[文章编号] 1007－7405(2017)03－0008－10

大黄鱼转录组水平 miRNA 的生物信息学分析

房路京[1,2]，肖世俊[1,2]，王志勇[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] 根据成熟 miRNA 序列在物种间高度保守的特性，miRNA 前体可以形成茎环状的二级结构，以及 miRNA 非编码的特性，搭建出了 miRNA 挖掘的生物信息学流程，并利用此流程在大黄鱼全转录组中进行 miRNA 的挖掘，得到 54 条成熟 miRNA 序列以及 54 条前体序列（pre－miRNA），根据其种子区的序列将其归为 29 个 miRNA 家族。通过对预测得到的 miRNA 前体序列同其他物种中相同家族的 miRNA 前体序列构建系统进化树，揭示出同一家族的 miRNA 前体在不同物种之间的保守性和多样性。利用两种靶基因预测软件，对成熟 miRNA 进行靶基因预测，共得到 2360 个靶基因，并对这些靶基因使用 BLAST2GO 做了进一步的 GO 注释和富集分析。

[关键词] miRNA；大黄鱼；转录组；生物信息

[中图分类号] S 917.4

Transcriptome-wide Bioinformatics Analysis of miRNA in Large Yellow Croaker *Larimichthys crocea*

FANG Lujing[1]，XIAO Shijun[1]，WANG Zhiyong[1]

（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 study，a miRNA prediction bioinformatics pipeline was designed according to three basic criteria，mature miRNA sequences are highly conserved among different species，miRNA precursors could form stem loop secondary structure and miRNAs are non-coding sequences. The large yellow croaker transcriptome sequences were used to predict miRNA by this pipeline. At last，54 mature miRNAs，which belong to 29 miRNA families according to seed region ，as well as 54 pre-miRNA sequences were gained. When compared with other animals' miRNA precursors，the phylogenetic analysis reveals conservation and diversification of large yellow croaker miRNA precursors among diverse animal species. 2360 miRNA target genes were gained by using two miRNA target gene prediction tools，and BLAST2GO was also used to GO annotation for target genes.

Keywords：miRNA；*Larimichthys crocea*；transcriptome；bioinformatics

0 引言

小 RNA（microRNA，miRNA）是一种长度在 22 个碱基左右的非编码、内源性的单链 RNA。它

[收稿日期] 2017－02－20 [修回日期] 2017－04－01
[基金项目] 厦门南方研究中心重大项目（2014GZY70NF34）；国家自然科学基金重点项目（U1205122）
[作者简介] 房路京（1990—），男，硕士生，从事生物信息学方面研究。通信作者：王志勇（1963—），男，教授，博导，从事水产动物遗传育种和分子生物学研究。

广泛存在于动植物以及病毒中，在转录后水平发挥着重要的作用[1-3]。自 20 世纪 90 年代初期发现 miRNA[4]以来，已有研究表明 miRNA 调控着一些基本的细胞功能，如：增殖、分化以及细胞的凋亡[5-8]，同时 miRNA 还发挥着一些生物学功能，包括生长、分化、应激以及某些疾病的发生等[9-10]。在大多数的情况下 miRNA 是对其靶基因进行负调控，作用的基本原理是通过成熟的 miRNA 与靶基因部分或全部结合，抑制靶基因的表达，从而产生调控靶基因的作用。生物体形成 miRNA 的生物学过程主要分为三步：第一步，发生在细胞核中，miRNA 基因被转录成为具有几百 bp 的初体 miRNA（pri-miRNA）；第二步，初体 miRNA 通过酶的剪切可以形成具有发夹结构的前体（pre-miRNA），随后前体被运送到细胞质中；最后一步，在细胞质中，前体被再次剪切形成成熟的 miRNA 序列[2,11-12]。

大黄鱼（*Larimichthys crocea*）隶属于鲈形目，石首鱼科，黄鱼属，是一种重要的海洋经济鱼类。在已知的 miRNA 公共数据库（miRBase）中并没有收录大黄鱼的 miRNA 序列，已发表的文献中仅有关于基因组水平的相关报道[13]，以及 poly I:C 刺激下 miRNA 变化的研究[14]，而没有针对全转录组水平的相关研究报道。预测新的 miRNA 的方法主要有：高通量测序法、克隆法以及利用生物信息预测法[15]。相较而言，前两种方法都比较耗时而且费用也较高，而生物信息学预测法则耗时短且方便。用生物信息学法挖掘 miRNA 可分为两种，一种是从头预测，其依据的原理是根据 miRNA 成簇的特点挖掘已经确定的 miRNA 周围可能的 miRNA；另一种是依据同源比对原理挖掘 miRNA。前者受到诸多因素的影响，预测的结果并不是很理想。在其他物种的研究[16-21]中，使用后者预测 miRNA 报道多于前者。本文旨在利用生物信息学方法，基于成熟 miRNA 序列在物种之间的高度保守性这一原理[22]，用笔者编写的脚本和软件去预测 miRNA 并对靶基因进行 GO（gene ontology）的注释和富集分析，从而为大黄鱼在全转录组水平上的 miRNA 研究提供可参考的数据，有助于后续大黄鱼 miRNA 的研究。

1　材料和方法

1.1　数据和软件

miRNA 公共数据库（miRBase）的版本为 Release 21（http://www.mirbase.org/）。大黄鱼的全转录组数据从 NCBI 数据库（project accession number PRJNA254539）以及蛋白质的非冗余数据库（nr database）中下载。使用的软件主要有比对软件 BLAST（版本为 2.2.31）、二级结构预测软件 RNAfold（版本为 2.1.9，http://www.tbi.univie.ac.at/RNA）、靶基因预测软件 RNAhybrid（http://bibiserv.techfak.uni-bielefeld.de）和 miRanda（http://www.microrna.org）。GO 基因功能注释分析使用软件 BLAST2GO（https://www.blast2go.com），系统进化树采用 MEGA7（http://www.megasoftware.net）构建。

1.2　生物信息预测 miRNA 的主要流程

根据 miRNA 序列在物种间高度保守的特性，搭建 miRNA 预测流程（见图 1），其主要包含以下几个步骤：

1）利用 BLAST 将大黄鱼全转录组的序列同 miRNA 数据库中的动物 miRNA 做比对，BLAST 参数中 *E* 值选择为 10，并设定 miRNA 序列与转录组序列错配不能超过 3。

2）将成熟 miRNA 序列比对到转录组序列上，在其位置上下游各取 100 个碱基（nt）作为可能的 miRNA 前体，根据现有的报道和 miRNA 数据库前体的长度，设置前体最短为 55 nt。

3）为去除 miRNA 中可能存在的编码序列，将步骤 2）中产生的 miRNA 前体序列同蛋白质数据库进行比对，将比对到蛋白质数据库中的序列去掉。

4）对步骤 3）中筛选得到的序列，用 RNAfold 软件预测二级结构，将其中不能形成二级结构、腺嘌呤和尿嘧啶比例大于 70% 或小于 30%、最小自由能和最小自由能系数分别大于 -84 kJ/mol 和 0.8 的序列去除。

5）用 SVM 进行 miRNA 前体的分类，先通过 SVM 进行模型的训练，训练出一个分类效果明显的

模型。从 miRBase 上下载得到的动物前体序列，用 RNAfold 将其生成二级结构，并从中选取具有代表性的最小自由能、最小自由能系数、GC 含量等 36 个特征值，将这些特征值的数据作为 SVM 输入的正集。从大黄鱼的编码区随机选择长度相似的序列使用 RNAfold 形成二级结构，同样将 36 个特征值输入作为负集。

6）将步骤 5）中产生的二级结构特征值输入到训练好的 SVM 模型中，最后再根据比对的位置得到最终的成熟的 miRNA。

1.3 miRNA 靶基因的预测

miRNA 一般是结合靶基因 mRNA 的 3′ UTR 发挥作用，不同的软件预测参数会有不同，这使得产生的结果在数量和范围上也存在不同，导致产生的假阳性结果增多。为降低假阳性结果，本研究使用 miRanda 和 RNAhybrid 两种不同的软件，将两种软件的结果取交集作为最后的结果。RNAhybrid 具体参数命令（仅在 linux 下运行）为：RNAhybrid -f 2，8 -c -u 2 -v 3 -e -23 -p 0.1 -d 1.65，0.21 -t 靶基因 3′UTR 序列 -q 成熟 miRNA 序列 > 预测的靶基因。miRanda 命令是：miRanda 成熟 miRNA 序列 3′UTR 序列 -sc 165 -en -23 -strict -quiet > 靶基因结果。

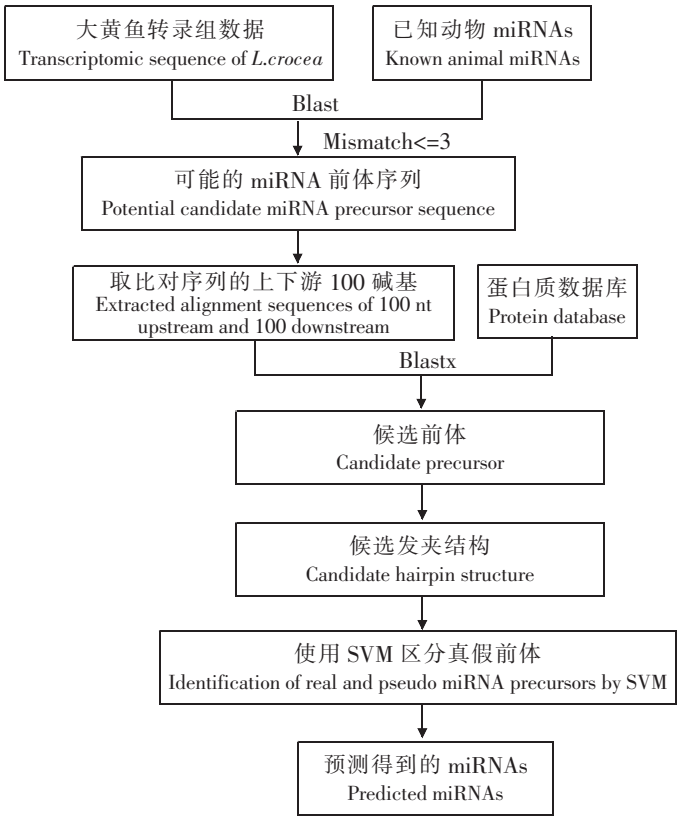


图 1 miRNA 挖掘流程

Fig.1 Pipeline of miRNA prediction

1.4 大黄鱼成熟 miRNA 分析和 miRNA 前体序列系统进化分析

对大黄鱼预测得到的 miRNA 进行初步的分析，包括长度、碱基频率等的初步统计，使用 Python 统计及用 R 语言绘制图形。分析 miRNA 前体序列时，挑选大黄鱼预测结果中家族成员较多的前体，并从 miRBase 中挑选了大西洋鲑（*Salmo salar*）、斑马鱼（*Danio rerio*）、红鳍东方鲀（*Fugu rubripes*）、金娃娃（*Tetraodon nigroviridis*）以及鲤鱼（*Cyprinus carpio*）进行进化分析。

1.5 靶基因的功能注释和分析

首先将靶基因使用 BLASTX 和 NR 数据库（蛋白质非冗余数据库）比对进行靶基因注释，然后将最优结果用 BLAST2GO 处理。在流程搭建以及 miRNA 靶基因预测中，均使用 Python2.7 写成的脚本进行处理，运行环境为集美大学超级计算机平台下的 Centos6 系统。

2 结果与讨论

2.1 预测得到的成熟 miRNA

按图 1 所示的 miRNA 生物信息学挖掘流程，将包含有 88 103 条大黄鱼转录组序列比对到 miRBase 中的动物 miRNA 上，其中 miRBase 中包含了 26 332 条动物的 miRNA 序列。成熟的 miRNA 在物种之间是保守的，因此在 miRBase 中存在着较多的名称不同但序列相同的 miRNA，为提高流程的运行速度和方便后期的分析，流程中使用的是去掉冗余序列后的 miRNA。本研究去掉冗余后的 miRNA 为14 990条，最后在大黄鱼全转录组水平上共挖掘预测得到 54 个成熟的 miRNA 序列（见表 1，其中名称相同但来源可能不同）。

表 1　预测得到的大黄鱼 miRNA

Tab. 1　Predicted miRNA in *L. crocea*

编号 No.	miRNA 名称 miRNA name	miRNA 序列 miRNA sequence	长度 Length /nt	所属家族 Family	位置 Position	前体长度 Length of precursor/nt	最小自由能 MFE /(kJ · mol^{-1})
1	lcr-let-7b-3p	CUAUACAAUCUAUUGCCUUCCC	22	let-7	3′	87	−167.16
2	lcr-miR-22-5p	AGUUCUUCACUGGCAAGCUUUA	22	mir-22	5′	176	−331.80
3	lcr-let-7g-5p	UGAGGUAGUAGAUUGUAUAGUU	22	let-7	5′	87	−167.16
4	lcr-miR-149-3p	GAGGGAGGGAGGGAGGCGGUGC	22	mir-149	3′	105	−231.00
5	lcr-miR-199a-5p	CCCAGUGUUCAGACUACCUGUUC	23	mir-199	5′	76	−165.48
6	lcr-miR-203-5p	AGUGGUUCUCGACAGUUCAACA	22	mir-203	5′	73	−136.50
7	lcr-miR-7-1-3p	CAACAAAUCACAGUCUGCCA	20	mir-7	3′	77	−120.96
8	lcr-miR-203a-5p	AGUGGUUCUCAACAGUUCAACAGUU	25	mir-203	5′	98	−150.36
9	lcr-miR-297a-5p	UGUAUGUGCGUGCAUGUGCAUGU	23	mi-297	5′	131	−232.26
10	lcr-miR-122-3p	AACGCCAUUAUCACACUAAAUA	22	mir-122	3′	88	−159.60
11	lcr-miR-297	AUGUAUGUGUGCAUGCAUGCCAUG	24	mir-297	3′	63	−107.94
12	lcr-miR-34a-3p	CAAUCAGCAAGUAUACUGCC	20	mir-34	3′	100	−163.80
13	lcr-miR-124b	UUAAGGCACGCGGUGAAUGCCA	22	mir-124	3′	101	−187.32
14	lcr-let-7j-3p	CUAUACAGUCUAUUGCCUUCC	21	let-7	3′	154	−309.12
15	lcr-miR-142-5p	CCCAUAAAGUAGAAAGCACUAC	22	mir-142	5′	143	−178.92
16	lcr-miR-16c-3p	UCCAGUAUUGAUCGUGCUGCUGA	23	mir-15	3′	65	−117.18
17	lcr-miR-92a-5p	GGUUGGGAGAGGUAGCAAUGCU	22	mir-25	5′	56	−118.86
18	lcr-miR-202-5p	UUCCUAUGCAUAUACCUUUUU	21	mir-202	5′	78	−124.74
19	lcr-miR-202-3p	AGAGGCAUAGGGCAUGGGAAAA	22	mir-202	3′	78	−124.74
20	lcr-miR-203b-3p	GUGAAAUGUUUAGGACCACUUG	22	mir-203	3′	64	−97.02
21	lcr-miR-219-3p	GGAGUUGUGGAUGGACAUCACGC	23	mir-219	3′	98	−181.02
22	lcr-miR-144	CAGGAUAUCAUCUUAUACUGU	21	mir-144	3′	72	−118.44
23	lcr-miR-145-3p	GGAUUCCUGGAAAUACUGUUCUU	23	mir-145	3′	70	−130.20
24	lcr-miR-150	CUCCCAAUCCUUGUACCAGUGU	22	mir-150	5′	95	−184.80
25	lcr-miR-1275	GUGGGGGAGGGGCUGUC	17	mir-1275	5′	64	−137.34
26	lcr-miR-184	UGGACGGAGAACUGAUAAGGGU	22	mir-184	3′	92	−159.60
27	lcr-miR-122b	AGUGUGACAAUGGUGUUU	18	mir-122	3′	89	−167.16
28	lcr-miR-7-5p	UGGAAGACUAGUGAUUUUGUUGUU	24	mir-7	5′	76	−113.82
29	lcr-miR-2723	CAGCAGUGGGCGUCUGUG	18	–	5′	72	−128.10
30	lcr-miR-1889-5p	AAUCACAUAUUGUAACAGUGG	21	mir-1889	5′	85	−126.00
31	lcr-miR-2985	GUGGGUGGAAUAGUAUAACAAU	22	mir-2985	5′	69	−172.20
32	lcr-miR-199b-3p	ACAGUAGUCUGCACAUUGGUUA	22	mir-199	3′	100	−198.66
33	lcr-miR-3529	GGCAGACUGUGAUUUGUUGU	20	mir-3529	5′	77	−110.04
34	lcr-miR-2985	UGUUAUAGUAUCCCACCUACCC	22	mir-2985	3′	69	−172.20
35	lcr-miR-4121-3p	GUGUUGGAGGUGGGUUU	17	–	3′	57	−117.18
36	lcr-miR-3966	AGCUGCCAGCUGAAGAACUGU	21	–	3′	181	−331.80
37	lcr-miR-29a-5p	ACUGAUUUCCCUUGGUGCUUAGA	23	mir-29	5′	94	−152.46
38	lcr-miR-129b	ACCUUCUGGGGUUUGAGCAAUCG	23	mir-129	5′	57	−91.14
39	lcr-miR-4957-3p	CAGCUCCAGCAGCGAGCGG	19	–	3′	58	−136.50
40	lcr-miR-21	UAGCUUAUCAGACUGGUGUUGGCUG	25	mir-21	5′	77	−178.92
41	lcr-miR-29c-5p	ACCGAUUUCUUUUGGUGUUCAGA	23	mir-29	5′	60	−105.42
42	lcr-miR-1788	GGCUUGUUUUAAGUUGCCUGCGA	23	mir-1788	5′	102	−177.66
43	lcr-miR-15a-5p	GUAGCAGCACGGAAUGGUUUGU	22	mir-15	5′	69	−140.70
44	lcr-miR-15d-3p	CAGGCCAUACUGUGCUGCCGCAG	23	mir-15	3′	69	−140.70
45	lcr-miR-219c-3p	AGGAGUUGUGGAUGGACAUC	20	mir-219	3′	94	−181.02
46	lcr-miR-430c-5p	ACCCUAACAGAAACAUUGACU	21	mir-430	5′	83	−131.88
47	lcr-let-7a-3-3p	CUAUUCAACCUACUGUCUUUC	21	let-7	3′	93	−165.06
48	lcr-miR-8160-3p	CAGCGACUGUUGUUAUUGGGA	21	–	3′	132	−213.78
49	lcr-miR-8862	UGCUGAGCGGUGGCCGGCCCUC	23	–	5′	58	−155.40
50	lcr-miR-29c	CUAGCACCAUUUGAAAUCGGUUAU	24	mir-29	3′	94	−152.46
51	lcr-let-7c	UGAGGUAGUAGGUUGUAUAGUUU	23	let-7	5′	93	−165.06
52	lcr-miR-9399-3p	UGGUUACAGAUUAAAAA	17	–	3′	63	−68.04
53	lcr-let-7a-3-3p	AACUAUACAGUCUAUUGCCU	20	let-7	3′	123	−253.26
54	lcr-miR-2985a-2-3p	UGUUAUGGUAUUCCACCUUCCC	22	mir-2985	3′	107	−199.08

由表 1 可见，大黄鱼 miRNA 序列长度 17 ~ 25 nt，平均长度为 21 nt。种子区是成熟 miRNA 序列 5′端第 2 至第 8 个碱基的区域，此区域的序列具有高度保守的特性。miRNA 依据种子区的序列进行家族归类，54 个成熟的 miRNA 序列最后归类为 29 个 miRNA 家族，其中 let - 7 家族是包含成熟 miRNA 最为丰富的 miRNA 家族，在预测的结果中包含了 6 个成熟的 miRNA。这一结果与其他很多研究是相同的，如 Huang 等[13]在大黄鱼全基因组上发现的 miRNA，其中最为丰富的家族是 let - 7 家族。在大黄鱼的全基因组和全转录组中 let - 7 家族都比较丰富，这说明 let - 7 家族无论是在大黄鱼的基因组水平还是全转录组的水平中都发挥着很重要的作用。

2.2 成熟 miRNA 长度及碱基分析

图 2 为预测 miRNA 和 miRBase 中其他动物（指的是 miRBase 中收录的包含从海洋腔肠动物到陆地高等脊椎动物）的比较结果。从长度分布来看，大体的趋势是相似的，预测得到大黄鱼 miRNA 长度主要为 22 nt，其次分别是 23 nt 和 21 nt。长度分布存在不同，主要的原因可能是计算的数量相差比较大，在 miRBase 中动物成熟 miRNA 去掉冗余的序列后有 14 990 条。

本研究发现每个碱基所占的比例分别为腺嘌呤（A）23.80%、胞嘧啶（C）19.78%、尿嘧啶（U）29.96%、鸟嘌呤（G）26.45%。在之前婆罗洲猩猩（*Pongo pygmaeus*）以及黑猩猩（*Pan troglodytes*）[21]中发现的 miRNA 中胞嘧啶所占总体的比例也是最低的。miRNA 每个位置的碱基频率如图 3 所示，说明不同物种之间 miRNA 碱基种类所占的比例也存在着相似之处。

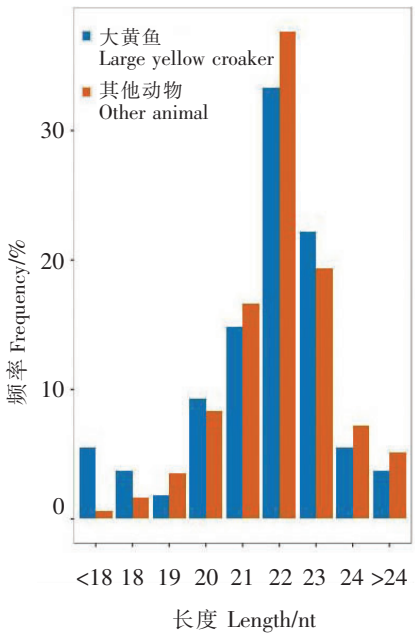
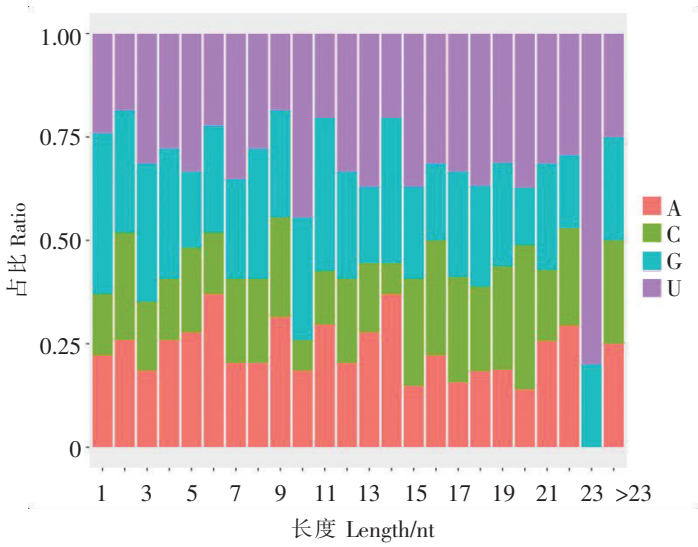


图 2 大黄鱼和其他动物 miRNA 序列长度分布比较

Fig.2 Comparison of miRNA length distribution in large yellow croaker and other animals



说明：横坐标为成熟 miRNA 序列 5′到 3′碱基位置，纵坐标为四种碱基在该位置的占比。

Notes:Horizontal ordinate represents each position in mature miRNA sequence from 5′ to 3′, and the vertical ordinate represents four kinds of nucleotides' ratio.

图 3 miRNA 序列每个位置碱基频率分布

Fig.3 Base distribution of each position in miRNA sequences

2.3 大黄鱼 miRNA 前体的分析

预测得到的前体长度为 56 ~ 176 nt，平均长度为 88.7 nt，形成具有颈环发夹状二级结构的 RNA 分子，其中除了 miRNA 的前体外还有其他种类的 RNA 分子如 mRNA、tRNA 以及 rRNA。在本流程结果中，miRNA 前体的最小自由能为 - 331.80 ~ - 68.4 kJ/mol，平均最小自由能为 - 162.02 kJ/mol。miRNA 前体在酶（RNase Ⅲ enzyme Dicer）的作用下剪切成为成熟的 miRNA，在生物体中有的前体可以产生多个 miRNA，但为了提高生物信息预测的准确性，本流程只考虑一个前体生成一个成熟的 miRNA。前体的最大特征就是可以自身形成茎环结构（见图 4），从本结果中随机选出的前体均能形成茎环结构，而且 miRNA 序列不在环上。

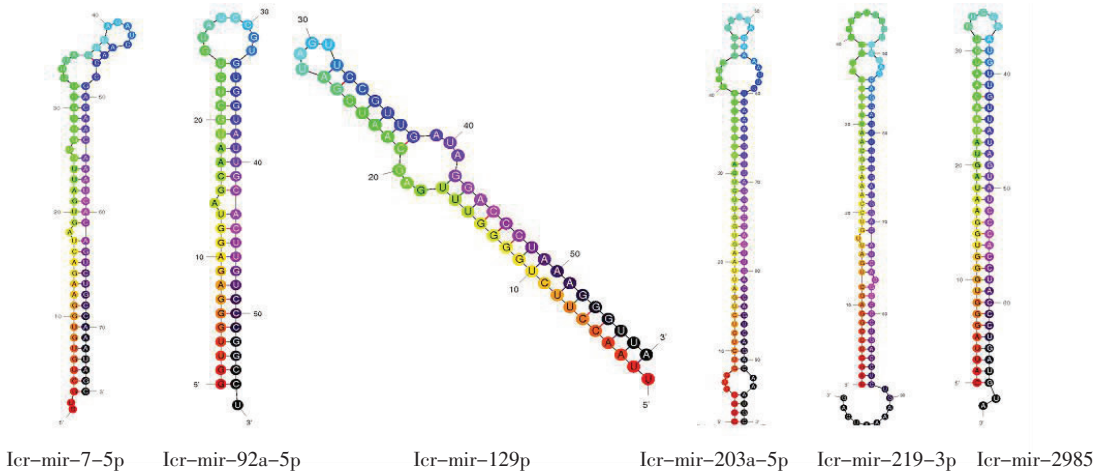


图 4　对预测出的 miRNA 前体进行二级结构检验

Fig.4　Test for the pre-miRNA secondary structure

2.4　miRNA 前体序列进化分析

从预测的结果中选取 4 个 miRNA 家族共 6 条 miRNA 的前体序列和其他物种相同的 4 个 miRNA 家族构建了进化树，结果见图 5。

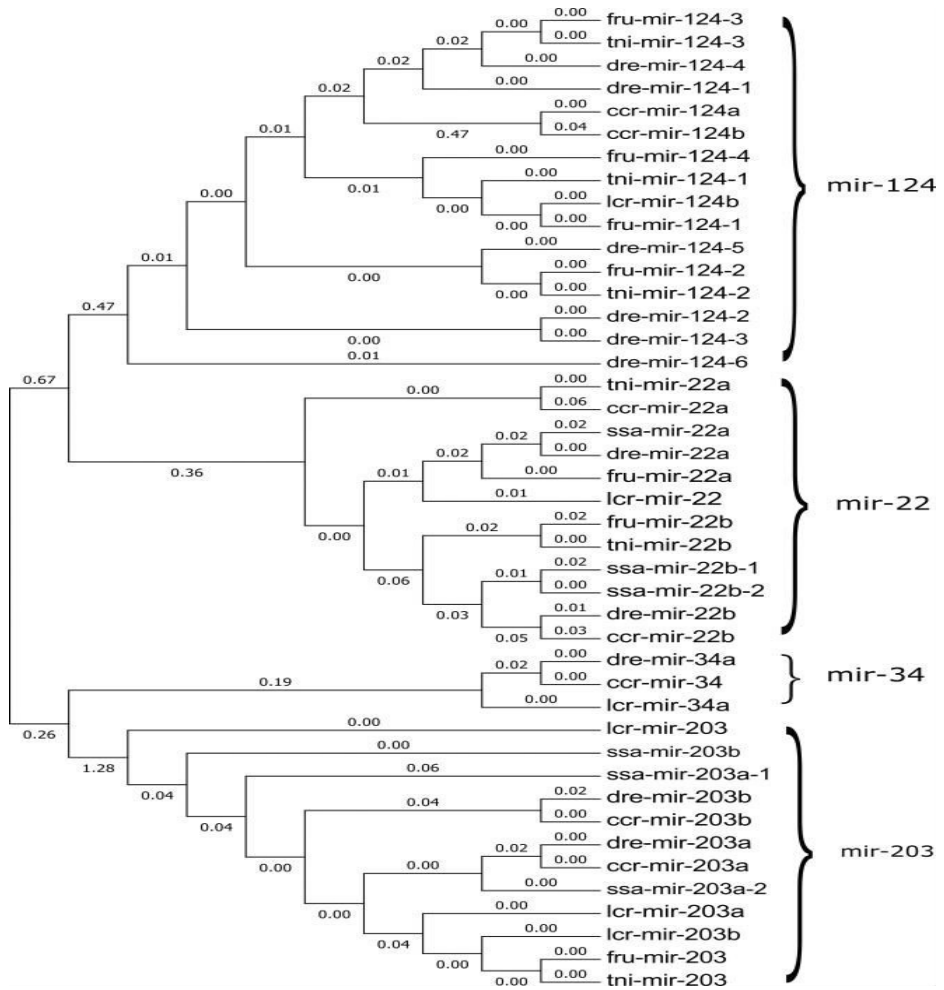


图 5　基于前体序列构建的进化树

Fig.5　The phylogenetic trees of precursor sequences

从图 5 可以看出：在 mir－12 家族中，大黄鱼的 lcr－mir－12 的前体和红鳍东方鲀 fru－mir－12 的前体是极其相似的，在进化树中被归为同一支，这体现出同一家族 miRNA 前体序列在不同物种之间的保守性；但 mir－22 家族和 mir－34 家族与其他物种比较则差距较大，表明这两个家族 miRNA 前体在不同物种之间体现了多样性；miRNA 前体在不同物种之间有保守性也同时兼具多样性，这一点能从 mir－203 家族中体现出来，大黄鱼的 lcr－mir－203a 和 lcr－mir－203b 同红鳍东方鲀的 fru－mir－203 以及金娃娃的 tni－mir－203 归为同一支，但与 lcr－mir－203 却是有着明显的差别。在之前其他物种的研究[23]中也有对 miRNA 前体保守性和多样性的证实。

2.5　miRNA 靶基因预测及功能注释

为降低靶基因预测的假阳性结果，本研究中使用 RNAhybrid 和 miRanda 两种软件，通过 RNAhybrid 预测得到 52 个 miRNA 调控23 348个基因（这里的基因来自大黄鱼全基因组的注释文件中的 unigene），通过 miRanda 得到 51 个 miRNA 调控 2373 个基因，取两结果的交集，最终结果为预测出 51 个 miRNA 调控 2360 个靶基因。

将预测得到的靶基因进行注释和富集分析，结果如图 6 所示。在生物学过程（biological process）中，注释得到的结果较多的是 cellular process（细胞组成）；在细胞组成中注释最多的是 cell（细胞）；在 molecular function（分子功能）中注释最多的前两位分别是 binding（结合）和 catalytic activity（催化活性）。这一结果同大黄鱼在基因组中预测得到的结果是相同的。

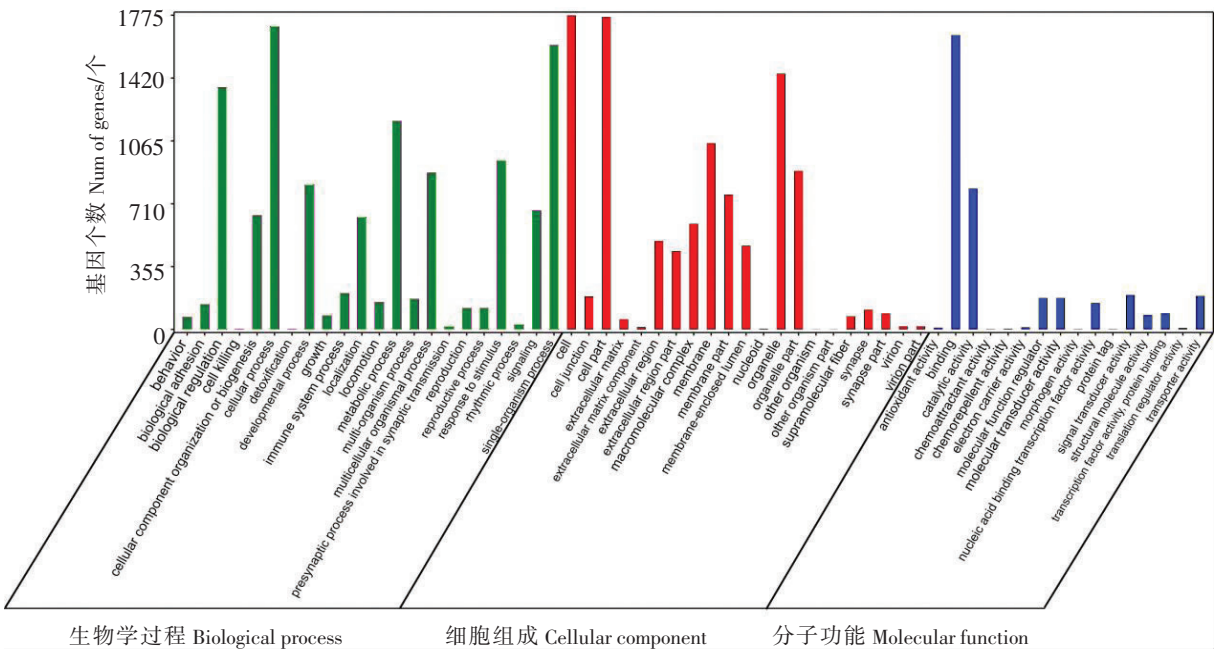


图 6　GO 注释结果

Fig.6　GO annotation result

miRNA 靶基因 GO 注释的结果是将靶基因归类（靶基因的 GO 注释结果见附录），利用超几何分布计算方法对基因进行 GO terms 富集度统计分析，计算出差异基因 GO term 的 P 值或 q 值，再定位差异基因最可能相关的 GO term。本研究按照三种分类分别做了 GO 富集分析，对每一类只选取了富集程度最高的前 10 个结果（p 值越小富集效果越明显）。从富集的结果（见表 2）看，在生物学过程中，富集到最多的是 negative regulation of smooth muscle cell migration（平滑肌细胞迁移的负调控），以及前 10 位中其他的几种。生物学过程都是有关于负向调控的，这和 miRNA 作用原理是相同的，很多 miRNA 是结合靶基因，从而抑制基因的表达，起到负向调控的作用。在细胞组成方面，富集结果最明显的是 extrinsic component of membrane（生物膜的外部组成），在分子功能方面，富集最明显的是 microtubule binding（微管结合）（见表 2）。

表 2　GO 富集结果

Tab. 2　The result of GO enrichment

GO 类别	GO ID	描述 Description	基因个数(比例) Gene(Ratio)	背景基因个数(比例) Bg(Ratio)	*P* 值 *P* value
生物学过程 Biological process	GO:0014912	Negative regulation of smooth muscle cell migration 平滑肌细胞迁移的负调控	8(0.41%)	14(0.07%)	0.000 012
	GO:0033119	Negative regulation of RNA splicing RNA 剪切负调控	11(0.56%)	34(0.16%)	0.000 199
	GO:0072718	Response to cisplatin 对顺氯氨铂的反应	6(0.31%)	11(0.05%)	0.000 220
	GO:0071417	Cellular response to organonitrogen compound 细胞对有机氮化合物的反应	73(3.72%)	510(2.47%)	0.000 247
	GO:0031333	Negative regulation of protein complex assembly 蛋白复合物合成的负向调控	25(1.27%)	126(0.61%)	0.000 294
	GO:0007264	Small GTPase mediated signal transduction 酶间接信号转换	102(5.2%)	776(3.75%)	0.000 438
	GO:1901699	Cellular response to nitrogen compound 细胞对含氮化合物的反应	76(3.87%)	547(2.65%)	0.000 466
	GO:0050794	Regulation of cellular process 细胞过程的调控	1265(64.48%)	12 623(61.07%)	0.000 592
	GO:0050686	Negative regulation of mRNA processing mRNA 处理负调控	12(0.61%)	44(0.21%)	0.000 609
	GO:0032271	Regulation of protein polymerization 蛋白聚合作用调控	37(1.89%)	224(1.08%)	0.000 612
细胞组成 Cellular component	GO:0019898	Extrinsic component of membrane 生物膜的外部组成	60(3.04%)	414(1.98%)	0.000 559
	GO:0005912	Adherens junction 附着结合部分	96(4.87%)	741(3.55%)	0.000 898
	GO:0017059	Serine C – palmitoyltransferase complex 丝氨酸棕榈酰转移酶 C 复合物	5(0.25%)	10(0.05%)	0.001 260
	GO:0031211	Endoplasmic reticulum palmitoyltransferase complex 内质网脂酰转移复合物	5(0.25%)	10(0.05%)	0.001 260
	GO:0070161	Anchoring junction 锚定链接部分	97(4.92%)	770(3.69%)	0.002 059
	GO:0001931	Uropod 尾肢	5(0.25%)	12(0.06%)	0.003 370
	GO:0030027	Lamellipodium 伪足	40(2.03%)	275(1.32%)	0.004 019
	GO:0005768	Endosome 内体	121(6.14%)	1016(4.87%)	0.004 466
	GO:0048471	Perinuclear region of cytoplasm 胞核周区	107(5.43%)	885(4.24%)	0.004 629
	GO:0002178	Palmitoyltransferase complex 棕榈酰转移酶复合物	5(0.25%)	13(0.06%)	0.005 053
分子功能 Molecular function	GO:0008017	Microtubule binding 微管结合	50(2.54%)	330(1.59%)	0.000 672
	GO:0008413	8 – oxo – 7,8 – dihydroguanosine triphosphate pyrophosphatase activity 腺苷三磷酸酶的活性	3(0.15%)	3(0.01%)	0.000 862
	GO:0035539	8 – oxo – 7,8 – dihydrodeoxyguanosine triphosphate pyrophosphatase activity 腺苷三磷酸酶的活性	3(0.15%)	3(0.01%)	0.000 862
	GO:0032403	Protein complex binding 蛋白复合结合	169(8.57%)	1413(6.83%)	0.000 974
	GO:0044877	Macromolecular complex binding 大分子复合物的结合	238(12.08%)	2091(10.1%)	0.001 569
	GO:0015630	Tubulin binding 微管蛋白结合	60(3.04%)	435(2.1%)	0.002 232
	GO:0016503	Pheromone receptor activity 信息素受体活性	4(0.2%)	7(0.03%)	0.002 265
	GO:0001069	Regulatory region RNA binding 调节区 RNA 结合	3(0.15%)	4(0.02%)	0.003 203
	GO:0008092	Cytoskeletal protein binding 细胞骨架蛋白结合	157(7.97%)	1339(6.47%)	0.003 243
	GO:0016408	C – acyltransferase activityC 酰基转移酶活性	6(0.3%)	17(0.08%)	0.003 635

3 讨论

3.1 miRNA 生物信息学预测流程的优化

miRNA 自被发现以来一直成为非编码 RNA 研究的焦点。从挖掘 miRNA 方面来看，从最初的克隆法到现在的高通量测序法和生物信息法，在不同的物种上都有对 miRNA 的相关研究，目前国内相关的报道中也有对 miRNA 的预测分析，如宋长年等[24]对 32 种果树的 miRNA 生物信息学预测，朱珊珊等[25]对美洲鲎的 miRNA 预测，以及柳承璋等[26]对淡水枝角水蚤 miRNA 的发掘，但从生物信息学预测的流程看还存在着许多的不足之处。本研究优化了 miRNA 生物信息预测流程，在比对方面，本研究使用 BLAST 比对尽可能多地预测得到 miRNA；BLAST 设置的 E 值一般都比较高，这会使误差增大，而本研究设定错配碱基个数不大于 3，则可以在很大程度上降低这种误差；对于预测中前体筛选，本研究使用了 36 个特征值进行 SVM 模型训练，并使用该模型筛选前体的序列，其中这 36 个特征值是文献［22，27－28］中报道的相关性比较大的特征值，其训练的模型的分类效果也更明显，从而降低了结果中的假阳性率，使结果更为准确。

3.2 miRNA 预测结果比较

本研究通过对预测得到的成熟 miRNA 序列同其他物种的碱基组成进行了比较，发现本文结果同之前的报道[21]是相似的；将预测得到的 miRNA 与 miRBase 中 miRNA 序列的长度分布进行比较，发现预测 miRNA 的长度分布范围和 miRBase 中的总体来说是相似的。对于得到的前体，分别进行了二级结构的检验，以及同其他几个物种前体序列一起构建了系统进化树，发现与其近缘物种同一家族的 miRNA 前体显示出较高的保守性以及多样性，这同文献［25］对于 miRNA 前体进化分析的结果也是相似的。通过这些比较，得出本流程预测得到的 miRNA 结果是较为准确的。

3.3 miRNA 靶基因预测以及 GO 注释分析

国内现有的关于 miRNA 预测分析的文献，很少有对 miRNA 进行靶基因的预测以及后续的分析，而对于 miRNA 的研究最终还是要归结为对 miRNA 调控的靶基因研究。本研究用两种靶基因预测软件对挖掘得到的 miRNA 预测了靶基因，并用 BLAST2GO 对靶基因进行了 GO 注释，将其注释到生物学过程、细胞组成以及分子功能中，再对 GO 注释的结果进行了富集分析。本研究预测的 miRNA 靶基因以及靶基因的 GO 注释都为大黄鱼 miRNA 后续研究奠定了基础。

［参考文献］

［1］ AMBROS V. The functions of animal microRNAs［J］. Nature, 2004, 431(7006): 350-355. DOI: 10.1038/nature02871.

［2］ BARTEL D P. MicroRNAs: genomics, biogenesis, mechanism, and function［J］. Cell, 2004, 116(2): 281-297. DOI: 10.1016/S0092-8674(04)00045-5.

［3］ HE L, HANNON G J. MicroRNAs: small RNAs with a big role in gene regulation［J］. Nature Reviews Genetics, 2004, 5(7): 522-531. DOI: 10.1038/nrg1379.

［4］ LEE R C, FEINBAUM R L, AMBROS V. The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14［J］. Cell, 1993, 75(5): 843-854. DOI: 10.1016/0092-8674(93)90529-Y.

［5］ HWANG H W, MENDELL J T. MicroRNAs in cell proliferation, cell death, and tumorigenesis［J］. Br J Cancer, 2006, 94(6): 776-780. DOI: 10.1038/sj.bjc.6603023.

［6］ LENKALA D, LACROIX B, GAMAZON E R, et al. The impact of microRNA expression on cellular proliferation［J］. Hum Genet, 2014, 133(7): 931-938. DOI: 10.1007/s00439-014-1434-4.

［7］ SHIVDASANI R A. MicroRNAs: regulators of gene expression and cell differentiation［J］. Blood, 2006, 108(12): 3646-3653. DOI: 10.1182/blood-2006-01-030015.

［8］ YAO S. MicroRNA biogenesis and their functions in regulating stem cell potency and differentiation［J］. Biol Proced Online, 2016, 18: 8. DOI: 10.1186/s12575-016-0037-y.

[9] FLEMING J L, GABLE D L, SAMADZADEH-TARIGHAT S, et al. Differential expression of miR-1, a putative tumor suppressing microRNA, in cancer resistant and cancer susceptible mice [J/OL]. Peer J, 2013; 1:e68. http://europepmc.org/articles/PMC3642704. DOI:10.7717/peerj.68.

[10] TÜFEKCI K U, MEUWISSEN R L J, GENç Ş. The role of microRNAs in biological processes [C] //miRNomics: MicroRNA Biology and Computational Analysis. Clifton, NJ: Humana Press, 2014: 15-31. DOI:10.1007/978-1-62703-748-8_2.

[11] HANNON G J. RNA interference [J]. Nature, 2002, 418(6894):244-251. DOI:10.1038/418244a.

[12] YANG Z, WANG L. Regulation of microRNA expression and function by nuclear receptor signaling [J]. Cell & bioscience, 2011, 1(1):1. DOI:10.1186/2045-3701-1-31.

[13] HUANG Y, CHENG J H, LUO F N, et al. Genome-wide identification and characterization of microRNA genes and their targets in large yellow croaker (*Larimichthys crocea*) [J]. Gene, 2016, 576(1/2):261-267. DOI:10.1016/j.gene.2015.10.044.

[14] QI P, GUO B, ZHU A, et al. Identification and comparative analysis of the *Pseudosciaena crocea* microRNA transcriptome response to poly(I:C) infection using a deep sequencing approach [J]. Fish Shellfish Immunol, 2014, 39(2): 483-491. DOI:10.1016/j.fsi.2014.06.009.

[15] ZHOU M, WANG Q, SUN J, et al. *In silico* detection and characteristics of novel microRNA genes in the *Equus caballus* genome using an integrated *ab initio* and comparative genomic approach [J]. Genomics, 2009, 94(2):125-131. DOI:10.1016/j.ygeno.2009.04.006.

[16] PATANUN O, LERTPANYASAMPATHA M, SOJIKUL P, et al. Computational identification of microRNAs and their targets in cassava (*Manihot esculenta* Crantz.) [J]. Mol Biotechnol, 2013, 53(3):257-269. DOI:10.1007/s12033-012-9521-z.

[17] HAN J, LI A, LIU H, et al. Computational identification of microRNAs in the strawberry (*Fragaria x ananassa*) genome sequence and validation of their precise sequences by miR-RACE [J]. Gene, 2014, 536(1):151-162. DOI:10.1016/j.gene.2013.11.023.

[18] HAN J, XIE H, KONG M L, et al. Computational identification of miRNAs and their targets in *Phaseolus vulgaris* [J]. Genet Mol Res, 2014, 13(1):310-322. DOI:10.4238/2014.January.17.16.

[19] BAEV V, DASKALOVA E, MINKOV I. Computational identification of novel microRNA homologs in the chimpanzee genome [J]. Comput Biol Chem, 2009, 33(1):62-70. DOI:10.1016/j.compbiochem.2008.07.024.

[20] TONG C Z, JIN Y F, ZHANG Y Z. Computational prediction of microRNA genes in silkworm genome [J]. J Zhejiang Univ Sci B, 2006, 7(10):806-816. DOI:10.1631/jzus.2006.B0806.

[21] WANG B. Base composition characteristics of mammalian miRNAs [J/OL]. Journal of Nucleic Acids, 2013, http://www.oalib.com/paper/3077084#WUCZG6FY9Yc. DOI:10.1155/2013/951570.

[22] BATUWITA R, PALADE V. microPred: effective classification of pre-miRNAs for human miRNA gene prediction [J]. Bioinformatics, 2009, 25(8):989-995. DOI:10.1093/bioinformatics/btp107.

[23] BARIK S, SARKARDAS S, SINGH A, et al. Phylogenetic analysis reveals conservation and diversification of micro RNA166 genes among diverse plant species [J]. Genomics, 2014, 103(1):114-121. DOI:10.1016/j.ygeno.2013.11.004.

[24] 宋长年，贾启东，王晨，等. 32 种果树 microRNA 的生物信息学预测与分析 [J]. 园艺学报，2010，37(6)：869-879.

[25] 朱珊珊，翁朝红，肖世俊，等. 美洲鲎 miRNA 的生物信息学挖掘与分析 [J]. 集美大学学报（自然科学版），2016，21(2)：107-114.

[26] 柳承璋，李富花，相建海. 淡水枝角水蚤（Daphnia pulex）微小 RNA（miRNA）的生物信息学发掘与分析 [J]. 海洋与湖沼，2013，44(4)：837-845.

[27] DING J, ZHOU S, GUAN J. MiRenSVM: towards better prediction of microRNA precursors using an ensemble SVM classifier with multi-loop features [J/OL]. BMC Bioinformatics, 2010(S11). http://link.springer.com/article/10.1186/1471-2015-11-S11-S11.

[28] NG K L, MISHRA S K. De novo SVM classification of precursor microRNAs from genomic pseudo hairpins using global and intrinsic folding measures [J]. Bioinformatics, 2007, 23(11):1321-1330. DOI:10.1093/bioinformatics/btm026.

（责任编辑　朱雪莲　英文审校　马　英）