



## 转座子来源的植物长链非编码RNA

王益军<sup>1,2\*</sup>, 王亚丽<sup>1,2</sup>, 陈煜东<sup>1,2</sup>

<sup>1</sup>扬州大学农学院, 江苏省作物遗传生理重点实验室/植物功能基因组学教育部重点实验室/江苏省作物基因组学和分子育种重点实验室, 扬州 225009; <sup>2</sup>扬州大学, 江苏省粮食作物现代产业技术协同创新中心, 扬州 225009

**摘要** 转座子是基因组的重要组成部分, 影响基因组的结构与稳定。长链非编码RNA (lncRNA)在转录及转录后水平调控多个生物学过程。转座子与lncRNA是物种进化的重要驱动力。含有转座子序列的lncRNA在自然界广泛存在。该文对植物lncRNA的发掘策略和功能研究进行概述, 围绕植物转座子来源lncRNA (TE-lncRNA)的分布和功能展开综述, 并对植物TE-lncRNA的调控机制、表观修饰及育种潜能等进行探讨与展望。

**关键词** 转座子, 长链非编码RNA, 转座子来源的长链非编码RNA, 植物

王益军, 王亚丽, 陈煜东 (2020). 转座子来源的植物长链非编码RNA. 植物学报 55, 768–776.

转座子是基因组中可移动的元件。根据不同的转座机制, 转座子可分为逆转座子(Class I)和DNA转座子(Class II) (Feschotte et al., 2002)。逆转座子通过RNA中介物, 在逆转座酶作用下进行转座。逆转座子主要有长末端重复序列(long terminal repeat, LTR)逆转座子和非LTR逆转座子。LTR逆转座子主要包括Copia和Gypsy类。非LTR逆转座子包括长散布重复元件(long interspersed nuclear element, LINE)和短散布重复元件(short interspersed nuclear element, SINE)。DNA转座子是以DNA到DNA的方式在原位置解离, 与新位置整合, 包括Ac/Ds、Spm/dSpm和Mutator转座子等(Dooner and Belachew, 1991; Bunkers et al., 1993; Cong and Li, 2020)。Helitron转座子是一种特殊类型的DNA转座子, 该转座子不具有末端反向重复序列, 也不产生靶点重复序列。Helitron以滚环复制的方式转座(Kapitonov and Jurka, 2001)。

RNA聚合酶II (Pol II)转录产生2种类型的RNA: 编码RNA与非编码RNA (noncoding RNA, ncRNA)。ncRNA也可以由RNA聚合酶I、III、IV、V (Pol I、III、IV、V)转录产生(Ulitsky, 2018)。根据成熟产物的大小,

ncRNA分为非编码小RNA (small noncoding RNA, sRNA)和长链非编码RNA (long noncoding RNA, lncRNA)。sRNA长度一般为20–24 nt。不同前体产生不同的sRNA。单链前体产生microRNA (miRNA)和非miRNA。双链和小干扰RNA (small interfering RNA, siRNA)前体产生次级siRNA和异源siRNA等(Axtell, 2013)。lncRNA的长度大于200 nt, 真核生物启动子、增强子和基因间等区域转录可以产生lncRNA。启动子上游区域、增强子、基因间区域及编码蛋白基因转录分别产生启动子上游转录本(promoter upstream transcript, PROMPT)、增强子RNA (enhancer RNA, eRNA)、基因间lncRNA (long intergenic noncoding RNA, lincRNA)和天然反义转录本(natural antisense transcript, NAT) (Wu et al., 2017)。

转座子和lncRNA皆为基因组中重要的调控元件。已有研究表明, 含有转座子序列的lncRNA在自然界广泛存在, 称为转座子来源的lncRNA (transposable element-derived lncRNA, TE-lncRNA) (Wang et al., 2017; Golicz et al., 2018)。在水稻(*Oryza sativa*)、玉米(*Zea mays*)和棉花(*Gossypium raimondii*)等植物中, TE-lncRNA主要来源于逆转座子序

收稿日期: 2020-05-26; 接受日期: 2020-09-03

基金项目: 国家自然科学基金(No.31571671)、扬州大学高端人才支持计划(No.18HTYZU12)、扬州大学青蓝工程(No.QLYZU201809)、扬州大学科技创新培育基金(No.2019CXJ097)和江苏高校优势学科建设工程(No.PAPD)

\* 通讯作者。E-mail: wyj@yzu.edu.cn

列(Wang et al., 2017; Zhao et al., 2018a; Lv et al., 2019)。与来源于编码蛋白基因反义转录的lncRNA相比, 植物TE-lncRNA的功能研究报道较少。拟南芥(*Arabidopsis thaliana*)和棉花TE-lncRNA的研究表明, TE-lncRNA调控株高和非生物胁迫应答等(Wang et al., 2016, 2017; Zhao et al., 2018a)。重要粮食作物(如水稻、小麦(*Triticum aestivum*)和玉米)中TE-lncRNA的功能研究报道较少。鉴于TE-lncRNA在植物中广泛分布, 亟待开展TE-lncRNA调控作物株型、产量、品质和抗性等重要性状的机制研究。本文综述了植物TE-lncRNA研究概况, 重点介绍TE-lncRNA在植物中的分布与功能, 并对植物TE-lncRNA的基础研究与应用潜势进行了展望。

## 1 植物基因组转座子构成与物种进化

转座子是基因组的重要组成部分, 影响基因组的稳定与进化。以基因组较小的物种拟南芥(~125 Mb)与基因组较大的物种玉米(~2.3 Gb)为例, 介绍植物基因组中转座子分布的特点以及转座子构成在基因组进化中的作用。至少10%的拟南芥基因组序列由转座子构成, 其中逆转座子、DNA转座子以及其它类型转座子的百分比分别约为38%、39%和23% (图1A)。在全部转座子中, LTR逆转座子约占28% (The *Arabidopsis* Genome Initiative, 2000)。基因组较大的物种玉米, 约85%的基因组为转座子所占据。在全部转座子中, 逆转座子和DNA转座子的百分比分别约为89%和11%。LTR逆转座子占全部转座子的86% (图1B) (Schnable et al., 2009)。通过分析拟南芥和玉米等物种转座子的组成, 发现不同植物中存在特有的转座子类型。转座子的作用方式也存在一定差异。逆转座子, 尤其是LTR逆转座子的“复制-粘贴”式扩增导致一些物种基因组大小增加。

转座子不仅影响基因组大小, 而且影响基因组功能。例如, 通过分析玉米z1C1位点的单倍型, 发现Helitron转座子携带胞苷脱氨酶基因ZmCDA3 (Xu and Messing, 2006)。Copia类LTR逆转座子插入MADS-box转录因子基因J2 (JOINTLESS2), 导致番茄(*Solanum lycopersicum*)果柄离区消失(Alonge et al., 2020)。可见, 转座子影响基因的结构与功能, 是物种进化的重要驱动力。

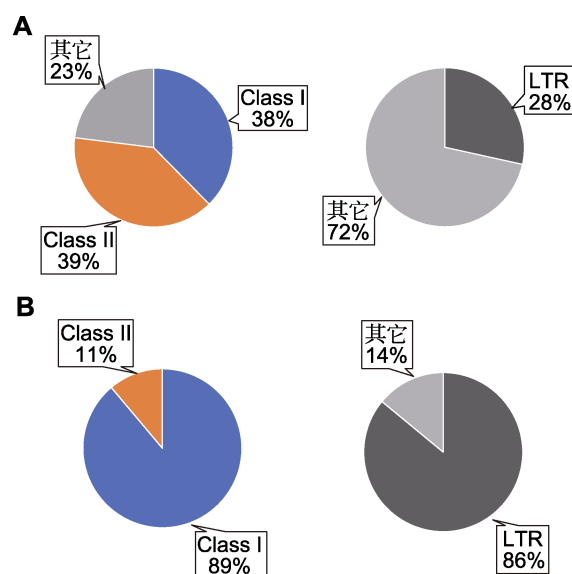


图1 拟南芥和玉米基因组中转座子的分布(数据来源: The *Arabidopsis* Genome Initiative, 2000; Schnable et al., 2009) (A) 拟南芥基因组中转座子分布(左图)和拟南芥长末端重复序列(LTR)逆转座子占全部转座子的比例(右图); (B) 玉米基因组中转座子分布(左图)和玉米LTR逆转座子占全部转座子的比例(右图)。Class I: 逆转座子; Class II: DNA转座子

Figure 1 Transposon composition in the genomes of *Arabidopsis* and maize (data source: The *Arabidopsis* Genome Initiative, 2000; Schnable et al., 2009) (A) Transposon composition in *Arabidopsis* genome (left) and the proportion of *Arabidopsis* long terminal repeat (LTR) retrotransposons (right); (B) Transposon composition in maize genome (left) and the proportion of maize LTR retrotransposons (right). Class I: Retrotransposons; Class II: DNA retrotransposons

## 2 植物lncRNA发掘

### 2.1 链特异性转录组测序发掘植物lncRNA

构建链特异性RNA文库并进行转录组测序(strand-specific RNA sequencing, ssRNA-seq), 对lncRNA进行高通量发掘。与普通RNA-seq相比, ssRNA-seq可以区分reads来源于正义链还是反义链, 在转录本定量时可以排除互补链转录本表达的干扰, 使表达定量更精确。此外, ssRNA-seq有助于确定lincRNA的转录方向(Zhao et al., 2015)。

采用ssRNA-seq, 已经对拟南芥、水稻和玉米的lncRNA进行了分离。细菌表面翻译延伸因子Tu (translation elongation factor Tu, EF-Tu)被受体识别结

合, 激活免疫反应。elf18是EF-Tu的保守肽段(Yang et al., 2020)。用elf18处理拟南芥幼苗, 通过ssRNA-seq发掘elf18诱导的lncRNA, 进而解析植物先天免疫的调控机制(Seo et al., 2017)。通过构建正常生长、干旱、低温和ABA处理等胁迫条件下的拟南芥全细胞、细胞核、细胞质组分的全核糖体、poly(A)和非poly(A)文库, 采用ssRNA-seq共发掘到6 510个拟南芥lncRNA (Zhao et al., 2018b)。选取白叶枯病菌感染不同时长的小稻叶片, 采用ssRNA-seq检测到567个白叶枯病菌响应的lncRNA (Yu et al., 2020)。对亚洲栽培稻(*O. sativa*)和普通野生稻(*O. rufipogon*)发育早期的幼穗进行ssRNA-seq, 发掘到3 363个lncRNA, 包括2 517个lncRNA、200个incRNA (intronic noncoding RNA)和646个NAT-lncRNA。分析亚洲栽培稻和普通野生稻lncRNA序列和表达变异, 有助于揭示稻属驯化进程中lncRNA的选择模式(Zheng et al., 2019)。通过去核糖体文库的ssRNA-seq, 发掘到玉米磷缺乏诱导和赤霉素响应的lncRNA (Du et al., 2018; Wang et al., 2018b)。随着高通量测序技术的发展, 通过ssRNA-seq已经发掘到大量的植物lncRNA。基于发掘的lncRNA进行功能验证和调控机制解析是植物lncRNA研究领域的重要挑战。

2.2 生物信息学分析发掘植物lncRNA

NCBI和EBI等数据库中已有海量的基因表达数据, 包括表达序列标签(expressed sequence tag, EST)、

tiling array和转录组数据等。基于数据库中的基因表达数据, 可以通过生物信息学方法发掘lncRNA。采用RepTAS (reproducibility-based tiling array analysis strategy)等方法, 整合拟南芥EST、cDNA、tiling array和基因组数据信息, 发掘到6 480个lncRNA (Liu et al., 2012)。通过分析EST和全长cDNA序列数据库, 发掘到水稻中的lncRNA (Liu et al., 2018)。整合数据库中的全基因组测序、转录组测序、含有全长cDNA的EST数据等信息, 发掘到20 163个玉米lncRNA (Li et al., 2014)。

随着多种组学的发展, 有些植物中已有丰富的公共数据可供lncRNA发掘。数据库中的数据质量直接影响到lncRNA发掘的可靠性。lncRNA发掘所用软件种类和阈值设定也会影响lncRNA的发掘结果。此外, 植物lncRNA具有序列保守性较低、丰度低以及组织特异性表达等特点(Huang et al., 2018), 对通过整合不同来源数据进行lncRNA发掘提出了新的挑战。在通过生物学信息学发掘植物lncRNA的过程中, 控制数据质量、权衡阈值设定、采集多维数据, 将有助于获得较为全面且可靠的植物lncRNA信息。

3 植物lncRNA的功能

基于ssRNA-seq和生物信息学分析等方法, 已发掘到大量的植物lncRNA, 但功能及其调控机制得以解析的lncRNA较少(表1)。植物lncRNA调控生长发

表1 植物长链非编码RNA (lncRNA)功能  
Table 1 Function of plant long noncoding RNA (lncRNA)

物种	lncRNA ID	lncRNA调控通路	参考文献
拟南芥	COLDAIR (COLD ASSISTED INTRONIC NONCODING RNA)	成花转变	Heo and Sung, 2011
	COOLAIR (cold induced long antisense intragenic RNA)	成花转变	Swiezewski et al., 2009
	DRIR (DROUGHT INDUCED lncRNA)	干旱和盐胁迫应答	Qin et al., 2017
	ELENA1 (ELF18-INDUCED LONG-NONCODING RNA1)	先天免疫反应	Seo et al., 2017
	MAS (lncRNA from MADS AFFECTING FLOWERING4)	春化响应	Zhao et al., 2018b
	T5120	NO <sub>3</sub> <sup>-</sup> 同化	Liu et al., 2019
	TE-lncRNA11195*	ABA响应	Wang et al., 2017
水稻	ALEX1 (An Leaf Expressed and Xoo-induced lncRNA 1)	白叶枯病抗性	Yu et al., 2020
	Ef-cd (Early flowering-completely dominant)	开花期和产量	Fang et al., 2019
	LAIR (LRK Antisense Intergenic RNA)	产量	Wang et al., 2018a
	TL (TWISTED LEAF)	叶片发育	Liu et al., 2018
玉米	PILNCR1 (Pi-deficiency-induced long non-coding RNA 1)	磷胁迫应答	Du et al., 2018
棉花	XLOC_409583*	苗高	Zhao et al., 2018a
	lncRNA-314*	果实成熟	Wang et al., 2016
番茄	lncRNA16397	晚疫病抗性	Cui et al., 2017
	lncRNA39026	晚疫病抗性	Hou et al., 2020

\* 转座子来源的lncRNA (TE-lncRNA)。\* Transposon-derived lncRNAs (TE-lncRNAs)。

育、生物/非生物逆境应答和养分平衡等多个生物学过程。

### 3.1 植物lncRNA调控生长发育

拟南芥lncRNA功能研究的开创性工作集中于调控 *FLC* (*FLOWERING LOCUS C*) 基因的lncRNA研究。*FLC*是春化途径关键基因, 调控拟南芥的成花转变。*FLC*受到多个lncRNA的调控。拟南芥 *COOLAIR* (lncRNA cold induced long antisense intragenic RNA) 是 *FLC* 的反义转录本, *COOLAIR* 与 polycomb 复合体互作, 对 *FLC* 进行表观抑制 (Swiezewski et al., 2009)。与 *COOLAIR* 类似, 来自内含子的拟南芥 *COLD AIR* (lncRNA COLD ASSISTED INTRONIC NONCODING RNA) 与 polycomb 复合体互作, 抑制 *FLC* 表达, 进而调控春化过程 (Heo and Sung, 2011)。拟南芥 *MAF4* (*MADS AFFECTING FLOWERING 4*) 反义转录产生 lncRNA MAS。lncRNA MAS 参与春化响应, 为 *MAF4* 转录激活所必需 (Zhao et al., 2018b)。水稻 lncRNA 调控叶片发育、开花期和产量等。水稻 *R2R3-MYB* 转录因子基因 *OsMYB60* 反向转录产生 *TL* (lncRNA TWISTED LEAF), *TL* 调控水稻叶片发育 (Liu et al., 2018)。*Ef-cd* (*Early flowering-completely dominant*) 座位编码 lncRNA, 该 lncRNA 来源于成花激活基因 *OsSOC1* 的反义转录。*Ef-cd* 变异可以促进水稻提早开花而不影响产量 (Fang et al., 2019; 张硕和吴昌银, 2019)。LRK (leucine-rich repeat receptor kinase) 簇反向转录产生 *LAIR* (lncRNA LRK Antisense Intergenic RNA), 过表达 *LAIR* 可提高水稻产量 (Wang et al., 2018a)。

### 3.2 植物lncRNA参与生物与非生物逆境应答

过表达拟南芥 *ELENA1* (lncRNA ELF18-INDUCED LONG-NONCODING RNA 1) 可增强细菌性疫病菌抗性, *ELENA1* 敲除对细菌性疫病菌敏感, 表明 *ELENA1* 正向调控细菌性疫病菌抗性。lncRNA *ELENA1* 与 MED19a (Mediator subunit 19a) 互作, 通过影响 *PR1* (*PATHOGENESIS-RELATED GENE 1*) 的表达调控先天免疫反应 (Seo et al., 2017)。过表达 *ALEX1* (lncRNA An Leaf Expressed and Xoo-induced lncRNA 1) 可激活茉莉酸途径, 增强水

稻白叶枯病的抗性 (Yu et al., 2020)。番茄 *SIGRX22* 基因反义转录产生 lncRNA16397。过表达 *lncRNA16397* 诱导 *SIGRX* 表达, 进而降低 ROS 的积累, 减轻细胞膜损伤, 增强对番茄晚疫病的抗性 (Cui et al., 2017)。番茄 lncRNA39026 含有 miR168a 的靶标。过表达 *lncRNA39026* 株系其 miR168a 表达量下降, miR168a 靶标基因 *SIAGO1* 表达量上升。*lncRNA39026* 可能通过“诱饵”策略抑制 miR168a, 诱导病程相关蛋白基因表达, 进而调控番茄晚疫病抗性 (Hou et al., 2020)。lncRNA 的非生物胁迫响应研究主要集中在拟南芥 lncRNA 对干旱和盐胁迫应答机制的解析。拟南芥 *DRIR* (*DROUGHT INDUCED lncRNA*) 正向调控干旱和盐胁迫应答 (Qin et al., 2017)。拟南芥 lncRNA *AtR8* 参与种子萌发过程中的盐胁迫响应 (张楠等, 2020)。

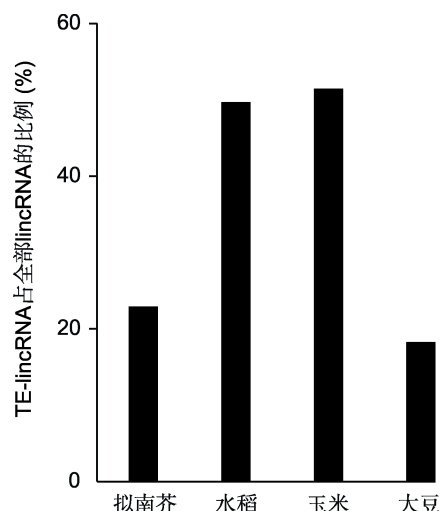
### 3.3 植物lncRNA调控养分平衡

植物 lncRNA 参与调控氮和磷等营养元素的平衡。拟南芥 NLP (NIN-like protein) 家族成员 AtNLP7 是氮响应的重要调控因子 (Wang et al., 2020)。lncRNA *T5120* 作用于 AtNLP7 的下游, 调控  $\text{NO}_3^-$  同化, 影响氮肥利用率 (Liu et al., 2019)。与拟南芥和水稻 lncRNA 研究相比, 玉米中功能解析的 lncRNA 较少。*miR399/PHO2* (*PHOSPHATE2*) 模块调控磷缺乏响应。磷缺乏诱导的玉米 *PILNCR1* (lncRNA Pi-deficiency-induced long non-coding RNA 1) 抑制 *miR399* 介导的 *PHO2* 剪切, *PILNCR1* 与 *miR399* 协同调控玉米对低磷的适应性 (Du et al., 2018)。

## 4 转座子来源的植物lncRNA

### 4.1 转座子来源植物lncRNA的分布

调控机制已得到解析的植物 lncRNA 多来源于编码蛋白基因的反义转录 (表1)。通过 ssRNA-seq 和生物信息学分析等方法, 在拟南芥、水稻、玉米、大豆 (*Glycine max*)、棉花和番茄等植物中发掘到一些 lncRNA 并非来自编码蛋白基因的反义转录本, 而是来源于转座子序列, 这些 lncRNA 被称为转座子来源的 lncRNA (transposable element-derived lncRNA, TE-lncRNA) (图2)。



**图2** 植物中转座子来源的基因间长链非编码RNA (TE-lincRNA) 占全部lincRNA的比例(数据来源: Wang et al., 2017; Golicz et al., 2018)

**Figure 2** The proportion of plant transposable element-derived long intergenic noncoding RNA (TE-lincRNA) (data source: Wang et al., 2017; Golicz et al., 2018)

通过ssRNA-seq, 在拟南芥、水稻和玉米中分别鉴定到47、611和398个TE-lincRNA, 约占全部lincRNA的23%、50%和51% (Wang et al., 2017)。约18%的大豆lincRNA含有转座子序列, 来源于逆转座子的TE-lincRNA多于来源于DNA转座子的TE-lincRNA (Golicz et al., 2018)。基于PacBio fl-cDNA和转录组等数据, 发掘玉米全基因组的lincRNA。结果表明, 约65%的lincRNA含有转座子序列, 属于TE-lincRNA。玉米TE-lincRNA含有的转座子序列主要来源于LTR逆转座子, LTR逆转座子来源的TE-lincRNA约占全部TE-lincRNA的86% (Lv et al., 2019)。整合ssRNA-seq、NCBI和EBI中的poly(A)转录组数据, 进行番茄lincRNA发掘。结果表明, 85%的番茄lincRNA含有转座子序列, 番茄基因组中存在LTR逆转座子来源的TE-lincRNA (Wang et al., 2016)。在棉花TE-lincRNA中, Gypsy类LTR逆转座子来源的TE-lincRNA占主要部分(Zhao et al., 2018a)。

TE-lincRNA的分布与基因组中转座子类型和数目相关。水稻、玉米、大豆、棉花和番茄中的TE-lincRNA主要来源于逆转座子(Wang et al., 2016, 2017; Golicz et al., 2018; Zhao et al., 2018a; Lv et al., 2019)。在这些物种中, 逆转座子数目的增加导致逆转座子来

源的lincRNA在整个TE-lincRNA占优势地位。

## 4.2 转座子来源植物lincRNA的功能

植物中lincRNA的功能研究集中于编码蛋白基因反义转录产生的lincRNA, 转座子来源lincRNA的功能研究报道较少, 主要集中在拟南芥、番茄和棉花等植物。TE-lincRNA调控株高、果实成熟和非生物胁迫应答等生物学过程(表1)。

棉花TE-lincRNA XLOC\_409583来源于去甲基化的LINE类逆转座子。与对照相比, 病毒诱导TE-lincRNA XLOC\_409583沉默的植株苗高增加, 表明TE-lincRNA XLOC\_409583调控棉花株高(Zhao et al., 2018a)。lincRNA-314来源于LTR逆转座子, 在番茄果实中特异表达。lincRNA-314与ABC转运基因共表达。在番茄果实成熟突变体rin和ful1/ful2中, lincRNA-314与ABC转运基因的表达被抑制。LTR逆转座子来源的TE-lincRNA lincRNA-314调控番茄果实成熟(Wang et al., 2016)。高温、低温、盐害、干旱和ABA处理等非生物胁迫诱导TE-lincRNA表达(Wang et al., 2017; Lv et al., 2019)。与野生型相比, 来源于LTR逆转座子的拟南芥TE-lincRNA TE-lincRNA11195突变体根长和幼茎鲜重增加, 对ABA敏感性下降, 表明TE-lincRNA TE-lincRNA11195调控ABA响应。TE-lincRNA11195中转座子的LTR序列缺失导致ABA敏感性下降, 表明转座子的LTR序列为TE-lincRNA TE-lincRNA11195调控非生物胁迫应答所必需(Wang et al., 2017)。

在整个植物lincRNA中, TE-lincRNA是重要组分(图2)。例如, 玉米基因组中, 超过一半的lincRNA属于TE-lincRNA (Wang et al., 2017; Lv et al., 2019)。然而, 功能得以解析的植物TE-lincRNA却较少。开展植物TE-lincRNA功能研究不仅有助于丰富lincRNA调控的理论基础, 而且能进一步揭示lincRNA的起源与进化。

## 5 结语与展望

### 5.1 植物转座子与lincRNA的相互作用

转座子是生物体的重要组成部分, 影响基因组的结构与稳定(Kidwell and Lisch, 2000)。lincRNA可以在转录和转录后水平对基因表达进行调控(Wu et al., 2017)。在物种进化过程中, 转座子与lincRNA相互作

用。一方面, 转座子影响植物lncRNA的产生。许多植物lncRNA含有转座子序列(图2) (Wang et al., 2016, 2017; Golicz et al., 2018; Zhao et al., 2018a; Lv et al., 2019)。由于非转座子来源的lncRNA比TE-lncRNA受到更强的选择压力, 因此在进化过程中, 植物lncRNA倾向于保留转座子序列(Khanduja et al., 2016; Zhao et al., 2018a)。此外, 植物转座子的表观修饰影响lncRNA的转录活性。例如, 棉花LINE类逆转座子去甲基化, 产生具有转录活性的TE-lncRNA (Zhao et al., 2018a)。TE-lncRNA中的转座子序列形成二级结构(与lncRNA功能相关) (Khanduja et al., 2016)。另一方面, lncRNA影响转座子活性。lncRNA可以作为small RNA前体, small RNA可导致转座子“沉默”(Kidwell and Lisch, 2000; Yoon et al., 2014)。植物转座子以及lncRNA的相互作用不仅增强了各自结构与功能的多样性, 也有助于驱动生物进化和提高适应性。

## 5.2 植物TE-lncRNA调控机制解析

转基因验证是TE-lncRNA功能解析的重要内容。基于CRISPR-Cas的基因编辑技术已成为植物基因功能验证的重要方法(Liu et al., 2020)。植物TE-lncRNA呈现序列保守性低和表达丰度低等特点, 给基因编辑工作带来了挑战。TE-lncRNA的低序列保守性, 可能导致在转基因受体中存在序列变异, 包括插入、缺失和碱基改变等, 从而影响gRNA (guide RNA)设计、编辑事件检测以及转基因结果的可靠性等。TE-lncRNA的低丰度影响编辑效率。低丰度的植物TE-lncRNA要求多靶点敲除并增加转化事件的数量。

基于功能验证解析lncRNA的调控机制, 是lncRNA研究的热点与难点。lncRNA多与靶标互作, 调控众多生物学过程。因此, 靶标发掘是lncRNA调控机制解析的关键。已报道调控机制的植物lncRNA多来源于已知编码蛋白基因的反义转录本, lncRNA调控机制的研究多围绕已知编码蛋白基因-已知编码蛋白基因反义转录形成的lncRNA这一模块展开。在该模块中, lncRNA互作靶标为其来源的已知编码蛋白基因(表1)。植物TE-lncRNA来源于转座子, 并非已知编码蛋白基因的转录本。可以通过RNA pull-down联合质谱分析来获得植物TE-lncRNA的靶标(Barnes and Kanhere, 2016)。发掘的植物TE-lncRNA靶标可通过

RIP (RNA immunoprecipitation) 进一步验证(Gagliardi and Matarazzo, 2016)。基于发掘的靶标, 围绕TE-lncRNA与靶标互作, 解析植物TE-lncRNA的调控机制。

## 5.3 植物TE-lncRNA表观修饰

转座元件的转座可能会产生突变, 负向的选择压力会限制转座活性, 但转座活性必须保持, 转座子进化就是这两种力量的一个动态平衡(Martin and Garfinkel, 2003)。植物中RNA依赖的DNA甲基化(RNA-directed DNA methylation, RdDM)是转座活性的重要调控方式(Deniz et al., 2019)。TE-lncRNA是来源于转座子的lncRNA, TE-lncRNA中转座子的DNA甲基化程度如何? DNA甲基化与TE-lncRNA活性的相关性如何? 通过比较棉花F<sub>1</sub>杂交种中全基因组转座子以及lncRNA来源转座子的DNA甲基化水平, 结果表明棉花F<sub>1</sub>杂交种中, 具有活性的TE-lncRNA主要来自去甲基化的转座子区域, 并且以去甲基化的LINE类逆转座子转录产生为主。此外, 棉花F<sub>1</sub>中DNA甲基化程度与TE-lncRNA的表达呈负相关(Zhao et al., 2018a)。可见, 转座元件的表观修饰影响植物TE-lncRNA的产生与功能。除了DNA甲基化, siRNA介导的共抑制和组蛋白H3赖氨酸9甲基化(histone H3 lysine-9 methylation, H3K9me)等表观修饰也可以调控植物转座子活性(Lippman et al., 2003)。RNA抑制和组蛋白修饰等表观调控对植物TE-lncRNA的产生、活性以及功能的影响, 尚待进一步研究。

## 5.4 植物TE-lncRNA的应用潜势

TE-lncRNA在植物中广泛存在(图2) (Wang et al., 2016, 2017; Golicz et al., 2018; Zhao et al., 2018a; Lv et al., 2019)。然而, 功能得以解析的植物TE-lncRNA较少。拟南芥、棉花和番茄TE-lncRNA分别调控ABA响应、苗高和果实成熟(表1)。不同植物以及全基因组范围功能性TE-lncRNA的数目有待系统研究。尤其是重要粮食作物, 如水稻、小麦和玉米中TE-lncRNA的功能研究亟待开展。

已有研究表明, lncRNA影响作物产量、品质和抗性等重要性状(表1)。因此, 植物lncRNA具有育种应用的潜势。鉴于TE-lncRNA在植物中广泛存在, 开展TE-lncRNA调控株型、产量、品质和抗性等重要性状



的机制研究, 发掘具有育种效用的TE-lncRNA, 具有重要的科学意义与应用价值。基于具有应用潜力的TE-lncRNA及其互作靶标, 通过单倍型发掘和基因编辑等策略, 寻找优良变异并进行材料创制, 有望为种质改良与品种选育等育种实践提供新的靶标和路径。

## 参考文献

- 张楠, 刘自广, 孙世臣, 刘圣怡, 林建辉, 彭凝芳, 张晓旭, 杨贺, 岑曦, 吴娟 (2020). 拟南芥 *AtR8* lncRNA对盐胁迫响应及其对种子萌发的调节作用. 植物学报 **55**, 421–429.
- 张硕, 吴昌银 (2019). 长链非编码RNA基因 *Ef-cd* 调控水稻早熟与稳产. 植物学报 **54**, 550–553.
- Alonge M, Wang XG, Benoit M, Soyk S, Pereira L, Zhang L, Suresh H, Ramakrishnan S, Maumus F, Ciren D, Levy Y, Harel TH, Shalev-Schlosser G, Amsellem Z, Razifard H, Caicedo AL, Tieman DM, Klee H, Kirsche M, Aganezov S, Ranallo-Benavidez TR, Lemmon ZH, Kim J, Robitaille G, Kramer M, Goodwin S, McCombie WR, Hutton S, Van Eck J, Gillis J, Eshed Y, Sedlazeck FJ, van der Knaap E, Schatz MC, Lippman ZB (2020). Major impacts of widespread structural variation on gene expression and crop improvement in tomato. *Cell* **182**, 145–161.
- Axtell MJ (2013). Classification and comparison of small RNAs from plants. *Annu Rev Plant Biol* **64**, 137–159.
- Barnes C, Kanhere A (2016). Identification of RNA-protein interactions through *in vitro* RNA pull-down assays. In: Lanzaolo C, Bodega B, eds. Polycomb Group Proteins. Methods in Molecular Biology, Vol. 1480. New York: Humana Press. pp. 99–113.
- Bunkers G, Nelson OE Jr, Raboy V (1993). Maize *bronze 1:dSpm* insertion mutations that are not fully suppressed by an active *Spm*. *Genetics* **134**, 1211–1220.
- Cong CS, Li YB (2020). Progress on *Mutator* superfamily. *Hereditas* **42**, 131–144.
- Cui J, Luan YS, Jiang N, Bao H, Meng J (2017). Comparative transcriptome analysis between resistant and susceptible tomato allows the identification of lncRNA-16397 conferring resistance to *Phytophthora infestans* by co-expressing glutaredoxin. *Plant J* **89**, 577–589.
- Deniz Ö, Frost JM, Branco MR (2019). Regulation of transposable elements by DNA modifications. *Nat Rev Genet* **20**, 417–431.
- Dooner HK, Belachew A (1991). Chromosome breakage by pairs of closely linked transposable elements of the *Ac-Ds* family in maize. *Genetics* **129**, 855–862.
- Du QG, Wang K, Zou C, Xu C, Li WX (2018). The *PILNCR1*-miR399 regulatory module is important for low phosphate tolerance in maize. *Plant Physiol* **177**, 1743–1753.
- Fang J, Zhang FT, Wang HR, Wang W, Zhao F, Li ZJ, Sun CH, Chen FM, Xu F, Chang SQ, Wu L, Bu QY, Wang PR, Xie JK, Chen F, Huang XH, Zhang YJ, Zhu XG, Han B, Deng XJ, Chu CC (2019). *Ef-cd* locus shortens rice maturity duration without yield penalty. *Proc Natl Acad Sci USA* **116**, 18717–18722.
- Feschotte C, Jiang N, Wessler SR (2002). Plant transposable elements: where genetics meets genomics. *Nat Rev Genet* **3**, 329–341.
- Gagliardi M, Matarazzo MR (2016). RIP: RNA immunoprecipitation. In: Lanzaolo C, Bodega B, eds. Polycomb Group Proteins. Methods in Molecular Biology, Vol. 1480. New York: Humana Press. pp. 73–86.
- Golicz AA, Singh MB, Bhalla PL (2018). The long intergenic noncoding RNA (lincRNA) landscape of the soybean genome. *Plant Physiol* **176**, 2133–2147.
- Heo JB, Sung S (2011). Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science* **331**, 76–79.
- Hou XX, Cui J, Liu WW, Jiang N, Zhou XX, Qi HY, Meng J, Luan YS (2020). LncRNA39026 enhances tomato resistance to *Phytophthora infestans* by decoying miR168a and inducing *PR* gene expression. *Phytopathology* **110**, 873–880.
- Huang L, Dong H, Zhou D, Li M, Liu YH, Zhang F, Feng YY, Yu DL, Lin S, Cao JS (2018). Systematic identification of long non-coding RNAs during pollen development and fertilization in *Brassica rapa*. *Plant J* **96**, 203–222.
- Kapitonov VV, Jurka J (2001). Rolling-circle transposons in eukaryotes. *Proc Natl Acad Sci USA* **98**, 8714–8719.
- Khandaia JS, Calvo IA, Joh RI, Hill IT, Motamedi M (2016). Nuclear noncoding RNAs and genome stability. *Mol Cell* **63**, 7–20.
- Kidwell MG, Lisch DR (2000). Transposable elements and host genome evolution. *Trends Ecol Evol* **15**, 95–99.
- Li L, Eichten SR, Shimizu R, Petsch K, Yeh CT, Wu W, Chetoor AM, Givan SA, Cole RA, Fowler JE, Evans MMS, Scanlon MJ, Yu JM, Schnable PS, Timmermans MCP, Springer NM, Muehlbauer GJ (2014). Genome-wide discovery and characterization of maize long non-coding RNAs. *Genome Biol* **15**, R40.
- Lippman Z, May B, Yordan C, Singer T, Martienssen R

- (2003). Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. *PLoS Biol* **1**, E67.
- Liu F, Xu YR, Chang KX, Li S, Liu ZG, Qi SD, Jia JB, Zhang M, Crawford NM, Wang Y (2019). The long non-coding RNA *T5120* regulates nitrate response and assimilation in *Arabidopsis*. *New Phytol* **224**, 117–131.
- Liu HJ, Jian LM, Xu JT, Zhang QH, Zhang ML, Jin ML, Peng Y, Yan JL, Han BZ, Liu J, Gao F, Liu XG, Huang L, Wei WJ, Ding YX, Yang XF, Li ZX, Zhang ML, Sun JM, Bai MJ, Song WH, Chen HM, Sun XA, Li WQ, Lu YM, Liu Y, Zhao JR, Qian YW, Jackson D, Fernie AR, Yan JB (2020). High-throughput CRISPR/Cas9 mutagenesis streamlines trait gene identification in maize. *Plant Cell* **32**, 1397–1413.
- Liu J, Jung C, Xu J, Wang H, Deng SL, Bernad L, Arenas-Huertero C, Chua NH (2012). Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in *Arabidopsis*. *Plant Cell* **24**, 4333–4345.
- Liu X, Li DY, Zhang DL, Yin DD, Zhao Y, Ji CJ, Zhao XF, Li XB, He Q, Chen RS, Hu SN, Zhu LH (2018). A novel antisense long noncoding RNA, *TWISTED LEAF*, maintains leaf blade flattening by regulating its associated sense R2R3-MYB gene in rice. *New Phytol* **218**, 774–788.
- Lv YD, Hu FQ, Zhou YF, Wu FL, Gaut BS (2019). Maize transposable elements contribute to long non-coding RNAs that are regulatory hubs for abiotic stress response. *BMC Genomics* **20**, 864.
- Martin SL, Garfinkel DJ (2003). Survival strategies for transposons and genomes. *Genome Biol* **4**, 313.
- Qin T, Zhao HY, Cui P, Albeshier N, Xiong LM (2017). A nucleus-localized long non-coding RNA enhances drought and salt stress tolerance. *Plant Physiol* **175**, 1321–1336.
- Schnable PS, Ware D, Fulton RS, Stein JC, Wei FS, Pasternak S, Liang CZ, Zhang JW, Fulton L, Graves TA, Minx P, Reilly AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du FY, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B, Chen WZ, Yan L, Higginbotham J, Cardenas M, Waligorski J, Applebaum E, Phelps L, Falcone J, Kanchi K, Thane T, Scimone A, Thane N, Henke J, Wang T, Ruppert J, Shah N, Rotter K, Hodges J, Ingenthron E, Cordes M, Kohlberg S, Sgro J, Delgado B, Mead K, Chinwalla A, Leonard S, Crouse K, Collura K, Kudrna D, Currie J, He RF, Angelova A, Rajasekar S, Mueller T, Lomeli R, Scara G, Ko A, Delaney K, Wissotski M, Lopez G, Campos D, Braidotti M, Ashley E, Golser W, Kim H, Lee S, Lin JK, Dujmic Z, Kim W, Talag J, Zuccolo A, Fan CZ, Sebastian A, Kramer M, Spiegel L, Nascimento L, Zutavern T, Miller B, Ambroise C, Muller S, Spooner W, Narechania A, Ren LY, Wei S, Kumari S, Faga B, Levy MJ, McMahan L, Van Buren P, Vaughn MW, Ying K, Yeh CT, Emrich SJ, Jia Y, Kalyanaraman A, Hsia AP, Barbazuk WB, Baucom RS, Brutnell TP, Carpita NC, Chaparro C, Chia JM, Deragon JM, Estill JC, Fu Y, Jeddelloh JA, Han YJ, Lee H, Li PH, Lisch DR, Liu SZ, Liu ZJ, Nagel DH, McCann MC, SanMiguel P, Myers AM, Nettleton D, Nguyen J, Penning BW, Ponnala L, Schneider KL, Schwartz DC, Sharma A, Soderlund C, Springer NM, Sun Q, Wang H, Waterman M, Westerman R, Wolfgruber TK, Yang LX, Yu Y, Zhang LF, Zhou SG, Zhu QH, Bennetzen JL, Dawe RK, Jiang JM, Jiang N, Presting GG, Wessler SR, Aluru S, Martienssen RA, Clifton SW, McCombie WR, Wing RA, Wilson RK (2009). The B73 maize genome: complexity, diversity, and dynamics. *Science* **326**, 1112–1115.
- Seo JS, Sun HX, Park BS, Huang CH, Yeh SD, Jung C, Chua NH (2017). ELF18-INDUCED LONG-NONCODING RNA associates with mediator to enhance expression of innate immune response genes in *Arabidopsis*. *Plant Cell* **29**, 1024–1038.
- Swiezewski S, Liu FQ, Magusin A, Dean C (2009). Cold-induced silencing by long antisense transcripts of an *Arabidopsis* polycomb target. *Nature* **462**, 799–802.
- The *Arabidopsis* Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815.
- Ulitsky I (2018). Interactions between short and long non-coding RNAs. *FEBS Lett* **592**, 2874–2883.
- Wang D, Qu ZP, Yang L, Zhang QZ, Liu ZH, Do T, Adelson DL, Wang ZY, Searle I, Zhu JK (2017). Transposable elements (TEs) contribute to stress-related long intergenic noncoding RNAs in plants. *Plant J* **90**, 133–146.
- Wang X, Ai G, Zhang CL, Cui L, Wang JF, Li HX, Zhang JH, Ye ZB (2016). Expression and diversification analysis reveals transposable elements play important roles in the origin of *Lycopersicon*-specific lncRNAs in tomato. *New Phytol* **209**, 1442–1455.
- Wang Y, Luo XJ, Sun F, Hu JH, Zha XJ, Su W, Yang JS (2018a). Overexpressing lncRNA *LAIR* increases grain yield and regulates neighbouring gene cluster expression



- in rice. *Nat Commun* **9**, 3516.
- Wang YC, Xu JY, Ge M, Ning LH, Hu MM, Zhao H** (2020). High-resolution profile of transcriptomes reveals a role of alternative splicing for modulating response to nitrogen in maize. *BMC Genomics* **21**, 353.
- Wang YJ, Wang YL, Zhao J, Huang JY, Shi YN, Deng DX** (2018b). Unveiling gibberellin-responsive coding and long noncoding RNAs in maize. *Plant Mol Biol* **98**, 427–438.
- Wu H, Yang L, Chen LL** (2017). The diversity of long non-coding RNAs and their generation. *Trends Genet* **33**, 540–552.
- Xu JH, Messing J** (2006). Maize haplotype with a *helitron*-amplified cytidine deaminase gene copy. *BMC Genet* **7**, 52.
- Yang DW, Li SP, Cui HT, Zou SH, Wang W** (2020). Molecular genetic mechanisms of interaction between host plants and pathogens. *Hereditas* **42**, 278–286.
- Yoon JH, Abdelmohsen K, Gorospe M** (2014). Functional interactions among microRNAs and long noncoding RNAs. *Semin Cell Dev Biol* **34**, 9–14.
- Yu Y, Zhou YF, Feng YZ, He H, Lian JP, Yang YW, Lei MQ, Zhang YC, Chen YQ** (2020). Transcriptional landscape of pathogen-responsive lncRNAs in rice unveils the role of ALEX1 in jasmonate pathway and disease resistance. *Plant Biotechnol J* **18**, 679–690.
- Zhao SR, Zhang Y, Gordon W, Quan J, Xi HL, Du S, von Schack D, Zhang BH** (2015). Comparison of stranded and non-stranded RNA-seq transcriptome profiling and investigation of gene overlap. *BMC Genomics* **16**, 675.
- Zhao T, Tao XY, Feng SL, Wang LY, Hong H, Ma W, Shang GD, Guo SS, He YX, Zhou BL, Guan XY** (2018a). LncRNAs in polyploid cotton interspecific hybrids are derived from transposon neofunctionalization. *Genome Biol* **19**, 195.
- Zhao XY, Li JR, Lian B, Gu HQ, Li Y, Qi YJ** (2018b). Global identification of *Arabidopsis* lncRNAs reveals the regulation of *MAF4* by a natural antisense RNA. *Nat Commun* **9**, 5056.
- Zheng XM, Chen J, Pang HB, Liu S, Gao Q, Wang JR, Qiao WH, Wang H, Liu J, Olsen KM, Yang QW** (2019). Genome-wide analyses reveal the role of noncoding variation in complex traits during rice domestication. *Sci Adv* **5**, eaax3619.

## Transposon-derived Long Noncoding RNA in Plants

Yijun Wang<sup>1, 2\*</sup>, Yali Wang<sup>1, 2</sup>, Yudong Chen<sup>1, 2</sup>

<sup>1</sup>Jiangsu Key Laboratory of Crop Genomics and Molecular Breeding/Key Laboratory of Plant Functional Genomics of the Ministry of Education/Jiangsu Key Laboratory of Crop Genetics and Physiology, College of Agriculture, Yangzhou University, Yangzhou 225009, China; <sup>2</sup>Jiangsu Co-Innovation Center for Modern Production Technology of Grain Crops, Yangzhou University, Yangzhou 225009, China

**Abstract** Transposable element (TE), core component of the genome, influences genome structure and stability. Long noncoding RNA (lncRNA) modulates diverse biological events at transcriptional and post-transcriptional levels. TE and lncRNA are major driving forces of evolution. Emerging evidence has revealed the wide distribution of lncRNA that harbors TE. In this review, we first briefly introduce the methodologies of identification and functional analysis of plant lncRNA. We focus on the distribution and function of transposable element-derived lncRNA (TE-lncRNA). Finally, we discuss the regulatory mechanism, epigenetic modification, and breeding potential of TE-lncRNA in plants.

**Key words** transposable element, long noncoding RNA, transposon-derived lncRNA, plant

**Wang YJ, Wang YL, Chen YD** (2020). Transposon-derived long noncoding RNA in plants. *Chin Bull Bot* **55**, 768–776.

\* Author for correspondence. E-mail: wyj@yzu.edu.cn

(责任编辑: 朱亚娜)