

线虫体内新 piRNA 研究

崔东亚^{1, 2, 3)} 王佳佳^{1, 2)} 孙旭滨^{1, 2)} 张 鹏²⁾ 何顺民^{2)*} 黄大卫^{1, 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 PIWI-interacting RNAs(piRNA)是一类内源性小 RNA, 负责抵御转座子和转基因对基因组的入侵. 已发现 1.6 万多种 piRNA, 在 piRNA 上游存在保守序列, 根据上游序列特征可以预测新的 piRNA. 将线虫同步化培养至 L4 时期, 分别提取野生和 *prg-1* 突变样本中的小 RNA, 并对其进行高通量测序. 基于 piRNA 上游保守序列特征, 在野生线虫 L4 时期中, 发现了 967 种新 piRNA, 这些新 piRNA 在 *prg-1* 突变后表达消失. 新 piRNA 的基因座集中分布在四号染色体的 2 个 piRNA 簇内, 首位碱基以 U 为主. 与已发表的成虫发育时期的 PRG-1 免疫共沉淀数据比对, 发现有 153 种 piRNA 存在于与 PRG-1 免疫沉淀的数据中. 同时还发现一些只在野生线虫中表达的 non-21nt 小 RNA, 它们与已知 piRNA 的基因座相同, 推测这些 non-21nt 小 RNA 可能是其 piRNA 前体加工的产物. 总之, 通过小 RNA 测序, 在线虫中发现了一些新的 piRNA.

关键词 线虫, piRNA, 预测, 新 piRNA, PRG-1

学科分类号 Q3

DOI: 10.3724/SP.J.1206.2014.00072

高通量测序是近年来发展的一项测序新技术, 测序通量远远高于传统的桑格尔测序技术(sanger method), 且测序成本低^[1-2], 这为研究低表达的 RNA 带来了方便, 也使得基因调控研究进入一个新的时代^[3-5]. 线虫(*Caenorhabditis elegans*)中存在多种 ncRNA (non-coding RNA), 如 lncRNA (long non-coding RNA)^[6]、miRNA^[7]、22G-RNA^[8]和 piRNA (PIWI-interacting RNA)^[9]等. 线虫 piRNA 也被称为 21U-RNA^[10], 因为长度为 21nt, 且首位碱基通常为尿嘧啶(U). 线虫 piRNA 与 PRG-1 形成复合物, 存在于生殖细胞中, 所以 piRNA 在幼虫发育后期和成虫期表达水平较高^[9]. 研究认为, 线虫 piRNA 是从 piRNA 簇(piRNA cluster)内被独立转录出来, 并通过允许错配的方式识别靶基因, 在靶位点附近募集 RdRP (RNA-dependent RNA polymerase)蛋白合成 22G-RNA, 在转录水平沉默外源基因, 但 piRNA 似乎只负责沉默的起始过程^[10-12].

Bagijn 等^[8]发现, 16 003 种线虫 piRNA, 但大部分 piRNA 自身的表达量很低, 其中测序量仅为 1~2 reads 的种类就高达 3 476 种, 因此推测线虫体内还存在表达偏低的未知 piRNA. 对野生型线

虫 L4 时期的小 RNA 测序发现 967 种新的 piRNA, 这些新 piRNA 与已知 piRNA 的序列特征和表达模式相同.

1 材料与方法

1.1 线虫培养与小 RNA 提取测序

野生型为 N2 线虫(N2), 由中国科学院生物物理所非编码重点实验室提供. 线虫 *prg-1* 株系是 *prg-1(tm872)*株系, 由美国明尼苏达大学线虫遗传学中心(Caenorhabditis Genetics Center, CGC)提供. *prg-1* 突变是一种序列片段缺失型突变, 缺失区域包含部分 Mid 区和部分 PIWI 区^[11]. 按照 wormbook 的方法进行培养^[12], 收集 L4 时期的虫体, 野生样本标记为 WT (wild type), *prg-1* 突变样本标记为 *prg-1*, 按照 Trizol (Life Technologies Co.Ltd.)的使用说明提取总 RNA, 由深圳华大科技公司利用

* 通讯联系人.

何顺民. Tel: 010-64807279, E-mail: hesm@ioz.ac.cn

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

收稿日期: 2014-05-01, 接受日期: 2014-09-10

Illumina 技术对约 18~30nt 小 RNA 进行测序. 线虫的使用符合国家及提供实验动物单位制订的有关实验动物福利的规则和制度.

1.2 新 piRNA 预测流程

线虫 piRNA 与其他物种的 piRNA 有所区别. 序列长度上, 线虫 piRNA 为 21nt 长度, 比果蝇^[13]和鼠^[14]piRNA 短一些; 在序列特征上, 线虫 piRNA 基因座的上游存在一段保守的序列(图 1)^[15], 该序列有两个特征, 一个是核心序列 CTGTTTCA 区, 另一个是核心序列两侧的 A/T 富集区. 该保守序列被认为是线虫 piRNA 转录的必需序列^[16]. 基于 piRNA 上游保守序列特征, Ruby 等^[15]最早用序列打分(motif score, Ms)的方法对线虫 piRNA 进行预测. Batista 等^[9]发表了该预测方法, 该方法可有效地将 piRNA 从小 RNA 中分离出来. 利用 Batista 等的方法来预测新的 piRNA, 并根据 Batista 和 Bagijn 等^[8-9]的研究成果设计新的 piRNA 发现流程(图 2), 具体数据分析过程分为 5 步:



Fig. 1 The conserved upstream sequence motif of *C. elegans* piRNA

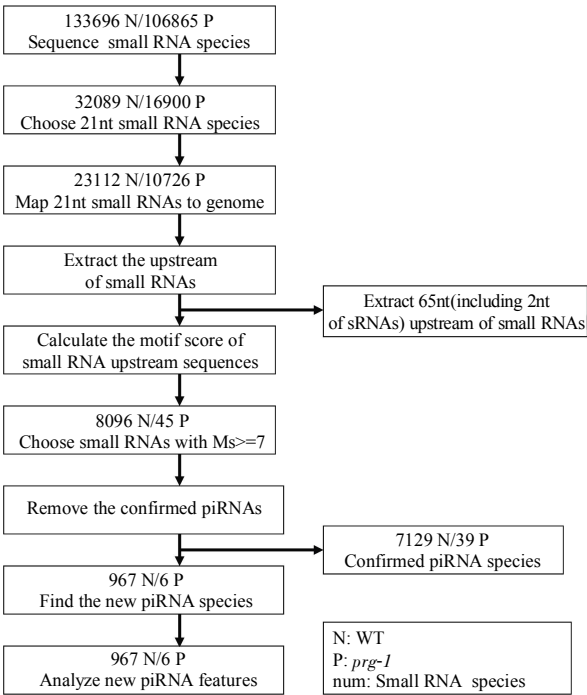


Fig. 2 The work flow of new piRNAs searching

第一, 从 wormbase 网站(<http://www.wormbase.org>)下载 WS220 版本的线虫基因组序列.

第二, 用 bowtie 软件把小 RNA 比对到基因组^[17](参数 -v 0-a).

第三, 利用 Galaxy 工具截取小 RNA 上游 65nt (包含小 RNA 前 2nt) 序列^[18], 利用 piRNA motif-scoring module 对上游序列打分^[9]. 筛选条件为: a. 21nt 长度; b. 完全匹配到 WS220 基因组; c. 上游序列得分 $Ms \geq 7$ ^[18-9]; d. 去除已知的 piRNA, 将剩余的小 RNA 定为新 piRNA.

第四, 编写 perl 脚本, 分析新 piRNA 的各项特征.

第五, 使用 MEME 工具^[19]和 WEBLOGO 工具^[20]对新 piRNA 序列特征进行分析.

2 结果与分析

2.1 线虫 N2L4 时期内含大量未知小 RNA

自从发现 piRNA 后^[13], piRNA 研究广受关注, piRNA 的种类也在不断增加. 线虫的 piRNA 种类也从最初的 5 454 种增加到 16 003 种^[8, 15]. 多数 piRNA 表达很低^[8], 因此推测还存在一些未发现的表达量低的 piRNA.

为了证实这个问题, 利用 illumina 测序技术, 对 L4 时期的 WT 和 *prg-1* 两个线虫样本进行小 RNA 测序, 分别获得 9 537 324 和 9 416 828 条小 RNA 序列. 将小 RNA 与基因组 WS220 比对, WT 和 *prg-1* 两个样本的匹配率基本相同(表 1), 但在 *prg-1* 中, 匹配到基因组的小 RNA 种类由 82 384 种减少到 54 908 种, 减大幅度为 33.4%, 其中减少幅度最多的是 21nt 长的小 RNA(图 3). PRG-1 是 piRNA 生成和积累的重要蛋白, 因此 *prg-1* 突变之后, 必然影响 piRNA 的积累^[9]. 对所有小 RNA 进行基因组匹配, 在 WT 样本中有 21nt 长度的小 RNA 23 112 种, 在 *prg-1* 样本中, 有 21nt 小 RNA 10 726 种(图 2, 3). 进一步统计发现, 在 L4 时期总共测到 7 148 种已知 piRNA, 根据小 RNA 种类变化, 推测在这些未知 21nt 小 RNA 中应该还有新 piRNA 存在.

Table 1 Summary of small RNAs annotated to genomic from two samples

Sample source	Total sequencing reads	Annotated to genome	Ratio/%
WT	9 537 324	9 139 731	95.8
<i>prg-1</i>	9 416 828	8 941 933	95.0

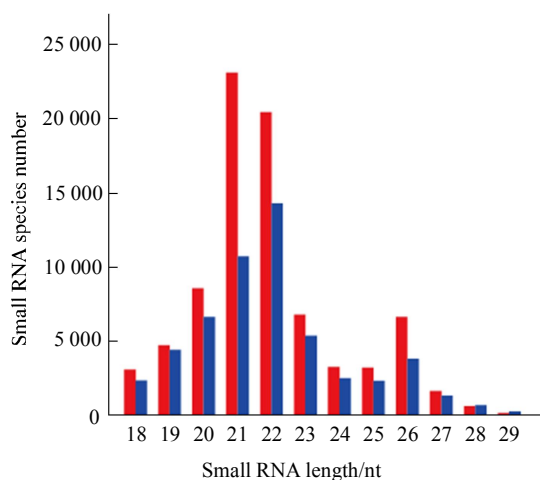


Fig. 3 The distribution of small RNA species annotated to genome
■: WT; ■: *prg-1*.

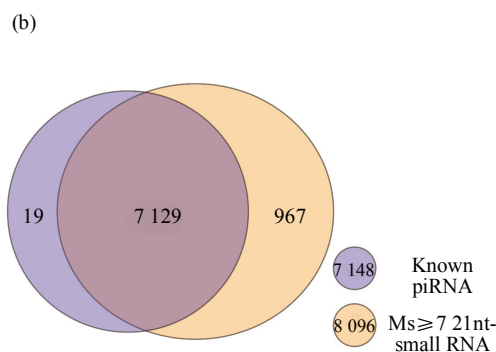
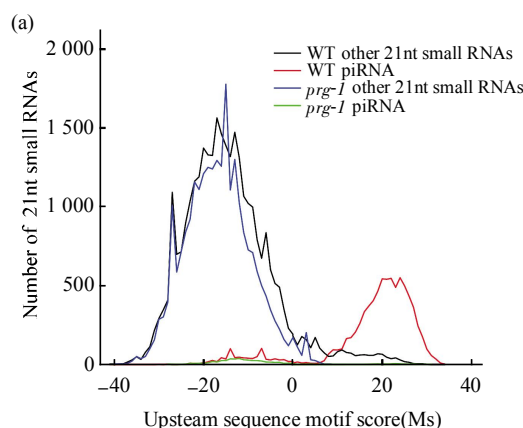


Fig. 4 The known piRNA can be found based on its upstream motif score

(a) Distribution of motif score at upstream of 21nt RNAs. Scores are rounded to the nearest integer. piRNAs have higher motif scores. (b) Most WT 21nt small RNAs with $Ms \geq 7$ are known piRNA.

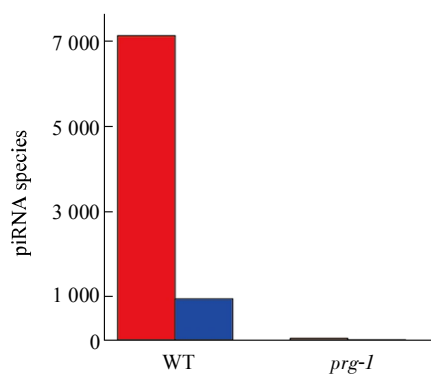


Fig. 5 Number of piRNAs in N2 vs *prg-1*
■: piRNA; ■: New piRNA.

2.2 新 piRNA 的表达依赖 PRG-1

线虫的 piRNA 与其他动物的 piRNA 有所不同, 线虫 piRNA 上游具有保守序列^[15-16], 可以将 piRNA 从众多小 RNA 中分离出来. 按照 Batista 等^[9]的序列打分算法, 对 WT 和 *prg-1* 的所有 21nt RNA 进行打分(图 4a), 将得分 $Ms \geq 7$ 的小 RNA 作为候选 piRNA. 在 WT 中共找到 8 096 种候选 piRNA, 经过注释, 发现其中 7 129 种是已知的 piRNA, 占 WT 中已知 piRNA 的 99.7%(7 129/7 148) (图 4b). 同时还发现 967 种 $Ms \geq 7$ 的候选 piRNA, 在 *prg-1* 突变后, 这些候选 piRNA 不再表达, 符合 piRNA 的表达依赖 PRG-1 的特点(图 5, 6b). 因此推测这 967 种 21nt 长度的小 RNA 有可能是新的 piRNA.

2.3 新 piRNA 具有 piRNA 的结构特征

piRNA 不仅依赖 PRG-1, 而且还有特殊的结构特征. 于是对新 piRNA 的结构特征进行分析.

2.3.1 新 piRNA 来自 piRNA 簇

根据文献报道, 线虫 piRNA 主要来自四号染色体(chromosome IV, chr IV)的 4.5~7.0 M 和 13.5~17.2 M 两个 piRNA 簇内^[15]. 在 WT 样本中发现约 89%(867/976)的新 piRNA 来自 chr IV(图 6a), 并且新 piRNA 在基因组上的分布与已知 piRNA 的分布相同(图 6b). 推测新 piRNA 和 piRNA 都是从 piRNA 簇转录而来.

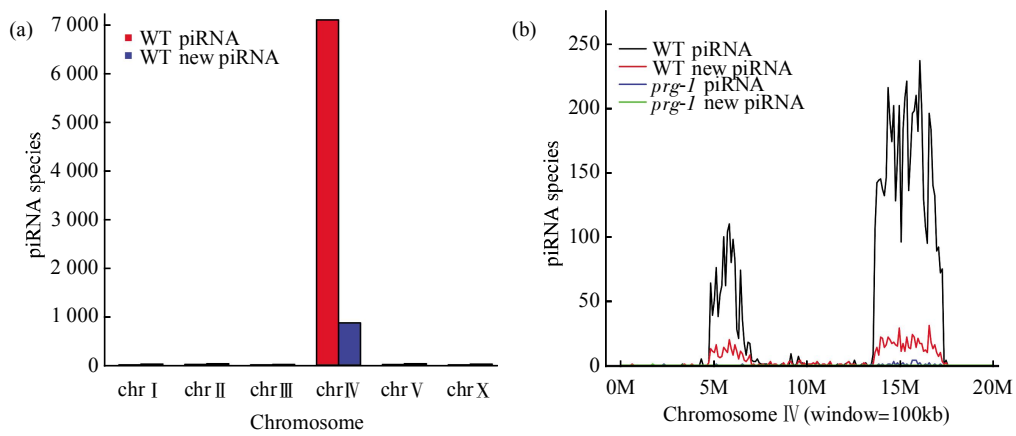


Fig. 6 New piRNAs and piRNAs share the same loci in chrIV

(a) Most of new piRNAs are from chrIV. (b) Majority of new piRNAs are settled in two piRNA clusters.

2.3.2 新 piRNA 具有保守上游序列

这 976 种新 piRNA 在基因组上共 1 079 个匹配位点, 截取所有匹配位点上游 65nt 序列(包含新 piRNA 的前 2nt), 用 MEME 工具绘制序列的特征^[19], 并与已知 piRNA 的上游序列特征进行比较. 发现新 piRNA 的上游序列特征与已知 piRNA 的上游特征非常相似, 不仅具有核心保守序列 CTGTTTCA 序列, 而且两侧是 A/T 富集的区域(图 7a, b)^[15].

2.3.3 新 piRNA 以 U 碱基开头

piRNA 首位碱基通常以 U 开头^[21-22], 线虫 piRNA 同样偏爱 U 作为首位碱基^[8]. WT 样本中, 约有 99.6%(7 104/7 129)的 piRNA 以 U 作为首位碱基, 同样约有 96.8%(932/967)新 piRNA 也以 U 开头, 利用 WEBLOGO 工具绘制新 piRNA 的序列特征^[20], 结果显示新 piRNA 具有一般 piRNA 首位碱基的特征(图 7c, d).

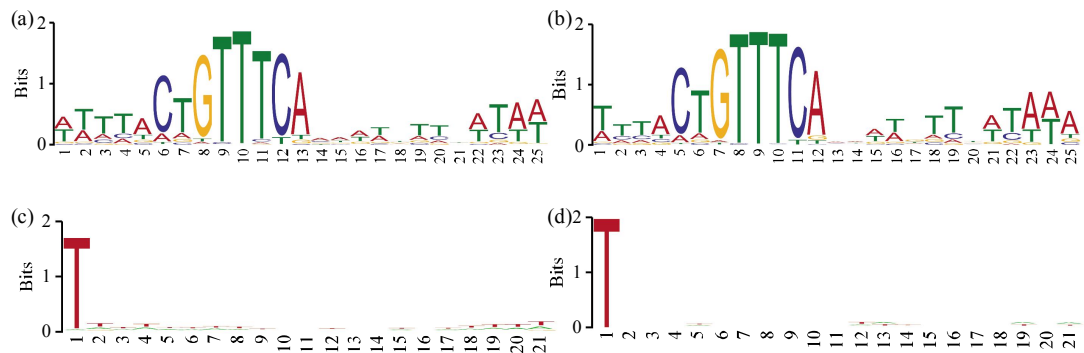


Fig. 7 New piRNAs have the same motif to piRNAs

(a) New piRNA upstream sequence motif. (b) piRNA upstream sequence motif. (c) New piRNA sequence motif. (d) piRNA sequence motif.

2.4 新 piRNA 与 PRG-1 蛋白存在相互作用

免疫共沉淀技术 (co-immunoprecipitation, Co-IP) 是研究蛋白质与 RNA 相互作用的常用实验技术, 利用抗体提取抗原蛋白, 提取纯化与抗原蛋

白结合的 RNA, 通过高通量测序技术可以得到与抗原蛋白相互作用的 RNA 序列. 线虫 piRNA 与 PRG-1 相互作用形成复合物, 利用 PRG-1 抗体将 PRG-1 沉淀下来, 提取其中的 RNA, 经过测序可

以知道与 PRG-1 相互作用的小 RNA. 以成虫为实验材料, Batista 等^[9]利用 IP 技术, 发现 7 876 种 piRNA 存在于 PRG-1 Co-IP 数据中. 将 967 种新 piRNA 与 Batista 等的 PRG-1 Co-IP 数据比对, 能发现有 153 种新 piRNA 可以随着 PRG-1 一起沉淀下来, 在对照的 PRG-1 INPUT 数据中, 仅能找到 5 种新 piRNA, 说明新 piRNA 与 PRG-1 之间存在相互作用.

3 结 论

基于 piRNA 的特征, 本文提取线虫 L4 时期的小 RNA 并测序, 发现了 967 种新的 piRNA, 这些新 piRNA 具有一般 piRNA 的相关特征, 主要体现在以下几个方面: a. 新 piRNA 在 *prg-1* 中不表达; b. 新 piRNA 绝大多数分布在 chrIV 的 2 个 piRNA 簇内; c. 新 piRNA 与已知 piRNA 有相同的上游保守序列; d. 新 piRNA 与 piRNA 序列长度都为 21nt, 且偏爱 U 作为首位碱基; e. 新 piRNA 与 PRG-1 存在直接相互作用关系.

4 讨 论

4.1 增大测序通量有助于发现低表达的小 RNA

Batista 等^[9]根据 piRNA 上游序列特征设计一套 piRNA 的预测方法, 依据该方法, 新发现的 piRNA 没有出现在 Batista 等的 piRNA 名单中, 推测这与 piRNA 表达量和测序通量有关. 新 piRNA 种类有 967 种, 其中 1~2 个测序 read 的种类就有 762 种, 可见新 piRNA 的表达丰度很低. Batista 等^[9]的测序通量为 4 516 215 read, 我们的测序通量为 9 537 324 read, 是其测序量的 2 倍. 虽然在 Batista 等^[9]的 PRG-1 IP 数据中存在 153 种新

piRNA, 但在小 RNA 测序数据中没有测到这类小 RNA, 这可能是 Batista 等没有将这 153 种小 RNA 归为 piRNA 的原因. 小 RNA 提取、测序建库等过程中都会引发不可避免的 RNA 损失, 一些低表达的小 RNA 极有可能会在 RNA 提取纯化过程中损失掉. 总之, 提高小 RNA 提取量, 可增加每种小 RNA 尽可能多的测序量, 便可以发现和测序到低表达的小 RNA.

4.2 新 piRNA 可能具有招募 22G RNA 合成的功能

一般认为 Piwi 蛋白在 piRNA 协助下, 识别和沉默与 piRNA 互补的转座子转录本, 以保持基因组结构的完整. Bagijn 等^[8]和 Lee 等^[23]认为 piRNA 可以通过允许错配的方式识别靶基因, 并招募 RdRP 蛋白合成与 WAGO 结合的 22G RNA, 通过 WAGO 22G RNA 复合物来调控 piRNA 靶基因的表达. 新 piRNA 具有一般 piRNA 的结构和表达特征, 同样应该具有招募 RdRP 在新 piRNA 靶位点处合成 22G 的功能, 22G RNA 是 5' 三磷酸小 RNA, 本研究的小 RNA 数据中缺乏这类三磷酸小 RNA, 因此还需要进一步测序验证.

4.3 piRNA 前体有多种加工产物

依据小 RNA 匹配位点的上游序列特征, 不仅能将已知的 piRNA 预测出来, 还发现一些 $Ms \geq 7$ non-21nt 小 RNA 也具有 piRNA 类似的特征, 如首位碱基为 U, 集中分布在 piRNA 簇内, *prg-1* 基因突变后这些 non-21nt 小 RNA 表达消失. 进一步分析发现, 大多数 non-21nt 小 RNA 与已知的 piRNA 匹配到同一基因组位置, 唯一不同之处是序列的长度不一样(图 8a), 这似乎提示 non-21nt 小 RNA 是 piRNA 前体的加工产物. piRNA 前体加工还有很多细节有待于进一步证实, 尤其是 piRNA 的 3'

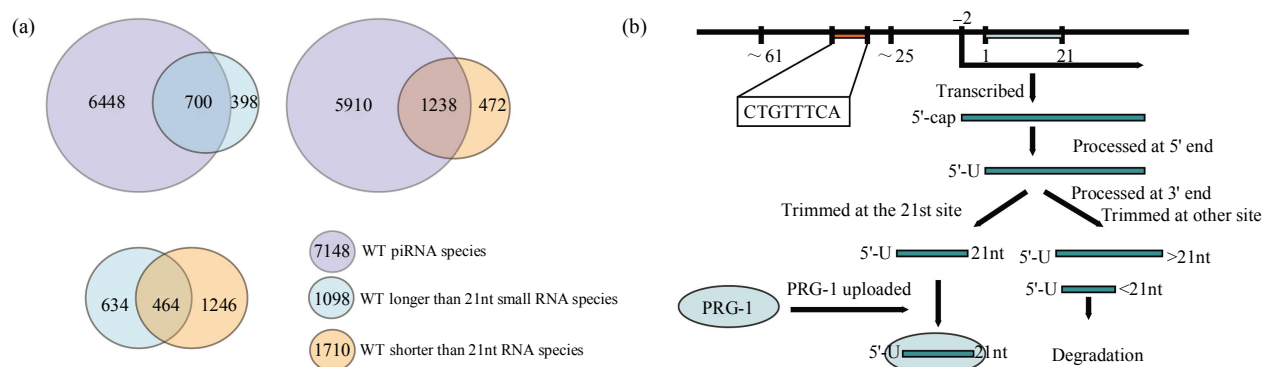


Fig. 8 Non-21nt small RNA may be some products of piRNA precursors

(a) Most of non-21nt small RNA with $Ms \geq 7$ share the same loci with piRNAs. (b) The model of processing of piRNA precursors.

端序列的加工^[16]。对 $Ms \geq 7$ 的 non-21nt 小 RNA 长度分析, 推测这些小 RNA 来自 piRNA 前体 3' 端的剪切, 即 piRNA 前体的 3' 端经过加工生产多种长度的小 RNA, 21nt 长度的 piRNA 与 PRG-1 形成复合物, non-21nt 长度的小 RNA 因为没有与之结合的蛋白, 随后被迅速降解(图 8b)。相对于 piRNA 的种类, $Ms \geq 7$ non-21nt 小 RNA 发现的种类偏少, 这可能与 non-21nt 小 RNA 降解有关。

致谢 感谢中国科学院生物物理研究所非编码重点实验室陈润生院士为本研究提供实验场所和技术指导。感谢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和危贵峰博士为本研究提供所需线虫和实验技术指导。感谢 CGC 为研究提供所需线虫。

参 考 文 献

- [1] Plocik A M, Graveley B R. New insights from existing sequence data: generating breakthroughs without a pipette. *Molecular Cell*, 2013, **49**(4): 605–617
- [2] Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 2009, **10**(1): 57–63
- [3] Sigova A A, Mullen A C, Molinier B, *et al.* Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. *Proc Natl Acad Sci USA*, 2013, **110**(8): 2876–2881
- [4] Lee J T. Epigenetic regulation by long noncoding RNAs. *Science*, 2012, **338**(6113): 1435–1439
- [5] Xu Q, Zhang Y, Chen Y, *et al.* Identification and differential expression of microRNAs in ovaries of laying and broody geese (*Anser cygnoides*) by solexa sequencing. *Plos One*, 2014, **9**(2): e87920
- [6] Nam J W, Bartel D P. Long noncoding RNAs in *C. elegans*. *Genome Research*, 2012, **22**(12): 2529–2540
- [7] Massirer K B, Perez S G, Mondol V, *et al.* The miR-35-41 family of microRNAs regulates RNAi sensitivity in *Caenorhabditis elegans*. *Plos Genetics*, 2012, **8**(3): e1002536
- [8] Bagijn M P, Goldstein L D, Sapetschnig A, *et al.* Function, targets, and evolution of *Caenorhabditis elegans* piRNAs. *Science*, 2012, **337**(6094): 574–578
- [9] 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
- [10] Das P P, Bagijn M P, Goldstein L D, *et al.* Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the *Caenorhabditis elegans* germline. *Molecular Cell*, 2008, **31**(1): 79–90
- [11] Wang G, Reinke V. A *C. elegans* Piwi, PRG-1, regulates 21U-RNAs during spermatogenesis. *Current Biology*, 2008, **18**(12): 861–867
- [12] 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
- [13] Brennecke J, Aravin A A, Stark A, *et al.* Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell*, 2007, **128**(6): 1089–1103
- [14] Vourekas A, Zheng Q, Alexiou P, *et al.* Mili and Miwi target RNA repertoire reveals piRNA biogenesis and function of Miwi in spermiogenesis. *Nature Structural & Molecular Biology*, 2012, **19**(8): 773–781
- [15] Ruby J G, Jan C, Player C, *et al.* Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell*, 2006, **127**(6): 1193–1207
- [16] Cecere G, Zheng G X, Mansisidor A R, *et al.* Promoters recognized by forkhead proteins exist for individual 21U-RNAs. *Molecular Cell*, 2012, **47**(5): 734–745
- [17] 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
- [18] Goecks J, Nekrutenko A, Taylor J. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology*, 2010, **11**(8): R86
- [19] Bailey T L, Boden M, Buske F A, *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research*, 2009, **37**(Web Server issue): W202–208
- [20] Crooks G E, Hon G, Chandonia J M, *et al.* WebLogo: A sequence logo generator. *Genome Research*, 2004, **14**(6): 1188–1190
- [21] Klattenhoff C, Theurkauf W. Biogenesis and germline functions of piRNAs. *Development*, 2008, **135**(1): 3–9
- [22] Vagin V V, Sigova A, Li C J, *et al.* A distinct small RNA pathway silences selfish genetic elements in the germline. *Science*, 2006, **313**(5785): 320–324
- [23] 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

The Research of *Caenorhabditis elegans* New piRNAs

CUI Dong-Ya^{1,2,3}, WANG Jia-Jia^{1,2}, SUN Xu-Bin^{1,2}, ZHANG Peng², HE Shun-Min^{2*}, HUANG Da-Wei^{1,2*}

⁽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 PIWI-interacting RNAs(piRNA) are endogenous small RNAs(sRNA), which play roles in resisting exogenous gene invasion and transposon mobility. More than 16 000 piRNAs are found in *Caenorhabditis elegans*, and the piRNA loci share a conserved upstream sequence. New piRNAs can be predicted based on its conserved upstream sequence. *C. elegans* are synchronized and small RNAs from wild type and *prg-1* strains at L4 developmental period are sequenced through illumina technology. Basing on piRNA conserved upstream sequence, we find 967 new piRNAs expressed at L4 period, and the expressions of new piRNAs disappear in *prg-1* mutant. Most of new piRNAs are mapped into two piRNA clusters of Chromosome IV, and have an overwhelming bias for a 5' uracil. Compared to the published co-immunoprecipitation data from adult worms, over 100 new piRNAs can interact with PRG-1. However, some non-21nt long sRNAs, which also share the same loci with piRNAs, are also found in wide type worms. These non-21nt long sRNA may be co-products of piRNA precursors. Together, these results indicate that our sRNA sequencing identified many new piRNAs in *C. elegans*.

Key words *Caenorhabditis elegans*, piRNA, prediction, new piRNA, PRG-1

DOI: 10.3724/SP.J.1206.2014.00072

*Corresponding author.

HE Shun-Min. Tel: 86-10-64807279, E-mail: hesm@ioz.ac.cn

HUANG Da-Wei. Tel: 86-10-64807235, E-mail: huangdw619@gmail.com

Received: May 1, 2014 Accepted: September 10, 2014