

[文章编号] 1007-7405(2017)06-0001-08

中国鲎保守 miRNA 靶基因预测及功能分析

翁朝红^{1,2}, 谢仰杰^{1,2}, 肖志群^{1,2}, 范赟颖^{1,2}

(1. 集美大学水产学院, 福建 厦门 361021; 2. 农业部东海海水健康养殖重点实验室, 福建 厦门 361021)

[摘要] 以中国鲎、美洲鲎和果蝇 EST 为靶标, 使用多个软件对中国鲎胚胎发育期高表达、高保守的 19 条 miRNA 进行靶基因预测。结果表明: 8 条 miRNA 在中国鲎 EST 中获得 18 条靶基因; 17 条 miRNA 在美洲鲎 EST 中获得 616 条靶基因; 12 条 miRNA 在果蝇 EST 中获得 543 条靶基因。这些保守 miRNA 潜在调控靶基因的功能, 不但可能与血液凝集反应以及抗菌功能相关, 而且还可能广泛参与生化调控、细胞组分组织架构、细胞组分空间定位、发育过程、多细胞器官构建、色素形成等生理过程。

[关键词] 中国鲎; miRNA; 靶基因; GO 功能注释

[中图分类号] Q 752

Target Genes Prediction and Their Functional Analysis of Conserved miRNAs from *Tachypleus tridentatus*

WENG Zhaohong^{1,2}, XIE Yangjie^{1,2}, XIAO Zhiqun^{1,2}, FANG Yunying^{1,2}

(1. Fisheries College, Jimei University, Xiamen 361021, China;

2. Key Laboratory of Healthy Mariculture for the East China Sea, Ministry of Agriculture, Xiamen 361021, China)

Abstract: In this paper, target genes of 19 high-expressed and conserved miRNAs from *Tachypleus tridentatus* embryos were predicted with ESTs from *T. tridentatus*, *Limulus polyphemus* and *Drosophila melanogaster* using several prediction softwares. 18 target genes matched with 8 miRNAs were obtained from *T. tridentatus* ESTs, and 616 target genes matched with 17 miRNAs were obtained from *L. polyphemus* ESTs. Meanwhile, 543 target genes matched with 12 miRNAs were obtained from *D. melanogaster* ESTs. The functions of these potential targets matched with the conserved miRNAs were not only potentially related to blood agglutination reaction and antibacterial function, but also intensively participate in biochemical regulation, cellular component organization, establishment of localization, developmental process, multicellular organismal process, pigmentation and so on.

Keywords: *Tachypleus tridentatus*; miRNA; target gene; gene ontology

0 引言

MicroRNAs (miRNA) 是一类内源性非编码小单链 RNA, 大小为 16 ~ 26 nt, 参与各种发育和生理过程, 通过转录后抑制来调节靶基因的表达^[1-2]。由于 miRNAs 是动物基因组中参与基因调控的重要分子^[3], 激发起人们的研究热情, 通过直接克隆、测序、RNA 印迹法、高通量测序技术甚至是同

[收稿日期] 2017-09-30

[修回日期] 2017-10-23

[基金项目] 国家水产种质资源平台项目 (2017DKA30470); 国家自然科学基金项目 (41106144); 福建省自然科学基金项目 (2014J01132)

[作者简介] 翁朝红 (1971—), 女, 教授, 博士, 从事水产动物学研究。通信作者: 谢仰杰 (1967—), 副教授, 博士, E-mail: yjxie@jmu.edu.cn。

源物搜索等方法,大量 miRNA 被成功鉴定。随着对 miRNA 的深入研究,其功能越来越受到关注。

中国鲎 (*Tachypleus tridentatus*) 为古老的大型海洋节肢动物,具有巨大的经济价值和药用价值。隶属节肢动物门 (Arthropoda), 肢口纲 (Merostomata), 剑尾目 (Xiphosura), 鲎科 (Limulidae)。20 世纪 70 年代以前,在中国长江口以南的海域大量分布有中国鲎。之后,由于潮间带、浅海栖息地不断缩小,以及过度捕捞,中国鲎资源面临枯竭^[4]。开展中国鲎人工繁殖是中国鲎资源恢复的一项重要工作。

中国鲎胚胎发育调控机理尚不清楚,有关 miRNA 对中国鲎的生理、发育等调控也未见报道。本研究利用中国鲎胚胎发育 miRNA 高通量测序获得的前期数据,挑选胚胎发育阶段高表达的一些 miRNA,利用生物信息学方法进行靶基因预测,并通过 GO 对靶基因功能进行深入分析。这些结果将为中国鲎胚胎发育 miRNA 的调控机理、人工繁殖生产实践提供理论基础。

1 材料与方法

1.1 数据搜集

实验用的 miRNA 来自中国鲎组织分化期胚胎,通过 Solexa 技术测序获得。选择胚胎发育期高表达 (表达量大于 10^4 拷贝数)、高保守性 (成熟序列与其他节肢动物完全一致) 的 19 条 miRNA 进行靶基因预测。19 条 miRNA 的名称、序列和长度信息见表 1。

表 1 中国鲎 19 条 miRNA 序列信息

Tab. 1 The information of 19 miRNAs from *T. tridentatus* used in this study

miRNA	miRNA 序列 Sequence of miRNA	长度 Length/nt	类型 Type
ttr-miR-10-5p	TACCCTGTAGATCCGAATTTGT	22	5'
ttr-miR-2b-3p	TATCACAGCCACCTTTGATGAGC	23	3'
ttr-miR-2a-3p	TATCACAGCCAGCTTTGATGAGC	23	3'
ttr-miR-8-3p	TAATACTGTCTAGGTAAGATGTC	23	3'
ttr-miR-12-5p	TGAGTATTACATCAGGTAAGATGTC	23	5'
ttr-miR-71-5p	TGAAAGACATGGGTAAGATGTC	23	5'
ttr-miR-87-3p	GTGAGCAAAGTTTCAGGTTGTC	21	3'
ttr-miR-96-5p	TTTGGCACTAGCACATTTTTGT	22	5'
ttr-miR-133-3p	TTGGTCCCCTTCAACCAGCTGT	22	3'
ttr-miR-153-3p	TTGCATAGTCACAAAAGTGATG	22	3'
ttr-miR-184-3p	TGGACGGAGAACTGATAAGGGC	22	3'
ttr-miR-252-5p	CTAAGTACTAGCGCCGAGGAG	22	5'
ttr-miR-276-3p	TAGGAACATTCATACCATGCTC	21	3'
ttr-miR-285-3p	TAGCACCATTTGAATTCAGTTC	22	3'
ttr-miR-5735-3p	TGGACAACAGGATAATGGCGT	21	3'

从 NCBI (<http://www.ncbi.nlm.nih.gov/>) 分别下载中国鲎和美洲鲎 (*Limulus polyphemus*) 的表达序列标签 (EST), 作为预测 miRNA 靶基因的靶标。去除重复序列以及未完全编码的序列后,共获取了 96 条中国鲎 EST、6248 条美洲鲎 EST。从 Gene Ontology 网站上下载 goa 数据包 (2015 年 4 月 16 日最新版),其中包含了 flybase 数据库中的基因 ID 以及相对应的 GO 号。

1.2 miRNA 靶基因预测

所采用靶基因预测软件有: miRanda^[5]、PicTar^[6]、DIANA^[7-9]、TargetScan^[9]、PITA^[10]。靶基因预测流程如图 1 所示。运算过程是在集美大学计算机中心 linux 系统下操作的。

1.2.1 以中国鲎和美洲鲎 EST 为靶标

采用 miRanda、PITA 两种靶基因预测算法, 以中国鲎和美洲鲎 EST 为靶标, 对中国鲎 miRNA 进行靶基因预测, 两个算法均属于初始算法。

miRanda 软件预测的参数如下: 1) miRNA 与靶基因 mRNA 匹配值 Score ≥ 140 ; 2) 双链形成自由能 $\Delta G \leq -83.6$ kJ/mol; 3) miRNA 5'端第 2 至第 8 个碱基与靶基因存在互补序列; 4) 其他参数为默认值。

PITA 算法关于种子区参数设置如下:

1) 最小种子序列大小为 8 nt; 2) 允许出现单个 G:U 配对; 3) 允许单个核苷酸不匹配; 4) 其他设置为默认值。

1.2.2 以果蝇 (*Drosophila melanogaster*) EST 为靶标

采用 miRanda、PicTar、DIANA、TargetScan 四个软件在线预测果蝇 EST 中 miRNA 靶基因。将四个软件预测的 miRNA 靶基因汇合、筛选, 取交集。

miRanda 互补参数分数 S 的计算 ($S \geq 80$): 1) G:C 或 A:U 分值为 +5 分; 2) G:U 分值为 +2 分; 3) 除这三种配对方式以外的分值为 -3 分; 4) 另外起始有空位分值为 -8 分; 5) 延伸空位分值为 -2 分, 热力学稳定性大小 $\text{mirSVG} \leq -0.1$, 最小自由能 $\Delta G \leq -58.5$ kJ/mol。

PicTar 参数: 1) seed 序列与 mRNA 3'端匹配; 2) 热力学稳定性大小 $\text{mirSVG} \leq 0.1$ 。

TargetScan 参数应满足下列三个条件之一: 1) 在成熟 miRNA 序列 (种子区域 + 位点 8) 第 2 ~ 第 8 nt 段有一个额外“A”匹配; 2) 在成熟 miRNA 序列 (种子区域 + 位点 8) 第 2 ~ 第 8 nt 段有一个额外核苷酸匹配; 3) 在种子区域第 2 ~ 第 7 nt 段, 有一个额外“A”核苷酸匹配。

靶基因在线预测软件: 1) Miranda 软件 (<http://www.microrna.org/microrna/home.do>); 2) PicTar 软件 (http://pictar.mdc.berlin.de/cgi-bin/pictar_fly.chi?species=fly); 3) DIANA 软件 (http://diana.imis.athenainnovation.gr/dianatool/index.php?r=microT_CDS/index); 4) TargetScan 软件 (http://www.targetscan.org/fly_12/); 5) PITA 软件 (http://genie.weizmann.ac.il/pubs/mir07/mir07_prediction.html)。

1.3 对 miRNA 的靶基因进行 GO 功能分析

利用 BLASTX 方法将 miRNA 靶基因序列与 NCBI 上非冗余蛋白质数据库进行比对 ($e < 10^{-5}$), 根据分值优先选取靶基因, 根据蛋白功能特征进行功能注释。从 GENE Ontology 下载 (最新版本 (2015 年 4 月 16 日)) 果蝇的数据包, 筛选靶基因 ID 以及 GO 号。使用 WEGO 在线功能注释分类。

2 结果

2.1 miRNA 靶基因预测结果

2.1.1 以中国鲎 EST 为靶标

通过 Miranda 软件预测结果表明, 仅有 8 条 miRNA 可获得相匹配的靶基因 (见表 2), 共获得靶基因 18 条, 平均每条 miRNA 获得 2.25 个靶基因, 其中 ttr-miR-34-5p 基因的靶基因数最多, 有 6 条。EST 与 miRNA 结合最小自由能范围为 $-120.59 \sim -89.95$ kJ/mol。

PITA 软件预测, 未获得结果。

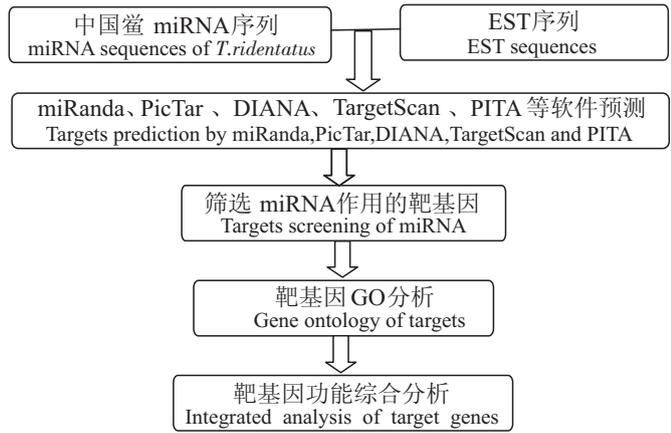


图 1 中国鲎 miRNA 的靶基因预测流程图

Fig.1 The flow diagram of miRNA target prediction of *T.tridentatus*

2.1.2 以美洲鲎 EST 为靶标

利用 miRanda 软件预测获得 616 条潜在靶基因, 平均每条 miRNA 匹配 36.23 个靶基因, 最多可达 129 条 (ttr-miR-2b), 大部分 miRNA (17 条) 至少有 2 个潜在的靶基因 (见表 2), 仅有 2 个 miRNA 未有潜在靶基因。58 个靶基因具有 2 个以上 miRNA 作用位点, 其中 gi283541495、gi283522801、gi283501022 三个靶基因分别具有 7、5、4 个靶位点。美洲鲎靶 EST-miRNA 的最小自由能范围分布在 -83.64 ~ -134.76 kJ/mol。结果表明 1 个 miRNA 可匹配多个靶基因, 1 个靶基因可能存在多个 miRNA 潜在靶位点, 每个 miRNA 均有 1 个及以上潜在的靶基因。

PITA 软件预测未得到结果。

2.1.3 以果蝇 EST 为靶标

用 Miranda、PicTar、DIANA、TargetScan 四个软件进行在线预测 miRNA 的靶基因。在果蝇 EST 中, 12 条 miRNA 共获得 543 条潜在靶基因, 平均每条 miRNA 有 45.2 个靶基因。大部分 miRNA (12 条) 至少可找到 4 个潜在靶基因, 仅有 7 个 miRNA 未找到 miRNA 靶基因, miR-8-3p、miR-252-5p、miR-2b-3p、miR-2b-3p 分别具有 109、103、89、89 个潜在靶基因 (见表 2)。有 4 个靶基因同时存在 3 个 miRNA 结合位点, 45 个靶基因同时存在 2 个 miRNA 结合位点。结果表明 1 个 miRNA 可能结合多个潜在的靶基因, 1 个靶基因上可能存在多个 miRNA 的靶位点。

表 2 miRNA 在中国鲎、美洲鲎、果蝇 EST 库中获得的靶基因数

Tab. 2 The number of miRNA targets from *T. tridentatus*, *L. polyphemus* and *D. melanogaster* ESTs

中国鲎 <i>T. tridentatus</i>		美洲鲎 <i>L. polyphemus</i>		果蝇 <i>D. melanogaster</i>	
miRNA	靶基因 ESTs Target ESTs /个	miRNA	靶基因 ESTs Target ESTs /个	miRNA	靶基因 ESTs Target ESTs /个
ttr-miR-34-5p	6	ttr-miR-2b	129	miR-8-3p	109
		ttr-miR-5735-5p	122	miR-252-5p	103
ttr-miR-12	3	ttr-miR-34-5p	105	miR-2b-3p	89
		ttr-miR-133	45	miR-2b-3p	89
ttr-miR-71	2	ttr-miR-87	41	miR-12-5p	53
		ttr-miR-285	33	miR-87-3p	35
ttr-miR-133	2	ttr-miR-10	31	miR-276a-3p	20
		ttr-miR-2a	30	miR-184-3p	14
ttr-miR-5735-3p	2	ttr-miR-252-5p	25	miR-133-3p	10
		ttr-miR-8	14	miR-2a-3p	9
ttr-miR-10	1	ttr-miR-5735-3p	13	miR-285-3p	8
		ttr-miR-276	12	miR-10-5p	4
ttr-miR-8-5p	1	ttr-miR-71	5		
		ttr-miR-184	5		
ttr-miR-184	1	ttr-miR-8-5p	2		
		ttr-miR-12	2		
		ttr-miR-96	2		

2.2 miRNA 靶基因 GO 功能分类

2.2.1 从中国鲎 EST 中获得的靶基因 GO 功能分类

预测结果显示, ttr-miR-34-5p 等 8 个基因所匹配的靶基因均与鲎血液凝固生化过程相关基因的相关性高 (见表 3), 与细胞内凝集抑制因子、丝氨酸蛋白水解酶、谷氨酰胺转移酶等有相互作用靶点, 可能调控这些蛋白因子的活性。

表 3 中国鲎靶基因功能预测
Tab. 3 The function predictions of target genes in *T. tridentatus*

编号 No.	miRNA	靶基因序列号 ID of target gene	功能 Function
1	ttr-miR-34-5p	gi 149404439 , gi 149401922 , gi 149401788 , gi 146770983 , gi 14622551 , gi 14622515	凝集抑制因子 II 型 Coagulation inhibitor type II 丝氨酸蛋白水解酶 Serine protease 谷氨酰胺转移酶 Glutamyltransferase
2	ttr-miR-12	gi 186857235 , gi 186775205 , gi 186749649	亲环素 G Cyclophilin G
3	ttr-miR-71	gi 157943878 , gi 146770983	凝集抑制因子 II 型 Coagulation inhibitor type II
4	ttr-miR-133	gi 14621915 , gi 37191865	谷氨酰胺转移酶 Glutamyltransferase 半胱氨酸蛋白酶抑制剂 Cysteine proteinase inhibitor
5	ttr-miR-5735	gi 113466725 , gi 109188886	凝集抑制因子 II 型 Coagulation inhibitor type II
6	ttr-miR-10	gi 160448873	亲环素 G Cyclophilin G
7	ttr-miR-8	gi 14621916	谷氨酰胺转移酶 Glutamyltransferase
8	ttr-miR-184	gi 151348252	谷氨酰胺转移酶 Glutamyltransferase

2.2.2 从美洲鲎 EST 中获得的靶基因 GO 功能分类

因美洲鲎 EST 数量较多, miRNA 预测获得靶基因也较多。ttr-miR-2b 基因获得的靶基因最多, 共 129 条, 其中 91 条 EST 没有相关已注释蛋白质, 其他 38 条 EST 序列与 11 种类型的蛋白质相关, 如 HDID 支架、精氨酸激酶、核糖体蛋白、GTP 结合蛋白、转录延长因子、钙离子 ATP 酶、凝固蛋白原、丝氨酸蛋白酶等等。可推测 ttr-miR-2b 主要参与鲎胚胎发育过程细胞内基因合成、转录、蛋白质表达等基础生命过程中分子功能和生物过程相关基因的调控。部分靶基因序列和功能见表 4。

表 4 ttr-miR-2b 靶基因功能预测
Tab. 4 The function predictions of target genes of ttr-miR-2b

编号 No.	靶基因序列号 ID of target gene	分值 Score	自由能 MFE	BLAST 基因 Genes from BLAST	功能 Function
1	gi 283543693	153	-29.21	emb LM384077.1	HDID 支架 HDID molder
2	gi 283543423	152	-26.16	gb FJ235415.1	精氨酸激酶 Arginine kinase
3	gi 283541947	152	-26.16	gb FJ235415.2	精氨酸激酶 Arginine kinase
4	gi 283541736	152	-26.16	gb DQ538932.1	精氨酸激酶 Arginine kinase
5	gi 283512327	152	-26.16	gb DQ538932.1	精氨酸激酶 Arginine kinase
6	gi 283500037	152	-26.16	gb DQ538932.1	精氨酸激酶 Arginine kinase
7	gi 283544004	142	-23.52	gb BC164084.1	核糖体蛋白 SA Ribosomal SA
8	gi 283543295	142	-23.52	gb BC164084.1	核糖体蛋白 SA Ribosomal SA
9	gi 283501427	142	-23.52	gb BC164084.1	核糖体蛋白 SA Ribosomal SA
10	gi 283500590	142	-23.52	gb BC164084.1	核糖体蛋白 SA Ribosomal SA
11	gi 283519775	142	-23.52	gb BC164084.1	核糖体蛋白 SA Ribosomal SA
12	gi 283515936	142	-23.52	gb BC164084.1	核糖体蛋白 SA Ribosomal SA
13	gi 283543178	165	-22.98	ref XM_009068014.1	假定蛋白质 Hypothetical protein
14	gi 283525782	146	-22.7	gb AY066022.1	SAP 正五聚蛋白 SAP pentaprotein
15	gi 283525425	146	-22.7	gb AY066022.1	SAP 正五聚蛋白 SAP pentaprotein
16	gi 283519744	153	-22.53	gb EU020904.1	GTP 结合蛋白 GTP binding protein
17	gi 283501355	143	-21.93	ref XM_006271600.1	转录延长因子 Transcriptimal elongationfactor
18	gi 283543606	151	-21.47	ref XM_012395469.1	转录激活蛋白 Pur-beta Transcription activation protren Pur-beta

续表

编号 No.	靶基因序列号 ID of target gene	分值 Score	自由能 MFE	BLAST 基因 Genes from BLAST	功能 Function
19	gi 283525831	141	-20.85	ref XM_010876575.1	RAN 结合蛋白 I RAN binding protein I
20	gi 283543678	151	-20.79	emb AJ298902.1	钙离子 ATP 酶 I Ca ²⁺ -ATPase I
21	gi 114107013	144	-20.73	emb X04424.1	凝固蛋白原 Coagulinogen
23	gi 114106960	144	-20.73	emb X04424.1	凝固蛋白原 Coagulinogen
24	gi 114106930	144	-20.73	emb X04424.1	凝固蛋白原 Coagulinogen
25	gi 114106915	144	-20.73	emb X04424.1	凝固蛋白原 Coagulinogen
26	gi 114106913	144	-20.73	emb X04424.1	凝固蛋白原 Coagulinogen
27	gi 283543880	144	-20.73	emb X04424.1	凝固蛋白原 Coagulinogen
28	gi 283543236	144	-20.73	emb X04424.1	凝固蛋白原 Coagulinogen
29	gi 283511448	144	-20.73	emb X04424.1	凝固蛋白原 Coagulinogen
30	gi 283511433	144	-20.73	emb X04424.1	凝固蛋白原 Coagulinogen
31	gi 283495787	144	-20.73	emb X04424.1	凝固蛋白原 Coagulinogen
32	gi 283525118	144	-20.73	emb X04424.1	凝固蛋白原 Coagulinogen
33	gi 283522957	144	-20.73	emb X04424.1	凝固蛋白原 Coagulinogen
34	gi 283519917	144	-20.73	emb X04424.1	凝固蛋白原 Coagulinogen
35	gi 283541879	149	-20.63	gb CP004014.1	凝固蛋白原 Coagulinogen
36	gi 283541917	140	-20.47	ref XM_006570204.1	丝氨酸蛋白酶 Serine protease
37	gi 283523012	148	-20.22	ref XM_011263664.1	DEAD-box 解旋酶 DEAD-box helicase

2.2.3 从果蝇 EST 中预测的靶基因 GO 功能分析

从果蝇 EST 中获得的靶基因, 经 GO 功能分析结果 (见图 2) 表明, 约 60% 靶基因参与生化进程, 21% 靶基因参与分子功能, 19% 靶基因参与细胞组分。WEGO 分析结果显示: 细胞组分功能分类中, 主要功能包括细胞结构组分、细胞器组分、大分子组分细胞外区等; 分子功能大类中, 涉及结合、催化、酶调节、转录调节、翻译调节等; 在生化进程分类中, 主要参与生化调控、细胞组分组织架构、细胞形成过程、细胞组分空间定位、发育过程、多细胞器官构建、色素形成、多细胞生物进程 (见图 2)。

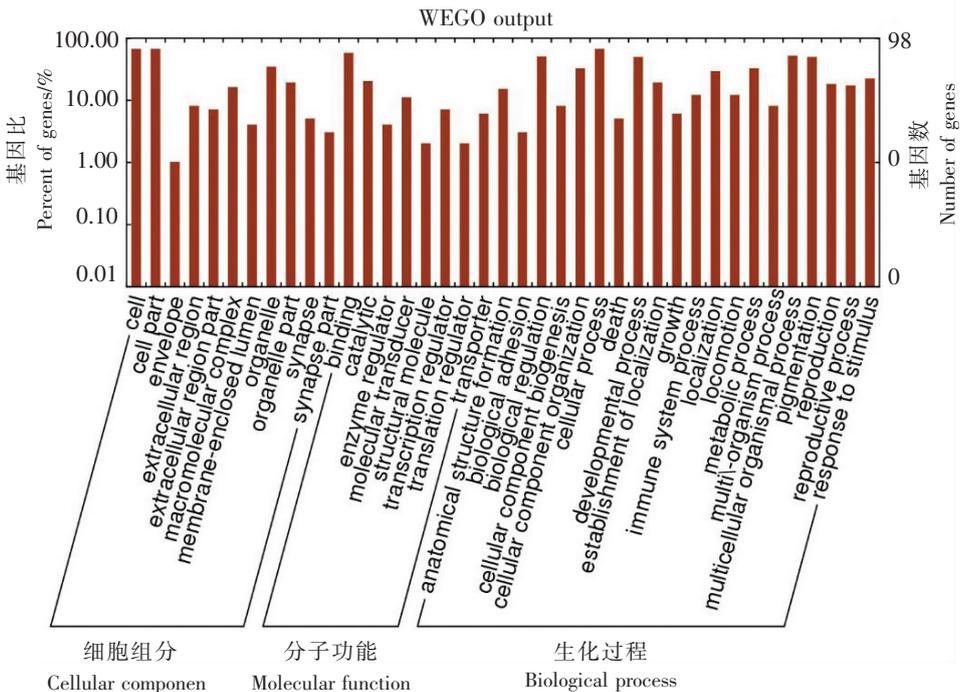


图 2 靶基因 GO 功能分类分布

Fig.2 Classification distribution of target genes in GO function

3 讨论

3.1 中国鲎胚胎发育期保守性 miRNA 潜在功能

中国鲎胚胎发育时间长, 从受精到孵化为 1 龄三叶虫幼体需要 50 多天时间^[11]。中国鲎的胚胎发育过程可分为 6 期: 卵裂期、囊胚期、原肠胚形成期、附肢原基形成期、组织分化期和快速生长期。鲎胚胎发育容易进入滞育状态, 尤其在原肠胚期。鲎发育时间长、滞育等问题是目前鲎人工繁育、恢复其自然资源等工作的主要技术瓶颈。而目前其胚胎发育的调控机理不甚清楚。

本文取中国鲎胚胎, 采用 Solexa 技术进行高通量测序, 获得大量胚胎阶段表达的 miRNA, 筛选出具有高表达高保守性的 19 条 miRNA 进行靶基因预测。这些 miRNA 表达量高, 推测在胚胎发育中发挥着重要功能, 而且保守性高, 与其他动物 miRNA 成熟序列完全相同, 在后生动物进化过程中序列高度保守。

由于已提交的中国鲎相关 EST 数量少, 且大多基因与鲎的血细胞凝集功能有关, 美洲鲎的 EST 则相对多一些, 但大多也是与血细胞凝集功能相关。利用这两物种的 EST 进行靶基因预测, 结果表明所选取的 miRNA 大多数与鲎血细胞凝集、抗菌等生理过程相关基因有匹配位点, 推测 miRNA 调控细胞生命活动的功能相当广泛, 除了与发育调控相关, 血液免疫功能也可能受到 miRNA 广泛调控, 但这些还需要实验验证。

果蝇为模式生物, 其研究较为深入。中国鲎与果蝇均属节肢动物, 具有较高的同源性。将中国鲎的 miRNA 与果蝇 EST 库进行靶基因预测, 结果表明, miRNA 不但序列具有保守性, 其功能也保守。这些保守的 miRNA 广泛参与细胞结构组分、细胞器组分、大分子组分等基础结构架构, 也广泛参与结合、催化、酶调节、转录调节、翻译调节等基础生物合成途径, 还广泛参与生化调控、细胞生理过程、发育过程、细胞成分组织、色素形成等多细胞生物进程。保守的 miRNA 在后生动物的进化过程中具有重要地位, 不但在动物发育期起着重要的调控作用, 而且可能比非保守的 miRNA 具有更高效的表达及更广泛的功能^[12]。

miRNA 在生命活动过程中起调控中心分子的作用, 其调控的广泛性和复杂性还体现在: 1 个 miRNA 可与多个靶基因结合, 调控多个靶基因的表达, 同时 1 个靶基因可受到多个 miRNA 调控。

3.2 靶基因预测需要多个软件结合分析

miRNA 调控功能主要源自于 miRNA 与其靶基因之间完美的或接近完美的互补作用。根据这一原理, 可鉴定 miRNA 的靶基因, 用以深入研究 miRNA 调控功能。然而, 目前没有一个确定的、可接受的标准来鉴定 miRNA 靶基因。利用生物信息学方法对 miRNA 进行靶基因预测是一个有效且快速的方法。

本试验利用多种预测软件, 如: miRanda、PITA、PicTar、DIANA、TargetScan 进行靶基因的预测。不同的软件参数不同, 通过取他们的交集来增加结果的可靠性。

miRanda 不受物种限制, 适用范围广, 可下载到本地进行离线分析。miRanda 算法通过碱基互补 (如 G=C、A=U 等) 构建打分系统, 通过对种子区的匹配参数来定义。PITA 是基于靶位点可接性来进行 miRNA 靶标预测的。PITA 算法遵循标准的种子参数设置, 可考虑从 miRNA 的第二个位点开始种子序列, 长度为 6~8 nt, 这段区域不允许有不匹配的碱基或环状结构, 但允许一个 G:U 匹配出现在 7-或 8-mers。PicTar 通过搜寻保守 3'UTR 中几乎或者完全匹配的区域中计算匹配上的双链自由能。DIANA 的算法需要计算潜在 miRNA 结合最小自由能, 需要至少有 6nt 的种子区域, 允许一定量的错误匹配, 并将结果与 3'UTR 100% 匹配的序列进行比较。四款软件各有自己的优势和局限, 只有结合使用才能提高结果的可靠性。

总之, 由于中国鲎和美洲鲎在 NCBI 上已注释的功能基因较少, 所以其靶基因功能分析进展甚微, 还需要更多的功能基因组学的工作。为了验证已预测的与胚胎发育相关 miRNA 靶基因作用通路

以及在转录水平上筛选更多差异表达基因, 后续靶基因研究将利用荧光定量 PCR 表达加以验证。本文初步分析了胚胎发育相关 miRNA 作用的靶基因在胚胎发育组织分化过程中参与的蛋白互作网络、分子调控通路和生物学功能, 从而为进一步深入地揭示 miRNA 及其调控靶基因参与胚胎发育的调控机理奠定基础。

[参 考 文 献]

- [1] BARTEL D P. MicroRNAs: target recognition and regulatory functions [J]. *Cell*, 2009, 136(2): 215-233.
- [2] KIM V N. MicroRNA biogenesis: coordinated cropping and dicing [J]. *Molecular Cell Biology*, 2005, 6: 376-385.
- [3] ROUX J, GONZALEZ-PORTA M, ROBINSON-RECHAVI M. Comparative analysis of human and mouse expression data illuminates tissue-special evolutionary patterns of miRNAs [J]. *Nucleic Acids Res*, 2012, 40: 5890-5900.
- [4] 翁朝红, 谢仰杰, 肖志群, 等. 福建及中国其他沿岸海域中国鲨资源分布现状调查 [J]. *动物学杂志*, 2012, 47(3): 45-53.
- [5] MARAGKAKIS M, ALEXIOU P, PAPADOPOULOS G L, et al. Accurate microRNA target prediction correlates with protein repression levels [J]. *BMC Bioinformatics*, 2009, 10: 295.
- [6] KREK A, GRUN D, POY M N, et al. Combinatorial microRNA target predictions [J]. *Nat Genet*, 2005, 37(5): 495-500.
- [7] ICHELLE B, FRANK S. A developmental timing microRNA and its target regulate life span in *C. elegans* [J]. *Science*, 2005, 310(5756): 1954-1957.
- [8] LEWIS B P, BURGE C B, BARTEL D P. Conserved seed pairing often flanked by adenosines, indicates that thousands of human genes are microRNA targets [J]. *Cell*, 2005, 120(1): 15-20.
- [9] FRIEDMAN R C, FARH K K, BURGE C B, et al. Most mammalian mRNAs are conserved targets of microRNAs [J]. *Genome Research*, 2009, 19: 92-105.
- [10] MICHAEL K, NICOLA I, ULRICH U, et al. The role of site accessibility in microRNA target recognition [J]. *Nature Genetics*, 2007, 39(10): 1278-1284.
- [11] 洪水根, 李祺福, 陈美华, 等. 中国鲨胚胎发育研究 [J]. *厦门大学学报 (自然科学版)*, 2002, 41(2): 239-246.
- [12] 罗艳, 张群, 梁宇君, 等. 动物中 microRNA 的保守性和进化历程 [J]. *中国科学: 生命科学*, 2012, 41(2): 96-106.

(责任编辑 朱雪莲 英文审校 黄力行)