

植物长链非编码 RNA 研究进展

黄小庆, 李丹丹, 吴娟

东北林业大学盐碱地生物资源环境研究中心, 东北油田盐碱植被恢复与重建教育部重点实验室, 哈尔滨 150040

摘要: 长链非编码 RNA(Long non-coding RNA, lncRNA)长度大于 200 个核苷酸, 大量存在于生物体中并具有多种生物学功能。目前, 植物中发现的 lncRNA 大多由 RNA 聚合酶 II 转录, 并通过目标模仿、转录干扰、组蛋白甲基化和 DNA 甲基化等多种机制介导基因的表达, 在植物开花、雄性不育、营养代谢、生物和非生物胁迫等生物过程中起着调节因子的作用。文章综述了近年来发现的植物 lncRNA 数据库、预测方法、表达及可能的生物学功能。

关键词: 长链非编码 RNA; 数据库; 基因表达调控; 生物学功能

Long non-coding RNAs in plants

Xiaoqing Huang, Dandan Li, Juan Wu

Key Laboratory of Saline-Alkali Vegetation Ecology Restoration in Oil Field of Ministry of Education, Alkali Soil Natural Environmental Science Center, Northeast Forestry University, Harbin 150040, China

Abstract: Long non-coding RNAs (lncRNAs), which are longer than 200 nucleotides in length, widely exist in organisms and function in a variety of biological processes. Currently, most of lncRNAs found in plants are transcribed by RNA polymerase II and mediate gene expression through multiple mechanisms, such as target mimicry, transcription interference, histone methylation and DNA methylation, and play important roles in flowering, male sterility, nutrition metabolism, biotic and abiotic stress and other biological processes as regulators in plants. In this review, we summarize the databases, prediction methods, and possible functions of plant lncRNAs discovered in recent years.

Keywords: long non-coding RNAs (lncRNAs); database; gene expression regulation; biological function

生物体内存在着两种不同的 RNA: 能翻译成蛋白质的编码 RNA(Coding RNA, 即 mRNA)和不翻译成蛋白质的非编码 RNA(Non-coding RNA, ncRNA)^[1]。ncRNA 种类繁多, 目前尚无规范的命名方法。根据

其表达特征, ncRNA 分为持家非编码 RNA(House-keeping non-coding RNA)和调控非编码 RNA(Regulatory non-coding RNA)^[2]。持家非编码 RNA 包括 rRNA、tRNA、snRNA(Small nuclear RNA)和 snoRNA

收稿日期: 2014-12-04; 修回日期: 2015-01-27

基金项目: 东北林业大学中央高校基本科研业务费专项资金项目(编号: DL12BA38), 长江学者和创新团队发展计划资助(编号: IRT13053), 东北林业大学中央高校基本科研业务费专项资金项目(编号: DL13EA04 - 02)和教育部留学回国人员科研启动基金(第 47 批)资助

作者简介: 黄小庆, 硕士研究生, 专业方向: 生物化学与分子生物学。E-mail: huangxiaoqing85@163.com

通讯作者: 吴娟, 博士, 副教授, 研究方向: 植物生理学和 RNA 分子生物学。E-mail: wuj1970@163.com

DOI: 10.16288/j.ycz.14-432

网络出版时间: 2015-3-3 16:58:43

URL: <http://www.cnki.net/kcms/detail/11.1913.R.20150303.1658.002.html>

(Small nucleolar RNA)等, 这些 RNA 在所有类型细胞中均要表达并且受环境因素影响较小, 对维持细胞的基本功能是必不可少的。调控非编码 RNA 在生物体的特定组织器官和发育阶段表达, 或者对应激环境产生应答反应后特异表达, 这种特异表达调控着各种生物过程^[3]。调控非编码 RNA 按照长度可分为短链非编码 RNA (Small ncRNA) 和长链非编码 RNA (Long non-coding RNA, lncRNA) 两大类^[2]。短链非编码 RNA 长度小于 200 nt, 包括 miRNA、siRNA 和 piRNA (Piwi-interacting RNA) 等^[4, 5]。短链非编码 RNA 的转录调控和作用机制已经比较清晰, 如 miRNA 首先经 RNA 聚合酶 II 转录形成 pri-miRNA, 然后经 RNA 核酸酶的剪切作用后成为 20~30nt 的成熟 miRNA, 它们在基因转录后水平通过与靶 mRNA 互补结合, 在 RNA 加工、转录及转录后的基因沉默、应激应答、发生及发育等生物过程中起着调节因子的作用。

lncRNA 是指长度在 200nt 以上的调控非编码 RNA^[6], 没有长的开放阅读框 (ORF), 不具备编码蛋白质的功能, 但是在特定的条件下有些 lncRNA 可以编码功能性寡肽^[7]。lncRNA 与 mRNA 有许多共同特征, 例如由 RNA 聚合酶 II 进行转录, 并有 5' 加帽、3' 聚腺苷酸化结构等^[8]。研究表明, 哺乳动物基因组中 80% 的转录产物为 lncRNA^[9]。lncRNA 的作用机制和生物学功能极其多样, 通过影响其他 RNA 或蛋白质的稳定性来调节基因的表达, 参与剂量补偿、基因组印记、X 染色体失活等生物过程^[10, 11]。哺乳动物中超过 1000 个 lncRNA 参与了基因表达调控, 在细胞周期调控、免疫监视和胚胎干细胞多能性过程发挥作用^[12]。大量研究证明, 多种疾病与 lncRNA 的异常表达或突变相关^[13]。lncRNA 的功能已经引起了人们广泛的关注。

与人类和动物相比, 植物 lncRNA 研究仍处于起步阶段。全基因组阵列分析和 RNA-seq 结果显示, 植物体内存在大量 lncRNA, 一些通过目标模仿、转录干扰、与多梳蛋白抑制复合体 2 (Polycomb repressive complex 2, PRC2) 有关的组蛋白甲基化和 DNA 甲基化等机制介导基因表达, 在植物开花、雄性不育、营养代谢、生物和非生物胁迫等生物过程中起着调节因子作用。本文主要对植物 lncRNA 的最新研究进展进行了综述。

1 长链非编码 RNA 概述

1.1 长链非编码 RNA 的来源和分类

真核生物基因组中非编码部分所占比例较大, 例如酿酒酵母 (*Saccharomyces cerevisiae*) 为 29%, 拟南芥 (*Arabidopsis thaliana*) 为 71%, 而人类基因组的非编码部分则占 98%^[14]。高等生物体内编码蛋白质的基因数目是相对保守的, 并且一些基因的同源性很高, 例如人类和小鼠有 99% 的蛋白质编码基因都是同源的^[15], 基因组测序发现人类不同个体之间的蛋白质编码序列可变性约为整个基因组的 0.3%^[16]。据上述研究结果推测, 个体间和物种间的差异可能是由于非编码序列的不同导致基因表达的不同而造成的^[15]。

由于限制 ncRNA 进化的因素较少, 导致物种间 lncRNA 序列保守性低、不能形成大的同源基因家族。Ponting 等^[2]认为 lncRNA 可能通过以下 5 种方式形成: (1) 蛋白质编码基因发生阅读框的插入, 插入的阅读框和之前的编码序列形成新的功能性 lncRNA, 如在哺乳动物进化过程中, 与雌性动物肌肉和成骨细胞分化和发育有关的 *Ln timer* 基因的部分序列形成了 X 染色体失活相关的 *Xist* lncRNA, *Xist* 基因启动子区域和 4 个外显子与 *Ln timer* 基因的部分外显子同源, 而其余 6 个外显子则来自不同的转座子^[17~19]; (2) 染色体重排后, 两个不转录的并且之前相隔很远的序列区域并列形成一个多外显子的 lncRNA, 如犬科动物经染色体重排形成了一个含有 ESTs 为 BM537447、C0597044 和 DN744681 的 lncRNA^[2]; (3) 非编码基因通过反转录转座进行复制, 形成有功能的非编码逆基因或非功能性的非编码反转录假基因, 如小鼠的一个 lncRNA 基因家族的形成^[20]; (4) 在 ncRNA 内部出现邻近的序列重复复制, 两个重复序列串联形成新的 lncRNA, 如 *Kcnq1ot1* 的 5' 区域和 *Xist* 部分序列的形成; (5) 插入转座因子产生有功能的 lncRNA, 如分别由不同的转座因子形成的存在于啮齿动物大脑细胞质中的 *BC1* lncRNA 和类人猿大脑细胞质中的 *BC200* lncRNA。

Ponting 等^[2]根据 lncRNA 在基因组中相对于蛋白质编码基因的位置将其分为 5 种类型: (1) 正义或 (2) 反义 lncRNA, 这种 lncRNA 分别与相同链或相反链的另一个蛋白质编码基因的一个或多个外显子相

重叠。小鼠 *Eyfl/2* lncRNA 与相同链的蛋白质编码基因的几个外显子部分序列重叠, *Dlx6as* lncRNA 形成于蛋白质编码基因 *Dlx6* 的反义链上; (3) 双向 (Bidirectional) lncRNA, 它的转录起始位点与相反链上编码蛋白质基因的转录起始位点非常接近, 但转录方向相反。拟南芥中最近发现的 *npc536* lncRNA 形成于编码高尔基体运输复合物相关蛋白基因的反义链上, 两者的转录方向相反^[7]; (4) 内含子 (intronic) lncRNA, 它来源于次级转录物的内含子区域 (有时可能为 mRNA 前体序列); (5) 基因间 (intergenic) lncRNA, 它产生于两个基因之间区域。小鼠 *Eyf2* lncRNA 即在编码基因 *Dlx5* 和 *Dlx6* 之间生成; 酵母的 *SRG1* lncRNA 也产生于 *AIM9* 和 *SER3* 编码基因之间。

1.2 长链非编码 RNA 的功能

lncRNA 起初被认为是基因组转录的“噪音”, 是 RNA 聚合酶 II 转录的副产物, 不具有生物学功能。然而, 研究表明, 许多 lncRNA 具有保守的二级结构及剪接形式^[21-25]。有些 lncRNA 在特异的组织、细胞、发育阶段表达并且表达受到调节。Dinger 等^[21]发现小鼠的胚胎干细胞中多个 lncRNA 有差异表达, 并且具有多能性; Mercer 等^[23]发现成年小鼠大脑不同区域特异表达的多个 lncRNA; Sone 等^[24]发现 *Gomafu* lncRNA 在小鼠神经系统的神经元中特异表达。这些特征预示着 lncRNA 是一种以 RNA 形式存在于生物体内的功能性分子。

lncRNA 功能分析表明, 它们在转录层面上调控编码基因的表达机制很大程度上是由其在基因组上的转录位点决定的。例如, 来源于蛋白质编码位点的顺式-NAT 通常采取“转录干扰”调控目标基因表达, 因为目标基因的启动子和 lncRNA 很接近, 两个启动子转录起始可以被共调节; 同时一些 lncRNA 可以与目标基因启动子 DNA 结合, 形成 RNA-dsDNA 三联体, 阻断转录起始复合物的形成。其他一些 lncRNA 也可以通过控制转录因子的亚细胞定位或抑制 RNA 聚合酶活性在转录水平调节目标基因^[8]。

lncRNA 还可以通过调节 pre-mRNA 的可变剪接、运输、翻译和降解等, 在转录后层面调节编码基因的表达。例如, lncRNA 通过反式作用方式调节目标

mRNA 的稳定性。当 lncRNA 与目标 mRNA 存在碱基互补时, 它就可以与目标 mRNA 形成双链 RNA 复合体, RNA 复合体可以被加工成 endo-siRNA, 使目标 mRNA 降解^[26]。

有些 lncRNA 能够招募染色质重构复合物到特定位点从而介导相关基因的表达沉默, 进而在表观遗传控制基因表达中发挥重要作用, 如 *Xist*、*Air*、*HOTAIR* 和 *COLDAR* 等。在人类基因组中, 来源于 *HOXC* 基因座的 *HOTAIR* 能够招募染色质重构复合物 PRC2 并将其定位到 *HOXD* 位点, 介导 *HOXD* 位点的表观遗传变化^[27]。

Wilusz 等^[28]总结了生物体内 lncRNA 的 8 种作用机制, 包括: (1) 在蛋白质编码基因上游的启动子区域发生转录, 上调或下调下游基因的表达。酵母 *SRG1* lncRNA 在 *SER3* 基因的启动子区域发生转录, 使 *SER3* 的表达受到抑制^[29]; (2) 通过抑制 RNA 聚合酶 II 活性或介导染色质重构以及组蛋白修饰, 影响下游基因表达。酵母中反义 RNA 的表达使 Hda1 组蛋白产生脱乙酰作用, 引起 *PHO84* 基因的表达沉默^[30]; (3) lncRNA 与蛋白质编码基因的转录物形成互补双链, 阻断剪接体对于剪接位点的识别, 从而形成可变剪接转录物。如哺乳动物的 A 型利尿钠肽前体 (NPPA) 的反义链可能与 NPPA mRNA 形成双链 RNA, 影响 NPPA mRNA 的剪接^[31]; (4) 与正义或者反义转录物杂交形成互补双链, 在 Dicer 酶作用下产生内源性的 siRNA, 调控基因的表达水平。在雌性哺乳动物中, 与 X 染色体失活相关的 *Xist* lncRNA 和其反义链转录物 *Tsix* 可形成互补双链, 并被 Dicer 酶加工为 siRNA, 参与 X 染色体失活^[32]; (5) 通过结合到特定蛋白质上, lncRNA 能够调节相应蛋白质的活性。如 *Eyf-2* 可与转录因子 *Dlx2* 形成复合物, 使 *Dlx2* 的活性增强^[33]; (6) 作为 RNA-蛋白质复合体结构的组成成分。Fox 等^[34]发现 *MENε/β* lncRNA 与 RNA 结合蛋白 PSP1、PSP2 和 p54/nrb 共同构成了一种新型核质结构域——paraspeckles^[25, 34-36]; (7) 结合到特定蛋白上, 改变蛋白质在细胞中的定位。如转录因子 NFAT 位于细胞质中, 钙依赖信号使其进入细胞核, 从而激活靶基因的转录, 而当 *NRON* lncRNA 与核质运输有关的蛋白结合后, 抑制了 NFAT 向细胞核的运输^[37, 38]; (8) lncRNA 能被加工产生小分子 RNA, 如 miRNA、piRNA 和其他小 RNA^[39, 40]。

2 植物中长链非编码 RNA 研究

植物在应对环境变化的过程中显示出非凡的发育可塑性。为了最大程度地减小环境的不利影响,植物与环境相互作用,逐渐形成了许多内在生理和外在形态方面的适应对策,出现了不同的植物性状,这些特殊生理现象的形成与其体内一些基因的表达调控息息相关。

植物 ncRNA 的研究起步较晚,且多数研究集中在短链非编码 RNA 上, lncRNA 的识别更是处于起步阶段,因此植物 lncRNA 的研究可能揭示控制植物生长和分化的未知新机制。

2.1 植物长链非编码 RNA 相关的数据库

在 ncRNA 研究过程中,随着 lncRNA 的分子特征、表达模式和功能数据增多,建立了一些针对人类和动物 lncRNA 的数据库,但这些数据库所含的植物 lncRNA 信息非常少。Yang 等^[41]首次通过收集人类、小鼠、狗、鸡、果蝇和线虫不同组织和细胞的 543 个 ChIP-seq 实验数据构建了 ChIPBase 数据库,主要包含了 848 834 个 TF-lncRNA(transcription factor-lncRNA)和 TF-miRNA(transcription factor-miRNA)调控关系,提供了 TF-lncRNA 和 TF-miRNA 调控关系图谱以及转录因子结合位点(TFBSs: transcription factor binding sites)等信息。Mattick 等^[42]于 2009 年通过芯片和原位杂交数据构建了哺乳类动物和人类的 NRED(Noncoding RNA Expression Database)数据库,提供了 lncRNA 的二级结构特征、进化保守性和表达等信息。中国科学院计算技术研究所 Zhao 等^[43]通过人类和小鼠芯片数据构建了包含 73 370 个 lncRNA 表达和功能注释的 NONCODEv3 数据库。2014 年, Zhao 等^[44]进一步收集 RNA-seq 数据,并更新 NONCODEv3 数据库到 NONCODEv4 版本,其中 lncRNA 数量已经扩增到 210 831 个,有 56 018 和 46 475 个 lncRNA 分别来自人类和小鼠。LNCipedia^[45]是一个专门提供人类 lncRNA 序列和结构全面注释的数据库。LncRNABase^[46, 47]提供了 miRNA 调控 lncRNA 的互作信息。LncRNADisease^[48]提供了疾病相关的 lncRNA 信息。

近年随着新的高通量测序技术的开发和运用,植物 lncRNA 的研究已经取得显著成果。拟南芥^[7, 49-52]、苜蓿(*Medicago truncatula*)^[53]、玉米(*Zea mays*)^[54]和

小麦(*Triticum aestivum* L.)^[55]等植物全基因组范围内 ncRNA 的研究中发现了大量 lncRNA。与人类和动物相比,植物 lncRNA 相关的专门数据库非常少,同时一些最新研究成果在公共数据库中并没有进行记录和注释^[56]。

TAIR(The *Arabidopsis* Information Resource)是专门的拟南芥基因数据库,是进行拟南芥生物学和生物模型研究的重要信息来源^[57]。TAIR 记载了拟南芥基因组序列及基因组图谱、各种基因的序列、结构、表达模式和功能注释及详尽的各种代谢途径,同时还有各种拟南芥种子库存数据等信息^[58]。从 2000 ~ 2010 年,TAIR 已经更新了 10 个版本。最新 TAIR10 中记载了拟南芥 478 条 lncRNA 信息,遗憾的是它们的生物学信息没有被记载。

2011 年, Mattick 等^[59]构建了所有真核生物的 lncRNAdb(lncRNAs Database)数据库,2014 年, Quek 等^[60]将其更新为第二版,这个数据库包含 lncRNA 的序列及结构特征、进化保守性、表达、基因组序列、亚细胞定位、功能证据和文献链接等其他相关信息,其中记载了拟南芥(6 个)、水稻(*Oryza sativa* L., 2 个)、大豆(*Glycine max*, 1 个)、苜蓿(3 个)、葡萄(*Vitis vinifera*, 1 个)、芜菁(*Brassica rapa*, 1 个)、番茄(*Lycopersicon esculentum*, 1 个)和杨树(*Populus tremula*, 1 个)共计 16 个 lncRNA 表达情况和可能的功能等相关信息。

PlantNATsDB^[61](Plant Natural Antisense Transcripts DataBase)是首个专门用于预测、查询植物天然反义转录物(Natural antisense transcript, NAT; 一种产生于编码基因或非编码基因反义链的 ncRNA, 参与调节各种生物和非生物胁迫响应过程^[62-64])及其调控功能的数据库,这个数据库大约包含 69 种植物的 200 万个 NAT,其中包括拟南芥的 7788 个 NAT(3005 个正义-NAT, 4783 个反义-NAT)。但是, PlantNATsDB 仅列出所有 NAT 对,不提供全基因组查询。

PLncDB^[56](Plant long non-coding RNA database)是目前唯一公开发表的植物 lncRNA 专用数据库,该数据库收集了 EST 分析、RepTAS 分析、芯片、RNA-seq 分析及 TAIR 中发现的 16 227 个拟南芥 lncRNA 的信息,记载了在不同组织、发育阶段、突变体和应激处理条件下这些 lncRNA 的表达特性,

介绍了编码 lncRNA 的基因座及其侧翼基因组区域的 DNA 甲基化和组蛋白修饰等表观遗传变化和可能的功能,同时也收集了全基因组 siRNA 信息。PLncDB 是目前较为全面的植物 lncRNA 查询数据库。

以上 lncRNA 数据库的相关信息见表 1。

2.2 植物长链非编码 RNA 的预测

植物 lncRNA 的预测方法主要分为计算 RNA 组学方法(Computational RNomics)和实验 RNA 组学方法(Experimental RNomics)^[65](表 2)。计算 RNA 组学方法是指采用计算机与生物学相结合的生物信息分

析(Bioinformatic analyse)进行 lncRNA 预测的方法。计算预测 lncRNA 方法中使用的原始数据可以是公开发表的 cDNA 序列、表达序列标签(EST)、各种全长 cDNA 克隆、tiling arrays 数据、RNA-seq 数据等。首先将这些数据与基因组序列进行比对,去除蛋白质编码基因重叠区域,通过 GeneMark.hmm^[66]、GenScan^[67]、ESTScan2^[68]、ANGLE^[69]和 ORF-Predictor^[70]等软件对剩余序列进行 ORF 预测,根据明显开放阅读框(ORF)存在与否,区分 mRNA 和 ncRNA,保留下来的 ncRNA 序列长度>200 nt 的被认为是 lncRNA。此外,预测 RNA 序列是否具有编码蛋白

表 1 lncRNA 数据库

数据库	lncRNA 数量	数据来源	生物体	网址	说明
ChIPBase	848 834	ChIP-seq	人、小鼠、狗、鸡、果蝇、线虫	http://deepbase.sysu.edu.cn/chipbase/	包括 lncRNA 表达图谱和转录调控全面鉴定和注释,TFBSs、TF-lncRNA 和 TF-miRNA 间调控关系。
lncRNAdb	287	文献	68 种真核生物,包括 16 个植物 lncRNA	http://www.lncrnadb.org/	包括 lncRNA 序列及结构特征、进化保守性、表达、亚细胞定位、功能证据和文献链接等。
NRED	13 213	Microarray、原位杂交	人、小鼠	http://jsm-research.imb.uq.edu.au/nred/	包括 lncRNA 表达信息,进化保守性,二级结构,基因组和反义关系。
NONCODE	210 831	RNA-seq、文献、其他数据库	人、小鼠,少数其他生物	http://www.bioinfo.org/noncode/ http://www.noncode.org/	包括 lncRNA 基本信息,如序列、长度等;生物学信息,如功能、细胞定位等;表达模式,如多组织表达模式和潜在功能。
LNCipedia	21 488	其他数据库、人类 lncRNA 图谱	人	http://www.lncipedia.org	包括 lncRNA 基本信息、蛋白质编码潜力和与 microRNA 结合位点等。
lncRNABase (starBase)	—	CLIP-seq、其他数据库	人、小鼠、线虫	http://starbase.sysu.edu.cn/	提供最全面的 miRNA 与 lncRNA 互作信息。
lncRNADisease	2 044	文献、其他数据库	人	http://cmbi.bjmu.edu.cn/lncrnadisease	提供与疾病相关 lncRNA 注释,开发一种预测与人类疾病相关新 lncRNA 的生物信息学方法。
TAIR10	478	文献、其他数据库、注释基因的计算预测、其他研究机构	拟南芥	http://arabidopsis.org	包括拟南芥基因组序列及基因组图谱,基因序列、结构、表达模式和功能注释及详尽的代谢途径,拟南芥种子库数据等信息。
PlantNATsDB	2 138 498	特定基因组测序计划、基因索引计划、基因表达综合(GEO)数据库	69 种植物	http://bis.zju.edu.cn/pnatdb/	提供 NAT 资源,包括浏览、检索、查看、下载等分析工具,可用于 NAT 预测和功能研究。
PLncDB	16 227	Tilling array、lncRNA array、文献	拟南芥	http://chualab.rockefeller.edu/gbrowse2/homepage.html	提供不同组织、发育阶段、突变体和胁迫处理的 lncRNA 表达特性,编码位点及其侧翼基因组区域表观遗传变化和全基因组 siRNA 信息。

表 2 植物 lncRNA 的预测

方法名称	物种	lncRNA	文献	
计算 RNA 组学方法	生物信息学分析	拟南芥(<i>Arabidopsis thaliana</i>)	<i>npc48</i> 和 <i>npc536</i> 等 76 个 lncRNA	[7]
		拟南芥(<i>Arabidopsis thaliana</i>)	<i>AtR8</i>	[22]
		拟南芥(<i>Arabidopsis thaliana</i>)	6480 个 lincRNA	[78]
		玉米(<i>Zea mays</i>)	1011 个 lncRNA	[54]
		玉米(<i>Zea mays</i>)	1704 个 lncRNA	[82]
比较基因组学	拟南芥(<i>Arabidopsis thaliana</i>)	21 个 ncRNA(12 个 lncRNA)	[52]	
实验 RNA 组学方法	RNA 测序	三角叶杨(<i>Populus trichocarpa</i>)	2542 个 lincRNA 候选基因	[80]
		玉米(<i>Zea mays</i>)	1704 个 lncRNA	[82]
		拟南芥(<i>Arabidopsis thaliana</i>)	15 个 lncNAT 和 20 个 lincRNA	[81]
	cDNA 数据库	玉米(<i>Zea mays</i>)	<i>Zm401</i>	[119]
		黄瓜(<i>Cucumis sativus</i> L.)	<i>CsM10</i>	[88]
		油菜(<i>Brassica campestris</i> L.)	<i>BcMF11</i>	[91]
		拟南芥(<i>Arabidopsis thaliana</i>)	<i>npc48</i> 和 <i>npc536</i> 等 76 个 lncRNA	[7]
	微阵列分析	小麦(<i>Triticum aestivum</i> L.)	125 个 lncRNA	[55]
		拟南芥(<i>Arabidopsis thaliana</i>)	6480 个 lincRNA	[78]
		拟南芥(<i>Arabidopsis thaliana</i>)	37 238 个 NAT 正义-反义对	[79]
	基因组 SELEX	—	—	—

质能力的生物信息学软件(如 CRITICAL^[71]、DIANA-EST^[72]、CSTminer^[73]、CONC^[74]、Coding Potential Calculator^[75]、integrated ncRNA finder^[76]和 RNA-code^[77])也已经被用于动、植物 lncRNA 探索研究中。

实验 RNA 组学方法是通过 RNA-seq、构建 cDNA 数据库、微阵列分析和基因组 SELEX 等发现 lncRNA^[65]。高通量测序技术特别是 RNA-seq 技术是发现 lncRNA 的有效方法,可以直接、快速地发现低丰度、新的 lncRNA。目前研究中更广泛采用了计算与实验相结合的方法,这种结合可以更准确、有效地预测 lncRNA。

Ben 等^[7]通过对拟南芥基因组全长 cDNA 数据库进行生物信息学分析,发现了 76 个 ncRNA 基因,其中 14 个为 NAT,5 个是 siRNA 前体,22 个 lncRNA 与非生物应激应答相关,其中 2 个 lncRNA 在非生物应激环境中异常表达影响了拟南芥的生长和分化。

最近, Liu 等^[78]开发了 RepTAS(Reproducibility-based tiling array analysis strategy)生物信息学方法,并对公开发表的 200 个拟南芥 tiling array 数据进行了全基因组范围综合分析,发现了 6480 个基因间长链非编码 RNA(Long intergenic non-coding RNA, lincRNA),后续的 RNA-seq 结果表明其中的

2708 个表现出组织特异性和非生物胁迫应答特性。Wang 等^[79]利用 RepTAS 生物信息学方法分析了公开发表的来自拟南芥不同器官、组织、不同激素处理、生物和非生物胁迫的 200 个 tiling array 数据,发现大量的新 lncRNA。这些 lncRNA 平均长度约 731nt,包括基因内转录物和 37 238 个 NAT 正义-反义对,实验分析证明大量光应答 NAT 对与组蛋白修饰有关。

Xin 等^[55]利用微阵列分析和高通量 SBS 测序相结合的方法确定了小麦中 125 个 lncRNA(其中部分为小 RNA 前体),它们具有组织特异性表达特征并对白粉病和热胁迫产生应激应答反应。

Shuai 等^[80]采用 RNA 测序方法在三角叶杨(*Populus trichocarpa*)中识别出 2542 个 lincRNA 候选基因,其中 504 个对干旱胁迫产生应激应答。

为了研究拟南芥对尖孢镰刀菌感染的应激应答反应, Zhu 等^[81]对来自感染细菌的拟南芥 RNA 进行深度测序,发现 15 个长链非编码 NAT 和 20 个新的 lincRNA 的表达受尖孢镰刀菌感染的调节,部分基因对真菌感染产生防御作用,表明某些 lncRNA 是拟南芥中抗真菌网络中的重要节点。

Wu 等^[22]侧重于 Pol III 的 III 型启动子结构特点,

对拟南芥基因组进行计算检索,发现了 20 种可能的 ncRNA 候选基因,通过体外无细胞转录系统活性检测,发现了 *AtR8* lncRNA(约 260nt),其在拟南芥幼苗中表现了低氧应激应答特性。

Boerner 等^[54]开发并优化了 Python Pipeline 计算机语言,采用 Python Pipeline 与 SVM(Support Vector Machine)相结合的生物信息分析方法筛选了玉米全长 cDNA 序列,发现 2492 个 lncRNA 候选基因,其中 19 个为 miRNA 前体,237 个为 shRNA(short hairpin)前体,1225 个 siRNA 前体,1011 个为 lncRNA,这些 lncRNA 中 572 个位于编码基因内,439 个位于基因间区域。

Li 等^[82]对已经发表的 30 个不同实验组玉米 EST 数据和玉米全基因组序列注释及 RNA-seq 数据进行生物信息学分析,首次发现玉米中的 1704 个 lncRNA,平均长度为 463nt,表达分析显示近 50% 的 lncRNA 具有组织特异性,其中一些来自于 *Tb1*、*Vgt1* 和 *B1*(与玉米生长发育和农业性状相关的关键基因)的调控区域。

研究表明,一些重要的 ncRNA 在许多生物基因组中是保守的。基于序列保守性的比较基因组学方法也是预测 ncRNA 的生物信息学方法^[52, 83]。Song 等^[52]运用比较基因组学方法比对了拟南芥、水稻、三角叶杨、木瓜(*Carica papaya*)、葡萄的全基因组序列,确定了拟南芥中 16 个基因家族的 21 个新 ncRNA 基因。

基因组 SELEX 方法已成功运用在 mRNA 结合蛋白筛选中,但 ncRNA 研究中的应用目前未见报道^[84]。

关于植物 lncRNA 的预测总结于表 2。

2.3 植物长链非编码 RNA 的表达及可能的生物学功能

2.3.1 植物长链非编码 RNA 的转录

真核生物细胞核中的 RNA 聚合酶分为三类: RNA 聚合酶 (Pol I: RNA polymerase I)、RNA 聚合酶 (Pol II: RNA polymerase II)和 RNA 聚合酶 (Pol III: RNA polymerase III)。Pol I 位于核仁中,合成 rRNA,如 28S rRNA、18S rRNA 和 5.8S rRNA,这些 rRNA 是细胞中主要的核糖核酸之一,与多种核糖体蛋白质共同构成核糖体,在蛋白质合成过程中起重要作用。除上述几种持家非编码 RNA

外,目前还没有发现 Pol I 合成的其他 ncRNA。Pol II 位于核质中,除合成编码蛋白质的 mRNA 外,还合成许多 ncRNA,如 snRNA、snoRNA 和 microRNA。此外,一些 lncRNA 也由 Pol II 合成,如大豆、苜蓿和水稻中保守的 *ENOD40*^[85-87]、黄瓜(*Cucumis sativus* L.)中的 *CsM10*^[88]、玉米中的 *Zm401*^[89]、拟南芥中的 *At4*^[90]、*npc48* 及 *npc536*^[71]和油菜(*Brassica campestris* L.)中的 *BcMF11*^[91]等。Pol III 位于核质中,合成 tRNA、5S rRNA、snoRNA 及 microRNA 等 ncRNA。最近研究发现,Pol III 也合成调节编码基因表达的新 ncRNA^[92]。另外拟南芥中 Pol III 的 III 型启动子合成了低氧应激应答型 *AtR8* lncRNA^[22]。Wierzbicki 等^[93]发现,拟南芥中 Pol II 进化而来的 RNA 聚合酶 IV 和 V(Pol IV/V)参与了短链非编码 RNA 的合成;同时 Pol V 也合成一类拟南芥 lncRNA,它们可以和一些抑制染色质修饰的沉默因子结合导致基因沉默^[94]。此外,拟南芥和玉米中 Pol IV 和 Pol V 可能也合成一类与 RNA 介导的 DNA 甲基化有关的 lncRNA^[54, 95]。

2.3.2 植物长链非编码 RNA 的表达及可能的生物学功能

lncRNA 的研究成果主要集中在人类和动物中。近年,随着对植物 ncRNA 研究的不断深入,发现了一些植物特有的 lncRNA。

TPSII/Mt4 是一个磷酸盐饥饿条件下表达被高度诱导的 lncRNA 家族^[8]。这类 lncRNA 除了具有短的、不保守的开放阅读框外,还具有 23nt 长度的保守核苷酸序列,根据上述特征推测它们可能作为核糖核酸调节子或编码一些信号肽在生物体内起作用^[96, 97]。番茄的 *TPSII* 是 *TPSII/Mt4* 家族中最早被发现的成员,其在基因组中为单拷贝,并且磷酸盐饥饿条件下在根和叶中被迅速诱导,表明 *TPSII* 可能与番茄中早期磷酸盐饥饿反应机制有关^[98]。苜蓿 *Mt4*^[99-101] 序列中含有许多短的开放阅读框,其中一个开放阅读框的部分序列与番茄中 *TPSII* 的开放阅读框重合。在磷酸盐饥饿条件下 *Mt4* 在苜蓿的根中被大量诱导,但在磷酸盐充足的条件下 *Mt4* 几乎不被转录。*OsPII* 是在水稻中发现的 *TPSII/Mt4* 家族成员,磷酸盐饥饿时 *OsPII* 在根和叶部被诱导表达,但在根中的诱导尤为显著^[102]。*AtIPSI*(Induced by phosphate star-

vation 1)是拟南芥中发现的又一个 *TPSII/Mt4* 家族成员,在高度磷酸盐饥饿条件下,*AtIPS1* 在叶和根部大量积累^[96]。*miR399* 是拟南芥中同样被磷酸盐饥饿诱导积累于地上部分和根中的 miRNA,它通过与 *PHO2* 的 5'UTR 区域的部分序列碱基互补配对介导 *PHO2*(*PHOSPHATE 2*)mRNA 的断裂,在转录后水平降低 *PHO2* 的表达^[97,103]。研究发现,*AtIPS1* 序列中 23 个核苷酸保守区域与 *miR399* 的部分序列也存在碱基互补,*AtIPS1* 通过模仿 *miR399* 的靶基因 *PHO2*,与 *PHO2* 竞争结合 *miR399*,抑制 *miR399* 的活性,使 *PHO2* mRNA 的积累增加,维持磷酸盐饥饿条件下拟南芥的正常生长^[96,97]。*AtIPS1* 通过“目标模仿”机制抑制 miRNA 的活性^[8],是首个明确作用机制的 *TPSII/Mt4* 家族成员。*AtIPS1* 作用机制表明,一些 lncRNA 可以通过自身部分序列与功能性 miRNA 碱基互补配对,从而模仿 miRNA 的靶基因,减弱 miRNA 的调节因子作用,在生物学过程中发挥重要作用。这为今后未知功能的 lncRNA 作用靶元件预测提供了一个新思路。

染色质修饰对动、植物发育过程中组织特异性基因的表达和基因组重组具有重要作用。lncRNA 和染色质重构复合物之间的相互作用是动物基因表观遗传沉默的普遍机制^[104]。植物开花调控基因 *FLC*(*FLOWERING LOCUS C*)可以通过染色质修饰影响植物开花时间。研究表明,冷处理诱导 *VIN3*(*Vernalization insensitive 3*)的表达,*VIN3* 可以招募染色质重构复合物 PRC2 到 *FLC* 基因座产生抑制型 H3K27me3 组蛋白修饰,从而改变 *FLC* 染色质活性状态,引起 *FLC* 基因的表观遗传沉默,调控植物开花时间^[8]。Swiezewski 等^[105] 采用 Single Nucleotide Resolution Array 技术对 *FLC* 双链及侧翼 50 kb 区域进行研究,发现了多个来自 *FLC* 基因座的冷诱导反义转录物,即 *COOLAIR*(Cold induced antisense intragenic RNA)。*COOLAIR* 具有 5'端加帽及 3'端聚腺苷酸化结构,并可进行剪接。*COOLAIR* 的冷应激反应比 *VIN3* 发生得更早,并且 *COOLAIR* 表达不受 *VIN3* 缺失影响。冷诱导条件下 *COOLAIR* 对 *FLC* 的表达有抑制作用,但环境条件由寒冷变为温暖时,*COOLAIR* 的抑制作用变得可逆,这表明 *COOLAIR* 只能暂时抑制 *FLC* 表达,而 *FLC* 基因失活的表观遗传模式是调控植物开花的必要途径。因为 *COOLAIR*

和 *FLC* 的表达水平呈相反趋势,并且在 *COOLAIR* 发生转录情况下,*FLC* 基因座上的 Pol II 含量减少,这意味着 *COOLAIR* 通过转录干扰的方式抑制 *FLC* 表达。所以冷胁迫下 *FLC* 的表达调控可能通过下面模式进行:冷胁迫诱导 *COOLAIR* 的表达,瞬时抑制 *FLC* 的转录,进一步冷胁迫处理诱导了 *VIN3* 的表达,*VIN3* 招募染色质重构复合物 PRC2 到 *FLC* 基因上,产生抑制型 H3K27me3 组蛋白修饰,引起 *FLC* 基因的表观遗传沉默,进而调控植物开花^[8]。另外,新的研究表明 *COOLAIR* 不是抑制 *FLC* 表达所必须的 ncRNA^[106]。

COLD AIR(Cold assisted intronic noncoding RNA)是来自拟南芥 *FLC* 基因第一个内含子的正义转录物,是不同于 *COOLAIR* 的又一类 Pol II 转录的 lncRNA^[107]。*COLD AIR* 长约 1100nt,具有 Pol IV 和 V 转录的 lncRNA 结构特点,如 5'端加帽,但 3'端没有聚腺苷酸化等特征。冷处理诱导 *COLD AIR* 表达,冷处理 20 d 后 *COLD AIR* 的表达达到峰值,冷处理 30 d 后其表达降到冷处理前水平。虽然 *COLD AIR* 的表达模式和 *COOLAIR* 相似,但它的表达峰值比 *COOLAIR* 滞后 10 d。研究表明,*COLD AIR* 缺失型拟南芥植株表现出冷处理后开花延迟现象,这可能是由于 *COLD AIR* 的缺失导致 *FLC* 基因的染色质没有富集抑制型 H3K27me3 组蛋白引起的。RNA 结合实验也进一步表明,*COLD AIR* 特异地与 CLF(*Curly Leaf*,与 H3K27me3 修饰有关的 PRC2 复合体的重要组成成分)相互作用。因此,冷处理条件下,*COLD AIR* 可能通过募集 PRC2 到 *FLC* 染色质引起表观遗传修饰,构建了 *FLC* 基因沉默机制,即通过 Polycomb-依赖模式负调控 *FLC* 的表达。*COLD AIR* 是首次发现的参与基因染色质表观遗传学修饰的植物 lncRNA^[8]。

Nongken 58S(NK 58S)是水稻 Nongken 58N(NK 58N)的自然突变品种,其光敏感引起雄性不育(Photoperiod-sensitive male sterility, PSMS)。NK 58S 在长日照条件下花粉完全不育,短日照则不会引起不育。Ding 等^[108]采用位置克隆方法(Position cloning strategy)发现长约 1236nt 的 lncRNA,即 *LDMAR*(Long-day-specific male-fertility-associated RNA)。长日照条件下,NK 58N 和 NK 58S 花粉的发育需要大量 *LDMAR*,而调查发现长日照条件下 NK 58S

LDMAR 的积累量少于 *NK 58N LDMAR*, 并且 *NK 58S LDMAR* 启动子区域甲基化程度高于 *NK 58N LDMAR*。进一步研究表明, 这个现象是由于 *NK 58S LDMAR* 序列的一个碱基 C 突变成 G 引起了 *LDMAR* 二级结构产生变化, 导致 *LDMAR* 启动子区域的甲基化, 抑制了 *NK 58S LDMAR* 表达, 使长日照条件下 *NK 58S* 雄花不育。同时, 来自 *LDMAR* 启动子区域正义链的 siRNA, 即 *Psi-LDMAR* 也对 *LDMAR* 启动子区域的甲基化起调节作用^[109]。Zhou 等^[110]证明 *LDMAR* 中 136nt 长度的小 RNA 是引起 PSMS 的重要序列。遗憾的是, *LDMAR* 基因和小 RNA 是否调控雄性花粉不育目前仍未阐明^[111]。

GmENOD40 是最初在豆科植物中发现的根瘤特异性表达并与共生固氮根瘤形成有关的 lncRNA^[86,112], 之后分别在水稻及苜蓿中发现了 *GmENOD40* 的同源基因 *OsENOD40*^[87]和 *MtENOD40*^[113]。*OsENOD40* 在水稻茎中特异表达, 对水稻的器官分化和维管束组织发育发挥关键作用^[87]。酵母三杂交实验发现苜蓿中 *MtENOD40* 与一种持续表达的 RNA 结合蛋白 MtRBP1 产生相互作用, 使细胞质中 MtRBP1 产生移位, 这种 RNA 结合蛋白的亚细胞定位改变可能代表 lncRNA 在细胞中的某种新功能^[85]。最近, 酵母三杂交实验进一步发现 *MtENOD40* 与两个结瘤素酸性 RNA 结合蛋白 MtSNARP1 和 MtSNARP2 相互作用^[114]。上述结果表明, 酵母三杂交技术可以有效应用于与 lncRNA 相互作用的生物大分子筛选研究。番茄 *ENOD40-GUS* 转基因植株的 GUS 表达模式显示, 番茄中 *ENOD40* 可能与消除乙烯的刺激反应有关^[115]。此外, *ENOD40* 在大豆中编码 12 和 24 个氨基酸长度的短肽, 这些短肽特定地与蔗糖合成酶结合, 表明 *ENOD40* 与根瘤中蔗糖利用率的调节有关。苜蓿 *ENOD40* 编码 13 和 27 个氨基酸长度的短肽, 这两个短肽与 *ENOD40* 生物活性相关^[116,117]。但是, *ENOD40* 的作用机制可能只是通过 *ENOD40* 的 RNA 分子本身而不是短肽起作用^[27]。

CsM10(1517nt) 是在黄瓜中发现的一种 lncRNA^[88], 其在黄瓜的不同组织、不同发育阶段和不同光照周期都有差异表达, 可能与黄瓜的雌雄分化有关。以前的研究表明, *CR20*、*GUT15* 和 *AtCR20-1*

这 3 种 ncRNA 是生物胁迫信号中至关重要的基因^[118]。*CsM10* 与 *CR20*、*GUT15* 和 *AtCR20-1* 有 179nt 的保守区域, 推测 *CsM10* 可能调节某些生物应激应答基因的表达。

Dai 等^[119]发现玉米中 *Zm401*(1149nt) 是一种花粉特异性表达的 lncRNA, 在单子叶植物中有高度保守的序列和稳定的 RNA 二级结构。*Zm401* 能调节 *ZmMADS2*、*MZm3-3*、*ZmC5* 这 3 种与玉米花粉发育有关的关键基因的表达, 抑制或者过表达该基因都会影响雄蕊和花药的的生长发育, 引起玉米小孢子空瘪、变形, 绒毡层出现退化延迟, 使花粉形态异常, 可用于玉米雄性不育植株的制作^[89]。在烟草 (*Nicotiana tabacum* L.) 的研究中也得到了相似的结果, 烟草中 *Zm401* 的异常表达扰乱了花的生长发育, 约有 80% 转基因烟草的花蕾在开花前花轴变黄, 最终大多数脱落^[120]。

Song 等^[91]发现了在油菜花粉发育的各个阶段持续表达的花粉特异基因 *BcMF11*(828nt)。在 *BcMF11* 转基因油菜中, 雌花发育正常, 而雄花产生不育的花粉。研究结果显示过表达植株的绒毡层和小孢子发育异常, 使花药发育不全, 颜色变为褐色或白色, 形态干瘪, 甚至不产生花粉, 引起油菜雄花不育^[91,121]。

Ben 等确定了拟南芥中 76 个 lncRNA^[7]: 14 个为蛋白质编码基因的反义转录物, 它们可能起着反式调控作用; 5 个是 24nt siRNA 的前体; *npc83* 经 DCL4 加工后形成 *miR869A*; 22 个 lncRNA 受非生物胁迫调节。生物学功能研究发现, *npc48* 过量表达改变了 *miR164* 积累并导致拟南芥出现锯齿状叶和开花延迟现象; 盐胁迫下 *npc536* 过量表达调节了拟南芥根的生长。

Wu 等^[22]用计算检索方法发现了 *Pol III* 合成的 *AtR8* lncRNA, 其在拟南芥根中及拟南芥由来的 MM2d 细胞的细胞质中特异地大量表达, 且在甘蓝 (*Brassica oleracea*) 和萝卜 (*Raphanus sativus* L.) 等十字花科植物中发现了它的同源物。低氧应激处理抑制 *AtR8* 表达, 表明它是低氧应激应答型 lncRNA。目前, *AtR8* 的作用机制尚不清楚。

植物中的 lncRNA 总结于表 3。

表 3 植物中的 lncRNA

名称	物种	长度(nt)	聚合酶转录	功能	可能的调节机制	文献
—	—	—	Pol I	—	—	—
<i>GmENOD40</i>	大豆(<i>Glycine max</i>)	700	Pol I	与根瘤形成有关	—	[112]
<i>MtENOD40</i>	苜蓿(<i>Medicago truncatula</i>)	700		与根瘤形成有关	—	[113]
<i>TPS11</i>	番茄(<i>Lycopersicon esculentum</i>)	474		与根中磷酸盐摄取有关	—	[98]
<i>Mt4</i>	苜蓿(<i>Medicago truncatula</i>)	约 500		与根中磷酸盐摄取有关	—	[99, 100]
<i>OsENOD40</i>	水稻(<i>Oryza sativa</i> L.)	约 640		与根瘤形成有关	—	[87]
<i>AtIPS1</i>	拟南芥(<i>Arabidopsis thaliana</i>)	542	Pol II	与根中磷酸盐摄取有关	目标模仿, 影响 miRNA 功能	[96, 97]
<i>OsP11</i>	水稻(<i>Oryza sativa</i> L.)	375		与根中磷酸盐摄取有关	—	[102]
<i>Zm401</i>	玉米(<i>Zea mays</i>)	1149		与雄花不育有关	—	[119]
<i>CsM10</i>	黄瓜(<i>Cucumis sativus</i> L.)	1517		与雌雄分化有关	—	[88]
<i>BcMF11</i>	油菜(<i>Brassica campestris</i> L.)	828		与雄花不育有关	—	[91]
<i>COOLAIR</i>	拟南芥(<i>Arabidopsis thaliana</i>)	—	Pol III	与开花时间有关	启动子转录干扰	[105]
<i>COLDAIR</i>	拟南芥(<i>Arabidopsis thaliana</i>)	约 1100		与开花时间有关	H3K27me3 组蛋白修饰	[107]
<i>LDMAR</i>	水稻(<i>Oryza sativa</i> L.)	1236		光敏感引起雄性不育	基因启动子 DNA 甲基化	[108]
<i>AtR8</i>	拟南芥(<i>Arabidopsis thaliana</i>)	260		根中特异表达, 与低氧应激应答有关	—	[22]
lncRNA	拟南芥(<i>Arabidopsis thaliana</i>)、玉米(<i>Zea mays</i>)	—		与 RNA 介导的 DNA 甲基化有关	—	[95]
lncRNA	拟南芥(<i>Arabidopsis thaliana</i>)、玉米(<i>Zea mays</i>)	—	Pol V	与 RNA 介导的 DNA 甲基化有关	—	[95]
lncRNA	拟南芥(<i>Arabidopsis thaliana</i>)	—		与基因沉默有关	和抑制染色质修饰的沉默因子结合	[94]

3 结语与展望

ncRNA 具有明确的生理功能, 但却不编码任何蛋白质, 仅以 RNA 分子形式发挥作用。从 ncRNA 在不同生物体细胞的所有转录物中所占的比例可以看出, RNA 的数量和多样性似乎与物种的复杂性密切相关, 其普遍性及重要性远远超出人们的想象, 在生物体的多种生命活动中起着至关重要的作用。到目前为止, 人们只看到部分 ncRNA 所具有的功能, 如果能进一步确定 ncRNA 的结构特征将会有更大的发现。

随着被发现的 ncRNA 增多, 越来越多的相关研究涉及到植物 lncRNA, 但目前植物 lncRNA 的研究仍处在初级阶段, 且研究主要集中在拟南芥、水稻、小麦和玉米这几种植物中, 已经发现的部分 lncRNA 不能被确定是功能性转录物还是转录“噪音”, 未知的 lncRNA 的详细数量仍不清楚。同时, 预测植物 lncRNA 作用靶元件的生物信息学工具相对较少,

这为准确推测其可能的作用机制增加了难度。目前, 已经证明了某些 lncRNA 是 microRNA 的前体, 部分 lncRNA 可以调节 microRNA 或蛋白质编码基因的功能, 但对其详细作用机制所知甚少, 而且不同 lncRNA 适用的研究方法很少可以相互借鉴。这些因素限制了植物 lncRNA 研究进展。在以后的研究中, 有必要开发针对植物 lncRNA 的研究方法, 从而更好地预测植物中 lncRNA 和其作用靶目标, 识别更多符合已知功能模式或未知功能模式的 lncRNA, 并阐明其作用机制, 为植物 ncRNA 研究提供新思路。

参考文献

[1] Eddy SR. Non-coding RNA genes and the modern RNA world. *Nat Rev Genet*, 2001, 2(12): 919–929. [\[DOI\]](#)

[2] Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. *Cell*, 2009, 136(4): 629–641. [\[DOI\]](#)

[3] Stoughton RB. Applications of DNA microarrays in biology. *Annu Rev Biochem*, 2005, 74: 53–82. [\[DOI\]](#)

- [4] Carthew RW, Sontheimer EJ. Origins and Mechanisms of miRNAs and siRNAs. *Cell*, 2009, 136(4): 642–655. [\[DOI\]](#)
- [5] Malone CD, Hannon GJ. Small RNAs as guardians of the genome. *Cell*, 2009, 136(4): 656–668. [\[DOI\]](#)
- [6] Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nat Rev Genet*, 2009, 10(3): 155–159. [\[DOI\]](#)
- [7] Ben Amor B, Wirth S, Merchan F, Laporte P, d'Aubenton-Carafa Y, Hirsch J, Maizel A, Mallory A, Lucas A, Deragon JM, Vaucheret H, Thermes C, Crespi M. Novel long non-protein coding RNAs involved in Arabidopsis differentiation and stress responses. *Genome Res*, 2009, 19(1): 57–69. [\[DOI\]](#)
- [8] Zhang J, Mujahid H, Hou YX, Nallamilli BR, Peng ZH. Plant long ncRNAs: a new frontier for gene regulatory control. *American J Plant Sci*, 2013, 4: 1038–1045. [\[DOI\]](#)
- [9] Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, Stadler PF, Hertel J, Hackermüller J, Hofacker IL, Bell I, Cheung E, Drenkow J, Dumais E, Patel S, Helt G, Ganesh M, Ghosh S, Piccolboni A, Sementchenko V, Tammana H, Gingeras TR. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science*, 2007, 316(5830): 1484–1488. [\[DOI\]](#)
- [10] Prasanth KV, Spector DL. Eukaryotic regulatory RNAs: an answer to the 'genome complexity' conundrum. *Genes Dev*, 2007, 21(1): 11–42. [\[DOI\]](#)
- [11] Amaral PP, Dinger ME, Mercer TR, Mattick JS. The eukaryotic genome as an RNA machine. *Science*, 2008, 319(5871): 1787–1789. [\[DOI\]](#)
- [12] Cao XW, Yeo G, Muotri AR, Kuwabara T, Gage FH. Noncoding RNAs in the mammalian central nervous system. *Annu Rev Neurosci*, 2006, 29: 77–103. [\[DOI\]](#)
- [13] Wapinski O, Chang HY. Long noncoding RNAs and human disease. *Trends Cell Biol*, 2011, 21(6): 354–361. [\[DOI\]](#)
- [14] 郑晓飞. 非编码 RNA. 北京: 化学工业出版社, 2008: 2–3. [\[DOI\]](#)
- [15] Mattick JS. Non-coding RNAs: the architects of eukaryotic complexity. *EMBO Rep*, 2001, 2(11): 986–991. [\[DOI\]](#)
- [16] Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huseon DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang JH, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng ZM, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan WN, Ge WM, Gong FC, Gu ZP, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke ZX, Ketchum KA, Lai ZW, Lei YD, Li ZY, Li JY, Liang Y, Lin XY, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun JT, Wang ZY, Wang AH, Wang X, Wang J, Wei MH, Wides R, Xiao CL, Yan CH, Yao A, Ye J, Zhan M, Zhang WQ, Zhang HY, Zhao Q, Zheng LS, Zhong F, Zhong WY, Zhu SC, Zhao SY, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An HJ, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkuch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigó R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays A, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu XJ, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu XH. The sequence of the human genome. *Science*, 2001, 291(5507): 1304–1351. [\[DOI\]](#)
- [17] Duret L, Chureau C, Samain S, Weissenbach J, Avner P.

- The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. *Science*, 2006, 312(5780): 1653–1655. [\[DOI\]](#)
- [18] Elisaphenko EA, Kolesnikov NN, Shevchenko AI, Rogozin IB, Nesterova TB, Brockdorff N, Zakian SM. A dual origin of the Xist gene from a protein-coding gene and a set of transposable elements. *PLoS One*, 2008, 3(6): e2521. [\[DOI\]](#)
- [19] Flynn M, Saha O, Young P. Molecular evolution of the LNX gene family. *BMC Evol Biol*, 2011, 11: 235. [\[DOI\]](#)
- [20] Ravasi T, Suzuki H, Pang KC, Katayama S, Furuno M, Okunishi R, Fukuda S, Ru K, Frith MC, Gongora MM, Grimmond SM, Hume DA, Hayashizaki Y, Mattick JS. Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. *Genome Res*, 2006, 16(1): 11–19. [\[DOI\]](#)
- [21] Dinger ME, Amaral PP, Mercer TR, Pang KC, Bruce SJ, Gardiner BB, Askarian-Amiri ME, Ru K, Soldà G, Simons C, Sunkin SM, Crowe ML, Grimmond SM, Perkins AC, Mattick JS. Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res*, 2008, 18(9): 1433–1445. [\[DOI\]](#)
- [22] Wu J, Okada T, Fukushima T, Tsudzuki T, Sugiura M, Yukawa Y. A novel hypoxic stress-responsive long non-coding RNA transcribed by RNA polymerase III in Arabidopsis. *RNA Biol*, 2012, 9(3): 302–313. [\[DOI\]](#)
- [23] Mercer TR, Dinger ME, Sunkin SM, Mehler MF, Mattick JS. Specific expression of long noncoding RNAs in the mouse brain. *Proc Natl Acad Sci USA*, 2008, 105(2): 716–721. [\[DOI\]](#)
- [24] Sone M, Hayashi T, Tarui H, Agata K, Takeichi M, Nakagawa S. The mRNA-like noncoding RNA Gomafu constitutes a novel nuclear domain in a subset of neurons. *J Cell Sci*, 2007, 120(Pt 15): 2498–2506. [\[DOI\]](#)
- [25] Clemson CM, Hutchinson JN, Sara SA, Ensminger AW, Fox AH, Chess A, Lawrence JB. An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Mol Cell*, 2009, 33(6): 717–726. [\[DOI\]](#)
- [26] Golden DE, Gerbasi VR, Sontheimer EJ. An inside job for siRNAs. *Mol Cell*, 2008, 31(3): 309–312. [\[DOI\]](#)
- [27] Zhu QH, Wang MB. Molecular functions of long non-coding RNAs in plants. *Genes (Basel)*, 2012, 3(1): 176–190. [\[DOI\]](#)
- [28] Wilusz JE, Sunwoo H, Spector DL. Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev*, 2009, 23(13): 1494–1504. [\[DOI\]](#)
- [29] Martens JA, Laprade L, Winston F. Intergenic transcription is required to repress the *Saccharomyces cerevisiae* SER3 gene. *Nature*, 2004, 429(6991): 571–574. [\[DOI\]](#)
- [30] Camblong J, Iglesias N, Fickentscher C, Dieppois G, Stutz F. Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in *S. cerevisiae*. *Cell*, 2007, 131(4): 706–717. [\[DOI\]](#)
- [31] Annilo T, Kepp K, Laan M. Natural antisense transcript of natriuretic peptide precursor A (NPPA): structural organization and modulation of NPPA expression. *BMC Mol Biol*, 2009, 10: 81. [\[DOI\]](#)
- [32] Ogawa Y, Sun BK, Lee JT. Intersection of the RNA interference and X-inactivation pathways. *Science*, 2008, 320(5881): 1336–1341. [\[DOI\]](#)
- [33] Feng JC, Bi CM, Clark BS, Mady R, Shah P, Kohtz JD. The *Evf-2* noncoding RNA is transcribed from the *Dlx-5/6* ultraconserved region and functions as a *Dlx-2* transcriptional coactivator. *Genes Dev*, 2006, 20(11): 1470–1484. [\[DOI\]](#)
- [34] Fox AH, Lam YW, Leung AKL, Lyon CE, Andersen J, Mann M, Lamond AI. Paraspeckles: a novel nuclear domain. *Curr Biol*, 2002, 12(1): 13–25. [\[DOI\]](#)
- [35] Sasaki YTF, Ideue T, Sano M, Mituyama T, Hirose T. MENε/β noncoding RNAs are essential for structural integrity of nuclear paraspeckles. *Proc Natl Acad Sci U S A*, 2009, 106(8): 2525–2530. [\[DOI\]](#)
- [36] Sunwoo H, Dinger ME, Wilusz JE, Amaral PP, Mattick JS, Spector DL. MEN epsilon/beta nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. *Genome Res*, 2009, 19(3): 347–359. [\[DOI\]](#)
- [37] Hogan PG, Chen L, Nardone J, Rao A. Transcriptional regulation by calcium, calcineurin, and NFAT. *Genes Dev*, 2003, 17(18): 2205–2232. [\[DOI\]](#)
- [38] Willingham AT, Orth AP, Batalov S, Peters EC, Wen BG, Aza-Blanc P, Hogenesch JB, Schultz PG. A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. *Science*, 2005, 309(5740): 1570–1573. [\[DOI\]](#)
- [39] Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B, Bartel DP. MicroRNAs in plants. *Genes Dev*, 2002, 16(13): 1616–1626. [\[DOI\]](#)
- [40] Hirsch J, Lefort V, Vankerschaver M, Boualem A, Lucas A, Thermes C, d'Aubenton-Carafa Y, Crespi M. Characterization of 43 non-protein-coding mRNA genes in *Arabidopsis*, including the *MIR162a*-derived transcripts. *Plant Physiol*, 2006, 140(4): 1192–1204. [\[DOI\]](#)
- [41] Yang JH, Li JH, Jiang S, Zhou H, Qu LH. ChIPBase: a database for decoding the transcriptional regulation of long non-coding RNA and microRNA genes from ChIP-

- Seq data. *Nucleic Acids Res*, 2013, 41(Database issue): D177–D187. [\[DOI\]](#)
- [42] Dinger ME, Pang KC, Mercer TR, Crowe ML, Grimmond SM, Mattick JS. NRED: a database of long noncoding RNA expression. *Nucleic Acids Res*, 2009, 37(Database issue): D122–D126. [\[DOI\]](#)
- [43] Bu D, Yu K, Sun S, Xie C, Skogerbø G, Miao R, Xiao H, Liao Q, Luo H, Zhao G, Zhao H, Liu Z, Liu C, Chen R, Zhao Y. NONCODE v3.0: integrative annotation of long noncoding RNAs. *Nucleic Acids Res*, 2012, 40(Database issue): D210–D215. [\[DOI\]](#)
- [44] Xie C, Yuan J, Li H, Li M, Zhao G, Bu D, Zhu W, Wu W, Chen R, Zhao Y. NONCODEv4: exploring the world of long non-coding RNA genes. *Nucleic Acids Res*, 2014, 42(Database issue): D98–D103. [\[DOI\]](#)
- [45] Volders PJ, Helsens K, Wang XW, Menten B, Martens L, Gevaert K, Vandesompele J, Mestdagh P. LNCipedia: a database for annotated human lncRNA transcript sequences and structures. *Nucleic Acids Res*, 2013, 41(Database issue): D246–D251. [\[DOI\]](#)
- [46] Li JH, Liu S, Zhou H, Qu LH, Yang JH. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res*, 2014, 42(Database issue): D92–D97. [\[DOI\]](#)
- [47] Yang JH, Li JH, Shao P, Zhou H, Chen YQ, Qu LH. starBase: a database for exploring microRNA-mRNA interaction maps from Argonaute CLIP-Seq and Degradome-Seq data. *Nucleic Acids Res*, 2011, 39(Database issue): D202–D209. [\[DOI\]](#)
- [48] Chen G, Wang ZY, Wang DQ, Qiu CX, Liu MX, Chen X, Zhang QP, Yan GY, Cui QH. LncRNADisease: a database for long-non-coding RNA-associated diseases. *Nucleic Acids Res*, 2013, 41(Database issue): D983–D986. [\[DOI\]](#)
- [49] MacIntosh GC, Wilkerson C, Green PJ. Identification and analysis of Arabidopsis expressed sequence tags characteristic of non-coding RNAs. *Plant Physiol*, 2001, 127(3): 765–776. [\[DOI\]](#)
- [50] Marker C, Zemann A, Terhøst T, Kiefmann M, Kastenmayer JP, Green P, Bachellerie JP, Brosius J, Hüttenhofer A. Experimental RNomics: identification of 140 candidates for small non-messenger RNAs in the plant Arabidopsis thaliana. *Curr Biol*, 2002, 12(23): 2002–2013. [\[DOI\]](#)
- [51] Rymarquis LA, Kastenmayer JP, Hüttenhofer AG, Green PJ. Diamonds in the rough: mRNA-like non-coding RNAs. *Trends Plant Sci*, 2008, 13(7): 329–334. [\[DOI\]](#)
- [52] Song D, Yang Y, Yu B, Zheng B, Deng Z, Lu BL, Chen X, Jiang T. Computational prediction of novel non-coding RNAs in Arabidopsis thaliana. *BMC Bioinformatics*, 2009, 10(Suppl 1): S36. [\[DOI\]](#)
- [53] Wen J, Parker BJ, Weiller GF. In Silico identification and characterization of mRNA-like noncoding transcripts in Medicago truncatula. *In Silico Biol*, 2007, 7(4–5): 485–505. [\[DOI\]](#)
- [54] Boerner S, McGinnis KM. Computational identification and functional predictions of long noncoding RNA in Zea mays. *PLoS One*, 2012, 7(8): e43047. [\[DOI\]](#)
- [55] Xin MM, Wang Y, Yao YY, Song N, Hu ZR, Qin DD, Xie CJ, Peng HR, Ni ZF, Sun QX. Identification and characterization of wheat long non-protein coding RNAs responsive to powdery mildew infection and heat stress by using microarray analysis and SBS sequencing. *BMC Plant Biol*, 2011, 11: 61. [\[DOI\]](#)
- [56] Jin JJ, Liu J, Wang H, Wong L, Chua NH. PLncDB: plant long non-coding RNA database. *Bioinformatics*, 2013, 29(8): 1068–1071. [\[DOI\]](#)
- [57] Swarbreck D, Wilks C, Lamesch P, Berardini TZ, Garcia-Hernandez M, Foerster H, Li D, Meyer T, Muller R, Ploetz L, Radenbaugh A, Singh S, Swing V, Tissier C, Zhang P, Huala E. The Arabidopsis Information Resource (TAIR): gene structure and function annotation. *Nucleic Acids Res*, 2008, 36(Database issue): D1009–D1014. [\[DOI\]](#)
- [58] Lamesch P, Berardini TZ, Li DH, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, Karthikeyan AS, Lee CH, Nelson WD, Ploetz L, Singh S, Wensel A, Huala E. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res*, 2012, 40(Database issue): D1202–D1210. [\[DOI\]](#)
- [59] Amaral PP, Clark MB, Gascoigne DK, Dinger ME, Mattick JS. lncRNADB: a reference database for long noncoding RNAs. *Nucleic Acids Res*, 2011, 39(Database issue): D146–D151. [\[DOI\]](#)
- [60] Quek XC, Thomson DW, Maag JL, Bartonicek N, Signal B, Clark MB, Gloss BS, Dinger ME. lncRNADB v2.0: expanding the reference database for functional long non-coding RNAs. *Nucleic Acids Res*, 2015, 43(Database issue): D168–D173. [\[DOI\]](#)
- [61] Chen DJ, Yuan CH, Zhang J, Zhang Z, Bai L, Meng YJ, Chen LL, Chen M. PlantNATsDB: a comprehensive database of plant natural antisense transcripts. *Nucleic Acids Res*, 2012, 40(Database issue): D1187–D1193. [\[DOI\]](#)
- [62] Lavorgna G, Dahary D, Lehner B, Sorek R, Sanderson CM, Casari G. In search of antisense. *Trends Biochem Sci*, 2004,

- 29(2): 88–94. [\[DOI\]](#)
- [63] Werner A. Natural antisense transcripts. *RNA Biol*, 2005, 2(2): 53–62. [\[DOI\]](#)
- [64] Charon C, Moreno AB, Bardou F, Crespi M. Non-protein-coding RNAs and their interacting RNA-binding proteins in the plant cell nucleus. *Mol Plant*, 2010, 3(4): 729–739. [\[DOI\]](#)
- [65] Hüttenhofer A. RNomics: identification and function of small non-protein-coding RNAs in model organisms. *Cold Spring Harb Symp Quant Biol*, 2006, 71: 135–140. [\[DOI\]](#)
- [66] Lukashin AV, Borodovsky M. GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res*, 1998, 26(4): 1107–1115. [\[DOI\]](#)
- [67] Burge CB, Karlin S. Finding the genes in genomic DNA. *Curr Opin Struct Biol*, 1998, 8(3): 346–354. [\[DOI\]](#)
- [68] Lottaz C, Iseli C, Jongeneel CV, Bucher P. Modeling sequencing errors by combining Hidden Markov models. *Bioinformatics*, 2003, 19(Suppl 2): ii103–ii112. [\[DOI\]](#)
- [69] Shimizu K, Adachi J, Muraoka Y. ANGLE: a sequencing errors resistant program for predicting protein coding regions in unfinished cDNA. *J Bioinform Comput Biol*, 2006, 4(3): 649–664. [\[DOI\]](#)
- [70] Jia H, Osak M, Bogu GK, Stanton LW, Johnson R, Lipovich L. Genome-wide computational identification and manual annotation of human long noncoding RNA genes. *RNA*, 2010, 16(8): 1478–1487. [\[DOI\]](#)
- [71] Badger JH, Olsen GJ. CRITICA: coding region identification tool invoking comparative analysis. *Mol Biol Evol*, 1999, 16(4): 512–524. [\[DOI\]](#)
- [72] Hatzigeorgiou AG, Fizev P, Reczko M. DIANA-EST: a statistical analysis. *Bioinformatics*, 2001, 17(10): 913–919. [\[DOI\]](#)
- [73] Mignone F, Grillo G, Liuni S, Pesole G. Computational identification of protein coding potential of conserved sequence tags through cross-species evolutionary analysis. *Nucleic Acids Res*, 2003, 31(15): 4639–4645. [\[DOI\]](#)
- [74] Liu J, Gough J, Rost B. Distinguishing protein-coding from non-coding RNAs through support vector machines. *PLoS Genet*, 2006, 2(4): e29. [\[DOI\]](#)
- [75] Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei LP, Gao G. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res*, 2007, 35(Web Server issue): W345–W349. [\[DOI\]](#)
- [76] Lu ZJ, Yip KY, Wang GL, Shou C, Hillier LW, Khurana E, Agarwal A, Auerbach R, Rozowsky J, Cheng C, Kato M, Miller DM, Slack F, Snyder M, Waterston RH, Reinke V, Gerstein MB. Prediction and characterization of noncoding RNAs in *C. elegans* by integrating conservation, secondary structure, and high-throughput sequencing and array data. *Genome Res*, 2011, 21(2): 276–285. [\[DOI\]](#)
- [77] Washietl S, Findeiß S, Müller SA, Kalkhof S, von Bergen M, Hofacker IL, Stadler PF, Goldman N. RNaCode: robust discrimination of coding and noncoding regions in comparative sequence data. *RNA*, 2011, 17(4): 578–594. [\[DOI\]](#)
- [78] Liu J, Jung C, Xu J, Wang H, Deng S, Bernad L, Arenas-Huertero C, Chua NH. Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in *Arabidopsis*. *Plant Cell*, 2012, 24(11): 4333–4345. [\[DOI\]](#)
- [79] Wang H, Chung PJ, Liu J, Jang IC, Kean MJ, Xu J, Chua NH. Genome-wide identification of long noncoding natural antisense transcripts and their responses to light in *Arabidopsis*. *Genome Res*, 2014, 24(3): 444–453. [\[DOI\]](#)
- [80] Shuai P, Liang D, Tang S, Zhang ZJ, Ye CY, Su YY, Xia XL, Yin WL. Genome-wide identification and functional prediction of novel and drought-responsive lincRNAs in *Populus trichocarpa*. *J Exp Bot*, 2014, 65(17): 4975–4983. [\[DOI\]](#)
- [81] Zhu QH, Stephen S, Taylor J, Helliwell CA, Wang MB. Long noncoding RNAs responsive to *Fusarium oxysporum* infection in *Arabidopsis thaliana*. *New Phytol*, 2014, 201(2): 574–584. [\[DOI\]](#)
- [82] Li L, Eichten SR, Shimizu R, Petsch K, Yeh CT, Wu W, Chetoor AM, Givan SA, Cole RA, Fowler JE, Evans MM, Scanlon MJ, Yu J, Schnable PS, Timmermans MC, Springer NM, Muehlbauer GJ. Genome-wide discovery and characterization of maize long non-coding RNAs. *Genome Biol*, 2014, 15(2): R40. [\[DOI\]](#)
- [83] McCutcheon JP, Eddy SR. Computational identification of non-coding RNAs in *Saccharomyces cerevisiae* by comparative genomics. *Nucleic Acids Res*, 2003, 31(14): 4119–4128. [\[DOI\]](#)
- [84] Hüttenhofer A, Vogel J. Experimental approaches to identify non-coding RNAs. *Nucleic Acids Res*, 2006, 34(2): 635–646. [\[DOI\]](#)
- [85] Campalans A, Kondorosi A, Crespi M. *Enod40*, a short open reading frame-containing mRNA, induces cytoplasmic localization of a nuclear RNA binding protein in *Medicago truncatula*. *Plant Cell*, 2004, 16(4): 1047–1059. [\[DOI\]](#)
- [86] van de Sande K, Pawlowski K, Czaja I, Wieneke U, Schell J, Schmidt J, Walden R, Matvienko M, Wellink J, van Kammen A, Franssen H, Bisseling T. Modification of phytohormone response by a peptide encoded by *ENOD40* of legumes and a nonlegume. *Science*, 1996, 273(5273): 370–373. [\[DOI\]](#)

- [87] Kouchi H, Takane K, So RB, Ladha JK, Reddy PM. Rice *ENOD40*: isolation and expression analysis in rice and transgenic soybean root nodules. *Plant J*, 1999, 18(2): 121–129. [\[DOI\]](#)
- [88] Cho J, Koo DH, Nam YW, Han CT, Lim HT, Bang JW, Hur Y. Isolation and characterization of cDNA clones expressed under male sex expression conditions in a monoecious cucumber plant (*Cucumis sativus* L. cv. Winter Long). *Euphytica*, 2005, 146(3): 271–281. [\[DOI\]](#)
- [89] Ma JX, Yan BX, Qu YY, Qin FF, Yang YT, Hao XJ, Yu JJ, Zhao Q, Zhu DY, Ao GM. *Zm401*, a short-open reading-frame mRNA or noncoding RNA, is essential for tapetum and microspore development and can regulate the floret formation in maize. *J Cell Biochem*, 2008, 105(1): 136–146. [\[DOI\]](#)
- [90] Shin H, Shin HS, Chen RJ, Harrison MJ. Loss of *At4* function impacts phosphate distribution between the roots and the shoots during phosphate starvation. *Plant J*, 2006, 45(5): 712–726. [\[DOI\]](#)
- [91] Song JH, Cao JS, Yu XL, Xiang X. BcMF11, a putative pollen-specific non-coding RNA from *Brassica campestris* ssp. *chinensis*. *J Plant Physiol*, 2007, 164(8): 1097–1100. [\[DOI\]](#)
- [92] Dieci G, Fiorino G, Castelnovo M, Teichmann M, Pagano A. The expanding RNA polymerase III transcriptome. *Trends Genet*, 2007, 23(12): 614–622. [\[DOI\]](#)
- [93] Wierzbicki AT, Haag JR, Pikaard CS. Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell*, 2008, 135(4): 635–648. [\[DOI\]](#)
- [94] Rowley MJ, Böhmendorfer G, Wierzbicki AT. Analysis of long non-coding RNAs produced by a specialized RNA polymerase in *Arabidopsis thaliana*. *Methods*, 2013, 63(2): 160–169. [\[DOI\]](#)
- [95] Haag JR, Pikaard CS. Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. *Nat Rev Mol Cell Biol*, 2011, 12(8): 483–492. [\[DOI\]](#)
- [96] Martín AC, del Pozo JC, Iglesias J, Rubio V, Solano R, de La Peña A, Leyva A, Paz-Ares J. Influence of cytokinins on the expression of phosphate starvation responsive genes in *Arabidopsis*. *Plant J*, 2000, 24(5): 559–567. [\[DOI\]](#)
- [97] Franco-Zorrilla JM, Valli A, Todesco M, Mateos I, Puga MI, Rubio-Somoza I, Leyva A, Weigel D, García JA, Paz-Ares J. Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat Genet*, 2007, 39(8): 1033–1037. [\[DOI\]](#)
- [98] Liu CM, Muchhal US, Raghothama KG. Differential expression of TPS11, a phosphate starvation-induced gene in tomato. *Plant Mol Biol*, 1997, 33(5): 867–874. [\[DOI\]](#)
- [99] Burleigh SM, Harrison MJ. Characterization of the *Mt4* gene from *Medicago truncatula*. *Gene*, 1998, 216(1): 47–53. [\[DOI\]](#)
- [100] Burleigh SH, Harrison MJ. The down-regulation of *Mt4*-like genes by phosphate fertilization occurs systemically and involves phosphate translocation to the shoots. *Plant Physiol*, 1999, 119(1): 241–248. [\[DOI\]](#)
- [101] Burleigh SH, Harrison MJ. A novel gene whose expression in *Medicago truncatula* roots is suppressed in response to colonization by vesicular-arbuscular mycorrhizal (VAM) fungi and to phosphate nutrition. *Plant Mol Biol*, 1997, 34(2): 199–208. [\[DOI\]](#)
- [102] Wasaki J, Yonetani R, Shinano T, Kai M, Osaki M. Expression of the *OsP11* gene, cloned from rice roots using cDNA microarray, rapidly responds to phosphorus status. *New Phytol*, 2003, 158(2): 239–248. [\[DOI\]](#)
- [103] Fujii H, Chiou TJ, Lin SI, Aung K, Zhu JK. A miRNA involved in phosphate-starvation response in *Arabidopsis*. *Curr Biol*, 2005, 15(22): 2038–2043. [\[DOI\]](#)
- [104] Nagano T, Fraser P. Emerging similarities in epigenetic gene silencing by long noncoding RNAs. *Mamm Genome*, 2009, 20(9–10): 557–562. [\[DOI\]](#)
- [105] Swiezewski S, Liu FQ, Magusin A, Dean C. Cold-induced silencing by long antisense transcripts of an *Arabidopsis* Polycomb target. *Nature*, 2009, 462(7274): 799–802. [\[DOI\]](#)
- [106] Helliwell CA, Robertson M, Finnegan EJ, Buzas DM, Dennis ES. Vernalization-repression of *Arabidopsis* FLC requires promoter sequences but not antisense transcripts. *PLoS One*, 2011, 6(6): e21513. [\[DOI\]](#)
- [107] Heo JB, Sung S. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science*, 2011, 331(6013): 76–79. [\[DOI\]](#)
- [108] Ding JH, Lu Q, Ouyang YD, Mao HL, Zhang PB, Yao JL, Xu CG, Li XH, Xiao JH, Zhang QF. A long noncoding RNA regulates photoperiod-sensitive male sterility, an essential component of hybrid rice. *Proc Natl Acad Sci U S A*, 2012, 109(7): 2654–2659. [\[DOI\]](#)
- [109] Ding JH, Shen JQ, Mao HL, Xie WB, Li XH, Zhang QF. RNA-directed DNA methylation is involved in regulating photoperiod-sensitive male sterility in rice. *Mol Plant*, 2012, 5(6): 1210–1216. [\[DOI\]](#)
- [110] Zhou H, Liu QJ, Li J, Jiang DG, Zhou LY, Wu P, Lu S, Li F, Zhu LY, Liu ZL, Chen LT, Liu YG, Zhuang CX. Photoperiod- and thermo-sensitive genic male sterility in rice

- are caused by a point mutation in a novel noncoding RNA that produces a small RNA. *Cell Res*, 2012, 22(4): 649–660. [\[DOI\]](#)
- [111] Zhu DM, Deng XW. A non-coding RNA locus mediates environment-conditioned male sterility in rice. *Cell Res*, 2012, 22(5): 791–792. [\[DOI\]](#)
- [112] Yang WC, Katinakis P, Hendriks P, Smolders A, de Vries F, Spee J, van Kammen A, Bisseling T, Franssen H. Characterization of *GmENOD40*, a gene showing novel patterns of cell-specific expression during soybean nodule development. *Plant J*, 1993, 3(4): 573–585. [\[DOI\]](#)
- [113] Crespi MD, Jurkevitch E, Poirer M, d'Aubenton-Carafa Y, Petrovics G, Kondorosi E, Kondorosi A. *enod40*, a gene expressed during nodule organogenesis, codes for a non-translatable RNA involved in plant growth. *EMBO J*, 1994, 13(21): 5099–5112. [\[DOI\]](#)
- [114] Laporte P, Satiat-Jeunemaître B, Velasco I, Csorba T, Van de Velde W, Campalans A, Burgya J, Arevalo-Rodriguez M, Crespi M. A novel RNA-binding peptide regulates the establishment of the *Medicago truncatula*-*Sinorhizobium meliloti* nitrogen-fixing symbiosis. *Plant J*, 2010, 62(1): 24–38. [\[DOI\]](#)
- [115] Vlegheels I, Hontelez J, Ribeiro A, Fransz P, Bisseling T, Franssen H. Expression of *ENOD40* during tomato plant development. *Planta*, 2003, 218(1): 42–49. [\[DOI\]](#)
- [116] Rohrig H, Schmidt J, Miklashevichs E, Schell J, John M. Soybean *ENOD40* encodes two peptides that bind to sucrose synthase. *Proc Natl Acad Sci USA*, 2002, 99(4): 1915–1920. [\[DOI\]](#)
- [117] Sousa C, Johansson C, Charon C, Manyani H, Sautter C, Kondorosi A, Crespi M. Translational and structural requirements of the early nodulin gene *enod40*, a short-open reading frame-containing RNA, for elicitation of a cell-specific growth response in the alfalfa root cortex. *Mol Cell Biol*, 2001, 21(1): 354–366. [\[DOI\]](#)
- [118] Erdmann VA, Szymanski M, Hochberg A, de Groot N, Barciszewski J. Non-coding, mRNA-like RNAs database Y2K. *Nucleic Acids Res*, 2000, 28(1): 197–200. [\[DOI\]](#)
- [119] Dai XY, Yu JJ, Zhao Q, Zhu DY, Ao GM. Non-coding RNA for *zm401*, a pollen-specific gene of *Zea mays*. *Acta Bot Sin*, 2004, 46(4): 497–504. [\[DOI\]](#)
- [120] Ma JX, Zhao Q, Yu JJ, Ao GM. Ectopic expression of a maize pollen specific gene, *zm401*, results in aberrant anther development in tobacco. *Euphytica*, 2005, 144(1-2): 133–140. [\[DOI\]](#)
- [121] Song JH, Cao JS, Wang CG. *BcMF11*, a novel non-coding RNA gene from *Brassica campestris*, is required for pollen development and male fertility. *Plant Cell Rep*, 2013, 32(1): 21–30. [\[DOI\]](#)

(责任编辑: 张宪省)