, 2007. The innate immune repertoire in Cnidaria-ancestral complexity and stochastic gene loss. *Genome Biology*, 8(4): R59
- Mortazavi A, Williams B A, McCue K *et al*, 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, 5(7): 621—628
- Nonaka M, Miyazawa S, 2001. Evolution of the initiating enzymes of the complement system. *Genome Biology*, 3(1): reviews1001.1
- Nonaka M, Yoshizaki F, 2004. Evolution of the complement system. *Molecular Immunology*, 40(12): 897—902
- Nonaka M, Kimura A, 2006. Genomic view of the evolution of the complement system. *Immunogenetics*, 58(9): 701—713
- Prado-Alvarez M, Rotllant J, Gestal C *et al*, 2009. Characterization of a C3 and a factor B-like in the carpet-shell clam, *Ruditapes decussatus*. *Fish & Shellfish Immunology*, 26(2): 305—315
- Schmidt B Z, Colten H R, 2000. Complement: a critical test of its biological importance. *Immunological Reviews*, 178(1): 166—176
- Smith L C, Clow L A, Terwilliger D P, 2001. The ancestral complement system in sea urchins. *Immunological Reviews*, 180(1): 16—34
- Tagawa K, Yoshihara T, Shibata T *et al*, 2012. Microbe-specific C3b deposition in the horseshoe crab complement system in a C2/factor B-dependent or-independent manner. *PLoS One*, 7(5): e36783
- Wang L L, Wang L L, Kong P F *et al*, 2012a. A novel C1qDC protein acting as pattern recognition receptor in scallop *Argopecten irradians*. *Fish & Shellfish Immunology*, 33(2): 427—435
- Wang L L, Wang L L, Zhang H *et al*, 2012b. A C1q domain containing protein from scallop *Chlamys farreri* serving as pattern recognition receptor with heat-aggregated IgG binding activity. *PLoS One*, 7(8): e43289
- Wang L L, Zhang H, Wang L L *et al*, 2017. The RNA-seq analysis suggests a potential multi-component complement system in oyster *Crassostrea gigas*. *Developmental & Comparative Immunology*, 76: 209—219
- Wang Z L, Wang B, Chen G *et al*, 2016. Transcriptome analysis of the pearl oyster (*Pinctada fucata*) hemocytes in response to *Vibrio alginolyticus* infection. *Gene*, 575(2): 421—428
- Wu C L, Noonin C, Jiravanichpaisal P *et al*, 2012. An insect TEP in a crustacean is specific for cuticular tissues and involved in intestinal defense. *Insect Biochemistry and Molecular Biology*, 42(2): 71—80
- Yoshizaki F Y, Ikawa S, Satake M *et al*, 2005. Structure and the evolutionary implication of the triplicated complement factor B genes of a urochordate ascidian, *Ciona intestinalis*. *Immunogenetics*, 56(12): 930—942
- Yu Y H, Huang H Q, Wang Y *et al*, 2008. A novel C1q family member of amphioxus was revealed to have a partial function of vertebrate C1q molecule. *The Journal of Immunology*, 181(10): 7024—7032
- Zhang G F, Fang X D, Guo X M *et al*, 2012. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature*, 490(7418): 49—54
- Zhang H, Song L S, Li C H *et al*, 2007. Molecular cloning and characterization of a thioester-containing protein from Zhikong scallop *Chlamys farreri*. *Molecular Immunology*, 44(14): 3492—3500
- Zhang H, Song L S, Li C H *et al*, 2008. A novel C1q-domain-containing protein from Zhikong scallop *Chlamys farreri* with lipopolysaccharide binding activity. *Fish & Shellfish Immunology*, 25(3): 281—289

ANALYSIS OF COMPLEMENT-LIKE COMPONENTS IN *PINCTADA FUCATA* BASED ON THE TRANSCRIPTOME DATASETS

WU Yong¹, ZHANG Ye², XU You-Hou³, LIANG Xue-Ru², WANG Zhong-Liang²

(1. Science and Technology Department, Guangdong Ocean University, Zhanjiang 524088, China; 2. Fisheries College, Guangdong Ocean University, Zhanjiang 524088, China; 3. Guangxi Key Laboratory of Beibu Gulf Marine Biodiversity Conservation, Beibu Gulf University, Qinzhou 535000, China)

Abstract The complement system connecting innate immunity and adaptive immunity plays an important role in humoral immunity. In this study, the complement components of *Pinctada fucata* were searched based on the RNA-Seq data, and the conserved domains and expression levels after *Vibrio alginolyticus* challenge were analyzed. Sixty-eight complement-like component unigenes were found in RNA-Seq data, encoding 14 C1q domain-containing proteins, 14 lectins, 4 fibrinogen-related proteins, 12 serine proteases, 10 thioester-containing proteins, 1 C6, 5 complement receptors, 1 complement factor, and 7 ficolins. The complement-like components of *P. fucata* contained corresponding conserved domains as revealed in domain analysis, which is consistent with the results of complement components from other mollusks. The expression levels of 21 unigenes were up-regulated by *V. alginolyticus* challenge, of which 12 were pattern recognition receptors. Combining the results of studies on the oyster and other mollusks, we infer that a primitive complement system containing multiple components exists in *P. fucata*, and performs biological functions by the lectin pathway or alternative pathway. The results provide a new basis for understanding the immune defense mechanisms of *P. fucata* and other marine invertebrates, and for exploring the evolution of the complement system.

Key words *Pinctada fucata*; RNA-Seq; complement-like component; conserved domain