利用拟南芥杂交组合研究 siRNA 与等位基因 DNA 甲基化调控中的联系

陈少霞, 何航*, 邓兴旺*

北京大学生命科学学院, 北京 100871

* 联系人, E-mail: hehang@pku.edu.cn; deng@pku.edu.cn

收稿日期: 2013-09-03; 接受日期: 2013-09-18

国家重点基础研究发展计划(批准号: 2012CB910900)和国家转基因生物新品种培育科技重大专项(批准号: 2010ZX08010-003)资助
doi: 10.1360/052013-282

摘要 近年来,越来越多的实验结果表明,表观遗传因子,如 DNA 甲基化、小 RNA、组蛋白修饰等在杂种优势形成中起到重要作用,然而对于这些表观遗传因子在 F_1 中遗传调控方式的认知仍很有限. 本实验室先前工作曾以拟南芥 C24 和 Ler 两种生态型及其正反交子一代为材料,运用新一代测序方法获得该杂交组合中 DNA 甲基化及小 RNA 单碱基分辨率的全基因组图谱. 本文进一步对这批数据中的等位基因 DNA 甲基化水平进行分析,区分 DNA 甲基化遗传过程中的顺式与反式调控方式,并发现这两种调控方式均有重要的贡献. 研究发现,siRNA 与 DNA 甲基化的两种调控方式有密切联系,尤其在 DNA 甲基化的反式调控中, F_1 中 DNA 甲基化变化程度越大,该区域内 siRNA 富集程度越强,二者可能存在某种调控机制. 通过等位基因表观遗传组的分析研究杂交过程中 DNA 甲基化和小 RNA 遗传调控的规律,为更好地理解杂种优势机制提供了帮助.

关键词

单核苷酸多态性(SNP)
DNA 甲基化
siRNA
顺式作用
反式作用
等位基因

DNA 甲基化作为一种重要的表观修饰,在基因表达、细胞分化以及系统发育过程中起着重要的调节作用,植物 DNA 甲基化模式的改变可以影响植物的花期、育性、花和叶的形态等^[1-8]. DNA 甲基化具有遗传效果, Morgan 等人^[1]发现,控制小鼠(*Mus musculus*)体色的等位基因 DNA 甲基化状态遗传到子一代;拟南芥(*Arabidopsis thaliana*)的研究也已证实了 DNA 甲基化可以忠实地遗传给子代^[2,3]. 同时大量研究发现^[4-9], DNA 甲基化在遗传过程中不是一成不

变的,在生殖发育过程中, DNA 甲基化状态是一个动态变化的调节过程,并且通过等位基因的 DNA 甲基化分析可知,这种调节包含顺式和反式的作用方式. 这种 DNA 甲基化的重塑方式在不同的生物中存在差异:小鼠中,两个亲本来源的基因组通过不同的去 DNA 甲基化机制,父本激活的去 DNA 甲基化,母本削弱的 DNA 甲基化,最后达到一个与两个配子都不同的高度 DNA 甲基化状态^[4-7];最新的对斑马鱼(*Danio rerio*)的研究表明,生殖发育过程中父本基因

引用格式: 陈少霞, 何航, 邓兴旺. 利用拟南芥杂交组合研究 siRNA 与等位基因 DNA 甲基化调控中的联系. 中国科学: 生命科学, 2013, 43: 897-904
Chen S X, He H, Deng X W. Allelic DNA methylation analysis in *Arabidopsis* hybrids Associated with siRNA. SCIENTIA SINICA Vitae, 2013, 43: 897-904, doi: 10.1360/052013-282

组 DNA 甲基化模式保持稳定, 依靠母本基因组 DNA 甲基化的动态变化调节该过程^[8,9]; 早期拟南芥胚胎发育过程中母本基因组甲基化占主导作用, 随着胚胎发育父本的贡献逐渐增大, 这一过程与 siRNA 相关^[10], 而生殖发育完成后, 亲本来源的基因组 DNA 甲基化状态在不同的组织器官中基本稳定; 小鼠等位基因差异 DNA 甲基化研究, 表明顺式作用元件很大程度上影响了 F₁ 中等位基因差异^[11]; 在拟南芥、水稻(*Oryza sativa*)、番茄(*Solanum lycopersicum*)等物种中也开展了等位基因差异 DNA 甲基化的研究, 并分析了影响这些差异的顺式作用和反式作用的比例, 指出了 siRNA 与这种差异的相关性^[12-18]。

DNA 甲基化常与 siRNA 通路共同作用, 以构建或保持染色体的螺旋或松弛状态, 一些 20~24 nt 的 siRNA 通常通过调节特定位点的 DNA 甲基化状态, 间接调控基因转录。RNA 指导的 DNA 甲基化(RNA-directed DNA methylation, RdDM)首次发现于植物系统^[19], 指内切核糖核酸酶 DICER LIKE3(DCL3)产生的 24 nt siRNA^[20], 通过与含有 PIWI 结构域的 ARGONAUTE4(AGO4)结合, 来指导 DRM2(domains rearranged methyltransferase 2)的活性^[21-23], 从而催化 DNA 从头甲基化。水稻的实验数据支持等位基因上 siRNA 的差异表达可能介导等位基因差异 DNA 甲基化^[12]; 对拟南芥和斑马鱼的研究, 也充分证实了 24 nt 的 siRNA 的表达与 DNA 甲基化正相关^[8,14]。

本实验室利用新一代测序方法, 在拟南芥 Ler 和 C24 的杂交组合中获得了全基因组 DNA 甲基化和小 RNA 图谱, 分析了基因组水平 DNA 甲基化和小 RNA 的变化以及它们的联系^[17]。本文通过进一步分析杂交子一代中等位基因上的 DNA 甲基化水平和小 RNA 水平, 研究调控 DNA 甲基化的顺式和反式作用, 并探索 siRNA 在这两种调控中扮演的作用。

1 材料与方法

1.1 材料

本文的实验材料和测序原始数据为本实验室的前期研究^[17]获得, 原始数据存于 NCBI 的测序数据库(登录号: GSE24658)。

1.2 C24 与 Ler 间核苷酸多态性鉴定

C24 的基因组被切割为 38983 个 3 kb 片段, 利用 Blastn^[24]比对到 Ler 基因组上, 得到 38921 个最佳比对结果, 其中 301 个片段的结果比对在非同一条染色体上, 1117 个片段比对在 Ler 基因组的重复序列区域, 在经过过滤后的 37503 个片段的比对结果中得到 586902 个单核苷酸多态性(single-nucleotide polymorphism, SNP)位点。将 Ler 上对应位置的碱基替换为 C24 的基因型, 作为 C24 生态型的参考基因组。

1.3 小 RNA 测序

获得 Illumina 测序数据后, 先进行质量控制: 去掉接头序列、N 的比例大于 10% 的序列、质量值小于 5 的碱基占全长比例超过 50% 的低质量序列, 剪切掉序列 3'端质量值小于 5 的部分, 最后保留 16 bp 以上的测序数据。用 bowtie 分别比对到 Ler 和 C24 的基因组上^[25], 不允许错配, 根据 SNP 标签分离出两个亲本来源的测序数据。本实验中, 小 RNA 的表达量以每 200 bp 内小 RNA 的序列数作为标准, 以 20 bp 步移(step)在基因组上移动, 计算出每 200 bp 的区间 Ler 和 C24 来源的小 RNA 的表达量。

1.4 DNA 甲基化测序

获得测序数据后, 采用 1.3 小节的方法进行质量控制, 去掉低质量和接头序列。使用 bismark 软件将序列比对到 Ler 和 C24 的基因组上^[26], 根据 SNP 标签分离出两个亲本来源的测序数据。本实验中, 以 DNA 甲基化序列数除以 DNA 甲基化序列和未 DNA 甲基化序列数的总和作为衡量单个 C/G 位点 DNA 甲基化程度的标准, 并计算出有 4 个以上序列覆盖的 C/G DNA 甲基化程度, 应用 R 语言中的“binom.test”负二项分布检验方法检测每个 C/G 在 Ler 和 C24 间的差异 DNA 甲基化情况, 挑取 $P < 0.05$, 且 DNA 甲基化程度差值大于 0.1 认为是差异 DNA 甲基化位点。应用“Wilcox.test”秩检验, 检测以 20 bp 步移在基因组上移动, 每 200 bp 的区间 Ler 和 C24 来源的 DNA 甲基化差异程度, 挑取 $P < 0.05$, 且 DNA 甲基化程度差值大于 0.1 认为是差异 DNA 甲基化区间(differential methylated regions, DMR)。

2 结果与讨论

2.1 利用 C24 与 Ler 基因组有效地分析等位基因 DNA 甲基化和小 RNA 水平

C24 和 Ler 是两种拟南芥生态型, 它们的基因组中存在大量 SNP^[27-29], 这是本实验设计的可行性基础. 将 C24 的基因组序列分为 38983 个 3 kb 长度的片段, 利用 Blastn 比对到 Ler 基因组上, 去除重复序列区域, 在 37503 片段中共发现 586902 个 SNP, 平均约 200 bp 存在一个 SNP, 而其中 53.1% 都为“C/T” (C/T, G/A, T/C, A/G) 型 SNP, 与水稻比例接近^[12].

运用新一代测序方法, 在拟南芥 C24 和 Ler 两种生态型及其正反交子一代 FLC 和 FCL 的组合中, 获得大量这一杂交组合中全基因组单碱基分辨率的 DNA 甲基化及小 RNA 数据(网络版附表 1). DNA 甲基化测序数据中落在 SNP 上的序列数为 C24: 7149819(4.81%), FCL: 8238223(5.52%), FLC: 7257946(5.29%) 和 Ler: 6931506(5.21%). 将 4 个材料中均不低于 4 条序列覆盖的 C/G 位点挑出进行研究, 共有 523278 个位点, 其中 CG, CHG 和 CHH 分别为 117928, 108727 和 296623 个. 而小 RNA 测序数据中落在 SNP 上的序列数为 C24: 613731(3.00%), FCL: 1825776(7.63%), FLC: 1872542(7.83%) 和 Ler: 785619(3.80%). 说明本实验获得了充分的测序数据以进行等位基因表观遗传分析.

利用基于 bowtie 比对算法的 bismark^[26], 将 F₁ 中的 DNA 甲基化测序序列分别比对到 C24 及 Ler 基因组上, 从而获得等位基因上的不同 DNA 甲基化水平. 同样将亲本中 DNA 甲基化测序序列分别比对到两个基因组上, 发现 C24 中有 17.33% (7149819 中的 1239415 条) 的测序序列比对到 Ler 基因组上, Ler 中有 6.60% (6931506 中的 457407 条) 的测序序列比对到 C24 基因组上, 说明本实验寻找等位基因 DNA 甲基化的差异方法具有较高的准确性(网络版附表 2). 亲本 C24 与 Ler 之间 CG 型 DNA 甲基化水平有一定差异, 而 CHG 和 CHH 型 DNA 甲基化水平基本一致. F₁ 中基因组整体 DNA 甲基化水平与亲本保持一致(图 1A).

2.2 子代 DNA 甲基化的顺式作用与反式作用

在遗传过程中, 不仅亲本的基因组序列信息会

保留到子代基因组中, 亲本的表观遗传信息也会以不同方式传递至子代中^[2,30]. 本研究中, 父母本中均检测到 DNA 甲基化修饰的位点, 绝大多数在 F₁ 中也遗传了 DNA 甲基化修饰 (FLC 中 94.5%, FCL 中 94.8%); 而对于仅在一个亲本中检测到 DNA 甲基化修饰的位点, 大约一半在 F₁ 中也遗传了 DNA 甲基化修饰: C24 中无 DNA 甲基化, Ler 中有 DNA 甲基化的位点, FLC 中 49.47%, FCL 中 49.4% 被 DNA 甲基化; C24 中有 DNA 甲基化, Ler 中无 DNA 甲基化的位点, FLC 中 37.26%, FCL 中 38.09% 被 DNA 甲基化(网络版附表 3); 至于在两个亲本中均未检测到 DNA 甲基化修饰的位点, 在 F₁ 中检测到 DNA 甲基化修饰 FLC 中仅为 3.64%, FCL 中为 3.62%, 而且考虑到亲本中检测深度不够的情况, 说明了在全基因组水平 DNA 甲基化修饰信息的遗传是基本稳定的. 进一步, 比较了亲本中 DNA 甲基化水平与 F₁ 中相同基因组来源的 DNA 甲基化水平的差异, 共有 2.77%~3.18% 的胞嘧啶位点在亲本与 F₁ 中 DNA 甲基化水平有差异, 在 CG, CHG 和 CHH 3 种 DNA 甲基化中, 亲本与 F₁ 中 DNA 甲基化水平有差异的位点占不低于 4 个读长 (reads) 覆盖的胞嘧啶位点比例分别为 2.19%~2.48%, 3.49%~4.00% 和 2.68%~3.16% (图 1B 和网络版附图 1).

在较为稳定地遗传亲本表观遗传组信息的前提下, F₁ 中的 DNA 甲基化水平会受众多因素的调控, 从而动态地影响个体的生长发育, 因此在一系列实验中都发现了 F₁ 中大量位点 DNA 甲基化水平相对于亲本的变化^[12,31,32]. 将任一 F₁ 与亲本比较 DNA 甲基化水平存在差异的胞嘧啶位点集合起来, 在 CG, CHG 和 CHH 3 种类型中分别发现了 28727, 24181 和 50555 个此类胞嘧啶位点. 其中一些 DNA 甲基化水平的差异是由顺式作用引起的, 即由父母本基因组序列的不同而导致的 DNA 甲基化水平差别, 并在 F₁ 中保留这种差别; 而另一些 DNA 甲基化水平的差异则是反式作用的结果, 即由反式作用因子的调控导致 F₁ 中 DNA 甲基化水平变化. 通过将 F₁ 中等位基因 DNA 甲基化的差异与父母本之间相同位点 DNA 甲基化的差异做比较, 可以有效地将这两种作用区分(图 1C). 将这 2 种差异值进行比较, 基于负二项分布检验, $P < 0.05$ 且 DNA 甲基化程度的差值大于 0.1 的认为有反式作用参与调控, 否则认为属于顺式作用调控. 通过此方法发现, FLC 和 FCL 中 CG 类型 60.19% 和 59.59% 的 DNA 甲基化差异是由顺式作用所引起, 而 39.81%

图1 Ler 与 C24 正反交子一代中的 DNA 甲基化调控模式

A: 亲本和子一代中 C, CG, CHG 及 CHH 型平均 DNA 甲基化水平; B: 亲本中 DNA 甲基化水平与 F₁ 中相同基因组来源的 DNA 甲基化水平的差异位点所占比例; C: CG, CHG 和 CHH 类型 DNA 甲基化差异位点的热图。P 示亲本中等位基因上 Ler 与 C24 的 DNA 甲基化水平差异; FCL 示 FCL 材料中等位基因上分别由 Ler 与 C24 来源的 DNA 甲基化水平差异; P-FLC 示亲本中的 DNA 甲基化水平差异与 FLC 中的差异相减。在 P-FCL 与 P-FLC 列中, 黑色部分示顺式作用, 红色或绿色部分示反式作用

和 40.41% 的 DNA 甲基化差异是反式作用参与的结果, 而 CHG 与 CHH 的 DNA 甲基化差异中, 反式作用所占比例更高, 分别在 FLC 和 FCL 中占 65.05% 与 64.23%, 75.66% 与 74.65% (图 1C, 网络版附表 4), 说明因为不同 DNA 甲基化类型调控方式不一样, 其在遗传过程中的调控模式也有区别. 网络版附图 2 显示了 F_1 中的顺式作用与反式作用的例子.

2.3 siRNA 可能参与子代 DNA 甲基化水平的调控

DMR 被认为与基因表达的调控有着密切的联系^[12], 以 20 bp 的步移在整个基因组上移动, 在每 200 bp 的区段(window)内采用 Wilcoxon 秩检验的方法来检测父母本之间以及 F_1 的等位基因之间的 DNA 甲基化差异区域. 通过这一方法, 获得父母本之间的差异 DNA 甲基化区域有 55759(28374 个 $C24 < \text{Ler}$; 27385 个 $C24 > \text{Ler}$) 个, 而 FLC 与 FCL 中等位基因之间的 DNA 甲基化差异区域数量分别为 39338(26244 个 $C24 < \text{Ler}$; 13094 个 $C24 > \text{Ler}$) 和 40561(28459 个 $C24 < \text{Ler}$; 12102 个 $C24 > \text{Ler}$). 将其中 3 类区域挑出进行研究: (i) 在父母本之间 DNA 甲基化差异, 同时在 F_1 的等位基因之间 DNA 甲基化差异, 并且 2 种差异有着相同的方向; (ii) 在父母本之间 DNA 甲基化差异, 而在 F_1 的等位基因之间 DNA 甲基化无差异; (iii) 在父母本之间 DNA 甲基化无差异, 而在 F_1 的等位基因之间 DNA 甲基化差异. 其中第一类仅受顺式作用调控, 而后两类仅受反式作用调控, 这样的区分为分析 2 种调控模式的差别提供了便利, 这 3 类区域的数量在 FCL 中分别为 9452, 37499 和 21078, FLC 中分别为 10171, 36635 和 21473 (网络版附表 5).

siRNA 属于影响 DNA 甲基化的反式作用因子之一, 它可以通过 RNA 介导的 RdDM 指导基因组的 DNA 甲基化修饰^[19], 因此可以通过小 RNA 测序的结果研究差异 DNA 甲基化区域与 siRNA 表达的联系. 选取上文所述 3 类区域中亲本 Ler 的 DNA 甲基化大于 C24 的 DNA 甲基化水平, 或者 F_1 中 Ler 等位基因的 DNA 甲基化水平大于 C24 等位基因的 DNA 甲基化水平的位点, 在每一类区域中按照 Ler 的 DNA 甲基化大于 C24 的 DNA 甲基化的位点所占比例分为 5 组, 5 组从 50%~60% 直到 90%~100%, 代表 Ler 等位基因 DNA 甲基化大于 C24 等位基因 DNA 甲基化的

程度依次增强(图 2A). 发现, 随着 DNA 甲基化差异程度的增加, 第一类区域内的 siRNA 数量无显著变化, 但都显著高于基因组平均水平, 说明在顺式作用调控中差异 DNA 甲基化区域总有大量 siRNA 富集, 并暗示这些 siRNA 可能并不影响差异 DNA 甲基化水平的遗传; 而在第二类、第三类区域内的 siRNA 数量随着差异程度的变高都显著上升, 暗示 F_1 中 DNA 甲基化的变化与 siRNA 的大量富集可能有因果关系(图 2A). 进一步, 对 F_1 中这三类区域内等位基因上 siRNA 的数量进行了比较, 观察 siRNA 丰度差异与 DNA 甲基化水平差异的关系, 图 2B 中统计了每类差异 DNA 甲基化区域中, Ler 来源 siRNA 占据比例大于 75% 的区域所占的百分比, 发现在顺式作用调控的区域内, 等位基因 siRNA 丰度的差异与等位基因 DNA 甲基化水平的差异呈正相关, 暗示在这样的区域内, siRNA 与 DNA 甲基化同样为顺式作用调控; 而在后两类受反式作用调控的区域中, 等位基因 siRNA 丰度的差异与等位基因 DNA 甲基化水平的差异无关, 并且与基因组等位基因 siRNA 差异的平均水平相当, 暗示在反式作用调控的 DNA 甲基化中, siRNA 可能是反式作用因子之一, 而其本身不受反式作用调控. 将第一类区域与第三类区域进行比较, 可以发现对于亲本中 DNA 甲基化水平差异的区域, 其 F_1 中是否仍保留一致的差异取决于亲本中 siRNA 丰度. 当亲本中 siRNA 丰度的差异与亲本中 DNA 甲基化水平差异一致时, 说明该区域受顺式作用调控, F_1 中的 DNA 甲基化水平将会保留这种差异; 当亲本中 siRNA 丰度的差异与亲本中 DNA 甲基化水平差异不一致时, 说明该区域受反式作用调控, F_1 中的 DNA 甲基化水平会受反式作用因子调控而改变. 图 2C 中分别显示了这两类区域受不同调控作用的例子.

3 结论

本实验重点研究了全基因组单核苷酸分辨率的 2 个拟南芥亲本 C24 和 Ler 以及正反交子一代中等位基因的 DNA 甲基化变化. 首先, 观察到全基因组平均水平的 CG, CHG 和 CHH 的 DNA 甲基化程度分别在大约 25%, 10% 和 5% 的水平(图 1A), 亲本与 F_1 中基本保持稳定. 而在 4 个材料中均不低于 4 条序列覆盖的胞嘧啶共有 523278 个, 并且亲本 C24 及 Ler 中

的 DNA 甲基化序列大部分都比对到自身基因组上 (80.50%和 96.12%), 说明本实验测序数据可以有效并准确地进行等位基因 DNA 甲基化水平分析。

通过等位基因 DNA 甲基化水平的分析, 发现 F_1 中遗传了亲本的大部分 DNA 甲基化修饰 (最低 96.00%, 最高 97.81%, 图 1B)。对 F_1 与亲本 DNA 甲基化水平差异的位点, 通过比较等位基因 DNA 甲基

化差异与父母本 DNA 甲基化差异程度的方法进行分析, 利用系统聚类方法, 获得 F_1 中 DNA 甲基化水平变化的模式, 可以显著地区分顺式作用与反式作用的模式, 在 CG 型 DNA 甲基化中顺式作用为主导 (60.19%), 而在 CHG 与 CHH 型 DNA 甲基化中反式作用占较高比例 (75.66%, 图 1C, 网络版附表 4)。

结合小 RNA 测序的数据发现, F_1 及亲本的 DMR

图 2 差异 DNA 甲基化区域与 siRNA 差异表达

A: 差异 DNA 甲基化区域内 siRNA 的丰度。X 轴示 Ler 来源 DNA 甲基化高于 C24 来源 DNA 甲基化的位点的区域所占比例, Y 轴示区域内每 200 bp 的 siRNA 丰度, 对照组示 DNA 甲基化没有差异的区域; B: 差异 DNA 甲基化区域内 siRNA 的差异表达。X 轴示 Ler 来源 DNA 甲基化高于 C24 来源 DNA 甲基化的位点的区域所占比例, Y 轴示 Ler 来源 siRNA 占据比例大于 75% 的区域所占比例, 蓝色与红色均示 F_1 中 siRNA 差异区域的比例, 绿色示亲本中 siRNA 差异区域的比例; C: 同样亲本中 DNA 甲基化水平差异的区域。因 siRNA 调控方式不同, F_1 中的 DNA 甲基化表现为顺式作用和反式作用。蓝色柱示甲基化程度, 红色柱示 siRNA 表达水平, 坐标示读长数量

内 siRNA 明显富集. 具体对顺式作用的区域而言, siRNA 总是具有较高丰度, 同样为顺式作用; 而对反式作用的区域而言, siRNA 的富集程度与 DNA 甲基化水平的变化程度正相关. 尤其发现, 对同样亲本中的差异 DNA 甲基化区域, F_1 的基因 DNA 甲基化水平

是否变化与亲本中的 siRNA 丰度是否差异是一致的, 提示 F_1 中的 DNA 甲基化调控方式可能取决于该区域内 siRNA 与 DNA 甲基化的作用模式, 说明 siRNA 可能通过 RNA 介导的 RdDM 在 F_1 的 DNA 甲基化形成中起到重要调控作用.

参考文献

- 1 Morgan H D, Sutherland H G, Martin D I, et al. Epigenetic inheritance at the agouti locus in the mouse. *Nat Genet*, 1999, 23: 314–318
- 2 Schmitz R J, Schultz M D, Lewsey M G, et al. Transgenerational epigenetic instability is a source of novel methylation variants. *Science*, 2011, 334: 369–373
- 3 Becker C, Hagmann J, Müller J, et al. Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome. *Nature*, 2011, 480: 245–249
- 4 Wossidlo M, Nakamura T, Lepikhov K, et al. 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nat Commun*, 2011, 2: 241
- 5 Gu T P, Guo F, Yang H, et al. The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature*, 2011, 477: 606–610
- 6 Inoue A, Zhang Y. Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. *Science*, 2011, 334: 194
- 7 Smith Z D, Chan M M, Mikkelsen T S, et al. A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature*, 2012, 484: 339–344
- 8 Jiang L, Zhang J, Wang J J, et al. Sperm, but not oocyte, DNA methylome is inherited by zebrafish early embryos. *Cell*, 2013, 153: 773–784
- 9 Potok M E, Nix D A, Parnell T J, et al. Reprogramming the maternal zebrafish genome after fertilization to match the paternal methylation pattern. *Cell*, 2013, 153: 759–772
- 10 Autran D, Baroux C, Raissig M T, et al. Maternal epigenetic pathways control parental contributions to *Arabidopsis* early embryogenesis. *Cell*, 2011, 145: 707–719
- 11 Blewitt M E, Gendrel A V, Pang Z, et al. SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. *Nat Genet*, 2008, 40: 663–669
- 12 Chodavarapu R K, Feng S, Ding B, et al. Transcriptome and methylome interactions in rice hybrids. *Proc Natl Acad Sci USA*, 2012, 109: 12040–12045
- 13 Greaves I K, Groszmann M, Ying H, et al. Trans chromosomal methylation in *Arabidopsis* hybrids. *Proc Natl Acad Sci USA*, 2012, 109: 3570–3575
- 14 Groszmann M, Greaves I K, Albertyn Z I, et al. Changes in 24-nt siRNA levels in *Arabidopsis* hybrids suggest an epigenetic contribution to hybrid vigor. *Proc Natl Acad Sci USA*, 2011, 108: 2617–2622
- 15 Greaves I, Groszmann M, Dennis E S, et al. Trans-chromosomal methylation. *Epigenetics*, 2012, 7: 800–805
- 16 He G, Zhu X, Elling A A, et al. Global epigenetic and transcriptional trends among two rice subspecies and their reciprocal hybrids. *Plant Cell*, 2010, 22: 17–33
- 17 Shen H, He H, Li J, et al. Genome-wide analysis of DNA methylation and gene expression changes in two *Arabidopsis* ecotypes and their reciprocal hybrids. *Plant Cell*, 2012, 24: 875–892
- 18 Shivaprasad P V, Dunn R M, Santos B A, et al. Extraordinary transgressive phenotypes of hybrid tomato are influenced by epigenetics and small silencing RNAs. *EMBO J*, 2012, 31: 257–266
- 19 Wassenegger M, Heimes S, Riedel L, et al. RNA-directed de novo methylation of genomic sequences in plants. *Cell*, 1994, 76: 567–576
- 20 Henderson I R, Zhang X, Lu C, et al. Dissecting *Arabidopsis thaliana* DICER function in small RNA processing, gene silencing and DNA methylation patterning. *Nat Genet*, 2006, 38: 721–725
- 21 Henderson I R, Jacobsen S E. Epigenetic inheritance in plants. *Nature*, 2007, 447: 418–424
- 22 Law J A, Jacobsen S E. Molecular biology. Dynamic DNA methylation. *Science*, 2009, 323: 1568–1569
- 23 Law J A, Jacobsen S E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet*, 2010, 11: 204–220
- 24 Camacho C, Coulouris G, Avagyan V, et al. BLAST+: architecture and applications. *BMC Bioinformatics*, 2009, 10: 421
- 25 Langmead B, Trapnell C, Pop M, et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*, 2009, 10: R25
- 26 Krueger F, Andrews S R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*, 2011, 27: 1571–1572
- 27 Schneeberger K, Ossowski S, Ott F, et al. Reference-guided assembly of four diverse *Arabidopsis thaliana* genomes. *Proc Natl Acad Sci USA*,

- 2011, 108: 10249–10254
- 28 Gan X, Stegle O, Behr J, et al. Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. *Nature*, 2011, 477: 419–423
- 29 Cao J, Schneeberger K, Ossowski S, et al. Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nat Genet*, 2011, 43: 956–963
- 30 Stouder C, Paoloni-Giacobino A. Transgenerational effects of the endocrine disruptor vinclozolin on the methylation pattern of imprinted genes in the mouse sperm. *Reproduction*, 2010, 139: 373–379
- 31 Schilling E, El Chartouni C, Rehli M. Allele-specific DNA methylation in mouse strains is mainly determined by *cis*-acting sequences. *Genome Res*, 2009, 19: 2028–2035
- 32 Xie W, Barr C L, Kim A, et al. Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. *Cell*, 2012, 148: 816–831

Allelic DNA Methylation Analysis in *Arabidopsis* Hybrids Associated with siRNA

CHEN ShaoXia, HE Hang & DENG XingWang

School of Life Sciences, Peking University, Beijing 100871, China

Recently, accumulating evidence suggests that epigenetic factors including DNA methylation, small RNA and histone modifications may play important roles in hybrid vigor in plants. However, knowledge about the mechanism of these epigenetic factors in hybrids is still limited. Based on genome-wide DNA methylation and small RNA maps of *Arabidopsis thaliana* Ler and C24 ecotypes and their reciprocal F₁ hybrids in previous work, we further analyzed allelic specific DNA methylation. Our study indicated that both *trans*- and *cis*-factors play roles in DNA methylation regulations in hybrids. In addition, we found that siRNAs correlate with DNA methylation *trans*-regulations. High degree *trans*-regulation DNA methylation is significantly concurrent with enriched siRNA regions. Our study analyzed DNA methylation regulations associated with siRNAs in hybrids, and shed light on understanding the mechanism of hybrid vigor.

SNP, DNA methylation, siRNA, *cis*-effect, *trans*-effect, allele

doi: 10.1360/052013-282

论文

附录

附图 1 顺式和反式作用 DNA 差异甲基化位点所占比例

附表 1 测序获得的原始序列数量

	sRNA-seq	methylC-seq
C24	20492294	148505088
FCL	23920380	149235643
FLC	23906626	137203310
Ler	20714708	133138314

附图 2 差异 DNA 甲基化位点中顺式和反式作用的实例
蓝色柱示甲基化程度

附表 2 比对到亲本及子一代中 C24 和 Ler 基因组上的序列数

		原始序列数	比对到 C24 基因组上的 总序列数	比对到 Ler 基因组上的 总序列数	落在 SNP 周围 ler 基因组 上的序列数	落在 SNP 周围 C24 基因组 上的序列数
DNA 甲基化	C24	148505088	53545405	52440953	1239415	5910404
	FCL	149235643	63801399	64261858	4753466	3485257
	FLC	137203310	54564953	55093914	4265871	2992075
	Ler	133138314	50830812	52484036	6474099	457407
siRNA	C24	20492294	11433703	11012892	119666	494065
	FCL	23920380	13559479	13798883	1053289	772487
	FLC	23906626	13753845	14005178	1071896	800646
	Ler	20714708	12619049	11725146	755102	30517

附表 3 亲本 C24 或者 Ler 上无甲基化位点在子一代中的甲基化情况

	C24 无甲基化, Ler 甲基化	C24 甲基化, Ler 无甲基化	C24 无甲基化, Ler 无甲基化
FCL	210432	143936	259374
FLC	210687	140792	260890

附表 4 子一代等位基因的差异 DNA 甲基化位点中受顺式反式作用的位点数

	FCL <i>cis</i>	FCL <i>trans</i>	FLC <i>cis</i>	FLC <i>trans</i>
C	71689	27512	70763	26426
CG	22204	13231	20841	12544
CHG	16666	5962	16325	5706
CHH	32819	8319	33597	8176

附表 5 子一代差异 DNA 甲基化区间数中顺式反式作用的数量

		Ler>C24	C24>Ler
FCL	cis	8815	637
	P diff	18655	18844
	F1 diff	9525	11553
FLC	cis	9510	661
	P diff	17882	18753
	F1 diff	10978	10459