



长牡蛎中 I84 蛋白酶抑制因子家族的扩张和功能分化

吴钱江¹, 刘 圣^{1,2}, 鲁佳丽^{1,3}, 范聪灵¹, 叶 静^{1,3},
阮紫燕¹, 林志华^{1,2}, 薛清刚^{1,2*}

(1. 浙江万里学院, 浙江省水产种质资源高效利用技术研究重点实验室, 浙江 宁波 315100;

2. 浙江万里学院, 宁海海洋生物种业研究院, 浙江 宁波 315100;

3. 上海海洋大学, 水产科学国家级实验教学示范中心, 上海 201306)

摘要: 蛋白酶抑制因子是极其多样的蛋白质或多肽, 能抑制蛋白酶的水解活性。研究发现蛋白酶抑制因子能通过抑制病原蛋白酶活性, 从而抑制病原的入侵。I84 蛋白酶抑制因子家族是 MEROPS 数据库中新增的一个家族, 其部分成员在免疫防御过程中的作用得到了一定的揭示。为探究 I84 蛋白酶抑制因子家族在长牡蛎中的分布和功能状况, 实验鉴定了长牡蛎中 23 个潜在的 I84 家族基因, 根据系统进化树分析, 挑选了 5 个同源基因进行时空表达和功能分析。首先, 利用克隆技术验证了长牡蛎中 5 个 I84 家族同源基因 *CgSi3*、*CgSi5*、*CgSi6*、*CgSi16* 和 *CgSi19* 可表达性。结果显示, 5 个基因在外套膜、闭壳肌、性腺、血淋巴细胞、肝胰腺和鳃等 6 个组织中均表达, 但肝胰腺中表达量显著高于其他组织。同时, 5 个基因在长牡蛎幼体不同发育阶段表达模式不同, 其中各基因在受精卵时期均不表达, *CgSi3* 表达量则在眼点幼虫期显著上升后下降, 而 *CgSi6* 在壳顶幼虫期开始表达后, 表达量随长牡蛎发育而持续增加。另外, 对牡蛎进行病原相关模式分子 (PAMPs) 注射后, 5 个基因也表现出不同的表达模式, 其中 LPS 和 PGN 注射后 *CgSi6* 表达量变化明显, 而 poly (I:C) 和 GLU 注射后 *CgSi3* 表达量变化明显, 且 *CgSi3* 和 *CgSi6* 在不同刺激下表达模式也存在差异。研究表明, I84 蛋白酶抑制因子家族在长牡蛎中发生了明显的家族扩张, 而且扩张形成的同源基因在功能上产生分化。本研究为全面认识 I84 家族蛋白酶抑制因子生物学功能及相关机理提供了依据。

关键词: 长牡蛎; I84 蛋白酶抑制因子家族; 基因鉴定; 基因表达; 功能分化

中图分类号: Q 786; S 917.4

文献标志码: A

蛋白酶抑制因子是一类广泛分布于各种生物体内的蛋白质或多肽。它们调节生物体内蛋白酶活性, 以防止蛋白酶水解作用对宿主细胞和组织造成破坏^[1]。大多数蛋白酶抑制因子选择性抑制某一类蛋白酶, 并依据其抑制特异性分为金属蛋白酶抑制因子、丝氨酸蛋白酶抑制因子、天冬氨

酸蛋白酶抑制因子和半胱氨酸蛋白酶抑制因子等^[2]。蛋白酶抑制因子的一项生物学功能是参与机体免疫反应。一方面, 它们通过调节内源蛋白酶活性, 在炎症反应、血液凝固、补体级联反应、抗菌肽产生、酚氧化酶激活和免疫细胞激活等自我防御过程中发挥作用^[3-6]。另一方面, 许多生物体内的

收稿日期: 2021-04-27 修回日期: 2021-11-03

资助项目: 国家自然科学基金 (32073010, 31672629); 宁波市“泛 3315”创新团队项目; 国家现代农业产业技术体系专项 (CARS-49); 浙江省属高校基本科研业务费专项 (SC1032080180110)

第一作者: 吴钱江 (照片), 从事生物工程研究, E-mail: 627597232@qq.com

通信作者: 薛清刚, 从事贝类免疫学研究, E-mail: qxue@zwu.edu.cn



蛋白酶抑制因子通过灭活病原毒力蛋白酶或直接抑制病原参与机体抗感染免疫过程^[7-8]。蛋白酶抑制因子被认为是各种生物的重要免疫反应因子。

MEROPS 蛋白酶与蛋白酶抑制因子数据库 (<https://www.ebi.ac.uk/merops/>) 是一个基于分子进化关系建立的蛋白酶和蛋白酶抑制因子分类系统。I84 家族是 MEROPS 中一个丝氨酸蛋白酶抑制因子家族。它是在美洲牡蛎 (*Crassostrea virginica*) 中发现的 2 个有独特丝氨酸蛋白酶抑制活性的新蛋白质分子基础上建立的^[9-10]。目前已经报道的 I84 家族成员包括美洲牡蛎的 cvSI-1、cvSI-2、cvSI-3 和缢蛏 (*Sinonovacula constricta*) 的 scSI-1、scSI-2^[11]。已鉴定的 I84 家族蛋白酶抑制因子均为分子量约 8 ku 的多肽, 含有 12 个保守的半胱氨酸残基, 推测可能形成 6 个分子内二硫键^[9-10, 12]。它们特异性抑制丝氨酸蛋白酶, 对其他类型蛋白酶无抑制活性^[13]。多项研究结果提示, I84 家族蛋白酶抑制因子参与宿主抗感染和抗胁迫反应。例如, 纯化的 cvSI-1、cvSI-2 能有效灭活牡蛎病原体派金虫 (*Perkinsus marinus*) 分泌的毒力蛋白酶, 并抑制寄生虫繁殖^[9]。同时, I84 家族蛋白酶抑制因子基因在病原刺激或环境胁迫下表达量显著增加^[11]。另外, 一些对病原体抵抗力高的牡蛎群体和物种的基因表达水平显著高于敏感群体或物种^[10]。

转录水平研究显示, I84 蛋白酶抑制因子家族在多个贝类物种中有多个成员, 并且物种之间可能存在差异^[9-12]。然而, 迄今还没有关于该基因家族在物种内扩张与进化的研究。长牡蛎 (*C. gigas*) 是最重要的海水养殖贝类之一, 也是在分子水平研究贝类各种功能与进化的重要代表种^[14]。同时, 作为第一个完成全基因组测序的贝类物种, 其基因组序列的确定为全基因组水平鉴定 I84 蛋白酶抑制因子家族创造了条件。本实验拟对长牡蛎全基因组包含的全部 I84 家族基因进行序列水平系统鉴定, 并在系统分子进化分析的基础上, 通过比较代表性基因的时空表达和针对病原相关分子模式物 (pathogen-associated molecular patterns, PAMPs) 刺激的反应性表达模式, 来探讨该家族成员在长牡蛎中的功能分化。

1 材料与方法

1.1 实验材料

长牡蛎 (壳高 7~9 cm、壳长 4~6 cm、壳宽 3~5 cm), 取自山东威海桑沟湾的牡蛎养殖场, 置

于盐度 30 的人工海水中暂养适应 5 d。用牡蛎刀将牡蛎撬开小口, 以手术剪剪断闭壳肌并去除右壳, 使用 5 mL 注射器从围心腔内抽取血淋巴液, 3 000×g 离心 3min, 弃上清液, 沉淀为血淋巴细胞。随后血淋巴细胞由 Trizol 吹散, 并于液氮中速冻, 转移至-80 °C 冰箱内保存备用。血淋巴液经取样后, 用解剖刀分别切取约 0.1 g 肝胰腺、闭壳肌、外套膜、鳃、性腺等组织, 分别置于液氮中快速冷冻, 并于-80 °C 冰箱内保存。

长牡蛎不同幼虫发育期样本于 2020 年人工繁殖期间, 取自宁波市海洋与渔业研究院咸祥试验基地育苗车间。所取样本包括受精卵、D 型幼虫、壳顶幼虫、眼点幼虫和稚贝。所有样本均在液氮中速冻, 并置于-80 °C 冰箱内保存。实验过程中操作人员严格遵守实验动物相关伦理规范。

1.2 同源基因查询、鉴定和系统进化分析

以已报道的 5 个 I84 家族成员氨基酸序列为查询序列, 在长牡蛎基因组 (GCF_902806645.1) 中进行在线 BLASTP 查询, 然后将查询结果中所有注释为 cvSI 类似基因序列进行下载分析。蛋白质的分子量、等电点、亲水性、不稳定指数等理化特性在 ExPASy 的 ProtParam 模块 (<https://web.expasy.org/protparam/>) 中预测, 氨基酸序列比对在 clustalX 软件中完成^[15]。其氨基酸相似性和一致性在 SMS 工具 (<http://www.bio-soft.net/sms/>) 中获得。所有候选多肽序列在 MEGA X 软件中进行系统发育分析, 最大似然法 (maximum likelihood method) 模型下, 使用离散的 Gamma 分布来模拟位点间的进化率差异, 允许 20% 的对齐间隙和模糊碱基, 自举值 (Bootstrap) 为 1 000^[16]。

1.3 cDNA 克隆

按照试剂盒说明书, 使用 TRIzol 法 (OMEGA, 美国) 提取长牡蛎幼虫及成体各组织总 RNA, 通过 1% 琼脂糖凝胶电泳鉴定完整性, 根据紫外分光光度计 (奥盛, 中国) 测定 OD_{260/280} 值。以 500 ng 总 RNA 为模板, 按照 PrimeScript™ RT Master Mix 反转录试剂盒 (TaKaRa, 日本) 说明书操作合成 cDNA 单链, 并以此为模板进行 PCR 扩增。依据基因组数据库获取的基因序列片段, 利用 primer 5 软件设计基因特异性引物克隆验证开放阅读框序列 (表 1)。PCR 反应体系包含上下游引物各 1 μL, 模板 cDNA 1 μL (850 ng/μL), 无菌水 9.5 μL, Taq DNA 聚合酶 12.5 μL (生工生物工程上海股份有限公司), 反应过程为预变性 94 °C 5 min,

按照变性 94 °C 45 s、退火 60 °C 45 s、延伸 72 °C 45 s, 35 个循环扩增, 最后 72 °C 延伸 10 min。PCR 产物经 1% 琼脂糖凝胶电泳鉴定、纯化后连接到 T1 质粒载体, 转化入 DH5 α 感受态细胞, 挑取单菌落进行序列测定 (生工生物工程上海股份有限公司)。

表 1 本实验所用引物

Tab. 1 Primers used for cloning and characterization in this study

基因 gene	用途 usage	序列 primer sequence (5'-3')
CgSi3	ORF克隆	F: TCCGCGCTAAGATTCTAGGTT R: GATTATTTTTGGTTTTGG
	qPCR	F: CTGTGTTCTTGTGGCAGT R: CATCATTGTTTACGCATCC
CgSi5	ORF克隆	F: TGCAGTGGACAACTTGAG R: GTATTTATTTATCAAGGATC
	qPCR	F: CCTGGCTGTGTGCT R: AGGGAATCGGCGTCA
CgSi6	ORF克隆	F: GCCCCATATAACTCACTAT R: CGTACAAAATTCAGAGGAGTC
	qPCR	F: ATTCTCCTCGCCGTTG R: CCAAACAGTTAGCAAGGG
CgSi16	ORF克隆	F: ATGCTGGTGTACGTATAGA R: ACAAACACCACGTGGCC
	qPCR	F: TCGTGTGTCTTTGGGG R: ATCCGTCTCTCCGTGG
CgSi19	ORF克隆	F: TCGTAAACTTCCACC R: CCGGTATAGCAAAATTTATT
	qPCR	F: CAAAAATGAGGACTGCC R: TAATGGTGGTCACACTCTATG
EF	qPCR	F: AGTCACCAAGGCTGCACAGAAAG R: TCCGACGTATTCTTTGCGATGT

1.4 病原相关分子模式物刺激

4 种病原相关分子模式物脂多糖 (LPS)、肽聚糖 (PGN)、葡聚糖 (GLU) 和聚胞肌酐酸 [poly(I:C)] 分别用无菌海水相应配成 1、2、1 和 2 mg/mL 溶液备用。无菌海水以自来水溶解海水晶 (蓝海星, 杭州) 至盐度 30, 并用 0.45 μ m 滤膜过滤除菌。取暂养 5 d 的长牡蛎 640 只, 平均分为 LPS、PGN、GLU 和 poly(I:C) 刺激组, 每组 160 只, 其中每组分为实验组和对照组, 各 80 只。将长牡蛎在加入 5% MgCl₂ (质量体积比) 的海水麻醉溶液中处

理 12 h 使其壳张开, 用微量进样器在每只牡蛎闭壳肌内注射 50 μ L 刺激物 (对照组注射等量的无菌海水)。注射后, 将牡蛎置于盐度 30 的海水中, 分别于 0、3、6、12、24、48 和 72 h, 每组随机取 6 只牡蛎肝胰腺组织, 液氮内冷冻后, -80 °C 低温冰箱保存直至提取总 RNA 进行基因表达分析。

1.5 实时荧光定量 PCR

依据 cDNA 克隆确定的核苷酸序列设计荧光定量引物 (表 1), 以延伸因子 (EF) 为内参基因, 以前述方法合成的总 cDNA 为模板, 利用实时荧光定量 PCR 测定有关 I84 家族基因表达。PCR 反应体系包括荧光染料 (iTagTM Universal SYBR[®] Green Supermix) 10 μ L, 上下游引物各 1 μ L, 1 : 50 稀释 cDNA 模板 8 μ L, 反应条件为 95 °C 预变性 30 s, 95 °C 变性 5 s, 55 °C 退火 30 s, 共 40 个循环。每个样本测定 3 次。基因相对表达量按 2^{- $\Delta\Delta$ C_t} 法计算。测定结果利用 IBM Spss 软件做单因素方差分析, Duncan 氏多重比较检验以确定差异显著性。

2 结果

2.1 长牡蛎基因组包含的 I84 家族基因及其编码的多肽分子

在长牡蛎基因组共鉴定出 23 个标注为 I84 家族类似基因同源序列, 并确认是染色体上的不同基因。按照其在染色体上的分布顺序, 分别命名为 CgSi1~CgSi23 (表 2)。其中 CgSi3 在基因组序列版本 GCF_902806645.1 中已被注销, 但在版本 GCF_000297895.1 中仍然存在, 而且 cDNA 克隆也证实了其可表达性。这些序列预测的开放阅读框编码 78~116 aa 的多肽, 相应分子量为 8.14~16.76 ku, 等电点 4.46~6.95。所有候选多肽 cvSI-2 之间的氨基酸序列一致性大多高于 25%, $E < 0.01$ (E 值是在特定数据库随机条件下发生得分大于或等于当前比对得分的序列数目的期望值)。其中 CgSi23 与 cvSI-2 之间只有 12.2% 的氨基酸序列一致性, 但与 CgSi2 的一致性高于 75%。另外, 预测不同分子之间不稳定指数 (21.44~75.01) 和亲水性 (-0.193~0.479) 差异显著。

2.2 长牡蛎 I84 家族基因的系统进化关系

使用最大似然法建立长牡蛎 I84 家族成员系统进化树, 将鉴定到的 23 个成员分为 4 支, 每支包含 4~7 个基因 (图 1), 其自举值大多数大于 70,

表 2 长牡蛎 I84 蛋白酶抑制因子家族多肽性预测

Tab. 2 Predicted properties of *C. gigas* I84 family proteins

命名 name	蛋白编号 protein number	基因编号 gene number	长度/aa length	分子量 ¹⁾ /ku MW	等电点 pI	亲水性 hydrophilia	不稳定指数 index of instability	一致性 ²⁾ identity
CgSi1	XP_034328111.1	LOC105319549	78	8.14	5.84	0.479	47.24	27.8
CgSi2	XP_034326758.1	LOC109619486	85	9.28	6.95	0.111	47.77	75.3
CgSi3	XP_011441581.1	LOC105338259	95	10.22	6.19	-0.093	43.34	27.1
CgSi4	XP_011430295.2	LOC105330358	84	9.24	6.47	0.308	75.01	27.9
CgSi5	XP_011430294.2	LOC105330357	84	9.37	6.37	-0.067	49.73	37.2
CgSi6	XP_011426647.2	LOC105327731	91	9.85	6.10	0.427	32.12	25.8
CgSi7	XP_011431397.2	LOC105331058	93	9.97	5.74	-0.023	44.84	26.6
CgSi8	XP_011454757.1	LOC105347386	86	9.55	5.37	-0.012	74.25	33.3
CgSi9	XP_011454756.2	LOC105347384	87	9.55	4.46	0.270	35.04	31.1
CgSi10	XP_034336481.1	LOC105332954	88	9.38	5.88	-0.193	64.27	22.6
CgSi11	XP_011415275.1	LOC105319431	79	8.73	5.59	-0.080	31.51	26.4
CgSi12	XP_019924056.1	LOC105331113	85	9.33	6.37	0.404	29.53	28.4
CgSi13	XP_011431471.1	LOC105331114	84	8.96	5.62	0.413	36.32	26.4
CgSi14	XP_011419556.2	LOC105322508	116	12.4	6.62	0.268	38.02	23.3
CgSi15	XP_034318868.1	LOC105319782	90	9.49	6.34	0.181	52.29	31.6
CgSi16	XP_011426376.2	LOC105327527	87	9.65	5.35	0.129	64.59	33.0
CgSi17	XP_011426375.3	LOC105327526	116	12.83	5.54	-0.095	53.97	26.3
CgSi18	XP_034299566.1	LOC117680973	82	8.44	6.62	-0.017	42.81	34.4
CgSi19	XP_011417812.2	LOC105321246	83	8.50	5.88	0.213	48.23	34.4
CgSi20	XP_011417808.2	LOC117680298	85	9.11	6.62	0.045	21.44	34.1
CgSi21	XP_034299569.1	LOC105321245	85	9.08	6.34	-0.140	36.59	28.6
CgSi22	XP_019920031.2	LOC109617762	82	8.66	6.33	0.179	40.44	28.1
CgSi23	XP_011426646.1	LOC105327730	155	16.76	6.07	-0.028	35.66	12.2

注: CgSi3在长牡蛎基因组(GCF_902806645.1)中被注销, 经验证存在且表达。1)分子量为预测的成熟蛋白结果, 2)一致性为目标多肽与cvSI-2 BLASTP比较结果, 使用BLOSUM62默认参数。

Notes: CgSi3 gene has been deleted in the database, but it has been verified that the gene does exist and express. 1) molecular weight is the predicted result of mature protein, 2) consistency is the result of comparing the target peptide with cvSI-2 BLASTP, using the BLOSUM62 default parameter.

显示分支确认同源关系的可靠性。

2.3 长牡蛎 I84 家族代表基因的 cDNA 克隆验证

根据系统进化关系, 在每个分支中挑选 1~2 个基因 (CgSi3、CgSi5、CgSi6、CgSi16 和 CgSi19) 作为长牡蛎 I84 家族代表基因, 进行基因可表达性和序列准确性验证。cDNA 克隆获得了所有 5 个代表基因的开放阅读框, 测序结果显示它们与基因组预测的序列一致。图 2 和表 3 分别显示 cDNA 克隆确定的 5 个代表基因所编码氨基酸序列与 5 个已报道 I84 家族蛋白酶抑制因子的比较。结果显示长牡蛎中 5 个代表性基因与已报道成员之间具有较高的相似性 (32%~86%) 和一致性 (21%~83%), 且都含有 12 个保守的半胱氨酸残基。

2.4 长牡蛎 I84 家族代表基因的组织表达

定量 PCR 检测 CgSi3、CgSi5、CgSi6、CgSi16 和 CgSi19 在成体牡蛎 6 种不同组织中的表达水平。所有 5 个基因的相对表达量均为肝胰腺组织显著高于鳃、闭壳肌、性腺、外套膜和血淋巴细胞 ($P<0.05$)。而肝胰腺组织以外的 5 种组织之间未检测到基因表达水平的显著差异 (图 3)。在肝胰腺中 CgSi6 表达量显著高于其他 5 个基因 ($P<0.05$), 其余基因之间并无显著差异。

2.5 长牡蛎 I84 家族代表基因在不同幼虫发育期的表达

5 个代表基因在受精卵时期均未检测到表达。CgSi5、CgSi16 和 CgSi19 在 D 型幼虫、壳顶幼虫、眼点幼虫和稚贝中低量表达。CgSi3 表达量则在

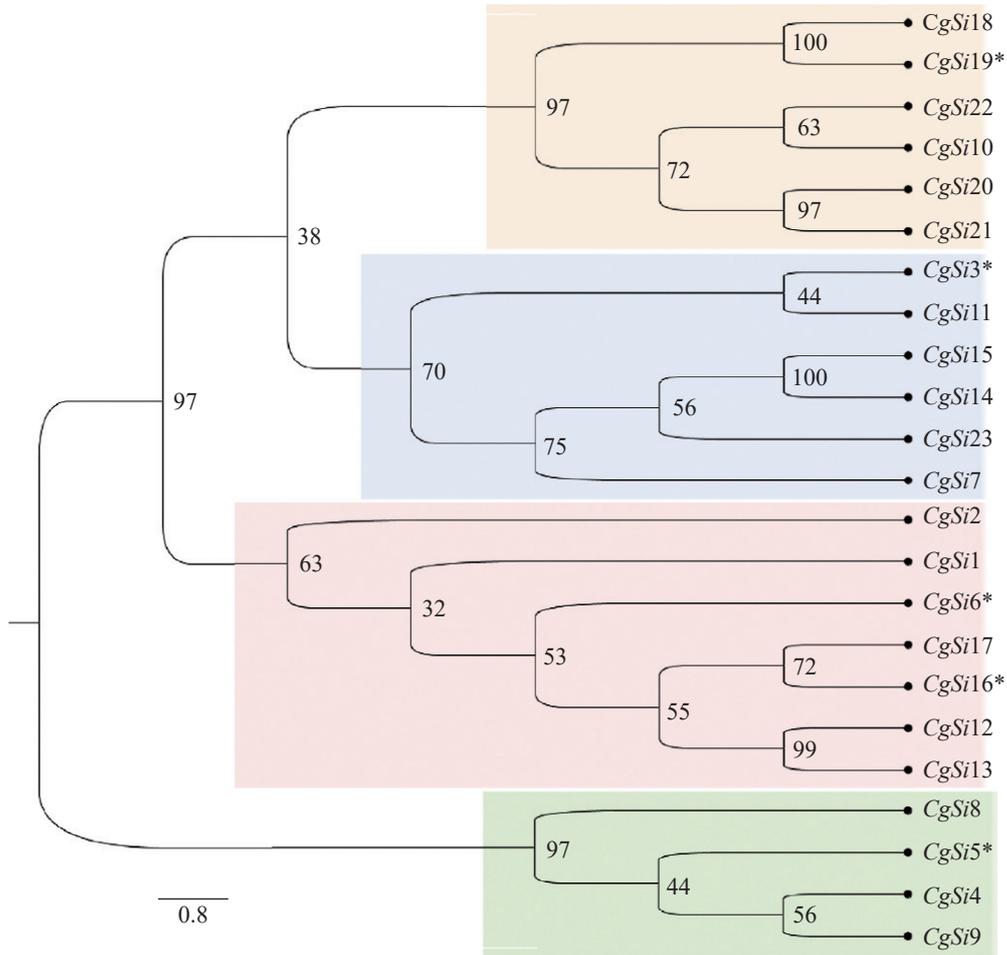


图 1 长牡蛎中 I84 蛋白酶抑制因子家族成员系统进化树

该系统进化树显示了长牡蛎中 I84 蛋白酶抑制因子家族基因间的同源性, 星号表示本研究验证的基因, 不同颜色代表不同分支。

Fig. 1 Phylogenetic tree of I84 family genes in *C. gigas*

Phylogenetic tree shows homology among genes of the I84 protease inhibitor family in *C. gigas*, with asterisks indicating the genes validated in this study; different colors represent different branches.

眼点幼虫期显著上升 ($P < 0.05$), 到稚贝期回落。而 CgSi6 表达量从壳顶幼虫时期开始显著增加 ($P < 0.05$), 并随幼虫发育而继续升高, 其 mRNA 水平在稚贝期达到最高 (图 4)。

2.6 长牡蛎 I84 家族代表基因在 PAMPs 刺激下的反应性表达

5 个检测基因在不同的 PAMPs 刺激下, 表现出不同的表达模式。在 poly(I:C) 刺激下, CgSi3 在 3 h 表达量显著上升 ($P < 0.01$), 之后在 6 h 表达量恢复正常水平, 其余基因表达量无显著变化 (图 5-a)。在 LPS 刺激下, CgSi6 在 3 h 表达量极显著上升 ($P < 0.01$), 在 6 h 表达量恢复正常水平, 其余基因表达量无显著变化 (图 5-b)。在 PGN 刺激下, CgSi6 表达量在 3 h 显著下降 ($P < 0.05$), 之后一直保持低水平稳定表达, 其余基因表达量无

显著变化 (图 5-c)。在 GLU 刺激下 (图 5-d), CgSi3 表达量在 3 h 显著上升 ($P < 0.05$), 在 6 h 恢复至初始水平, 在 12 h 显著上升 ($P < 0.05$) 到最高, 24 h 后恢复正常水平, 其余基因表达量无显著变化。

3 讨论

许多有重要生物学功能的蛋白质通过不断分子进化形成多成员家族, 以适应其在不同物种中的功能需求。在物种水平上系统研究这些多成员家族, 对于认识其功能和家族形成机理有重要意义。I84 蛋白酶抑制因子家族是一个新的蛋白质家族, 部分前期研究结果已经提示了它在一些软体动物中以多基因形式存在^[9-12]。本研究从全基因组水平全面分析了 I84 家族基因在长牡蛎中的分布和基本结构特征, 为深入认识该家族蛋白质的

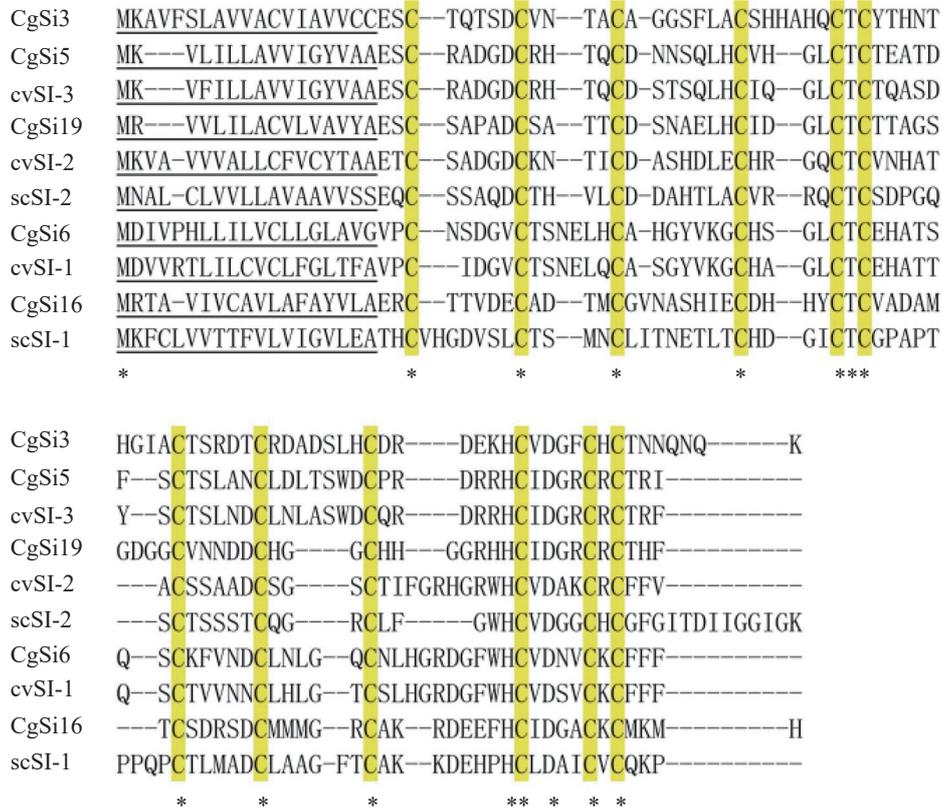


图 2 长牡蛎中 184 家族代表成员氨基酸序列比对

高亮标记为保守的半胱氨酸残基, 下划线部分为预测信号肽区域, *代表相同的氨基酸。

Fig. 2 Amino acid sequence alignment of 184 family representative members in *C. gigas*

The conserved cysteine residues are highlighted, the underlined region is the region of the predicted signal peptide, * stands for the same amino acid.

表 3 长牡蛎中 184 蛋白酶抑制因子家族氨基酸比对

Tab. 3 Amino acid sequences alignment of 184 family members with the target genes

	CgSi5	CgSi16	CgSi3	CgSi6	CgSi19	cvSI-1	cvSI-2	cvSI-3	scSI-1
CgSi5		4.00E-12	2.00E-12	2.00E-08	2.00E-19	1.00E-10	4.00E-18	2.00E-53	2.00E-07
CgSi16	23.4/35.1		6.00E-07	4.00E-08	5.00E-05	2.00E-07	2.00E-14	2.00E-11	1.00E-09
CgSi3	37.9/49.5	23.2/35.4		1.00E-02	1.00E-06	2.00E-04	3.00E-09	4.00E-14	4.00E-04
CgSi6	30.8/42.9	20.0/34.7	23.2/37.4		5.00E-04	7.00E-50	1.00E-15	5.00E-14	3.00E-08
CgSi19	48.3/63.2	24.5/33.0	31.3/41.8	28.0/39.8		2.00E-06	9.00E-17	1.00E-25	9.00E-07
cvSI-1	35.2/42.9	21.1/32.6	24.2/37.4	74.7/81.3	26.9/38.7				
cvSI-2	39.5/51.2	36.6/37.2	28.4/42.1	28.6/44.0	37.1/51.7				
cvSI-3	83.3/86.9	22.3/36.2	35.8/46.3	34.1/46.2	51.7/67.8				
scSI-1	28.0/39.8	29.8/47.9	23.2/32.3	25.5/37.2	25.5/40.4				

注: 对角线以下为序列间的一致性/相似性, 对角线以上为一致性的E值。

Notes: below the diagonal line is the consistency/similarity between the sequences, above the diagonal line is the E value.

生物学功能和分子进化提供了信息。

基因家族是具有共同祖先的一类基因集合, 其家族成员识别主要是根据氨基酸序列的一致性^[17]。本研究通过序列比对, 发现在长牡蛎基因组内有 23 个 184 家族同源基因, 其预测的物理性

质和氨基酸序列与 184 家族成员高度相似。其中, 12 个保守的半胱氨酸尤其具有特征性意义。氨基酸序列中半胱氨酸排列以及由此形成的分子内二硫键模式保守性是多个丝氨酸蛋白酶抑制因子家族所具有的特征, 这一特征曾作为蛋白酶抑制因

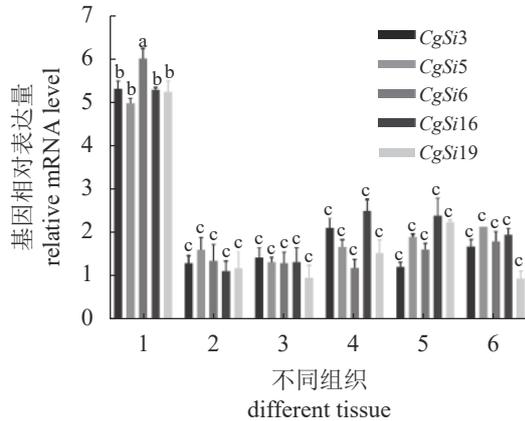


图3 长牡蛎中 I84 家族代表成员在不同组织中的相对表达量

1.肝胰腺, 2.鳃, 3.闭壳肌, 4.性腺, 5.外套膜, 6.血淋巴细胞。小写字母表示各组织间的显著性差异 ($P < 0.05$)。图中相对表达量数值为 lg 转化后的值。

Fig. 3 Relative expression levels of I84 family members in different tissues in *C. gigas*

1. digestive gland, 2. gills, 3. adductor muscle, 4. gonad, 5. mantle, 6. hemolymph cell. Different letters represent significant differences between tissues ($P < 0.05$). The relative expression values in the figure were lg transformed.

子分类的重要依据^[2]。但 I84 家族蛋白酶抑制因子的半胱氨酸保守性特征与已知的其他蛋白酶抑制因子家族不同。在鉴定到的 23 个 I84 家族同源基因中, 挑选了 5 个代表性基因进行 cDNA 克隆, 结果显示它们是能够表达的真基因。组织表达分析结果表明, 它们也与美洲牡蛎和缢蛭中确定的 I84 家族成员一致, 都在肝胰腺内高表达^[9-12]。这些均与已知 I84 家族的分子特征相符。

当一个物种基因组中包含同一个基因家族的 23 个同源基因时, 则该基因家族必定在相应物种发生了显著扩张。本团队近期的另一项研究发现, I84 家族蛋白酶抑制因子似乎只存在于软体动物 (Mollusca) 中^[18]。除长牡蛎外, 在欧洲扇贝 (*Pecten maximus*)、虾夷扇贝 (*Mizuhopecten yessoensis*)、美洲牡蛎基因组中分别发现了 16、5 和 29 个 I84 家族同源基因, 也证实了 I84 家族在贝类中的扩张现象。至于 I84 蛋白酶抑制因子家族在长牡蛎中的扩张究竟是发生在物种形成之前还是之后, 需要进一步研究。研究发现, 很多与物种抗感染免疫等相关的基因家族都发生扩张, 以适应物种对基因相关功能的需求。例如, 海胆 Toll 样受体 (TLR)、补体系统的关键组件、肽聚糖识别蛋白 (PGRP) 和革兰氏阴性菌结合蛋白 (GNBP) 等基因

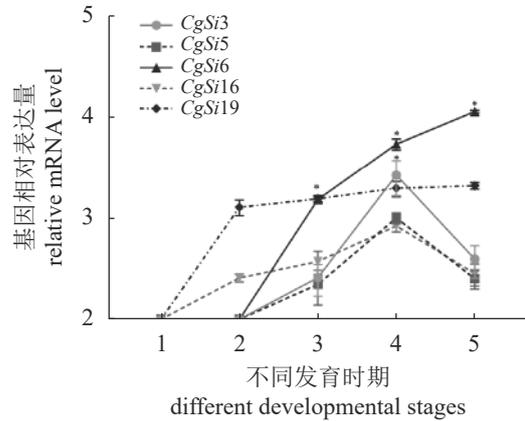


图4 长牡蛎中 I84 家族代表成员不同发育时期的相对表达量

1.受精卵, 2.D型幼虫, 3.壳顶幼虫, 4.眼点幼虫, 5.稚贝。星号表示组间差异显著 ($P < 0.05$), 下同。

Fig. 4 Relative expression levels of I84 family in *C. gigas* at different developmental stages

1. fertilized eggs, 2. D-shaped larvae, 3. umbo larvae, 4. pediveliger larvae, 5. juvenile oysters. Asterisks represent significant difference in the expression of each gene at different developmental stages ($P < 0.05$), the same below.

家族均发现有大量扩张^[19]。同样, 长牡蛎基因组中的 Toll 样受体、热休克蛋白 70 和细胞凋亡抑制因子等也发生了极为显著的基因家族扩张^[13, 20]。这也间接提示, I84 蛋白酶抑制因子家族的功能与相关物种的生存适应有关。

显然, I84 蛋白酶抑制因子家族在长牡蛎中的扩张并非简单基因复制形成的基因数目增加。本研究中, 长牡蛎 I84 家族基因被分成不同的支, 说明基因复制产生的新基因之间发生了序列变化, 或者基因结构上的分化。同时, 实验观察到不同家族基因表达模式存在显著差别。这些差别既表现在不同表达与不同幼体发育阶段, 也存在于成体牡蛎相同组织内不同基因的表达量, 说明相关基因发生了功能分化。基因功能的分化是基因在扩张的过程中编码区出现差异所导致的, 其重要特征表现为同一物种内旁系同源基因间表达模式的差异^[17]。从进化原理上讲, 基因复制导致基因家族扩张后, 新形成的家族基因在基因相关生物学功能的选择压力下发生结构和功能分化。而在物种水平上, 这种分化则构成相应物种在面临刺激和胁迫时的一种进化响应机制^[21]。因此, 尽管迄今的研究结果尚不足以对 I84 蛋白酶抑制因子家族的生物学功能做出全面系统的判断, 但基因家族在长牡蛎物种内的快速扩张及其后的功能分

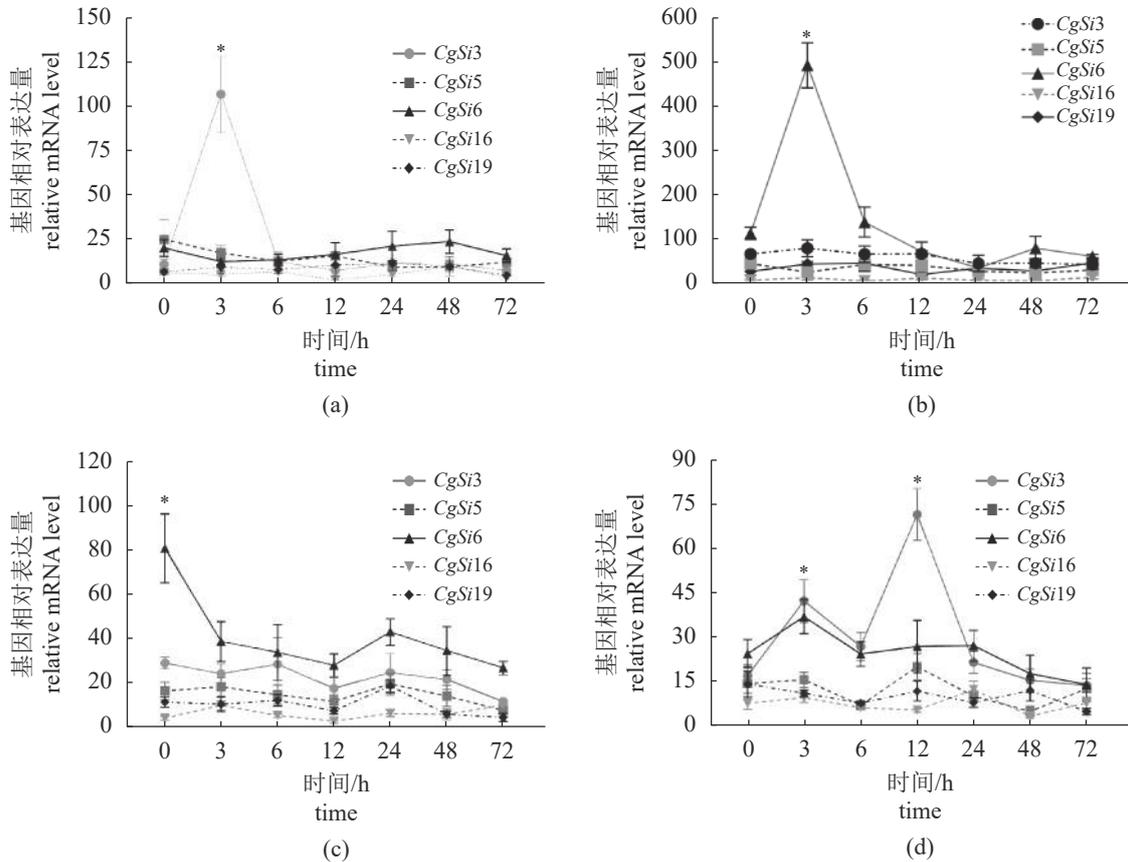


图 5 病原相关分子模式物刺激下长牡蛎中 I84 家族代表基因的相对表达量

(a) 长牡蛎分别在聚肌胞苷酸, (b) 脂多糖, (c) 肽聚糖, (d) 葡聚糖, 刺激下肝胰腺中 I84 家族代表基因相对表达量。

Fig. 5 Relative expression levels of representative genes of I84 family in *C. gigas* stimulated by PAMPs

Expression of the gene was changed by poly(I:C) (a), LPS (b), PGN (c) and GLU (d) stimulation.

化足以说明它们参与的生物学过程对于相应物种的生存适应至关重要, 值得深入研究。

在一个充满病原微生物的环境中, 生物的自我防御是其维持生存的最基本条件之一。有关 I84 家族蛋白酶抑制因子的一些前期研究发现, 部分家族成员有抑制病原体的体外活性, 并在病原微生物刺激后的响应性表达量增加^[11]。本研究发现, 4 种保守病原体相关分子模式 (PAMPs) 物刺激后, 长牡蛎部分 I84 家族基因表达量诱导性增加, 从而支持该家族蛋白质参与宿主抗病原反应的功能性。对于无脊椎动物所具有的先天免疫而言, 宿主的病原体相关模式识别受体 (PAMR) 对病原微生物 PAMPs 的识别, 是涉及细胞内信号传导途径激活、触发促炎症和抗菌效应物产生等防御反应的开始^[22]。一般认为, 蛋白酶抑制因子的高表达可能有助于机体清除侵入的病原体。而长牡蛎 I84 家族蛋白酶抑制因子基因应对 PAMPs 刺激所表现表达量上调, 提示有关蛋白质分子可能

在调节宿主免疫识别过程中也发挥作用。而这进一步为 I84 家族蛋白酶抑制因子作为免疫反应因子参与宿主自我防御过程提供了新的依据。

本实验证实了长牡蛎基因组包含 23 个 I84 家族同源基因, 说明该家族发生了显著扩张。同时, 旁系同源基因之间存在的表达模式差异提示家族扩张后的基因发生了结构与功能分化。而部分 I84 家族基因针对 PAMPs 刺激出现的反应性表达量增加进一步证实该家族的生物学功能与宿主防御反应有关。下一步将在本实验基础上深入探讨 I84 家族蛋白酶抑制因子参与的具体反应过程及相关机理, 为全面认识其生物学功能提供依据。

(作者声明本文无实际或潜在的利益冲突)

参考文献 (References):

[1] Rawlings N D, Barrett A J, Bateman A. Asparagine peptide lyases: a seventh catalytic type of proteolytic

- enzymes[J]. *Journal of Biological Chemistry*, 2011, 286(44): 38321-38328.
- [2] Laskowski M Jr, Kato I. Protein inhibitors of proteinases[J]. *Annual Review of Biochemistry*, 1980, 49: 593-626.
- [3] 朱玲. 扇贝丝氨酸蛋白酶及其抑制剂基因的克隆与表达 [D]. 青岛: 中国科学院研究生院 (海洋研究所), 2006. Zhu L. Cloning and expression of serine protease and its inhibitor gene in scallop[D]. Qingdao: Institute of Oceanology, Chinese Academy of Sciences, 2006 (in Chinese).
- [4] Miura Y, Kawabata S I, Iwanaga S. A *Limulus* intracellular coagulation inhibitor with characteristics of the serpin superfamily. Purification, characterization, and cDNA cloning[J]. *The Journal of Biological Chemistry*, 1994, 269(1): 542-547.
- [5] Agarwala K L, Kawabata S I, Miura Y, *et al.* *Limulus* intracellular coagulation inhibitor type 3. Purification, characterization, cDNA cloning, and tissue localization[J]. *Journal of Biological Chemistry*, 1996, 271(39): 23768-23774.
- [6] Ashton-Rickardt P G. An emerging role for serine protease inhibitors in T lymphocyte immunity and beyond[J]. *Immunology Letters*, 2013, 152(1): 65-76.
- [7] Abraham E G, Pinto S B, Ghosh A, *et al.* An immune-responsive serpin, SRPN6, mediates mosquito defense against malaria parasites[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2005, 102(45): 16327-16332.
- [8] Vertyporokh L, Wojda I. Expression of the insect metalloproteinase inhibitor IMPI in the fat body of *Galleria mellonella* exposed to infection with *Beauveria bassiana*[J]. *Acta Biochimica Polonica*, 2017, 64(2): 273-278.
- [9] Xue Q G, Waldrop G L, Schey K L, *et al.* A novel slow-tight binding serine protease inhibitor from eastern oyster (*Crassostrea virginica*) plasma inhibits perkinsin, the major extracellular protease of the oyster protozoan parasite *Perkinsus marinus*[J]. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, 2006, 145(1): 16-26.
- [10] Xue Q G, Itoh N, Schey K L, *et al.* Evidence indicating the existence of a novel family of serine protease inhibitors that may be involved in marine invertebrate immunity[J]. *Fish & Shellfish Immunology*, 2009, 27(2): 250-259.
- [11] Wang X R, Xue Q G, Mao X W *et al.* Two I84 family protease inhibitors from Chinese razor clams *Sinonovacula constricta* expressed in response to environmental challenges[J]. *Fish & Shellfish Immunology*, 2018, 75: 149-157.
- [12] Xue Q G, Beguel J P, Gauthier J, *et al.* Identification of cvSI-3 and evidence for the wide distribution and active evolution of the I84 family of protease inhibitors in mollusks[J]. *Fish & Shellfish Immunology*, 2017, 62: 332-340.
- [13] Xue Q G. Pathogen proteases and protease inhibitors in molluscan infectious diseases[J]. *Journal of Invertebrate Pathology*, 2019, 166: 107214.
- [14] Rosenthal P J. Cysteine proteases of malaria parasites[J]. *International Journal for Parasitology*, 2004, 34(13-14): 1489-1499.
- [15] Le S Q, Gascuel O. An improved general amino acid replacement matrix[J]. *Molecular Biology and Evolution*, 2008, 25(7): 1307-1320.
- [16] 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.
- [17] 曾燕舞. 基于基因家族序列分析研究基因的倍增、分化和多效性 [D]. 上海: 复旦大学, 2010. Zeng Y W. Study of the duplication, divergence and pleiotropy of genes by analysis of gene family sequences[D]. Shanghai: Fudan University, 2010 (in Chinese).
- [18] Lu J L, Liu S, Ruan Z Y, *et al.* Family I84 protease inhibitors likely constitute a mollusca-specific protein family functioning in host defense[J]. *Diseases of Aquatic Organisms*, 2021, 145: 89-100.
- [19] Huang S F, Yuan S C, Guo L, *et al.* Genomic analysis of the immune gene repertoire of amphioxus reveals extraordinary innate complexity and diversity[J]. *Genome Research*, 2008, 18(7): 1112-1126.
- [20] Zhang G F, Fang X D, Guo X M, *et al.* The oyster genome reveals stress adaptation and complexity of shell formation[J]. *Nature*, 2012, 490(7418): 49-54.
- [21] Innan H, Kondrashov F. The evolution of gene duplications: Classifying and distinguishing between models[J]. *Nature Reviews Genetics*, 2010, 11(2): 97-108.

- [22] Zhang L L, Li L, Guo X M, *et al.* Massive expansion and functional divergence of innate immune genes in a proto-stome[J]. *Scientific Reports*, 2015, 5: 8693.

I84 protease inhibitor family show significant expansion and functional differentiation in *Crassostrea gigas*

WU Qianjiang¹, LIU Sheng^{1,2}, LU Jiali^{1,3}, FAN Congling¹, YE Jing^{1,3},
RUAN Ziyang¹, LIN Zhihua^{1,2}, XUE Qinggang^{1,2*}

(1. Zhejiang Key Laboratory of Aquatic Germplasm Resources, Zhejiang Wanli University, Ningbo 315100, China;

2. Institute of Mariculture Breeding and Seed Industry, Zhejiang Wanli University, Ningbo 315100, China;

3. National Demonstration Center for Experimental Fisheries Science Education, Shanghai Ocean University, Shanghai 201306, China)

Abstract: Protease inhibitors are extremely diverse proteins or peptides that inhibit the proteolytic activities of proteases. Protease inhibitors could inhibit the invasion of pathogens by inactivating pathogen proteases. The I84 protease inhibitor family is a new family in MEROPS database, and the role of its members in host defense has been studied to some extent. To investigate the distribution and function of I84 protease inhibitor family in *Crassostrea gigas*, in this study, 23 potential I84 family genes were identified in *C. gigas*. Based on phylogenetic analysis, 5 homeotic genes, *Cgsi3*, *Cgsi5*, *Cgsi6*, *Cgsi16* and *Cgsi19*, were selected for gene expression and functional assessment. They were confirmed by cDNA cloning to represent expressible real genes and expressed in digestive gland, gills, adductor muscle, gonad, mantle, and hemocytes, with the highest level of expression detected in the digestive gland ($P < 0.05$). In the larval oysters, none of the 5 selected genes was detected in zygote, and they showed different expression patterns in the remaining developmental stages. The expression level of *CgSi3* increased significantly in the pediveliger stage, and then decreased. The expression level of *CgSi6* increased significantly from the umbo larvae onwards. In oysters challenged with pathogen associated molecular patterns (PAMPs), gene expression pattern also differed with genes. In oysters injected with LPS and PGN, the expression level of *CgSi6* changed significantly, while in those injected with Poly (I:C) and GLU the expression level of *CgSi3* showed significant changes. This study revealed the gene expansion and functional differentiation of the I84 family protease inhibitors in *C. gigas*, and provides a basis for the comprehensive understanding of the biological functions of the I84 family protease inhibitors and the related mechanisms.

Key words: *Crassostrea gigas*; family I84 Protease inhibitors; gene identification; gene expression; functional differentiation

Corresponding author: XUE Qinggang. E-mail:qxue@zww.edu.cn

Funding projects: National Natural Science Foundation of China (32073010, 31672629); “3315” Innovative Team of Ningbo City; China Agriculture Research System (CARS-49); Fundamental Research Foundation for Provincial University of Zhejiang Province (SC1032080180110)