

DOI:10.12113/j.issn.1672-5565.201812006

植物长链非编码 RNA 的生物信息学预测与分析研究进展

蔡媛, 钟灿, 刘浩, 金剑, 王勇庆, 张水寒*

(湖南省中医药研究院 中药研究所, 长沙 410013)

摘要:长链非编码 RNA (Long non-coding RNAs, lncRNAs) 是一类广泛存在于真核生物中, 长度大于 200 个核苷酸、无蛋白编码功能, 具有调控基因转录后表达的 RNA 转录本。新近研究表明, lncRNA 在多种生物途径中起着重要调节作用。生物信息学由生物、数学、计算机科学, 统计学等多学科交叉产生, 能从全局和系统水平对大数据信息进行深入挖掘与分析。采用生物信息学方法预测与分析 lncRNA 是当前发现和鉴定植物 lncRNA 的重要策略之一。本文梳理和总结了近年来采用生物信息学预测植物 lncRNA 及其靶基因的方法策略, 以期在今后深入认知植物 lncRNA 在植物的生长发育过程、抗逆境胁迫及系统进化等过程中的作用研究提供一定参考。

关键词:长链非编码 RNA; 生物信息学; 功能预测

中图分类号:Q522 **文献标志码:**A **文章编号:**1672-5565(2019)03-151-10

Progress of bioinformatics prediction and analysis of long non-coding RNA in plants

CAI Yuan, ZHONG Can, LIU Hao, JIN Jian, WANG Yongqing, ZHANG Shuihan*

(Institute of Traditional Chinese Medicine, Hunan Academy of Traditional Chinese Medicine, Changsha 410013, China)

Abstract: Long non-coding RNAs (lncRNAs) widely exist in eukaryotes, which are RNA transcripts expressed by the transcription of regulatory genes with more than 200 nucleotides in length and no protein-coding ability. Numerous studies have shown that lncRNAs play an important role in regulating a variety of biological pathways. Bioinformatics is generated by multiple disciplines including biology, mathematics, computer science, and statistics, which can deeply mine and analyze big data information from the global and system levels. Currently, using bioinformatic method to predict and analyze lncRNA is one of the important strategies for the discovery and identification of plant lncRNA. This paper summarizes and discusses the methodological strategies of bioinformatics for predicting plant lncRNA and its target genes, so as to provide a reference for the future research on the role of plant lncRNAs in plants' growth and development, stress resistance, and phylogenetic evolution.

Keywords: Long non-coding RNA; Bioinformatics; Functional prediction

多年来, 研究人员对基因组的研究主要聚焦在蛋白质编码基因, 其研究思维也一直遵从经典中心法则“DNA-mRNA-蛋白质”。随着人类基因组及其他物种海量基因组的不断解析和深入研究, 以及蛋白组学和转录组学的蓬勃发展, 促进了 RNA 组学研究的日趋成熟, 揭示高等真核生物的遗传物质只有极小一部分编码蛋白质, 绝大部分都不编码蛋白质和多肽, 这部分非编码蛋白基因一直被当做“噪音”

或者是“垃圾”(Junk)分子^[1]。能够编码蛋白的 mRNA 已经不再独占转录组鳌头, 许多不同类型的非编码调控 RNA 逐一进入研究者的视线, 并陆续被证明具有重要的生物学意义。

长链非编码 RNA (long non-coding RNA, lncRNA) 是在真核生物中新发现的一类长度大于 200 个核苷酸、没有阅读框架, 但往往具有 mRNA 结构特征(帽式结构和 polyA 尾巴)的 RNA^[2-3]。大多

收稿日期: 2018-12-29; 修回日期: 2019-05-06.

基金项目: 国家自然科学基金青年基金项目 (No. 81503197); 湖南省自然科学基金项目 (No. 2018JJ3310); 湖南省中医药管理局项目 (No. 201810); 湖南省中医药研究院科研计划项目 (No. 201705).

作者简介: 蔡媛, 女, 助理研究员, 研究方向: 分子生物学. E-mail: tcmuanyuan@163.com.

* 通信作者: 张水寒, 女, 研究员, 研究方向: 中药学. E-mail: zhangshuihan@126.com.

数 lncRNA 由 RNA 聚合酶 II 转录翻译而来,少数由 RNA 聚合酶 III 转录而来,以 RNA 的形式在多种层面上调控基因的表达。lncRNA 在基因组中普遍存在转录现象,但较之 mRNA 往往表达水平比较低,其自身的表达水平也受到转录及转录后调控机制的严密调节。相对于长链非编码 RNA 在哺乳动物上的研究,植物长链非编码 RNA 的研究才刚刚开始^[4-6]。生物信息学的迅猛发展,改变了传统 lncRNA 的研究方式,极大地促进了植物 lncRNA 的研究发展。长链非编码 RNA 在植物生命活动中发挥着重要的作用,不同类的长链非编码 RNA 在植物中发挥的功能也不尽相同,主要包括调节生长发育、影响转录调控、染色体结构、mRNA 的稳定性与翻译、RNA 加工与修饰等^[7-9]。最近获得的基因组序列呈爆炸性地增加,从而激发了用快速、有效和精确的方法组织和获取重要序列以及结构元件的能力需求。新获得的数据显示,迄今为止,一类潜在的重要基因类型我们还没有检测到,有一大类功能 RNA 分子或隐藏在蛋白质编码之间或位于编码蛋白质区内(内含子中),至今未被注释。然而,任何功能分析中,都不应该遗漏在基因组范围寻找非编码 RNA (non-coding RNA)。如何能像发现读码框一样发现没有强烈结构特征的调控 RNA,本文就近年来植物中已发现的 lncRNA 的种类、参与的生物学过程、发挥功能的分子机制及其生物信息学预测与分析进行综述和展望,以期为更深入认识植物 lncRNA 提供借鉴。

1 植物 lncRNA 的分类及其生物学功能

1.1 植物 lncRNA 的分类

1.1.1 按其相对位置分类

长链非编码 RNA 根据其在基因组中与蛋白质编码基因的相对位置,一般将其分为正义 lncRNA (Sense long non-coding RNA)、反义 lncRNA (Antisense long non-coding RNA)、双向 lncRNA (Bidirectional long non-coding RNA)、基因内 lncRNA (Intronic long non-coding RNA) 和基因间 lncRNA (Intergenic long non-coding RNA) 五类^[10]。其中,基因间 lncRNA 也被称为大型介入性非编码 RNA,即 lincRNA (Large intervening noncoding RNA),位置关系对于推测其功能具有重要的作用。

1.1.2 按其作用机制分类

按照 lncRNA 发挥作用的分子机制,Wang 等将 lncRNA 分为了以下四类,即信号分子 (Signals)、诱

饵分子 (Decoys)、引导分子 (Guides)、骨架分子 (Scaffolds)^[11]。Wilusz 等^[12]总结了生物体内 lncRNA 的具体机制,包括:(1)在编码蛋白基因的上游启动子区转录,从而干扰邻近蛋白编码基因的表达(如酵母 *SER3* 基因)^[13];(2)抑制 RNA 聚合酶 II,或介导染色质重构和组蛋白修饰,而影响基因表达^[14];(3)lncRNA 与编码蛋白基因的转录本形成互补双链,干扰 mRNA 的剪切,进而产生不同的剪切形式^[15];(4)lncRNA 与编码蛋白基因的转录本形成互补双链,在 Dicer 酶作用下产生内源性的 siRNA,调控基因的表达水平^[16];(5)lncRNA 结合在特定蛋白质上调节相应蛋白的活性^[17];(6)作为结构组分与蛋白质形成核酸蛋白质复合物^[18];(7)结合在特定蛋白上从而改变该蛋白的胞质定位^[19],研究者发现,*MtEnod40* 能够与 MtRBP1 蛋白结合,引导 MtRBP1 从细胞核的核小点到细胞质颗粒的重定位^[20];(8)可作为小分子 RNA(如 miRNA)的前体分子^[21]。在植物中,有研究表明 lncRNA 作为小 RNA 生物合成前体。研究发现水稻光敏雄性不育关键调控基因 *LDMAR* 最终被发现通过剪切加工形成了长 21 nt 的小 RNA *osa-smR5846w/m*。Ding 和 Zhu 等研究显示,1 236 nt 长的 *LDMAR* 可能是初级转录本,该初级转录本会被加工为一条长 136 nt 的中间转录本,最终才形成 21 nt 的小 RNA^[22-24]。

1.2 植物 lncRNA 的生物学功能

长链非编码 RNA 的基因不像编码基因那样编码蛋白质,但是它们具有丰富强大的生物学功能。大量数据表明,这些 lncRNA 可能在从蛋白质分泌到广泛的基因调节细胞过程中起重要作用。它们在诸如剂量补偿、基因印迹、转录调控、前 mRNA 剪接和 mRNA 翻译控制等不同的途径中发挥功能^[25]。lncRNA 在植物生长发育过程中扮演着重要角色,处于基因调控网络的核心位置,调控各种重要生物途径,包括发育、代谢、抗病、胁迫、应激反应、激素信号和维护基因组的完整性等^[26-30]。目前,在拟南芥、小麦、水稻、玉米、黄瓜等植物中发现了大量 lncRNAs,但是明确其功能机制的不到 1%。研究比较透彻的主要集中在生长发育、逆境胁迫、生殖发育等方面。

1.2.1 lncRNA 参与植物成花过程

长链非编码 RNA 引起植物研究人员的兴趣始于 lncRNA 能够调控春化作用的关键蛋白 FLC 来实现开花的转变。并且,研究者发现有两种 lncRNA 参与调节 *FLC* 基因,从而参与调节植物的开花过程^[28, 31]。在拟南芥中,*COOLAIR* 通过吸引相关蛋白清除 *FLC* 上激活型组蛋白甲基标记,引起 *FLC* 正向

转录本的沉默。*COLDAIR* 则通过结合 PcG 蛋白复合体使 *FLC* 染色质组蛋白抑制型甲基化,引起 *FLC* 沉默。这两种长链非编码 RNA 都通过沉默 *FLC* 而参与调节春化过程,影响植物开花的时间。

1.2.2 lncRNA 参与植物的胁迫响应

在植物的生长发育过程中,不可避免地受到病害、干旱、磷、盐等胁迫。研究者发现,在逆境胁迫中,lncRNA 扮演重要角色。研究者发现 *Npc536* 在拟南芥根和叶中调控响应盐胁迫。在盐胁迫下,*Npc536* 的 T-DNA 插入突变体无明显表型,过表达可以促进在盐胁迫条件下的根系生长,提高初生根的生长和次生根的伸长^[32]。另外,在拟南芥中,*IPS1* 和 *At4* 是由磷饥饿诱导产生,其可以阻止 *miR399* 对靶基因 *PHO2* 的抑制作用而调节磷含量的动态平衡^[33-35]。

1.2.3 lncRNA 参与植物的生殖发育

研究者在水稻中筛选到调控水稻光敏性雄性不育的 lncRNA-LD MA,该基因是长日照植物特有的雄性生殖相关的 lincRNA,在长日照条件下转录表达,是水稻花粉正常发育所需。自发突变系水稻植株改变 *LDMAR* 的二级结构,提高 *LDMAR* 启动子区域的甲基化程度,从而降低 *LDMAR* 转录,尤其是在长日照调节下,导致发育中花药过早的程序性死亡,从而导致 PSMS,阐明 lncRNA 在水稻生殖进程中的重要调控作用^[22-24]。

2 植物 lncRNA 的生物信息学预测及策略

目前,植物长链非编码 RNA 的预测方法主要分为生物信息学方法和实验 RNA 组学方法。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^[36]。高通量测序技术是发现 lncRNA 的有效方法,可以直接、快速地发现低丰度、新的 lncRNA。目前研究中一般首先采用生物信息获得目标序列,然后再进行下一步的功能验证及其机制研究。

生物信息学方法主要是利用某种算法,通过设计筛选标准,建立评分系统,对候选序列进行选择。伴随着物种测序工作的逐步开展和序列信息的日益丰富,利用生物信息学发现和预测 lncRNA 序列的方法已经成为简单、高效的策略之一。但是大部分策略仅能预测得到保守的 lncRNA 序列。同时也会出现假阳性序列,需要通过进一步的实验验证,来完善预测和研究。

在众多发现和研究 lncRNA 的方法中,生物信息学方法以其方便、快速和经济的优势受到许多研究者的青睐^[37]。事实证明,生物信息学方法是预测和发现新 lncRNA 的有效方法,是以基因组序列和计算机程序鉴定为基础^[38]。目前,通过各种计算机软件以及其他计算工具已经成功地预测和鉴定了动植物中大多数 lncRNA^[39-40]。

在过去十几年中,人们通过生物信息学手段和分子克隆方法从拟南芥 (*Arabidopsis thaliana* L.)^[21, 28, 31-34, 41-53]、水稻 (*Oryza sativa* L.)^[22-24, 54-57]、玉米 (*Zea mays* L.)^[58-61]、小麦 (*Triticum aestivum* L.)^[30]、黄瓜 (*Cucumis sativus* L.)^[62]、苜蓿 (*Medicago truncatula*)^[63-66]、番茄 (*Solanum lycopersicum*)^[67]、大豆 (*Glycine max*)^[68]、油菜 (*Brassica campestris* L.)^[69-71] 等植物中发现了大量的多种类型的 lncRNA (见表 1)。

2.1 植物 lncRNA 靶基因预测及策略

lncRNA 具有多种重要功能,寻找 lncRNA 靶基因并挖掘它对基因表达、蛋白合成各方面的调控作用成了 lncRNA 研究的关键。lncRNA 靶基因的预测,大致可分为两种情况:已知 gene symbol 的 lncRNA 和新发现的 lncRNA。对已知 gene symbol 的 lncRNA,可以直接通过软件数据库例如 starBase、CHIPBase、NONCODE 等,利用 gene symbol 搜索其相关信息。对于新发现的 lncRNA 来说,目前现有的数据库物种信息十分有限,数据信息也必将少,所以对于转录本预测出的没有 symbol 号的新 lncRNA,由于 lncRNA 对靶基因没有固定的作用模式,基因调控可能以顺式 (Cis) 或反式 (Trans) 作用发生。所以,可根据不同作用方式分别进行预测,包括顺式作用靶基因预测及反式作用靶基因预测。

顺式作用靶基因预测,认为 lncRNA 的功能与其坐标临近的蛋白编码基因相关,位于编码蛋白上下游的 lncRNA 可能与启动子或者共表达基因的其他顺式作用元件有交集,从而在转录或者转录后水平对基因的表达进行调控。判断一个 lncRNA 具有顺式调控作用通常要同时满足以下几个条件:(1)附近的基因表达情况与其保持一致;(2)该基因失活后会影周围基因的表达;(3)会影响附近同一位点的基因表达。对于满足以上条件的 lncRNA,首先找出位于其上游或者下游附近 (10 k) 的编码蛋白基因,通过对编码蛋白的功能富集分析,从而预测 lncRNA 的主要功能,为后续顺式作用分析打下基础。

表 1 植物中已报道的 lncRNA

Table 1 The reported lncRNAs in plants

LncRNA 名称	物种来源	生物学功能	参考文献
<i>OOLAIR</i>	拟南芥 (<i>Arabidopsis thaliana</i>)	调控春化途径	[28]
<i>COLDAIR</i>	拟南芥 (<i>Arabidopsis thaliana</i>)	调控春化途径	[31]
<i>AtIPS1</i>	拟南芥 (<i>Arabidopsis thaliana</i>)	与磷酸胁迫相关	[33]
<i>AtR8</i>	拟南芥 (<i>Arabidopsis thaliana</i>)	根中特异表达,与低氧应激应答有关	[48]
<i>Npc48</i>	拟南芥 (<i>Arabidopsis thaliana</i>)	参与叶和花的发育	[21]
<i>Npc536</i>	拟南芥 (<i>Arabidopsis thaliana</i>)	应对生物和非生物胁迫	[32]
<i>TERRA</i>	拟南芥 (<i>Arabidopsis thaliana</i>)	端粒酶甲基化相关	[49]
<i>At4</i>	拟南芥 (<i>Arabidopsis thaliana</i>)	与磷酸胁迫相关	[34]
<i>SVALKA</i>	拟南芥 (<i>Arabidopsis thaliana</i>)	与冷冻胁迫相关	[50]
<i>MAS</i>	拟南芥 (<i>Arabidopsis thaliana</i>)	冷诱导条件下与开花相关	[51]
<i>AsHSFB2a</i>	拟南芥 (<i>Arabidopsis thaliana</i>)	与热应激调控相关	[52]
<i>AsDRIR</i>	拟南芥 (<i>Arabidopsis thaliana</i>)	与干旱胁迫相关	[46]
<i>ELENA1</i>	拟南芥 (<i>Arabidopsis thaliana</i>)	与免疫应答相关	[53]
<i>APOLO</i>	拟南芥 (<i>Arabidopsis thaliana</i>)	与生物素应答相关	[72]
<i>HID1</i>	拟南芥 (<i>Arabidopsis thaliana</i>)	与其红光形态相关	[73]
<i>LDMAR</i>	水稻 (<i>Oryza sativa</i>)	光敏感引起雄性不育	[22-24]
<i>OsPI1</i>	水稻 (<i>Oryza sativa</i>)	磷酸胁迫相关	[54]
<i>OsENOD40</i>	水稻 (<i>Oryza sativa</i>)	与根瘤形成相关	[55]
<i>TWISTED LEAF (TL)</i>	水稻 (<i>Oryza sativa</i>)	与水稻叶片平整相关	[56]
<i>PMS1T</i>	水稻 (<i>Oryza sativa</i>)	光敏感引起雄性不育	[57]
<i>MiENOD40</i>	苜蓿 (<i>Medicago truncatula</i>)	与根瘤形成相关	[64]
<i>Mi4</i>	苜蓿 (<i>Medicago truncatula</i>)	磷酸胁迫相关	[65, 66]
<i>TPS1</i>	番茄 (<i>Solanum lycopersicum</i>)	磷酸胁迫相关	[67]
<i>GmENOD40</i>	大豆 (<i>Glycine max</i>)	与根瘤形成相关	[68]
<i>BcMF11</i>	油菜 (<i>Brassica campestris L.</i>)	与胸花不育相关	[69-71]
<i>CsM10</i>	黄瓜 (<i>Cucumis sativus L.</i>)	与雌雄分化相关	[62]
<i>Zm401</i>	玉米 (<i>Zea mays</i>)	与雄花不育相关	[58-61]

反式作用靶基因预测基本原理认为 lncRNA 的功能与编码基因的位置关系没有关系,而与其共表达的蛋白编码基因相关。当 lncRNA 与一些距离较远的基因在表达量上存在正相关或者负相关的情况时,可以通过样本间 lncRNA 与蛋白编码基因的表达量相关性分析或共表达分析方法来预测其靶基因。当样本数 ≥ 6 时,使用 Pearson 相关系数法分析样本间 lncRNA 与蛋白编码基因的相关性,对相关性最高的编码基因蛋白进行功能富集分析进而预测 lncRNA 功能;当样本数 ≥ 24 时,使用 WGCNA 方法将表达模式相似的基因聚类得到不同的共表达模块,根据模块内的已知的编码基因功能预测 lncRNA 的功能。

除了顺式作用和反式作用调控,lncRNA 参与调控许多转录后进程时,与 miRNA 和 snoRNA 等小

RNA 类似,这些调控往往与碱基的互补配对有关。一部分反义 lncRNA 可能因为与正义链的 mRNA 结合而调控基因沉默、转录及 mRNA 的稳定性。所以,lncRNA 的反义分析,可以利用软件(如 RNAplex)预测反义 lncRNA 与 mRNA 之间的互补配对关系,根据热力学结构计算最小自由能来预测最佳碱基配对关系。

生物信息学预测还可以通过比较基因组学策略。尽管 lncRNA 一级和二级结构不完整,但小部分具有保守性,可根据其保守性推测功能。通过与 miRNA 或蛋白相互作用预测,如通过 miRcode 算法通过 miRNA 推测 lncRNA,也可通过评估其潜在互作蛋白推测其功能如 catRAPID 算法。

2.2 植物 lncRNA 相关数据库及预测工具

随着高通量测序技术的发展,植物 lncRNA 的研

究已经取得显著进展。随着与植物相关的新 lncRNA 出现,收集归类植物非编码 RNA 基因相关信息的数据库也开始出现,建立了系列针对植物的 lncRNA 数据库,如 TAIR10、PlantNATsDB, PLncdb 等(见表 2)。这些数据库信息的来源主要是通过汇总 ChIP-seq、RNA-seq、Tilling array、文献,等其他数据库信息。LncRNAdb(lncRNAs Database)数据库收录了 289 条真核生物的 lncRNA 相关信息,包含 lncRNA 特征、进化保守性、表达、功能、种类、相关组件、序列及对应的文献。与植物相关的包括 *At4*、*COOLAIR*、*COLDAIR*、*AtIPS1*、*Npc48*、*Npc536*、*TERRA*、*OsPI1*、*OsENOD40*、*MtENOD40*、*TPS11*、*GmENOD40* 等 16 个 lncRNA 表达情况及相关信息^[74]。TAIR10 (The *Arabidopsis* Information Resource)是拟南芥专属数据库^[75-76],汇总了拟南芥基因组序列及其基因图谱、序列、表达、功能等,共收录 478 条 lncRNA 信息。PlantNATsDB (Plant Natural Antisense Transcripts DataBase)数据库包括 2 138 498 条反义转录对,其主要用于预测和查询植物天然反转录及其调控功能的数据库^[77]。PLncDB (Plant long non-coding RNA database)是目前收录植物 lncRNA 最齐全的数据库。该数据库通过收集 Tilling array、RNA-seq、文献过滤等信息,获得 16 227 条 lncRNA 数据,能够提供不同组织、发育阶段、突变体和胁迫处理等应激条件下 lncRNA 的表达特征,编码位点及其侧翼基因组区域表观遗传变化和功能^[78]。NONCODE 数据库是专注于分析非编码 RNA 基因的综合知识平台,所有收入的非编码 RNA 基因都是采取计算机自动过滤 GenBank 数据和文献报道中的信息,且经过人工检查确认,共收录 423 976 条 lncRNA 信息。可查阅 lncRNA 的长度、序列信息、生物功能及其表达模式、基因组上下游序列调控元件等^[79]。PNRD 数据库目前数量约 28 214,来自 166 种物种。包括 lncRNA, tRNA, rRNA, tasiRNA, snRNA 和 snoRNA 等。PNRD 是一个植物 ncRNA 综合分析平台,能够提供许多功能搜索和分析工具,涉及 ncRNA 关键词搜索,基于文献的功能搜索,miRNA-target 搜索和在线分析,包括新的 miRNA 预测工具包,编码潜在计算器工具包, Blast 工具和定制的 UCSC 基因组浏览器^[80]。GREENC 数据库是基于 Wiki 数据建立的,可作为植物 lncRNA 的注释和存档。该数据库于 2015 年开始提供有关 lncRNA 的序列,基因组坐标,编码潜力和折叠能量的信息。其中包含来自 37 种植物和 6 种藻类的超过 19 万份转录本的信息^[81]。CANTATAdb 数据库目前有 45 117 条 lncRNA 数据,来源于 10 个

物种。提供序列、RNA-Seq 文库中的表达值,基因组位置等信息^[82]。而 PLNlncRbase 数据库是基于系列实验得到的 lncRNA 形成的数据库^[83]。除了不断完善的数据库外,研究者们还开发了系列 lncRNA 的分析与预测工具,比如 PhyloCSF 能够使用 CSF 评分来计算 lncRNA 编码潜力^[84];CPS 使用序列特征和 SVM 计算 lncRNA 的蛋白编码潜力^[85];CNCl 通过分析相邻核苷酸计算 lncRNA 的编码潜力^[86];CPAT 使用逻辑回归模型计算 lncRNA 的蛋白编码潜力,DeepLNC 利用深度神经网络预测 lncRNAs^[87];iSeeRNA 使用 SVM 算法预测 lncRNAs^[88]。LncRNATargets 能够基于核酸热力学预测 lncRNAs 靶标^[89]。

3 结语与展望

近年来调节转录和翻译的长链非编码 RNA 的不断发现,使基因组调节的全貌已发生了彻底改变。随着研究的不断深入以及相关理论技术的完善,发现植物中存在一系列特异 lncRNA,其数量也在以惊人的速度增加。虽然目前已确定的 lncRNAs 很多,但对绝大部分 lncRNA 在生命活动过程中的具体调控机制及功能模式仍不清晰。与哺乳动物 lncRNA 的研究报道相比,有关植物 lncRNA 的研究还比较落后,目前仅在拟南芥、水稻、小麦、玉米、黄瓜等中对 lncRNA 进行了初步的系统识别和功能研究,距离完全解释 lncRNA 的作用机制和生物学功能仍有较大距离。

生物信息学分析作为一种强有力的技术手段,在从实验设计到结果分析等各个层面发挥着不可替代的作用,既能够启迪研究人员设计阶段的预判以少走弯路,也可以从结果分析中挖掘大量的有用信息,起到事半功倍的作用。

总体来说,目前植物 lncRNA 的研究还处于初级探索阶段,主要存在以下问题:(1) lncRNA 数据库不够完善。与植物 lncRNA 相关的基因组和蛋白质组数据库寥寥无几;(2) lncRNA 功能预测工具不多,针对 lncRNA 的生物信息学工具少,难以对 lncRNA 二级结构等进行有效地预测, lncRNA 作用机理还不清楚,应用研究领域有限;(3)整体上有关植物 lncRNA 的新研究方法和预测功能的工具不够,其功能研究的思路和技术不成熟,不能很好地注释 lncRNA 的调节机制和生物学功能,也不能系统深入地研究 lncRNA。在未来的几年里, lncRNA 表型的分析和解释将是生物信息学工作者的主要挑战。

表 2 植物 lncRNA 数据库及预测工具
Table 2 Databases and prediction tools of plant lncRNAs

数据库名称	主要特点	数据来源	链接
LncRNAdb	数量 287, 包含 68 种真核生物, 其中包括 16 个植物 lncRNA; 提供 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。	文献	http://www.lncrnadb.org/
TAIR10	数量 478, 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, 拟南芥种子库存储数据等信息。	文献、其他数据库、注释基因的计算机预测、其他研究机构	http://arabidopsis.org
PlantNATsDB	数量 2 138 498, 包括 69 种植物, 提供 NAT 资源, 包括浏览、检索、查看、下载等分析工具, 可用于 NAT 预测和功能分析。	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	http://bis.zju.edu.cn/pnatdb/
PLncDB	数量 16 227, 提供不同组织、发育阶段、突变体和胁迫处理的 lncRNA 表达特异性, 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。	Tilling array、lincRNA array、拟南芥文献	http://chualab.rockefeller.edu/gbrowse2/homepage.html
NONCODE	数量 423 976, 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。	RNA-seq、文献、其他数据库	http://www.bioinfo.org/noncode/ http://www.noncode.org
PNRD	数量约 28 214, 包括 166 种植物, 植物 ncRNA 综合分析平台, 包括预测分析。	已发表文献、其他数据库	http://structuralbiology.cau.edu.cn/PNRD/
GREENC	数量 19 万, 包括 37 种植物和 6 个藻类; 提供 lncRNA 基因组坐标, 编码潜力和折叠能量的信息	wiki 数据、其他数据库、已发表文献、基因组序列	http://greenc.sciencedesigners.com
CANTATAdb	数量 45 117, 来源于 10 种植物中预测的 lncRNA; 提供序列, RNA-Seq 文库中的表达值, 基因组位置等信息	RNA-seq、文献、其他数据库	http://yeti.amu.edu.pl/CANTATA/
PLNlncRbase	搜集与整理通过实验得到的 lncRNA 数据	文献, 实验数据	http://bioinformatics.ahau.edu.cn/PLNlncRbase/pcsb
PhyloCSF	使用 CSF 评分计算 lncRNA 编码潜力	——	https://github.com/mlin/PhyloCSF
CPS	使用序列特征和 SVM 计算编码蛋白质的 lncRNA	——	http://cpc.cbi.pku.edu.cn/
CNCI	通过分析相邻核苷酸计算编码蛋白质的 lncRNA	——	https://github.com/www-bioinfo-org/CNCI
CPAT	使用逻辑回归模型计算编码蛋白质的 lncRNA	——	http://rna-cpat.sourceforge.net/
DeepLNC	利用深度神经网络预测 lncRNAs	——	http://bioserver.iita.ac.in/deeplnc/
iSeeRNA	使用 SVM 算法预测 lncRNAs	——	http://137.189.133.71/iSeeRNA/
lncRNATargets	基于核酸热力学预测 lncRNAs 靶标	——	http://www.herbboi.org;8001/1rt/index.php
PLIT ^[90]	基于 L ₁ 正则化和迭代随机森林分类、序列和密码子偏好特征	——	——

针对以上问题,未来在植物中开展 lncRNA 的相关研究,首先需要不断完善和逐步建立有共同特征的长链非编码 RNA 文库,它将有助于确认和预测非编码 RNA 的功能特征;其次需要通过生物信息学结合实验组学的方法深入阐明 lncRNAs 基因调控机制,挖掘新的植物 lncRNAs,并探索其在植物生长发育、逆境胁迫、系统发育、品种改良等方面的功能及机制,这将对成功破解基因组的各种遗传信息,深入了解植物的生命代谢、系统发育等具有十分积极的意义。

参考文献(References)

- [1] CONSORTIUM E P, BIRNEY E, STAMATOYANNOPOULOS J A, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project[J]. *Nature*, 2007, 447 (7146): 799–816. DOI: 10.1038/nature05874.
- [2] WANG K C, YANG Y W, LIU B, et al. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression[J]. *Nature*, 2011, 472 (7341): 120–124. DOI: 10.1038/nature09819.
- [3] SCHAUKOWITCH K, KIM T K. Emerging epigenetic mechanisms of long non-coding RNAs[J]. *Neuroscience*, 2014, 264:25–38. DOI: 10.1016/j.neuroscience.2013.12.009.
- [4] XIAO B, ZHANG X, LI Y, et al. Identification, bioinformatic analysis and expression profiling of candidate mRNA-like non-coding RNAs in *Sus scrofa*[J]. *Journal of Genetics and Genomics*, 2009, 36 (12): 695–702. DOI: 10.1016/S1673-8527(08)60162-9.
- [5] LIU J, JUNG C, XU J, et al. Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in *Arabidopsis*[J]. *Plant Cell*, 2012, 24(11):4333–4345. DOI: 10.1105/tpc.112.102855.
- [6] HEO J B, LEE Y S, SUNG S. Epigenetic regulation by long noncoding RNAs in plants[J]. *Science*, 2013, 21(6–7):685–693. DOI:10.1007/s10577-013-9392-6.
- [7] RINN J L, CHANG H Y. Genome regulation by long noncoding RNAs[J]. *Annual Review of Biochemistry*, 2012, 81 (1): 145–166. DOI: 10.1146/annurev-biochem-051410-092902.
- [8] GUTTMAN M, RINN J L. Modular regulatory principles of large non-coding RNAs[J]. *Nature*, 2012, 482 (7385): 339–346. DOI:10.1038/nature10887.
- [9] TANG T H, POLACEK N, ZYWICKI M, et al. Identification of novel non-coding RNAs as potential antisense regulators in the archaeon *Sulfolobus solfataricus*[J]. *Molecular Microbiology*, 2005, 55 (2): 469–481. DOI:10.1111/j.1365-2958.2004.04428.x.
- [10] PONTING C P, OLIVER P L, REIK W. Evolution and functions of long noncoding RNAs[J]. *Cell*, 2009, 136 (4): 629–641. DOI: 10.1016/j.cell.2009.02.006.
- [11] WANG K C, CHANG H Y. Molecular mechanisms of long noncoding RNAs[J]. *Molecular Cell*, 2011, 43 (6): 904–914. DOI:10.1016/j.molcel.2011.08.018.
- [12] WILUSZ J E, SUNWOO H, SPECTOR D L. Long noncoding RNAs: Functional surprises from the RNA world[J]. *Genes & Development*, 2009, 23 (13): 1494–1504. DOI: 10.1101/gad.1800909.
- [13] MARTENS J A, LAPRADE L, WINSTON F. Intergenic transcription is required to repress the *Saccharomyces cerevisiae* SER3 gene[J]. *Nature*, 2004, 429 (6991): 571–574. DOI: 10.1038/nature02538.
- [14] CAMBLONG J, IGLESIAS N, FICKENTSCHER C, et al. Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in *S. cerevisiae*[J]. *Cell*, 2007, 131 (4): 706–717. DOI:10.1016/j.cell.2007.09.014.
- [15] ANNILO T, KEPP K, LAAN M. Natural antisense transcript of natriuretic peptide precursor A (NPPA): Structural organization and modulation of NPPA expression[J]. *BMC Molecular Biology*, 2009, 10 (1): 81. DOI: 10.1186/1471-2199-10-81.
- [16] OGAWA Y, SUN B K, LEE J T. Intersection of the RNA Interference and X-Inactivation Pathways[J]. *Science*, 2008, 320 (5881): 1336–1341. DOI: 10.1126/science.1157676.
- [17] FENG J, BI C, CLARK B S, et al. The Ebf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator[J]. *Genes & Development*, 2006, 20 (11): 1470–1484. DOI: 10.1101/gad.1416106.
- [18] FOX A H, LAM Y W, LEUNG A K, et al. Paraspeckles: A novel nuclear domain[J]. *Current Biology*, 2002, 12 (1): 13–25. DOI:10.1016/S0960-9822(01)00632-7.
- [19] WILLINGHAM A T, ORTH A P, BATALOV S, et al. A strategy for probing the function of noncoding RNAs finds a repressor of NFAT[J]. *Science*, 2005, 309 (5740): 1570–1573. DOI:10.1126/science.1115901.
- [20] 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*[J]. *Plant Cell*, 2004, 16 (4): 1047–1059. DOI:10.1105/tpc.019406.
- [21] HIRSCH J, LEFORT V, VANKERSCHAUVER M, et al. Characterization of 43 non-protein-coding mRNA genes in *Arabidopsis*, including the MIR162a-derived transcripts[J]. *Plant Physiology*, 2006, 140 (4): 1192. DOI: 10.1104/pp.105.073817.
- [22] DING J, LU Q, OUYANG Y, et al. A long noncoding RNA regulates photoperiod-sensitive male sterility, an essential component of hybrid rice[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2012, 109 (7): 2654–2659. DOI:10.1073/pnas.1121374109.

- [23] ZHOU H, LIU Q, LI J, et al. 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 [J]. *Cell Research*, 2012, 22(4): 649–660. DOI: 10.1038/cr.2012.28.
- [24] ZHU D, DENG X W. A non-coding RNA locus mediates environment-conditioned male sterility in *rice* [J]. *Cell Research*, 2012, 22(5): 791–792. DOI: 10.1038/cr.2012.43.
- [25] HEO J B, SUNG S. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA [J]. *Science*, 2011, 331(6013): 76–79. DOI: 10.1126/science.1197349.
- [26] SECCO D, BAUMANN A, POIRIER Y. Characterization of the *rice PHO1* gene family reveals a key role for *OsPHO1;2* in phosphate homeostasis and the evolution of a distinct clade in dicotyledons [J]. *Plant Physiology*, 2010, 152(3): 1693–1704. DOI: 10.1104/pp.109.149872.
- [27] JABNOUNE M, SECCO D, LECAMPION C, et al. A Rice cis-natural antisense RNA acts as a translational enhancer for its cognate mRNA and contributes to phosphate homeostasis and plant fitness [J]. *Plant Cell*, 2013, 25(10): 4166–4182. DOI: 10.1105/tpc.113.116251.
- [28] SWIEZEWSKI S, LIU F, MAGUSIN A, et al. Cold-induced silencing by long antisense transcripts of an *Arabidopsis* Polycomb target [J]. *Nature*, 2009, 462(7274): 799–802. DOI: 10.1038/nature08618.
- [29] ZHU Q H, STEPHEN S, TAYLOR J, et al. Long noncoding RNAs responsive to *Fusarium oxysporum* infection in *Arabidopsis thaliana* [J]. *New Phytologist*, 2014, 201(2): 574–584. DOI: 10.1111/nph.12537.
- [30] XIN M, YU W, YAO Y, et al. 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 [J]. *BMC Plant Biology*, 2011, 11(1): 1–13. DOI: 10.1186/1471-2229-11-61.
- [31] HELLIWELL C A, ROBERTSON M, FINNEGAN E J, et al. Vernalization-repression of *Arabidopsis FLC* requires promoter sequences but not antisense transcripts [J]. *Plos One*, 2011, 6(6): 240–247. DOI: 10.1371/journal.pone.0021513.
- [32] AMOR B B, WIRTH S, MERCHAN F, et al. Novel long non-protein coding RNAs involved in *Arabidopsis* differentiation and stress responses [J]. *Genome Research*, 2012, 19(1): 57. DOI: 10.1101/gr.080275.108.
- [33] MARTIN A C, DEL P J J, RUBIO V, et al. Influence of cytokinins on the expression of phosphate starvation-responsive genes in *Arabidopsis* [J]. *Plant Journal*, 2010, 24(5): 559–567. DOI: 10.1046/j.1365-313x.2000.00893.x.
- [34] HEUNGSOP S, HWA-SOO S, RUJIN C, et al. Loss of *At4* function impacts phosphate distribution between the roots and the shoots during phosphate starvation [J]. *Plant Journal for Cell & Molecular Biology*, 2010, 45(5): 712–726. DOI: 10.1111/j.1365-313x.2005.02629.x.
- [35] FRANCO-ZORRILLA J M, VALLI A, TODESCO M, et al. Target mimicry provides a new mechanism for regulation of microRNA activity [J]. *Nature Genetics*, 2007, 39(8): 1033–1037. DOI: 10.1038/ng2079.
- [36] LUKASHIN A, BORODOVSKY M. GeneMark.hmm: New solutions for gene finding [J]. *Nucleic Acids Research*, 1998, 26(4): 1107–1115. DOI: 10.1093/nar/26.4.1107.
- [37] LETIZIA D S, ANTONELLA B, ANDREA M. Bioinformatics tools and novel challenges in long non-coding RNAs (lncRNAs) functional analysis [J]. *International Journal of Molecular Sciences*, 2011, 13(1): 97–114. DOI: 10.3390/ijms13010097.
- [38] CHARNY P, NAMHEE Y, IKJUNG C, et al. lncRNAtor: A comprehensive resource for functional investigation of long non-coding RNAs [J]. *Bioinformatics*, 2014, 30(17): 2480–2485. DOI: 10.1093/bioinformatics/btu325.
- [39] YANG X, GAO L, GUO X, et al. A network based method for analysis of lncrna-disease associations and prediction of lncRNAs implicated in diseases [J]. *Plos One*, 2014, 9(1): e87797. DOI: 10.1371/journal.pone.0087797.
- [40] YE S, YANG L, ZHAO X, et al. Bioinformatics method to predict two regulation mechanism: TF-miRNA-mRNA and lncRNA-miRNA-mRNA in pancreatic cancer [J]. *Cell Biochemistry & Biophysics*, 2014, 70(3): 1849–1858. DOI: 10.1007/s12013-014-0142-y.
- [41] KHANG A P C, DENNIS E S, WANG M B. Analysis of argonaute 4-associated long non-coding RNA in *Arabidopsis thaliana* sheds novel insights into gene regulation through RNA-directed DNA methylation [J]. *Genes*, 2017, 8(8): 198. DOI: 10.3390/genes8080198.
- [42] SEVERING E, FAINO L, JAMGE S, et al. *Arabidopsis thaliana* ambient temperature responsive lncRNAs [J]. *BMC Plant Biology*, 2018, 18(1): 145. DOI: 10.1186/s12870-018-1362-x.
- [43] 鲁志, 吴玥, 邸超. 拟南芥全基因组范围的 non-polyA lncRNA 检测 [J]. *清华大学学报(自然科学版)*, 2014, 54(8): 1117–1121. DOI: 10.16511/j.cnki.qhdxxb.2014.08.003.
LU Zhi, WU Yue, DI Chao. Genome-wide identification of non-polyA lncRNAs in *Arabidopsis thaliana* [J]. *Journal of Tsinghua University*, 2014, 54(8): 1117–1121. DOI: 10.16511/j.cnki.qhdxxb.2014.08.003.
- [44] ROWLEY M J, BÖHMDORFER G, WIERZBICKI A T. Analysis of long non-coding RNAs produced by a specialized RNA polymerase in *Arabidopsis thaliana* [J]. *Methods*, 2013, 63(2): 160. DOI: 10.1016/j.ymeth.2013.05.006.
- [45] YAMAGUCHI A, ABE M. Regulation of reproductive development by non-coding RNA in *Arabidopsis*: To flower or not to flower [J]. *Journal of Plant Research*, 2012, 125(6): 693–704. DOI: 10.1007/s10265-012-0513-7.
- [46] QIN T, ZHAO H, CUI P, et al. A nucleus-localized long

- non-coding RNA enhances drought and salt stress tolerance [J]. *Plant Physiology*, 2017, 175(3):1321. DOI:10.1104/pp.17.00574.
- [47] YUAN J, YE Z, DONG J, et al. Systematic characterization of novel lncRNAs responding to phosphate starvation in *Arabidopsis thaliana* [J]. *BMC Genomics*, 2016, 17(1):655. DOI:10.1186/s12864-016-2929-2.
- [48] WU J, OKADA T, FUKUSHIMA T, et al. A novel hypoxic stress-responsive long non-coding RNA transcribed by RNA polymerase III in *Arabidopsis* [J]. *RNA Biology*, 2012, 9(3):302-313. DOI:10.4161/rna.19101.
- [49] VRBSKY J, AKIMCHEVA S, WATSON J M, et al. siRNA-Mediated methylation of *Arabidopsis* Telomeres [J]. *Plos Genetics*, 2010, 6(6):e1000986. DOI: 10.1371/journal.pgen.1000986.
- [50] KINDGREN P, ARD R, IVANOV M, et al. Transcriptional read-through of the long non-coding RNA SVALKA governs plant cold acclimation [J]. *Nature Communications*, 2018, 9(1):4561. DOI:10.1038/s41467-018-07010-6.
- [51] ZHAO Xinyue, LI Jingrui, LIAN Bi, et al. Global identification of *Arabidopsis* lncRNAs reveals the regulation of *MAF4* by a natural antisense RNA [J]. *Nature Communications*, 2018, 9(1):5056. DOI: 10.1038/s41467-018-07500-7.
- [52] WUNDERLICH M, GROß-HARDT R, SCHÖFFL F. Heat shock factor HSF2a involved in gametophyte development of *Arabidopsis thaliana* and its expression is controlled by a heat-inducible long non-coding antisense RNA [J]. *Plant Molecular Biology*, 2014, 85(6):541-550. DOI:10.1007/s11103-014-0202-0.
- [53] SEO J S, SUN H X, PARK B S, et al. ELF18-Induced long noncoding RNA associates with Mediator to enhance expression of innate immune response genes in *Arabidopsis* [J]. *Plant Cell*, 2017, 29(5):1024-1038. DOI: 10.1105/tpc.16.00886.
- [54] WASAKI J, YONETANI R, SHINANO T, et al. Expression of the *OsPI1* gene, cloned from rice roots using cDNA microarray, rapidly responds to phosphorus status [J]. *New Phytologist*, 2003, 158(2):239-248. DOI: 10.1046/j.1469-8137.2003.00748.x.
- [55] KOUCHI H, TAKANE K, SO R B, et al. Rice *ENOD40*: Isolation and expression analysis in rice and transgenic soybean root nodules [J]. *Plant Journal*, 1999, 18(2):121-129. DOI:10.1046/j.1365-313X.1999.00432.x.
- [56] LIU X, LI D, ZHANG D, et al. A novel antisense long noncoding RNA, *TWISTED LEAF*, maintains leaf blade flattening by regulating its associated sense R2R3-MYB gene in rice [J]. *New Phytologist*, 2018, 218(D1):774-788. DOI:10.1111/nph.15023.
- [57] FAN Y, YANG J, MATHIONI S M, et al. *PMS1T*, producing phased small-interfering RNAs, regulates photoperiod-sensitive male sterility in rice [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2016, 113(52):15144. DOI:10.1073/pnas.1619159114.
- [58] BOERNER S, MCGINNIS K M. Computational identification and functional predictions of long noncoding RNA in *Zea mays* [J]. *Plos One*, 2012, 7(8):e43047. DOI: 10.1371/journal.pone.0043047.
- [59] LI L, EICHTEN S R, SHIMIZU R, et al. Genome-wide discovery and characterization of maize long non-coding RNAs [J]. *Genome Biology*, 2014, 15(2):R40. DOI:10.1186/gb-2014-15-2-r40.
- [60] DAI X Y, YU J J, ZHAO Q, et al. Non-coding RNA for *ZM401*, a pollen-specific gene of *Zea mays* [J]. *Acta Botanica Sinica*, 2004, 46(4):497-504. DOI:10.1002/jcb.21807.
- [61] HAAG J R, PIKAARD C S. Multisubunit RNA polymerases IV and V: Purveyors of non-coding RNA for plant gene silencing [J]. *Nature Reviews Molecular Cell Biology*, 2011, 12(8):483-492. DOI:10.1038/nrm3152.
- [62] CHO J, KOO D H, NAM Y W, et al. Isolation and characterization of cDNA clones expressed under male sex expression conditions in a monoecious cucumber plant (*Cucumis sativus* L. cv. Winter Long) [J]. *Euphytica*, 2005, 146(3):271-281.
- [63] WEN J, PARKER B J, WEILLER G F. In silico identification and characterization of mRNA-like noncoding transcripts in *Medicago truncatula* [J]. *Silico Biology*, 2007, 7(4-5):485. DOI:10.1617/s11527-009-9500-4.
- [64] CRESPI M D, JURKEVITCH E, POIRET M, et al. *enod40*, a gene expressed during nodule organogenesis, codes for a non-translatable RNA involved in plant growth [J]. *Embo Journal*, 1994, 13(21):5099-5112. DOI:10.1002/j.1460-2075.1994.tb06839.x.
- [65] BURLEIGH S H, HARRISON M J. The down-regulation of *Mt4*-like genes by phosphate fertilization occurs systemically and involves phosphate translocation to the shoots [J]. *Plant Physiology*, 1999, 119(1):241-248. DOI:10.2307/4278617.
- [66] BURLEIGH S M, HARRISON M J. Characterization of the *Mt4* gene from *Medicago truncatula* [J]. *Gene*, 1998, 216(1):47-53. DOI:10.1016/s0378-1119(98)00326-6.
- [67] LIU C, MUCHHAL U S, RAGHOTHAMA K G. Differential expression of *TPS11*, a phosphate starvation-induced gene in *tomato* [J]. *Plant Molecular Biology*, 1997, 33(5):867-874. DOI:10.1023/A:1005729309569.
- [68] YANG W C, KATINAKIS P, HENDRIKS P, et al. Characterization of *GmENOD40*, a gene showing novel patterns of cell-specific expression during soybean nodule development [J]. *Plant Journal*, 1993, 3(4):573-585. DOI:10.1046/j.1365-313X.1993.03040573.x.
- [69] SONG Jianguhua, CAO Jiashu, YU Xiaolin, et al. *BcMF11*, a putative pollen-specific non-coding RNA from *Brassica campestris ssp. chinensis* [J]. *Journal of Plant Physiology*, 2007, 164(8):1097-1100. DOI:10.1016/j.jplph.2006.10.

- 002.
- [70] SONG Jianhua, CAO Jiashu, WANG Chenggang. *BcMF11*, a novel non-coding RNA gene from *Brassica campestris*, is required for pollen development and male fertility[J]. *Plant Cell Reports*, 2013, 32(1): 21–30. DOI: 10.1007/s00299-012-1337-6.
- [71] ZHANG Fang, DONG Heng, LIU Yanhong, et al. *BcMF11* and its homologous sequences may form a lncRNA family in *Brassica diploids* [J]. *Acta Physiologiae Plantarum*, 2018, 40(4): 65. DOI: 10.1007/s11738-018-2640-9.
- [72] ARIEL F, JEGU T, LATRASSE D, et al. Noncoding transcription by alternative RNA polymerases dynamically regulates an auxin-driven chromatin loop [J]. *Molecular Cell*, 2014, 55(3): 383–396. DOI: 10.1016/j.molcel.2014.06.011.
- [73] WANG Yuqiu, FAN Xiuduo, LIN Fang, et al. *Arabidopsis* noncoding RNA mediates control of photomorphogenesis by red light [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2014, 111(28): 10359–10364. DOI: 10.1073/pnas.1409457111.
- [74] AMARAL P P, CLARK M B, GASCOIGNE D K, et al. lncRNAdb: A reference database for long noncoding RNAs [J]. *Nucleic Acids Research*, 2011, 39(Database issue): 146–151. DOI: 10.1093/nar/gkq1138.
- [75] LAMESCH P, BERARDINI T Z, LI D, et al. The *Arabidopsis* information resource (TAIR): Improved gene annotation and new tools [J]. *Nucleic Acids Research*, 2012, 40(Database issue): 1202–1210. DOI: 10.1093/nar/gkr1090.
- [76] SWARBRECK D, WILKS C, LAMESCH P, et al. The *Arabidopsis* Information Resource (TAIR): Gene structure and function annotation [J]. *Nucleic Acids Research*, 2008, 36(Database issue): D1009. DOI: 10.1093/nar/gkm965.
- [77] CHEN Dijun, YUAN Chunhui, ZHANG Jian, et al. PlantNATsDB: A comprehensive database of plant natural antisense transcripts [J]. *Nucleic Acids Research*, 2012, 40(Database issue): D1187–1193. DOI: 10.1093/nar/gkr823.
- [78] JIN Jingjing, LIU Jun, WANG Huan, et al. PLncDB: Plant long non-coding RNA database [J]. *Bioinformatics*, 2013, 29(8): 1068–1071. DOI: 10.1093/bioinformatics/btt107.
- [79] BU Dechao, YU Kuntao, SUN Silong, et al. NONCODE v3.0: Integrative annotation of long noncoding RNAs [J]. *Nucleic Acids Research*, 2012, 40(Database issue): D210–D215. DOI: 10.1093/nar/gkr1175.
- [80] YI Xin, ZHANG Zhenhai, LING Yi, et al. PNRD: A plant non-coding RNA database [J]. *Nucleic Acids Research*, 2015, 43(Database issue): D982–989. DOI: 10.1093/nar/gku1162.
- [81] PAYTUVÍ GALLART A, HERMOSO PULIDO A, ANIAR MARTINEZ D L I, et al. GREENC: A Wiki-based database of plant lncRNAs [J]. *Nucleic Acids Research*, 2016, 44(Database issue): D1161–D1166. DOI: 10.1093/nar/gkv1215.
- [82] SZCZEŚNIAK M W, ROSIKIEWICZ W, MAKALOWSKA I. CANTATAdb: A collection of plant long non-coding RNAs [J]. *Plant & Cell Physiology*, 2015, 57(1): e8. DOI: 10.1093/pcp/pcv201.
- [83] XUAN Hongdong, ZHANG Linzhong, LIU Xueshi, et al. PLnlncRbase: A resource for experimentally identified lncRNAs in plants [J]. *Gene*, 2015, 573(2): 328–332. DOI: 10.1016/j.gene.2015.07.069.
- [84] LIN M F, JUNGREIS I, KELLIS M. PhyloCSF: A comparative genomics method to distinguish protein-coding and non-coding regions [J]. *Bioinformatics*, 2011, 27(13): i275–i282. DOI: 10.1093/bioinformatics/btr209.
- [85] KONG Lei, ZHANG Yong, YE Zhiqiang, et al. CPC: Assess the protein-coding potential of transcripts using sequence features and support vector machine [J]. *Nucleic Acids Research*, 2007, 35(Web Server issue): W345. DOI: 10.1093/nar/gkm391.
- [86] SUN Liang, LUO Haitao, BU Dechao, et al. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts [J]. *Nucleic Acids Research*, 2013, 41(17): e166–e166. DOI: 10.1093/nar/gkt646.
- [87] TRIPATHI R, PATEL S, KUMARI V, et al. DeepLNC, a long non-coding RNA prediction tool using deep neural network [J]. *Network Modeling Analysis in Health Informatics & Bioinformatics*, 2016, 5(1): 21. DOI: 10.1007/s13721-016-0129-2.
- [88] SUN Kun, CHEN Xiaona, JIANG Peiyong, et al. iSeeRNA: Identification of long intergenic non-coding RNA transcripts from transcriptome sequencing data [J]. *BMC Genomics*, 2013, 14(S2): S7. DOI: 10.1186/1471-2164-14-S2-S7.
- [89] HU Ruifeng, SUN Xiaobo. lncmatargets: A platform for lncRNA target prediction based on nucleic acid thermodynamics [J]. *Journal of Bioinformatics and Computational Biology*, 2016, 14(04): 1650016. DOI: 10.1142/S0219720016500165.
- [90] DESHPANDE S, SHUTTLEWORTH J, YANG J, et al. PLIT: An alignment-free computational tool for identification of long non-coding RNAs in plant transcriptomic datasets [J/OL]. <http://xueshu.baidu.com/usercenter/paper/j.complbiomed>. 2018. 12. 014. DOI: 10.1016/j.complbiomed.2018.12.014.