

# 淡水枝角水蚤(*Daphnia pulex*)微小 RNA (miRNA) 的生物信息学发掘与分析\*

柳承璋<sup>1,2</sup> 李富花<sup>1</sup> 相建海<sup>1</sup>

(1. 中国科学院海洋研究所 青岛 266071; 2. 中国科学院大学 北京 100049)

**摘要** 微小 RNA(microRNA 或 miRNA), 是广泛存在于多细胞生物及其病毒的一类由 22 个左右核苷酸组成的内源性非编码 RNA, 具有转录后水平上的基因表达调控功能。本文在淡水枝角水蚤的全基因组序列的基础上, 通过生物信息学方法对淡水枝角水蚤(*Daphnia pulex*) miRNA 进行了发掘, 得到 miRNA 前体(pre-miRNA) 252 个、可编码 262 个功能 miRNA; 发现了 6 个 miRNA 簇。8 号染色体 miRNA 的出现频率显著高于其他染色体, 而 4 号染色体 miRNA 频率明显显著较低。通过与几种两侧对称动物的聚类分析将 pre-miRNA 归为 191 个 miRNA 家族, 其中 105 个家族为淡水枝角水蚤所特有, 15 个家族为节肢动物门特有。在几种两侧对称动物中, 淡水枝角水蚤的 miRNA 与果蝇最为相似。

**关键词** 淡水枝角水蚤; 甲壳动物; microRNA; miRNA; 微小 RNA; 生物信息学

**中图分类号** Q753

自 1958 年 Crick 提出分子生物学的中心法则以来, 对遗传信息的研究主要聚焦在 DNA 和蛋白质, 因为前者是遗传信息的存储和承载者, 后者是遗传信息的表达和执行者。然而令人惊奇的是, 近年来研究者发现, 哺乳动物转录组的最大组成部分却是非编码 RNA(Okazaki *et al*, 2002), 也就是说, DNA 传递的遗传信息有相当部分没有到达蛋白质, 而是仅供非编码 RNA 使用, 这暗示着有大量的遗传信息需要在转录之后、翻译之前得到处理和传递。认识到这一问题之后, 在转录后水平上的基因表达调控受到越来越多的重视。微小 RNA(microRNA 或 miRNA)是近年来发现的最重要的一类转录后调控因子。

MiRNA, 广泛存在于病毒、多细胞动物和植物中, 是一类由 22 个左右核苷酸组成的内源性非编码 RNA(ncRNA), 它们是由 mRNA 上长约 80nt、形似发夹环的 miRNA 前体(miRNA precursor, pre-miRNA)经酶切而产生的(Gregory *et al*, 2004)。miRNA 与

argonaute 蛋白等结合可形成 RNA 诱导的沉默复合体(RNA induced silencing complex, RISC), RISC 可与靶标 mRNA 的 3'非翻译区(3')进行序列特异性结合, 从而抑制翻译的起始(Humphreys *et al*, 2005; Nilsen, 2007; Standart *et al*, 2007)或降解靶标 mRNA(Bagga *et al*, 2005; Iorio *et al*, 2005), 是一类重要的转录后基因表达调控因子。据估计, 动物 miRNA 基因数量约相当于其编码蛋白基因数量的十分之一。人类 miRNA 的数量超过一千种(Bentwich *et al*, 2005; Griffiths-Jones *et al*, 2008), 每一种 miRNA 可以调控数十至数百个靶标基因(Lim *et al*, 2005; Friedman *et al*, 2009)。据估计, 动物中约 30%—60%编码蛋白的基因受到 miRNA 的调控, 涉及已知的所有细胞生物学过程(Bushati *et al*, 2007; Rana, 2007; Friedman *et al*, 2009)。

miRNA 在进化方面具有两个非常明显的特点: 首先, 在多细胞动物进化过程中, miRNA 数量不断增

\* 国家自然科学基金重大项目, 30730071 号; 国家 863 项目, 2012AA10A404 号, 2012AA092205 号。柳承璋, 助理研究员, E-mail: liu@qdio.ac.cn

通讯作者: 相建海, 研究员, E-mail: jhxiang@qdio.ac.cn

收稿日期: 2011-06-27, 收修改稿日期: 2011-11-15

加,部分关键的 miRNA 因经历了高强度的负选择,序列变异频率较低,这使得研究者有可能准确地追溯某一类群动物共同祖先的 miRNA 组成(Sempere *et al*, 2006; Ruby *et al*, 2007; Stark *et al*, 2007)。同源且功能相近的 miRNA 类群被称为 miRNA 家族,其特征是具有相同的种子序列(miRNA 5'端第 2 到 7 位核苷酸)。对 miRNA 家族进化的研究已经广泛地应用于生物系统进化分析(Xie *et al*, 2004; Farh *et al*, 2005; Berezhikov *et al*, 2006; Gardner *et al*, 2008; Wheeler *et al*, 2009)。由于 miRNA 与其调控的靶标即 mRNA 的 3'UTR 之间只要通过单向、独立的序列变异就有可能建立新的调控关系,因此其进化速度应该比蛋白转录因子介导的调控要高得多。对 miRNA 家族进化史的研究发现,miRNA 的进化可能是脊椎动物结构复杂性迅速提高的推动力(Heimberg *et al*, 2008)。因此,对特定类群生物的 miRNA 进行研究对于理解 miRNA 的系统发生以及该生物类群的进化具有重要意义。

甲壳动物隶属于节肢动物门,包含了具有重要学术研究价值的化石种类(鲎)和某些世界上养殖产值最大的水产动物(虾、蟹),然而对甲壳动物 miRNA 的研究尚处于起步阶段,迄今为止,只对个别物种开展了相关研究:Ruan 等(2011)利用小 RNA 克隆的方法鉴定了 35 个日本囊对虾(*Marsupenaeus japonicus*) miRNA,发现其中 22 个的表达与白斑杆状病毒(WSSV)感染相关(Ruan *et al*, 2011);Yang 等通过 Illumina/Solexa 测序技术鉴别了 24 个与先天性免疫相关的日本囊对虾 miRNA (Yang *et al*, 2012);Ou 等(2012)同样利用高通量测序技术对螺原体(*Spiroplasma eriocheiris*)感染前后的中华绒螯蟹(*Eriocheir sinensis*)血细胞 miRNA 进行研究,鉴定了 735 个 miRNA,包括 36 个已知的甲壳动物 miRNA,134 个已知的节肢动物 miRNA 以及 565 个新的 miRNA,其中 228 个在感染过程中有显著的表达变化 (Ou *et al*, 2012);Blythe 等通过 SOLiD 高通量测序技术对夏威夷钩虾(*Parhyale hawaiiensis*)不同发育胚胎阶段进行了 miRNA 检测,发现了 55 个保守的 miRNA (Blythe *et al*, 2012)。目前在 miRNA 注册数据库中只有 53 个甲壳动物的 miRNA,远少于节肢动物门的昆虫纲(2633 个 miRNA)以及其他重要模式生物。

淡水枝角水蚤(*Daphnia pulex*)为浮游生活的甲壳动物,隶属鳃足亚纲、枝角目,广布于全球淡水静止水体,常成为该生态系统的优势种,在物质能量循环中占据着中心位置,具有重要的生态学地位;同时由

于对人类活动造成的环境污染非常敏感,又是理想的环境指示生物和毒性实验材料;其世代短,繁殖快,具有孤雌生殖能力;既便于野外采样调查,又非常适合实验室操作。淡水枝角水蚤已经成为研究甲壳动物的生物学、生态学、生理学、毒理学、种群遗传学和进化的重要模式生物。其全基因组序列于 2007 年完成测序。虽然基因组序列大小只有人类基因组大小的十分之一(200Mb),但基因数量却比人类多 8000 多个,而且超过三分之一的基因是该物种所特有的,其基因组中存在大量的基因重复现象,可能是在进化中对多变环境压力进行适应的结果(Cristescu *et al*, 2006; Colbourne *et al*, 2011)。淡水枝角水蚤基因组全序列的公布使其迅速发展成为利用基因组学进行生态和进化研究的模式生物。

本文利用生物信息学方法,对淡水枝角水蚤的 miRNA 进行了预测和分析,研究结果可为了解其进化和对环境的适应提供新的视角。

## 1 材料与方法

### 1.1 数据来源

淡水枝角水蚤基因组从 The Daphnia Genomics Consortium (DGC)的网站(<http://daphnia.cgb.indiana.edu/>)下载。已知的 miRNA 以及 pre-miRNA 序列来自 miRBase (<http://microrna.sanger.ac.uk/sequences/>)。

### 1.2 miRNA 的预测算法

用 GENSCAN 软件(<http://genes.mit.edu/GENSCAN.html>)对淡水枝角水蚤基因组序列进行开放阅读框分析 (Burge *et al*, 1997),提取所有非编码序列,包括基因间隔区(intergenic region)、内含子(intron)和非翻译区(UTR),作为可能编码 miRNA 的源区域;利用 miRbase 注册的所有成熟 miRNA 序列(~22nt)作为查询(query)序列,用 BLAST 软件比对淡水枝角水蚤基因组的所有非编码基因组序列,筛选可能的成熟 miRNA 序列。筛选标准要求序列相似度大于 90% (Norden *et al*, 2007),并且必须包含高度保守的已知 miRNA 的种子序列(第 2 至第 7 位碱基);对符合条件的可能含有 miRNA 的序列提取其侧翼序列,用软件 MIREAP(<http://sourceforge.net/projects/mireap/>)检查是否存在具有 pre-miRNA 特征的二级结构(采用 MIREAP 缺省参数:miRNA 序列长度 = 18—26nt, pre-miRNA 最小自由能 < -18 kcal/mole, miRNA 与其互补序列 miRNA\*之间的距离 < 35nt, miRNA-miRNA\*双链错配碱基数 < 4nt)。

利用基于支持向量机(SVM)算法的 triplet-svm 软件对类似 pre-miRNA 样的 RNA 二级结构进行第二次筛选。此算法的准确率在 90%以上, 而且不依赖于 pre-miRNA 在物种间的保守性(Xue *et al*, 2005)。

### 1.3 miRNA 的富集分析

以不同编码区之间的 miRNA 富集分析为例: 将编码区长度除以可能产生 pre-miRNA 的最小长度 100nt, 得到可能产生 miRNA 的位点总数, 用确切概率法检验预测发生 miRNA 的位点与未发生 miRNA 的位点的分布在某种编码区(如内含子)和其余编码区之间有无显著差异。

不同染色体的 miRNA 富集分析首先需要根据淡水枝角水蚤的遗传连锁图将各个 scaffold 定位到各染色体上; 并用各染色体上定位的 scaffold 长度之和代表染色体序列长度。其余步骤同上段。

### 1.4 miRNA 簇分析

多数 miRNA 单独分布于不同的开放阅读框(ORF)或者基因间隔区中, 然而也有少数不同的 miRNA 位于同一个 ORF 或者基因间隔区内, 说明它们之间可能因为具有功能上的联系而共同转录。我们将这类 miRNA 归纳为 miRNA 簇。

### 1.5 miRNA 家族聚类分析

选择淡水枝角水蚤、拟暗果蝇(*Drosophila pseudoobscura*)、两种线虫(*Caenorhabditis elegans* 和 *C. briggsae*)、淡水三角涡虫(*Schmidtea mediterranea*)、斑马鱼(*Danio rerio*)6 个物种, 分别代表甲壳动物、昆虫、线形动物、冠轮动物、后口动物类群进行 miRNA 家族聚类分析。由于成熟 miRNA 序列太短, 通过对 miRNA 前体(pre-miRNA)序列进行比较来确定 miRNA 的家族归类是较为可靠的方法(Griffiths-Jones *et al*, 2003)。本文分析针对 pre-miRNA 的二级结构序列进行, 其中除淡水枝角水蚤之外的其他各物种的 pre-miRNA 序列均从 miRbase 数据库下载; 聚类方法采用基于 bootstrap 算法的软件 PBC(the Phylogeny-Bootstrap-Cluster pipeline)(Huang *et al*, 2007); 分析之后, 对部分聚类结果进行人工核对, 以保证结果的可靠性。

## 2 结果与分析

### 2.1 miRNA 预测

利用本文的 miRNA 预测算法, 共预测了淡水枝角水蚤 pre-miRNA 结构 252 个, 可以产生 miRNA 262 个。表 1 列示了 miRNA 家族序号 < 200 的预测结

果(以相似 miRNA 为准; miRNA 家族序号近似代表 miRNA 家族被发现的次序)。图 1 展示了部分 miRNA 前体的二级结构。预测的 miRNA 来自淡水枝角水蚤的所有 12 条染色体上的 123 个 scaffold。淡水枝角水蚤 miRNA 在编码蛋白基因的间隔区出现的次数最多(165/262=63.0%), 编码蛋白基因的内含子次之(66/262=25.2%), 其它的 miRNA 位于 5'UTR(16/262=6.1%)或 3'UTR(15/180=5.7%)。然而如果考虑到序列长度的差异, 则 miRNA 在不同编码区的发生频率没有显著差异(确切概率法,  $P>0.19$ )。此外我们还分析了淡水枝角水蚤各染色体上 miRNA 的分布(图 2), 确切概率法检验发现, 已定位到 12 条染色体上的 scaffold 上的 miRNA 发生频率显著低于没有染色体定位的 scaffold( $P=0$ ); 而在已定位的 12 条染色体中, 8 号染色体的 miRNA 频率显著高于其他染色体( $P=0.007$ ), 而 4 号染色体的 miRNA 频率显著低于其他染色体( $P=0.004$ )。

通过分析基因组空间结构发现了淡水枝角水蚤 miRNA 簇 6 个(表 2)。其中 4 个 miRNA 簇由序列相似的同家族 miRNA 组成, 包括 miR-1661、122 簇各 1 个, miR-1603 簇 2 个, 此外还有混合 miRNA 簇两个, 分别由 miR-317 和 miR-34, 以及 miR-2 和 miR-71 组成。图 3 显示了两个位于编码蛋白基因内含子中的 miRNA 簇。

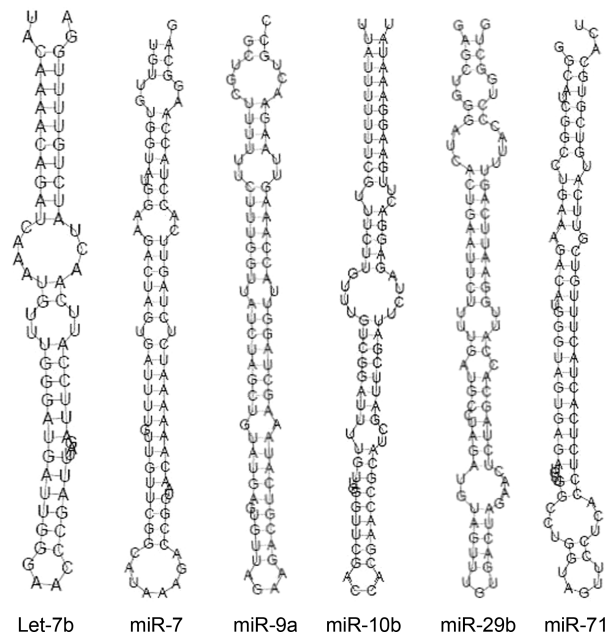


图 1 部分预测 pre-miRNA 二级结构  
Fig.1 Some of the pre-miRNA structures

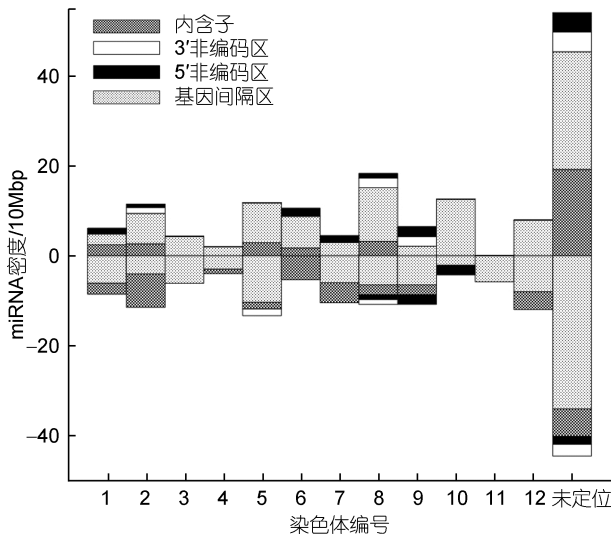


图 2 miRNA 的染色体和编码区分布

Fig.2 Distribution of miRNAs among chromosomes and different coding regions

横轴下方的柱形表示基因组序列互补链上的 miRNA

### 2.2 miRNA 家族聚类分析

通过与几种两侧对称动物 miRNA 前体序列进行 PBC bootstrap 聚类分析, 将 252 个淡水枝角水蚤 miRNA 聚类为 191 个 miRNA 家族; 其中 105 个家族为淡水枝角水蚤所特有; 15 个家族为节肢动物门特有, 包括 miR-12、miR-276、miR-306、miR-317、miR-959、miR-965、miR-2502、miR-2503、miR-2548、miR-2547、

miR-2567 等; 24 个家族与蜕皮动物总门的线虫共有; 只有少数 miRNA 家族跨越后口动物、蜕皮动物和冠轮动物等两侧对称动物各类群, 包括 miR-1、miR-9、miR-137、miR-190 等。与昆虫相比, 淡水枝角水蚤缺少 miRNA 家族 miR-11、miR-14、miR-274 等。图 5 展示了部分重要 miRNA 家族在几种两侧对称动物的分布情况。就淡水枝角水蚤与其他几种两侧对称动物之间共有 miRNA 家族的数量来说, 淡水枝角水蚤与果蝇的 miRNA 相似性最高, 有 44 个共有的 miRNA 家族, 其次是斑马鱼(29), 然后依次是线虫(24)和涡虫(23)。

### 3 讨论

#### 3.1 miRNA 预测方法

在本研究之前, miRbase 注册的淡水枝角水蚤 miRNA 共有 45 个, 全部由 Wheeler 等通过与其他后口动物 miRNA 的序列比对结合 RNA 二级结构分析预测得到(Wheeler *et al*, 2009), 本研究预测的 262 个 miRNA 大幅增加了该物种预测 miRNA 的数量。

在本文发现的两个混合 miRNA 簇当中, miR-2/miR-71 簇在扁形动物门的日本血吸虫、曼氏血吸虫和淡水涡虫以及节肢动物门昆虫纲的意大利蜜蜂中都有报道(Weaver *et al*, 2007; Huang *et al*, 2009; Lu *et al*, 2009; Gomes *et al*, 2011)。本研究发现 miR-2/miR-71 簇

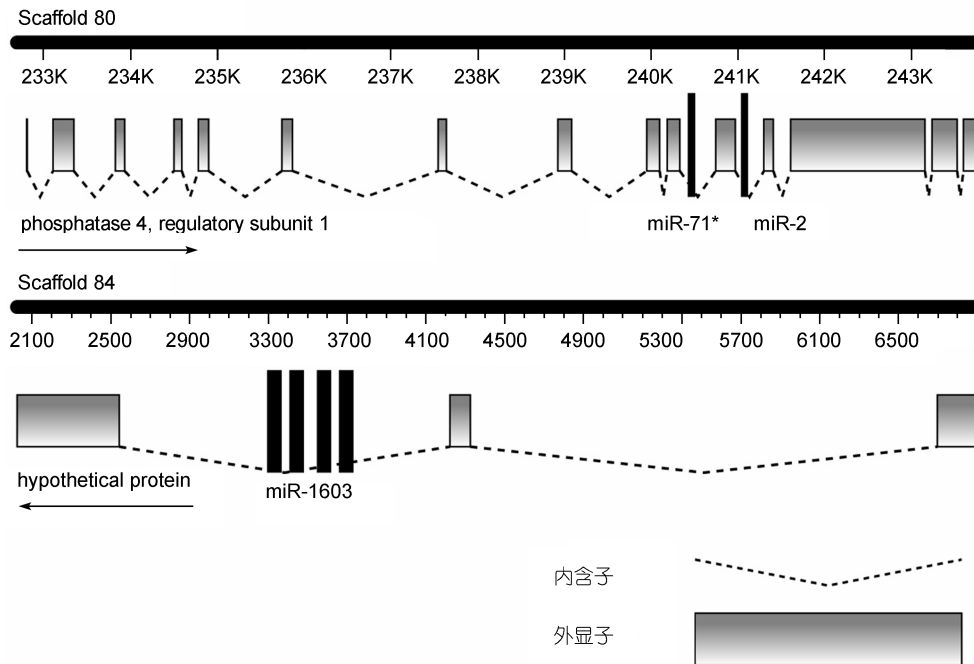


图 3 淡水枝角水蚤的两个 miRNA 基因簇

Fig.3 Two miRNA clusters of *Daphnia pulex*

表 1 淡水枝角水蚤 miRNA 预测部分结果(miRNA 家族序号<200)  
Tab.1 Part of the *Daphnia pulex* miRNA candidates (family ID < 200)

Scaffold	链	区域	Pre-miR 起始	Pre-miR 终止	臂	miRNA 序列	相似 miRNA
1	+	基因间	332500	332586	3	UAGCACCAUUGGAAUUCAGUUU	ssc-miR-29b
1	-	内含子	1720954	1720873	3	UGGAAUGUAAAAGAAGUAUGGAG	cel-miR-1
1	-	内含子	1720954	1720873	5	CCGUGCUUCCUUACUCCCAUA	bmo-miR-1*
1	-	内含子	1847915	1847834	5	UGAGUAAUACAUCAGGUACUGG	dme-miR-12
2	+	基因间	1526203	1526282	3	GUCAUAAAGCUAGGUUACCAAA	bmo-miR-79
2	+	基因间	1526203	1526282	5	UCUUUGGUUAUCUAGCUGUAUG	dme-miR-9a
2	-	基因间	1526281	1526202	5	ACUUUGGUAACCUAGCUUUAUG	ame-miR-9b
3	+	内含子	1681559	1681641	5	AGAUUUGUUUGAUUUCUUGGU	cbr-miR-50
3	-	内含子	3560719	3560636	3	UUGCAUAGUCACAAAAGUGAUG	mmu-miR-153
3	+	内含子	3408836	3408931	5	AAAUACUGUCUGGUGAUCAGAU	spu-miR-200
4	-	基因间	1242124	1242040	5	UGGCAGUGUGGUUAGCUGGUUG	cel-miR-34
9	+	3UTR	936554	936638	5	AGUUUUGCUGGUUGGCAUGAAU	dre-miR-19b*
14	-	基因间	1421810	1421723	3	CGAUUCGAUUCUAGAGGACUUG	oan-miR-10b*
21	+	基因间	613566	613639	3	AUCACUUUCACUGCAUCUCAUC	aly-miR161.2*
38	+	基因间	876130	876225	3	UAUUGCACUCGUCCCGGCCUGU	hsa-miR-92a
74	+	内含子	114750	114848	3	UCAGUCUUUUUUUCUCCUAU	dme-miR-14
80	+	内含子	240420	240504	3	CUCUCACUACUUUGUCGUUCAU	bmo-miR-71*
80	+	内含子	240420	240504	5	UGAAAGACAUGGGUAGUGAGAU	cel-miR-71
80	+	内含子	241033	241114	3	UAUCACAGCCAGCUUUGAUGAG	cel-miR-2
85	+	内含子	171386	171484	3	CCAUUCAACUAUCUGUUUUGGA	sme-let-7b*
91	-	基因间	232317	232219	3	CAUCACAGCGAUGCUACAGACA	sme-miR-2f*
92	+	内含子	410923	411005	3	UUAUUGCUUGAGAAUACACGUU	mmu-miR-137
107	-	基因间	321262	321192	3	AGUGUGACACUGGUUUGACACA	gga-miR-122b
107	-	基因间	321262	321192	3	AGUGUGACACUGGUUUGACACA	gga-miR-122b
115	+	基因间	370158	370240	3	UGAGAUCAUUGUGAAAGCUGAU	cel-miR-81
120	+	内含子	76885	76970	3	AUAAGGCACGCGGUGAAUGCCA	gga-miR-124a
154	+	基因间	94532	94630	3	CAUCACAGCGAUGCUACAGACA	sme-miR-2f*
167	-	基因间	85527	85447	3	UACUGGCCUGCUAAGUCCCAAA	dme-miR-193
191	-	内含子	112623	112541	5	UGGAAGACUAGUGAUUUUGUUG	dme-miR-7
225	+	基因间	106055	106146	3	GACUAAAUGUCAAACGUAAAUU	rno-miR-190b*

表 2 淡水枝角水蚤基因组中的 6 个 miRNA 簇

Tab.2 Six miRNA clusters in the *Daphnia pulex* genome

序号	Scaffold	区域	链	Pre-miR 起始	Pre-miR 终止	臂	miRNA 序列	相似 miRNA
1	4	基因间	-	1244034	1243956	3	GGUGAACACAGCUGGUGGUAUC	bmo-miR-317
1	4	基因间	-	1242124	1242040	5	UGGCAGUGUGGUUAGCUGGUUG	cel-miR-34
2	158	基因间	+	39604	39676	5	GUGGUUUUUUGUCAUUAACU	bta-miR-1603
2	158	基因间	+	39352	39424	5	GUGGUUUUUUGUCAUUAACU	bta-miR-1603
2	158	基因间	+	39718	39790	5	GUGGUUUUUUGUCACAAACU	bta-miR-1603
3	44	基因间	-	484349	484283	5	GAAUGGAGGCUGGUAGGUUUA	zma-miR1661*
3	44	基因间	-	487378	487312	5	GAAUGGAGGCUGGUAGGUUUA	zma-miR1661*
4	107	基因间	-	325266	325196	3	AGUGUGACACUGGUUUGACACA	gga-miR-122b
4	107	基因间	-	321262	321192	3	AGUGUGACACUGGUUUGACACA	gga-miR-122b
5	84	内含子	-	3618	3546	5	GUGGUUUUUUGUCACAAACU	bta-miR-1603
5	84	内含子	-	3479	3407	5	GUGGUUUUUUGUCAUUAACU	bta-miR-1603
5	84	内含子	-	3731	3659	5	GUGGUUUUUUGUCAUUAACU	bta-miR-1603
5	84	内含子	-	3366	3294	5	GUGGUUUUUUGUCACAAACU	bta-miR-1603
6	80	内含子	+	241033	241114	3	UAUCACAGCCAGCUUUGAUGAG	cel-miR-2
6	80	内含子	+	240420	240504	3	CUCUCACUACUUUGUCGUUCAU	bmo-miR-71*



图 4 不同物种 miR-29 家族 pre-miRNA 的序列联配

Fig.4 Pre-miRNA alignments of miR-29 family from different species

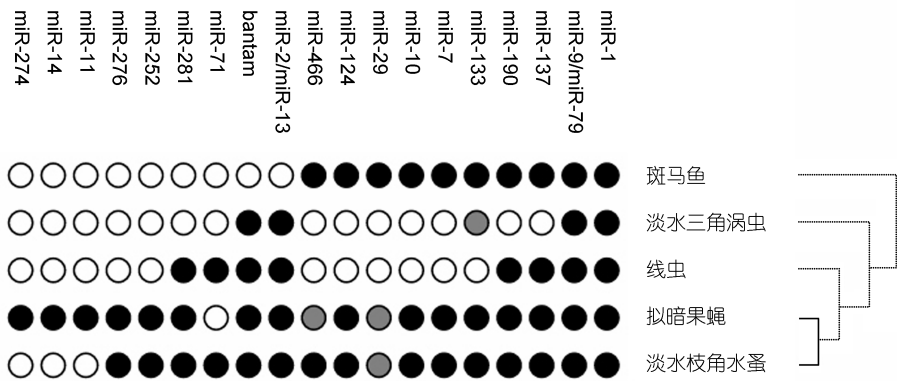


图 5 几种两侧对称动物 miRNA 家族的分布情况

Fig.5 Distribution of miRNA across Bilateria

黑色表示 miRNA 家族存在, 白色表示 miRNA 家族缺失, 灰色表示不确定(序列保守性低)

存在于淡水枝角水蚤磷酸酶亚基的内含子中, 这与 Weaver(2007)在意大利蜂的发现相符, 在果蝇当中此 miRNA 簇随同宿主基因一起丢失了(Weaver *et al.*, 2007)。miR-317/miR-34 簇尚未见于前人的报道。

3.2 miRNA 家族聚类

miR-29 家族此前仅在脊索动物中定义过, 而 miR-285 家族此前只发现于昆虫纲双翅目(Lai *et al.*, 2003), 我们的结果中淡水枝角水蚤 miR-0276、果蝇 miR-285 与几种脊索动物的 miR-29 聚类在一起。图 4 展示了几个物种 miR-29 家族的 pre-miRNA 序列联配情况。其中果蝇与其他物种相似性较低。究竟淡水枝角水蚤 miR-2076 属于 miR-285 还是 miR-29 家族, 以及昆虫纲的 miR-285 家族与脊索动物 miR-29 家族是

否同源, 需要通过更多物种进行研究来解答。

淡水枝角水蚤与斑马鱼和线虫之间共有 miRNA 家族的数量是很有趣的数据, 因为这关系到两侧对称动物的系统发生中存在的争议, 即体腔动物假说(Coelomata hypothesis)和蜕皮动物(Ecdysozoa hypothesis)假说的对立: 前者将脊索动物、节肢动物、环节动物和软体动物等具有由中胚层构成的体腔的类群归纳为真体腔动物, 而将线形动物等归为假体腔动物, 因此节肢动物与脊索动物的关系比与线形动物近(Blair *et al.*, 2002; Wolf *et al.*, 2004; Philip *et al.*, 2005; Rogozin *et al.*, 2007; Zheng *et al.*, 2007); 后者将线形动物门、节肢动物门和其他几个小的门归为一个单源的蜕皮动物类群, 认为其共同特征是具有定期

的蜕皮现象, 而节肢动物与线形动物的关系比与脊索动物近(Aguinaldo *et al*, 1997; Philippe *et al*, 2004; Philippe *et al*, 2005; Dunn *et al*, 2008; Budd *et al*, 2009)。我们的聚类结果略倾向于前者: 淡水枝角水蚤与斑马鱼共有的 miRNA 家族数量(29 个)超过与线虫共有的家族数量(24 个)。鉴于在生物进化中有可能发生 miRNA 家族丢失和相似 miRNA 家族发生的现象(Fu *et al*, 2008), 更准确的结果需要通过对更多物种进行大量分析, 增加对节肢动物中昆虫纲之外类群的研究, 包括甲壳动物。

本文研究内容限于具有物种间保守性的 miRNA。对 miRNA 的生物信息学预测结果的进一步利用可以从验证和扫描其组织表达入手。研究 miRNA 表达以及发现物种特异的 miRNA 可以通过小 RNA 高通量测序进行。

#### 4 结语

以基因组序列为基础的 miRNA 生物信息学预测和分析被证实是一种有效的方法(Grad *et al*, 2003; Lai *et al*, 2003; Lim *et al*, 2003; Pfeffer *et al*, 2005; Norden-Krichmar *et al*, 2007), 其优点是不受 miRNA 表达时空变化的影响, 因而与测序策略相比不易产生假阴性结果。淡水枝角水蚤作为淡水静水生态系统的一种优势物种、理想的环境指示生物和毒性实验材料, 是目前唯一完成基因组测序的甲壳动物, 已经成为甲壳动物生物学、生态学、生理学、毒理学、种群遗传学和进化等研究的重要模式生物(Colbourne *et al*, 2005); miRNA 所介导的转录后基因表达调控系统具有调控反应快、可逆性、进化速度快等特点, 在有机体对环境压力的响应方面具有特殊的作用(Hobert, 2008); 这些特点恰好与淡水枝角水蚤在适应多样和多变的环境时所需要的高度适应性相一致。本文在淡水枝角水蚤的全基因组序列的基础上, 通过生物信息学方法对淡水枝角水蚤 miRNA 进行了发掘, 得到 miRNA 前体 252 个、可编码 262 个功能 miRNA; 发现了 6 个 miRNA 簇。通过与几种两侧对称动物的聚类分析将 miRNA 前体归为 191 个 miRNA 家族, 其中 105 个家族为淡水枝角水蚤所特有, 15 个家族为节肢动物门特有。在几种两侧对称动物中, 淡水枝角水蚤的 miRNA 与果蝇最为相似, 而与斑马鱼的相似程度略高于线虫。这些结果为进一步理解淡水枝角水蚤 miRNA 的进化及其在水蚤对环境适应中的作用打下了良好的基础。

#### 参 考 文 献

- Aguinaldo A M A, Turbeville J M, Linford L S *et al*, 1997. Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature*, 387(6632): 489—493
- Bagga S, Bracht J, Hunter S *et al*, 2005. Regulation by let-7 and lin-4 miRNAs results in target mRNA degradation. *Cell*, 122(4): 553—563
- Bentwich I, Avniel A, Karov Y *et al*, 2005. Identification of hundreds of conserved and nonconserved human microRNAs. *Nature Genetics*, 37(7): 766—770
- Berezikov E, Thuemmler F, van Laake L W *et al*, 2006. Diversity of microRNAs in human and chimpanzee brain. *Nature Genetics*, 38(12): 1375—1377
- Blair J E, Ikeo K, Gojobori T *et al*, 2002. The evolutionary position of nematodes. *Bmc Evolutionary Biology*, 2: 1—7
- Blythe M J, Malla S, Everall R *et al*, 2012. High through-Put sequencing of the parhyale hawaiiensis mRNAs and microRNAs to aid comparative developmental studies. *Plos One*, 7(3): e33784
- Budd G E, Telford M J, 2009. The origin and evolution of arthropods. *Nature*, 457(7231): 812—817
- Burge C, Karlin S, 1997. Prediction of complete gene structures in human genomic DNA. *Journal of Molecular Biology*, 268(1): 78—94
- Bushati N, Cohen S M, 2007. MicroRNA functions. *Annual review of cell and Developmental Biology*, 23: 175—205
- Colbourne J K, Pfrender M E, Gilbert D *et al*, 2011. The ecoresponsive genome of *Daphnia pulex*. *Science*, 331(6017): 555—561
- Colbourne J K, Singan V R, Gilbert D G, 2005. wFleaBase: the *Daphnia* genome database. *Bmc Bioinformatics*, 6: 45
- Cristescu M E A, Colbourne J K, Radivojic J *et al*, 2006. A micro satellite-based genetic linkage map of the waterflea, *Daphnia pulex*: On the prospect of crustacean genomics. *Genomics*, 88(4): 415—430
- Dunn C W, Hejnol A, Matus D Q *et al*, 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature*, 452(7188): 745—U745
- Farh K K H, Grimson A, Jan C *et al*, 2005. The widespread impact of mammalian microRNAs on mRNA repression and evolution. *Science*, 310(5755): 1817—1821
- Friedman R C, Farh K K H, Burge C B *et al*, 2009. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Research*, 19(1): 92—105
- Fu X H, Adamski M, Thompson E M *et al*, 2008. Altered miRNA repertoire in the simplified chordate, *Oikopleura dioica*. *Molecular Biology and Evolution*, 25(6): 1067—1080
- Gardner P P, Vinther J, 2008. Mutation of miRNA target sequences during human evolution. *Trends in Genetics*, 24(6): 262—265
- Gomes M D, Muniyappa M K, Carvalho S G *et al*, 2011.



- Genome-wide identification of novel microRNAs and their target genes in the human parasite *Schistosoma mansoni*. *Genomics*, 98(2): 96—111
- Grad Y, Aach J, Hayes G D *et al*, 2003. Computational and experimental identification of *C-elegans* microRNAs. *Molecular Cell*, 11(5): 1253—1263
- Gregory R I, Yan K P, Amuthan G *et al*, 2004. The Microprocessor complex mediates the genesis of microRNAs. *Nature*, 432(7014): 235—240
- Griffiths-Jones S, Bateman A, Marshall M *et al*, 2003. Rfam: an RNA family database. *Nucleic Acids Research*, 31(1): 439—441
- Griffiths-Jones S, Saini H K, van Dongen S *et al*, 2008. miRBase: tools for microRNA genomics. *Nucleic Acids Research*, 36: D154—D158
- Heimberg A M, Sempere L F, Moy V N *et al*, 2008. MicroRNAs and the advent of vertebrate morphological complexity. *Proceedings of the National Academy of Sciences of the United States of America*, 105(8): 2946—2950
- Hobert O, 2008. Gene regulation by transcription factors and microRNAs. *Science*, 319(5871): 1785—1786
- Huang J, Hao P, Chen H *et al*, 2009. Genome-wide identification of *Schistosoma japonicum* microRNAs using a deep-sequencing approach. *PloS One*, 4(12): e8206
- Huang Y, Gu X, 2007. A bootstrap based analysis pipeline for efficient classification of phylogenetically related animal miRNAs. *Bmc Genomics*, 8: 66
- Humphreys D T, Westman B J, Martin D I K *et al*, 2005. MicroRNAs control translation initiation by inhibiting eukaryotic initiation factor 4E/cap and poly(A) tail function. *Proceedings of the National Academy of Sciences of the United States of America*, 102(47): 16961—16966
- Iorio M V, Ferracin M, Liu C G *et al*, 2005. MicroRNA gene expression deregulation in human breast cancer. *Cancer Research*, 65(16): 7065—7070
- Lai E C, Tomancak P, Williams R W *et al*, 2003. Computational identification of *Drosophila* microRNA genes. *Genome Biology*, 4(7): 1—20
- Lim L P, Lau N C, Garrett-Engele P *et al*, 2005. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature*, 433(7027): 769—773
- Lim L P, Lau N C, Weinstein E G *et al*, 2003. The microRNAs of *Caenorhabditis elegans*. *Genes & Development*, 17(8): 991—1008
- Lu Y C, Smielewska M, Palakodeti D *et al*, 2009. Deep sequencing identifies new and regulated microRNAs in *Schmidtea mediterranea*. *RNA—A Publication of the Rna Society*, 15(8): 1483—1491
- Nilsen T W, 2007. Mechanisms of microRNA-mediated gene regulation in animal cells. *Trends in Genetics*, 23(5): 243—249
- Norden-Krichmar T M, Holtz J, Pasquinelli A E *et al*, 2007. Computational prediction and experimental validation of *Ciona intestinalis* microRNA genes. *Bmc Genomics*, 8: 445
- Okazaki Y, Furuno M, Kasukawa T *et al*, 2002. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature*, 420(6915): 563—573
- Ou J T, Meng Q G, Li Y *et al*, 2012. Identification and comparative analysis of the *Eriocheir sinensis* microRNA transcriptome response to *Spiroplasma eriocheiris* infection using a deep sequencing approach. *Fish & Shellfish Immunology*, 32(2): 345—352
- Pfeffer S, Sewer A, Lagos-Quintana M *et al*, 2005. Identification of microRNAs of the herpesvirus family. *Nature Methods*, 2(4): 269—276
- Philip G K, Creevey C J, McInerney J O *et al*, 2005. The Opisthokonta and the Ecdysozoa may not be clades: Stronger support for the grouping of plant and animal than for animal and fungi and stronger support for the Coelomata than Ecdysozoa. *Molecular Biology and Evolution*, 22(5): 1175—1184
- Philippe H, Lartillot N, Brinkmann H, 2005. Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia. *Molecular Biology and Evolution*, 22(5): 1246—1253
- Philippe H, Snell E A, Baptiste E *et al*, 2004. Phylogenomics of eukaryotes: Impact of missing data on large alignments. *Molecular Biology and Evolution*, 21(9): 1740—1752
- Rana T M, 2007. Illuminating the silence: Understanding the structure and function of small RNAs. *Nature Reviews Molecular Cell Biology*, 8(1): 23—36
- Rogozin I B, Wolf Y I, Carmel L *et al*, 2007. Ecdysozoan clade rejected by genome-wide analysis of rare amino acid replacements. *Molecular Biology and Evolution*, 24(4): 1080—1090
- Ruan L W, Bian X F, Ji Y C *et al*, 2011. Isolation and identification of novel microRNAs from *Marsupenaeus japonicus*. *Fish & Shellfish Immunology*, 31(2): 334—340
- Ruby J G, Stark A, Johnston W K *et al*, 2007. Evolution, biogenesis, expression, and target predictions of a substantially expanded set of *Drosophila* microRNAs. *Genome Research*, 17(12): 1850—1864
- Sempere L F, Cole C N, McPeck M A *et al*, 2006. The phylogenetic distribution of metazoan microRNAs: Insights into evolutionary complexity and constraint. *Journal of Experimental Zoology Part B—Molecular and Developmental Evolution*, 306B(6): 575—588
- Standart N, Jackson R J, 2007. MicroRNAs repress translation of m(7) Gppp-capped target mRNAs in vitro by inhibiting initiation and promoting deadenylation. *Genes & Development*, 21(16): 1975—1982
- Stark A, Kheradpour P, Parts L *et al*, 2007. Systematic discovery and characterization of fly microRNAs using 12 *Drosophila*



- genomes. *Genome Research*, 17(12): 1865—1879
- Weaver D B, Anzola J M, Evans J D *et al*, 2007. Computational and transcriptional evidence for microRNAs in the honey bee genome. *Genome Biology*, 8(6): R97
- Wheeler, B M, Heimberg A M, Moy V N *et al*, 2009. The deep evolution of metazoan microRNAs. *Evolution & Development*, 11(1): 50—68
- Wolf Y I, Rogozin I B, Koonin E V *et al*, 2004. Coelomata and not ecdysozoa: Evidence from genome-wide phylogenetic analysis. *Genome Research*, 14(1): 29—36
- Xie Z X, Johansen L K, Gustafson A M *et al*, 2004. Genetic and functional diversification of small RNA pathways in plants. *Plos Biology*, 2(5): 642—652
- Xue C H, Li F, He T *et al*, 2005. Classification of real and pseudo microRNA precursors using local structure-sequence features and support vector machine. *Bmc Bioinformatics*, 6: 310
- Yang G, Yang L, Zhao Z *et al*, 2012. Signature miRNAs Involved in the innate immunity of invertebrates. *PloS One*, 7(6): e39015.
- Zheng J, Rogozin I B, Koonin E V *et al*, 2007. Support for the coelomata clade of animals from a rigorous analysis of the pattern of intron conservation. *Molecular Biology and Evolution*, 24(11): 2583—2592

## BIOINFORMATIC ANALYSIS OF MICRORNA GENES IN *DAPHNIA PULEX*

LIU Cheng-Zhang<sup>1, 2</sup>, LI Fu-Hua<sup>1</sup>, XIANG Jian-Hai<sup>1</sup>

(1. *Institute of Oceanology, Chinese Academy of Sciences, Qingdao, 266071;*

2. *University of Chinese Academy of Sciences, Beijing, 100049*)

**Abstract** MicroRNAs (miRNAs) are ~22nt long endogenous non-coding RNAs that regulate gene expression post-transcriptionally. Using a bioinformatic approach to analyze the whole genome sequence of *Daphnia pulex*, we predicted 252 pre-miRNAs which may produce 262 functional miRNAs, including 6 miRNA clusters. MiRNA occurrence frequency is significantly high on chromosome 8, while is significantly low on chromosome 4. Phylogenetic study of pre-miRNAs from 6 bilateria species was used to classify the *D. pulex* pre-miRNAs into 191 miRNA families, among which 105 are specific to *D. pulex*, while 15 are specific to Arthropoda. Among the compared bilateria species, *D. pulex* is most similar to Zebra fish in miRNA content.

**Key words** *Daphnia pulex*; crustacean; microRNA; miRNA; bioinformatics