

线虫 *prg-1* 基因对小 RNA 表达的影响

黄大卫^{1, 2)*, **} 崔东亚^{1, 2, 3)*} 王佳佳^{1, 2)} 何顺民²⁾ 张鹏²⁾

¹⁾ 河北大学生命科学学院, 保定 071000; ²⁾ 中国科学院动物研究所系统与进化重点实验室, 北京 100101;

³⁾ 运城学院生命科学系, 运城 044000)

摘要 以秀丽隐杆线虫为材料发现, *prg-1* 基因突变不仅影响 piRNA 的表达, 还影响部分 miRNA 的表达, 同时还发现 ncRNA-like 型小 RNA, 对 ncRNA-like 序列比较, 认为 ncRNA-like 与 piRNA 或 miRNA 序列极为相似; 对 ncRNA-like 与 miRNA 或 piRNA 的基因座比较, 发现 ncRNA-like 与 miRNA 或 piRNA 基因座完全相同. 推测这些 ncRNA-like 型小 RNA 可能与 miRNA 或 piRNA 有着相同的 RNA 前体来源.

关键词 秀丽隐杆线虫, *prg-1*, 小 RNA, 表达

学科分类号 Q3

DOI: 10.16476/j.pibb.2015.0297

秀丽隐杆线虫(*C. elegans*)是生命科学研究中的重要模式动物. *prg-1* 基因调控线虫的生殖发育, 同时也影响线虫 piRNA 生成^[1]. 线虫 piRNA 是目前研究的一个热点, 被认为是监视和调控基因表达的小 RNA, piRNA 还调控转座子的转座^[2-3]. 已发现的线虫 piRNA 种类非常多. *prg-1* 蛋白影响 piRNA 的表达. 本研究通过对 *prg-1* 基因突变处理发现, *prg-1* 蛋白除了影响 piRNA 的表达, 还影响部分 miRNA 的表达, 同时还发现众多与 piRNA 或 miRNA 相关小 RNA 也受到 *prg-1* 蛋白的调控.

1 材料与方法

1.1 线虫培养与小 RNA 提取测序

选用 N2 野生型和 *prg-1* 突变型线虫作为培养对象, 依据 Bianchi 等^[4]报道的培养方法对 2 种线虫同步化培养, 收集 L4 时期的虫体^[5], 分别标记为 N2L4 和 *prg-1*L4, 用 Trizol 试剂提取总 RNA, 由深圳华大科技公司测定长度, 18~30nt 范围内是小 RNA (small RNA, sRNA). N2 品种由中国科学院生物物理研究所非编码重点实验室肖腾飞博士提供, *prg-1* 突变品种 *prg-1(tm972)* 购自美国线虫遗传中心 (Caenorhabditis Genetics Center, CGC).

1.2 数据分析流程

对线虫中测序发现的 sRNA 数据进行分析, 主要包括 sRNA 序列注释、基因组座注释、表达差异分析等(图 1), 具体分析过程如下: a. 序列标准.

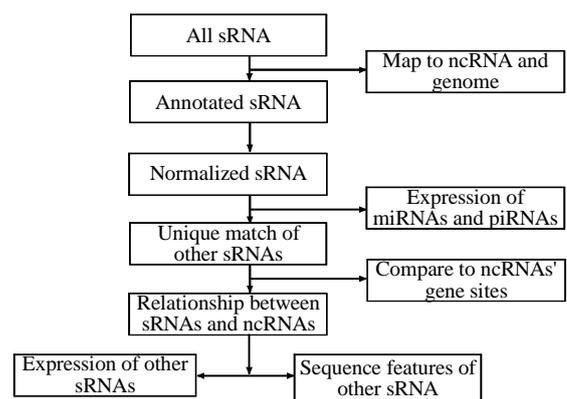


Fig. 1 The work flow of sRNA database

sRNA: Small RNA; ncRNA: Non-coding RNA; piRNA: PIWI interacting RNA.

* 共同第一作者.

** 通讯联系人.

Tel: 010-64807235, E-mail: huangdw619@gmail.com

收稿日期: 2016-01-16, 接受日期: 2016-02-02

自 Wormbase 网站下载 WS220 版本的基因组序列和非编码 RNA 序列(non-coding RNA, ncRNA).

b. sRNA 序列注释. 依据序列, 用 ncRNA 对 sRNA 注释. c. sRNA 位置注释. 比较 sRNA 与 ncRNA 在基因组位置上的分布^[6]. d. 归一化处理. 计算每种 sRNA 的 RPM (reads per million mapped reads, RPM). e. sRNA 表达分析. 利用 DEGseq 对 miRNA、piRNA 和未知 sRNA 进行表达分析^[7].

2 结果与分析

2.1 测序量统计

通过 illumina 技术对 N2L4 和 prgL4 两个样本中测到 18~30nt 范围内的 sRNA 进行测序, 最终测序量分别为 9 537 324 reads 和 9 416 828 reads,

其中分别有 95.8% 和 94.9% 的 sRNA 可以和基因组完全匹配. 从 sRNA 在基因组的匹配位点来看, 绝大多数的 sRNA 具有单一匹配位点, 说明 sRNA 来源比较单一(表 1). 从 sRNA 的长度来看, 序列长度集中在 21~23nt(图 2).

2.2 *prg-1* 影响多种小 RNA 的表达

2.2.1 piRNA 和部分 miRNA 受 *prg-1* 的影响

sRNA 种类分析发现, 在 N2L4 样本中存在 133 696 种 sRNA, 在 prgL4 样本中存在 106 865 种 sRNA, 与已公布的线虫 ncRNA 比较, 已知的 ncRNA 仅有 24 176 种^[8], 由此可见, 测序发现的 sRNA 中大多数是未知的 sRNA. 同时发现, *prg-1* 突变之后 sRNA 种类减少量达到 26 831 种(表 1), 因此推测, 这些未知的 sRNA 中可能存在受 *prg-1* 基因调控的 sRNA.

Table 1 The statistics of sRNA

Sample ID	Total reads	Total mapped reads	Unique match	Unique match sRNA species
N2L4	9 537 324(100%)	9 139 731(95.8%)	8 596 578(94.1%)	133 696
PrgL4	9 416 828(100%)	8 941 933(94.9%)	8 257 347(92.3%)	106 865

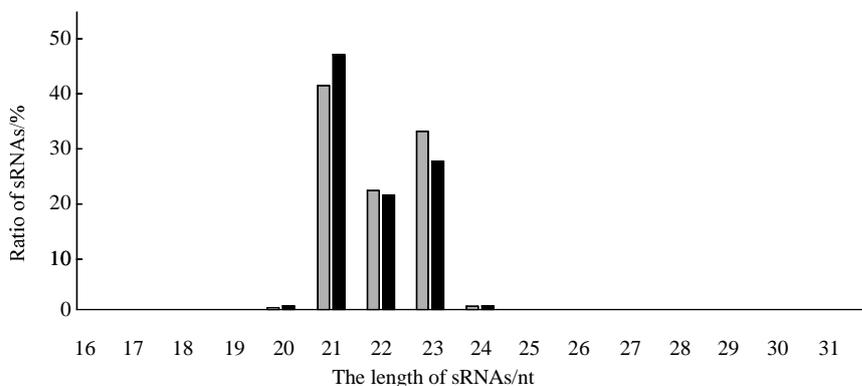


Fig. 2 The length of sRNA

□: n2L4; ■: prgL4.

为了进一步比较 *prg-1* 突变对 sRNA 表达的影响, 对 sRNA 的测序量进行归一化处理, sRNA 归一化通常用 RPM 来表示, 即平均百万匹配到基因组 read 中的单一 sRNA 的测序量. N2L4 和 prgL4 两个样本中分别测到 9 139 731 read 和 8 941 933 read 可匹配到基因组, 以此数据分别对两个样本中的 sRNA 进行归一化. 同时根据序列是否相同, 对 sRNA 与 ncRNA 比较, 发现测序样本中 sRNA 主要有 3 类, 既 miRNA、piRNA 和未知 sRNA. 从

表达量来看, *prg-1* 基因突变影响 sRNA 的组成变化, 尤其是影响 piRNA 的表达, 但 miRNA 的表达也受到一定程度的影响(表 2).

利用已报道 ncRNA 注释了 127 种 miRNA, 在 N2L4 和 prgL4 样本中, 分别发现 116 和 121 种 miRNA, 且有 110 种 miRNA 在突变前后均有表达. 利用 DEGseq 软件发现有 25 种 miRNA 在突变前后存在表达差异, 表达下降的有 18 种, 表达上升的有 7 种. 同样用 ncRNA 注释了 6 801 种

Table 2 The statistics of sRNA's expression

Sample ID	miRNA(species)	piRNA(species)	Other ncRNA(species)	Other sRNA(species)	All sRNAs(species)
N2L4	420 382.6(116)	33 640.7(6 801)	3.8(1)	545 972.9(13 540)	1 000 000(20 458)
PrgL4	338 248.4(121)	4.5(35)	0(0)	661 747.1(10 360)	1 000 000(10 516)

piRNA, 其中在 N2L4 中共发现 6 801 种 piRNA, 在 prgL4 中发现 35 种 piRNA, 即 *prg-1* 基因突变后所有 piRNA 均表达下降, 表明 *prg-1* 基因突变影响 piRNA 的表达. 经过 DEGseq 对 piRNA 表达量进行统计, 发现 807 种 piRNA 在 *prg-1* 突变前后存在统计学表达差异.

2.2.2 未知小 RNA 由差异表达的小 RNA 组成

从表达量看, 未知 sRNA 的表达只与 piRNA 或 miRNA 有关(表 3). 为了更好地发现具有明显表达差异的 sRNA, 利用 DEGseq 对未知 sRNA 进行表达差异统计, 发现 91 种 sRNA 在 *prg-1* 突变前

后存在表达变化. 值得注意的是, 这 91 种 sRNA 的表达量占有未知 sRNA 测序量的 95%, 也就是说, 未知 sRNA 的序列由表达有差异的未知 sRNA 组成(表 4). 利用 miRBase 报道的 miRNA[®]和 Bagijn 等[®]报道的 piRNA 对这 91 种差异表达的未知 sRNA 再次基因组位置注释, 发现这 91 种 sRNA 中, 19 种被 miRBase 被定义为 miRNA, 21 种被定义为 piRNA, 29 种与 miRNA 序列相关, 17 种与 piRNA 序列相关, 即这 91 种差异表达的 sRNA 中, 有 86 种未知 sRNA 与 miRNA 或 piRNA 有关.

Table 3 ncRNA-like has the same gene sites to ncRNAs¹⁾

Sample ID	All unknown sRNA	Around piRNA	Around miRNA	Around other ncRNA
N2L4	75 467(545 977)	3 612(2 250.7)	771(520 995)	347(338.6)
	RPM, 100%)	RPM, 0.4%)	RPM, 95.4%)	RPM, 0.1%)
prgL4	54 752 (661 747)	162(52.8)	778(637 697)	338(401.3)
	RPM, 100%)	RPM, 0.0%)	RPM, 96.4%)	RPM, 0.1%)

¹⁾ The figure means the species, the expression and the ratio of sRNAs expression, respectively.

Table 4 Differentially expressed sRNAs were the main sRNA species¹⁾

sRNA ID	N2L4	prgL4
All sRNA	75 467(545 977)	54 752 (661 747)
Differentially expressed sRNA	91(516 286)	91(631 918)

¹⁾ The figure means the species and the expression of sRNA.

2.3 ncRNA-like 型小 RNA 来自 miRNA 或 piRNA 的前体

2.3.1 未知 sRNA 与 ncRNA 表达相同

序列分析发现, 未知 sRNA 主要与 miRNA 或 piRNA 序列相关, 这些未知 sRNA 是否与 miRNA 或 piRNA 来源相同呢? 将未知 sRNA 与 ncRNA 进行基因组位置比较, 记录 ncRNA 基因座周围未知

sRNA 的分布. 从测序量来看, 在 N2L4 样本中, 95.9% 的未知 sRNA 分布在 ncRNA 的周围, 同样的结果也出现在 prgL4 样本中, 即这些未知 sRNA 与 piRNA 和 miRNA 的基因座相同, 说明未知的 sRNA 可能与 ncRNA 有来源和表达上的关联(表 3), 故将与 piRNA 和 miRNA 基因座相同的 sRNA 分别命名为 piRNA-like 和 miRNA-like, 统称为 ncRNA-like.

需要注意的是, 从测序量上来看, 虽然绝大多数的未知 sRNA 基因组位点分布在 ncRNA 周围, 但种类并不多(表 3), 将 2 个样本中所有 sRNA 进行汇总, 同样说明, 样本中未知 sRNA 的种类虽多, 但大部分种类 sRNA 表达量很低, 应属于测序背景 RNA(表 5), 也说明未知 sRNA 与 miRNA 或 piRNA 表达有关.

Table 5 The small RNA species in *C. elegans*

ID	piRNA	miRNA	piRNA-like	miRNA-like	Back ground sRNA	Total
Species	6 801	127	3 901	966	93 954	105 749
Ratio	6.4%	0.1%	3.7%	0.9%	88.8%	100.0%
Expression	33 644	758 631	3 278	1 159 129	45 318	2 000 000
Ratio	1.7%	37.9%	0.2%	58.0%	2.3%	100.0%

2.3.2 未知 sRNA 与 ncRNA 来源相同

未知 sRNA 在序列上和表达上与 ncRNA 相同, 是否意味着二者的来源相同呢? 对 ncRNA-like 与 ncRNA 两者的基因座位置进行比较. 发现绝大多数 miRNA-like 的 5'端匹配位置与 miRNA 的 5'端匹配位置几乎完全相同, 前后差距在 1 个碱基范围

内. 从 miRNA-like 的种类来看, miRNA-like 与 miRNA 之间的差异主要体现在 3'端序列上, 这说明 miRNA-like 与 miRNA 可能来自同一转录本^[10], 推测是在 miRNA 前体加工过程中, 有些 miRNA 前体被加工成 miRNA, 有些 miRNA 前体被加工成 miRNA-like.

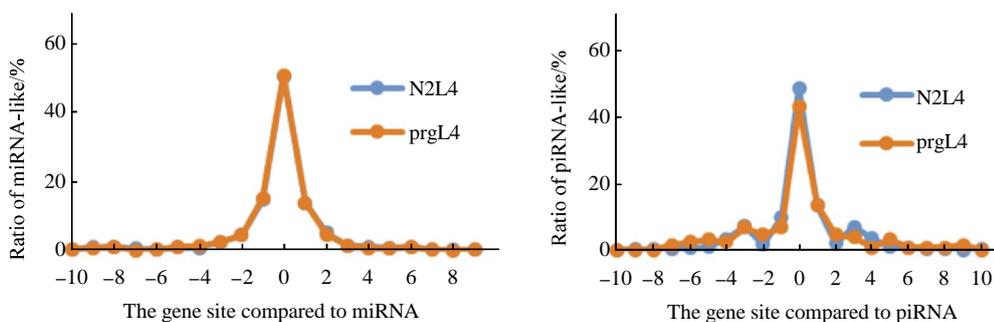


Fig. 3 ncRNA-like has the same gene site compared to miRNAs and piRNAs

同样 piRNA-like 5'端的匹配位置与 piRNA 5'端的匹配位置几乎完全相同, piRNA-like 与 piRNA 的差异也主要体现在 3'端的序列上. 从表达量和表达种类的统计结果来看 piRNA-like 还明显受 prg-1 蛋白的影响, 也说明 piRNA-like 与 piRNA 来源相同, 而且与 piRNAs 一样还受到 prg-1 蛋白表达的影响(表 3). 综合分析认为, miRNA-like 是 miRNA 前体加工而来, piRNA-like 是 piRNA 前体加工而来.

3 结 论

a. prg-1 蛋白影响部分 miRNA 和 piRNA 的表达. 高通量测序是 RNA 表达研究的重要技术, 对线虫 2 个样本的 sRNA 进行测序, 利用 WS220 ncRNA 数据对测序 sRNA 进行注释, 用 DEGseq 分析 sRNA 的表达, 发现 prg-1 蛋白对 25 种 miRNA 表达有影响, 利用 miRbase 注释后, 又发现 19 种受到 prg-1 蛋白调控的 miRNA, 前后共发

现有 44 种 miRNA 受到 prg-1 蛋白的调控. 同时也发现 prg-1 蛋白影响全部 piRNA 的表达, 利用 wormbase 数据库^[11]和 Bagijn 等^[9]报道的 piRNA 共同注释了 6 800 多种 piRNA, prg-1 基因突变后, piRNA 几乎处于不表达状态, 说明 prg-1 蛋白对 piRNA 的表达影响非常大.

b. ncRNA-like 与 miRNA 和 piRNA 有相同的 RNA 前体. 对 sRNA 进行序列和基因组位置统计后发现, 多数 sRNA 是未知的 sRNA, 去除背景 sRNA 后, 发现大量的未知 sRNA 通常与 ncRNA 共用一段 DNA 序列, 从表达量来看, 与 miRNA 和 piRNA 相关的序列最多, 说明这些未知的 sRNA 与 miRNA 或 piRNA 来自于同一 RNA 前体, 将这些 sRNA 统一命名为 ncRNA-like. 利用 DEGseq 对所有未知 sRNA 进行表达差异统计, 发现存在表达差异的 sRNA 中, 同样以 ncRNA-like 为主(表 3).

4 讨 论

从 miRNA-like 和 piRNA-like 的序列在基因组上的分布来看, 基本可以确定这些小 RNA 与 miRNA 或 piRNA 有相同的来源. 那这些小 RNA 是如何产生的呢? 文献报道 miRNA 是通过前体 RNA 经过 2 次剪切后生成的, 该过程包含 5 个主要部分: a. 从基因组转录出 1 条初级 miRNA (primary miRNA, pri-miRNA); b. pri-RNA 在 Drosha 复合体的作用下被剪切, 保留一个茎环状的 miRNA 前体 (precursor miRNA, pre-miRNA); c. pre-miRNA 被运输到细胞质中; d. pre-miRNA 在 Dicer 复合物的作用下, 再次被剪切, 去掉 pre-miRNA 中未配对的环状结构, 仅保留 pre-miRNA 中互补的那部分 RNA, 保留下来的即为 miRNA, 其互补链为 miRNA*; e. miRNA 与 AGO 类蛋白形成复合物, 参与基因的调控^[12]. 序列分析过程中发现一些 miRNA-like 保留 d 部分的

特征, 如在 mir-1832.2 和 mir-2208.2 的 5' 端和 3' 端同时发现了相应的 miRNA-like (图 4), 显示 miRNA-like 的加工来源方式与 miRNA 的来源也是非常相似的, 说明 miRNA-like 很有可能来自于 pre-miRNA 的剪切. 综合分析认为, pri-miRNA 在加工为成熟的 miRNA 过程中, 可能存在多种加工方式, 即 pre-miRNA 通过 Dicer 复合物加工成 miRNA 和 miRNA*, 同时 pre-miRNA 也可以被加工成 miRNA-like 和 miRNA-like*.

基于线虫 piRNA 来源的报道, piRNA 转录起始于 piRNA 上游的第 2 个碱基, 通过 RNA 聚合酶生成一条 piRNA 前体, 前体通过 5' 加工和 3' 加工, 成为成熟的 piRNA^[13]. 同样对 piRNA-like 序列分析, 推测 piRNA-like 很有可能也是 piRNA 前体的一种加工产物. 绝大多数的 piRNA-like 与 piRNA 有着相同的 5' 端, 仅在 3' 端存在几个碱基的差异 (图 5), 说明 piRNA-like 极有可能是 piRNA 前体 3' 加工过程中的副产物, 即 piRNA 前体 3' 端的加工



Fig. 4 The two kinds of miRNAs and their miRNA-likes in chromosome IV

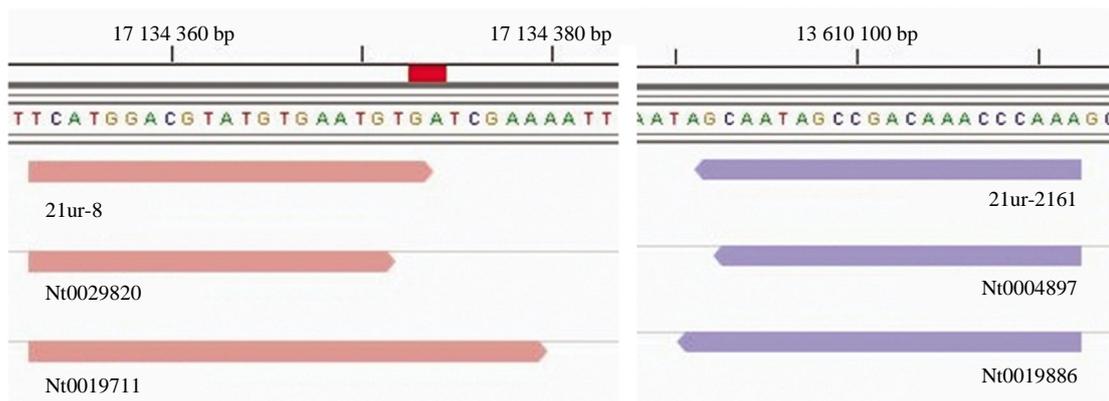


Fig. 5 The two kinds of piRNAs and their piRNA-likes in chromosome IV

可能存在多种加工形式，其中以 21nt 长度的 piRNA 为主，同时还伴随着其他长度的 piRNA-like 副产物生成。

参 考 文 献

- [1] Batista P J, Ruby J G, Claycomb J M, *et al.* *prg-1* and 21U-RNAs interact to form the piRNA complex required for fertility in *C. elegans*. *Molecular Cell*, 2008, **31**(1): 67–78
- [2] Shirayama M, Seth M, Lee H C, *et al.* piRNAs initiate an epigenetic memory of nonself RNA in the *C. elegans* germline. *Cell*, 2012, **150**(1): 65–77
- [3] Lee H C, Gu W, Shirayama M, *et al.* *C. elegans* piRNAs mediate the genome-wide surveillance of germline transcripts. *Cell*, 2012, **150**(1): 78–87
- [4] Bianchi L, Driscoll M. Culture of embryonic *C. elegans* cells for electrophysiological and pharmacological analyses. *WormBook: the online review of C. elegans biology*, 2006: 1–15
- [5] Deng W, Zhu X P, Skogerbo G, *et al.* Organization of the *Caenorhabditis elegans* small non-coding transcriptome: Genomic features, biogenesis, and expression. *Genome Research*, 2006, **16**(1): 20–29
- [6] Langmead B, Trapnell C, Pop M, *et al.* Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 2009, **10**(3): R25
- [7] Wang L, Feng Z, Wang X, *et al.* DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics*, 2010, **26**(1): 136–138
- [8] Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Research*, 2011, **39**: D152–D157
- [9] Bagijn M P, Goldstein L D, Sapetschnig A, *et al.* Function, targets, and evolution of *Caenorhabditis elegans* piRNAs. *Science*, 2012, **337**(6094): 574–578
- [10] Kutter C, Svoboda P. miRNA, siRNA, piRNA Knowns of the unknown. *Rna Biology*, 2008, **5**(4): 181–188
- [11] Harris T W, Antoshechkin I, Bieri T, *et al.* WormBase: a comprehensive resource for nematode research. *Nucleic Acids Research*, 2010, **38**(Database issue): D463–467
- [12] Nishimasu H, Ishizu H, Saito K, *et al.* Structure and function of Zucchini endoribonuclease in piRNA biogenesis. *Nature*, 2012, **491**(7423): 284–287
- [13] Gu W, Lee H C, Chaves D, *et al.* CapSeq and CIP-TAP Identify Pol II Start Sites and Reveal Capped Small RNAs as *C. elegans* piRNA Precursors. *Cell*, 2012, **151**(7): 1488–1500

The Effect of *prg-1* on Small RNAs Expression of *Caenorhabditis elegans*

HUANG Da-Wei^{1,2)*,**}, CUI Dong-Ya^{1,2,3)*}, WANG Jia-Jia^{1,2)}, HE Shun-Min²⁾, ZHANG Peng²⁾

¹⁾ College of Life Science, Hebei University, Baoding 071002, China;

²⁾ Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China;

³⁾ College of Life Science, Yuncheng University, Yuncheng 044000, China)

Abstract In *Caenorhabditis elegans*, the expression of all piRNAs, small of miRNAs were influenced by the mutation of *prg-1*. Especially, some ncRNA-like, a kinds of small RNA, were found in *C. elegans*. The ncRNA-like had the similar sequence with piRNAs or miRNAs. The results of the analysis in the gene sites between the ncRNA-like and piRNAs or between ncRNA-like and miRNAs, showed that ncRNA-like had the same gene sites with piRNAs or miRNAs. All of these showed that the ncRNA-like were a kind of small RNA and may be processed from the RNA precursor same to piRNAs or miRNAs.

Key words *Caenorhabditis elegans*, *prg-1*, small RNAs, expression

DOI: 10.16476/j.pibb.2015.0297

*These authors contributed equally to this work.

**Corresponding author.

Tel: 86-10-64807279, E-mail: huangdw619@gmail.com

Received: January 16, 2016 Accepted: February 2, 2016