



盾叶秋海棠基因组微型反向重复转座子MITEs的分布模式和表达分析

黄泽铖, 郭微, 王龙远, 李凌飞, 吴伟

引用本文:

黄泽铖, 郭微, 王龙远, 等. 盾叶秋海棠基因组微型反向重复转座子MITEs的分布模式和表达分析[J]. 热带亚热带植物学报, 2025, 33(5): 566–576.

在线阅读 View online: <https://doi.org/10.11926/jtsb.4930>

您可能感兴趣的其他文章

Articles you may be interested in

铁皮石斛WOX转录因子的鉴定和分析

Identification and Analysis of WOX Transcription Factor in *Dendrobium officinale*
热带亚热带植物学报. 2021, 29(3): 301–310 <https://doi.org/10.11926/jtsb.4294>

蔓花生PEPC基因家族的生物信息学分析

Bioinformatics Analysis of PEPC Gene Family in *Arachis duranensis*
热带亚热带植物学报. 2018, 26(2): 107–115 <https://doi.org/10.11926/jtsb.3804>

毛竹APX家族基因鉴定和表达分析

Identification and Expression Analysis of the APX Gene Family in *Phyllostachys edulis*
热带亚热带植物学报. 2020, 28(3): 255–264 <https://doi.org/10.11926/jtsb.4155>

三裂叶薯NBS-LRR类抗病基因的筛选鉴定与结构分析

Screening and Identification and Structural Analysis of NBS-LRR Family Genes in *Ipomoea triloba*
热带亚热带植物学报. 2020, 28(5): 495–504 <https://doi.org/10.11926/jtsb.4184>

益智不同组织多糖含量及其生物合成途径分析

Analysis of Polysaccharide Content and Biosynthesis Pathway in Different Tissues of *Alpinia oxyphylla*
热带亚热带植物学报. 2021, 29(6): 669–677 <https://doi.org/10.11926/jtsb.4374>

向下翻页, 浏览PDF全文

盾叶秋海棠基因组微型反向重复转座子 MITEs 的分布模式和表达分析

黄泽铖¹, 郭微¹, 王龙远¹, 李凌飞², 吴伟^{1*}

(1. 仲恺农业工程学院园艺园林学院, 广州 510250; 2. 仙湖植物园, 亚热带植物多样性重点实验室, 广东 深圳 518004)

摘要: 利用已发表盾叶秋海棠(*Begonia peltatifolia*)基因组和不同组织的转录组数据, 对基因组中微型重复转座子元件(MITEs)进行鉴定和分类, 并对其在基因组上的分布特征、插入偏好和表达状态进行了分析。结果表明, 共鉴定出 925 个 MITEs 拷贝, 分属于 *hAT*、*Mutator*、*PIF/Harbinger* 和 *Tc1/Mariner* 等 4 个超家族, 其中 *hAT* 家族占比最高, 达到 70.49%。插入位置分析发现, 有 59.5% 的 MITEs 元件位于基因侧翼(上下游各 5 kb)或基因内, 其它约 40% 的 MITEs 则位于端粒、着丝粒或其它基因稀疏区域; 且位于基因 5'侧翼的 MITEs 拷贝大约是 3'侧翼的 MITEs 的 2 倍, 插入热点在 1 000~3 000 碱基对距离内, 这些共转录的基因主要与次生代谢、能量合成和转运功能有关; 插入到基因内部的则主要分布在非编码区(UTRs 和内含子), 在外显子上没有观察到 MITEs 的插入, 插入基因功能富集主要与功能、蛋白质合成和遗传信息的表达有关。对盾叶秋海棠不同组织的转录组数据共检测到 68 次 MITEs 共转录, 在营养器官花梗和叶中极少表达。在生殖器官花和活跃分生组织的根中表达较多, 且含有 MITEs 的转录本中, 80.8% 和其异构体相比, 表达量降低或差别不显著, 仅有 19.1% 的表达量增加。这为开展其耐旱的转座子表观遗传调控机制奠定了基础, 为耐旱秋海棠新品种选育提供了基因资源。

关键词: 盾叶秋海棠; MITEs 转座子; 基因插入偏好; 基因功能

doi: 10.11926/jtsb.4930

CSTR:32235.14.jtsb.4930

Distribution of MITEs (Miniature Inverted-repeat Transposable Elements) and Their Expression Patterns in the Genome of *Begonia peltatifolia*

HUANG Zecheng¹, GUO Wei¹, WANG Longyuan¹, LI Lingfei², WU Wei^{1*}

(1. College of Horticulture and Landscape Architecture, Zhongkai University of Agriculture and Engineering, Guangzhou 510000, China; 2. Key Laboratory of Southern Subtropical Plant Diversity, Fairy Lake Botanical Garden, Shenzhen & Chinese Academy of Sciences, Shenzhen 518004, Guangdong, China)

Abstract: By using the published genome of *Begonia peltatifolia* and the transcriptome data of different tissues, the miniature inverted repeat transposable elements (MITEs) in the genome were identified and classified, and their distribution characteristics, insertion preferences and expression states in the genome were analyzed. The results showed that a total of 925 MITEs copies were identified, belonging to four superfamilies: *hAT*, *Mutator*, *PIF/Harbinger* and *Tc1/Mariner*, among which the *hAT* family accounted for the highest proportion, reaching 70.49%. The analysis of insertion positions revealed that 59.5% of the MITEs elements were located in the flanking regions (upstream and downstream 5 kb, respectively) or within genes, while the other approximately 40% of the MITEs were located in telomeres, centromeres or other gene-sparse regions; moreover, the number of MITEs copies located in the 5' flanking region of genes was approximately twice that in the 3' flanking region,

收稿日期: 2024-05-21 接受日期: 2024-07-20

基金项目: 岭南特色园林观赏植物种质创新与开发应用创新团队项目(2023KCXTD017); 广东省现代农业产业技术体系创新团队建设项目(2024CXTD12); 深圳市城管局科研项目(202408)资助

This work was supported by the Project for Lingnan Characteristic Garden Ornamental Plant Germplasm Innovation, Development and Application Innovation Team (Grant No. 2023KCXTD017), the Project for Modern Agricultural Industry Technology System Innovation Team Construction in Guangdong (Grant No. 2024CXTD12), and the Project for Scientific Research of Shenzhen Urban Management Bureau (Grant No. 202408).

作者简介: 黄泽铖(1999 年生), 男, 硕士研究生, 研究方向为观赏园艺。E-mail: 924386599@qq.com

* 通讯作者 Corresponding author. E-mail: wuei06@zhku.edu.cn

and the insertion hotspots were within 1 000 to 3 000 base pairs. These co-transcribed genes were mainly related to secondary metabolism, energy synthesis and transport functions; those inserted into the genes were mainly distributed in non-coding regions (UTRs and introns), and no MITEs insertions were observed in exons. The functional enrichment of the genes with MITEs insertions was mainly related to functions, protein synthesis and the expression of genetic information. A total of 68 MITEs co-transcriptions were detected in the transcriptome data of different tissues of *B. peltatifolia*, with very little expression in vegetative organs such as pedicels and leaves. They were expressed more in reproductive organs such as flowers and active meristems of roots. Among the transcripts containing MITEs, 80.8% had lower expression levels or no significant difference compared with their isoforms, while only 19.1% had increased expression levels. Therefore, these would lay the foundation for studying the epigenetic regulation mechanism of transposons in drought tolerance and provides genetic resources for the breeding of drought-tolerant *Begonia* varieties.

Key words: *Begonia peltatifolia*; MITEs; Gene insertion preference; Gene function

自 20 世纪 50 年代芭芭拉·麦克林托克在玉米基因组上发现可移动的遗传元件以来, 转座子被发现广泛存在于真核生物基因组, 尤其是植物基因组, 与植物基因组大小有很强的相关性, 如小麦基因组中转座元件达 85%^[1]。根据转座的中间体类型是核糖核酸和脱氧核糖核酸, 将转座子分为逆转座子(Class I 型)和 DNA 转座子(Class II 型)^[2], 分别以“复制-粘贴”或“剪切-粘贴”的方式完成扩增。在纲以下, 根据复制和整合机制以及系统发育关系, 转座子依次可以分为不同的目、超家族和家族^[3]。尽管长期被认为是基因组上的‘垃圾元件’, 但越来越多的研究表明, 转座子不仅是遗传变异的重要来源^[4], 也是植物生长发育和适应外界环境, 响应各种生物和非生物胁迫的重要调控元件, 是适应性分化和遗传多样性形成的重要进化力量^[5-7]。

微型反向重复转座元件(miniature inverted-repeat transposable elements, MITEs)是一类非自主性转座元件, 其扩增方式类似于 Class I 转座子通过“复制-粘贴”, 但没有 RNA 中间体产生, 而往往借助于 Class II 中与其同源的自主性末端反向重复(terminal inverted-repeats, TIRs)的转座酶完成^[3]。在植物中, 依据其与 TIRs 的同源性分为 *Tc1/Mariner*、*hAT*、*Mutator*、*PIF-Harbinger*、*CACTA* 类。MITEs 广泛存在于动植物的基因组中, 长度从几十到几百不等, 通常 50~800 bp, 其两端具有末端反向重复序列 TIR, 同时还具有 TSD (TIR 外两侧的正向重复的靶位点重复序列)^[8]。尽管 MITEs 是非自主性转座子, 但通常情况下, 它们的拷贝数量比对应的自主型转座子的拷贝数量多, 单倍体基因组上可以达到 1 000~1 500 拷贝^[9]。MITEs 另外一个重要特征

是插入位点偏向于低拷贝基因内或邻近区域, 从而导致基因表达沉默和基因表达水平的改变^[10-12]。目前, 关于 MITEs 的研究集中在水稻(*Oryza sativa*)、玉米(*Zea mays*)等主要作物中, 随着基因组测序技术发展和测序成本的降低, 对园艺作物在内的许多物种开展了基因组水平的 MITEs 鉴定和功能分析, 研究表明高等植物比低等植物含有更多的 MITEs 组成^[13]。通常 MITEs 的鉴定, 除了通用预测工具 RepeatMasker (<http://www.repeatmasker.org>)外, 还有一些基于同源性和结构特征的预测工具, 如 MITE-hunter^[14]、DetectMITE^[15]、MITE-Tracker^[16], 为在更多物种中鉴定 MITEs 提供了强有力的分析工具。

盾叶秋海棠(*Begonia peltatifolia*)是秋海棠科(Begoniaceae)秋海棠属的草本植物, 分布于海南海拔 900 m 左右的干燥石灰岩上。叶片肉质, 盾形、宽卵形到圆形, 具有较高的观赏价值^[17]。作为秋海棠属植物少数耐旱的种, 是秋海棠耐旱新品种选育的重要种质资源, 也是理解植物干旱适应性分子机制的理想材料。利用已发表的高质量染色体水平组装基因组^[18], 本研究对盾叶秋海棠基因组 MITEs 进行鉴定分类, 分析其分布特征, 插入位置和插入附近基因功能, 为解析盾叶秋海棠基因组进化和适应性性状产生机制提供线索。

1 材料和方法

1.1 MITEs 转座子的鉴定与注释

从国家基因组数据库下载盾叶秋海棠基因组序列(<https://db.cngb.org/cnsa>, 序列号 CNA0013976), 使用 MITE-Tracker^[16]软件进行鉴定, 参数设置为

'-tsd min len 2; -tsd max len 10; -mite min len 50; -mite max len 800'. 对搜寻到的 MITEs 进行筛选、过滤和聚类到不同的家族。对不同家族序列使用 MUSCLE-3.8^[19] 软件进行序列比对, 鉴定 MITE-Tracker 软件无法识别的转座子边界。对于存在边界缺失的序列, 扩展 50 bp 后再次进行序列比对, 并使用 BioEdit-7.0^[20] 软件修剪对齐。根据不同超家族 TSD 和 TIRs 序列特征对获得的盾叶秋海棠 MITEs 进行分类注释^[21]。对于无法正确分类的 MITEs, 使用深度学习软件 DeepTE^[22] 进一步注释。计算 MITEs 序列中腺嘌呤脱氧核糖核酸(A)和胸腺嘧啶脱氧核糖核酸(T)比例。

1.2 MITEs 转座子各家族进化树的构建

选择盾叶秋海棠 MITEs 不同家族的代表序列, 利用 MEGA v.11^[23] 软件进行序列比对, 然后构建各家族一致序列的邻接树, 自展性检验 Bootstrap 参数设置为 1 000 次, 使用软件 FigTree v1.4 (<http://tree.bio.ed.ac.uk/software/figtree/>) 对进化树进行可视化。

1.3 MITEs 转座子插入位置和密度分布分析

合并已发表盾叶秋海棠的基因组注释文件和 MITE-Tracker 生成的 MITEs 注释文件。根据基因和 MITEs 转座子的起始位置、终止位置和正链方向, 统计每个 MITEs 拷贝在基因组中的插入位置, 包括基因的 5'端(在 5 kb 范围内)、3'端(在 5 kb 范围内)、内含子和外显子。最后, 统计所有 MITEs 的插入位置信息和插入次数。使用 R 语言程序包 RIdogram^[24] 对秋海棠基因组编码基因和 MITEs 的分布进行可视化。

1.4 MITEs 转座子在基因附近的插入偏好分析及功能富集分析

根据盾叶秋海棠编码基因和 MITEs 的注释信息, 对于有 MITEs 插入或侧翼 5 kb 有 MITEs 分布的基因, 按照 500 bp 步移大小, 统计其 5'端和 3'端 5 kb 内的 MITEs 插入情况。使用 R 语言程序包 ggplot2^[25] 展示 MITEs 与邻近基因在染色体上的分布热图。

使用软件包 KOBAS v.3.0^[26] 对盾叶秋海棠在 5 kb 的范围内有 MITEs 插入的基因进行 GO 和 KEGG 富

集分析。使用 ggplot2 对最显著的前 20 个条目进行展示。

1.5 MITEs 转座子转录组分析

从美国国家生物技术信息中心(NCBI)下载已经发表的盾叶秋海棠 5 个不同组织转录组原始读长(序列号分别为花梗: SRR17381838~SRR17381839; 花: SRR17381840~SRR17381841; 茎: SRR17381842~SRR17381843; 根: SRR17381844~SRR17381845; 叶: SRR17381909~SRR17381910)。使用 fastp^[27] 对上述不同组织的原始读长按照默认参数进行质量过滤。

为了比较含有 MITEs 的转录本和不含有 MITEs 的同源异构体的表达情况, 使用 STAR v. 2.7.3^[28] 将过滤的读长比对到盾叶秋海棠的参考基因组上, 参数设置为: outSAMmapqUnique: 50; outFilterMultiMapNmax: 20; alignSJoverhangMin: 8; alignSJDBoverhangMin: 1; outFilterMismatchNmax: 999; outFilterMismatchNoverLmax: 0.04; alignIntronMin: 20; alignIntronMax: 1 000 000; alignMatesGapMax: 1 000 000。使用 StringTie^[29] 统计含有 MITEs 的转录本和不含 MITEs 同源异构体的表达水平, 标准为: 最小比对长度: 200; 接合处覆盖度: 10; 转录本最小覆盖度: 10。挑选 3 个代表基因, 对比有 MITEs 的转录本和无 MITEs 同源异构体在不同组织中的表达量, 并使用 R 语言程序包 ggplot2 对结果进行绘制和可视化。

2 结果和分析

2.1 MITEs 转座子的鉴定和分类

在盾叶秋海棠基因组上初步筛选到来自 5 个家族的 18 条一致性序列, 通过这些一致性序列比对基因组, 按照 MITE-Tracker 默认的参数过滤, 最后得到 925 个完整 MITEs 拷贝, 序列总长度为 145 kb, 占基因组总长的 0.046% (盾叶秋海棠基因组总长度为 310.46 Mb)。与其他植物基因组相比^[13], 盾叶秋海棠 MITEs 含量偏低。对边界不完整的 MITEs 拷贝进行手动校正(图 1)。

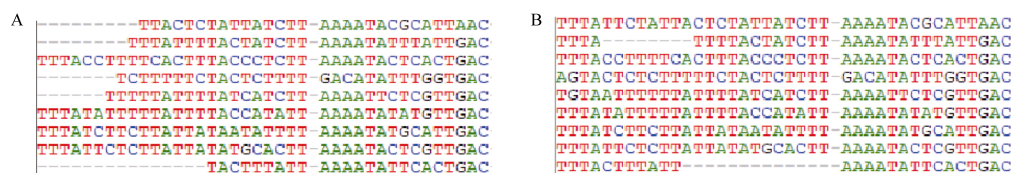


图 1 MITEs 转座子手工校正末端序列对比图。A: 矫正前不完整的末端; B: 矫正后完整的末端。

Fig. 1 Manual correction of end sequences of MITEs transposon based on alignments. A: Incomplete boundaries before correction; B: Intact boundaries after correction.

依据超家族特征并结合 DeepTE 注释结果, 将 MITEs 元件分为 *hAT*、*Tc1/Mariner*、*PIF/Harbinger* 和 *Mutator* 等 4 个超家族, 其余未能分类的则被归类为 Unknown (UN)。在基因组中, 不同超家族含量分布不均匀(表 1), 最大的超家族是 *hAT*, 共有 652 个拷贝, 占整个 MITEs 含量的 70.49%; 其他 3 个超家族含量相当, 拷贝数目为 71~103, 含量为 7.67%~11.13%。

表 1 盾叶秋海棠基因组中 MITEs 超家族

Table 1 MITEs superfamily in <i>Begonia peltatifolia</i> genome			
超家族 Superfamily	拷贝数 Copy number	%	
<i>hAT</i>	652	70.49	
<i>Tc1/Mariner</i>	103	11.13	
<i>PIF/Harbinger</i>	93	10.05	
<i>Mutator</i>	71	7.67	
UN	6	0.66	
合计 Total	925	100.00	

2.2 MITEs 中 AT 含量和超家族系统发育关系

在盾叶秋海棠基因组中, MITEs 的 AT 含量相对较高, 平均组成比率为 76.62%。*Mutator* 超家族具有最高的 AT 含量, 其次是 *PIF/Harbinger* 超家族, 而最低为 *hAT* 家族。结果表明(表 2), 盾叶秋

表 2 盾叶秋海棠基因组中 MITEs 转座子 AT 含量

Table 2 AT content of MITEs transposons in the <i>Begonia peltatifolia</i> genome			
超家族 Superfamily	AT 含量 AT content	总碱基数 Total number of base	%
<i>hAT</i>	59 856	81 424	73.51
<i>Tc1/Mariner</i>	12 423	15 888	78.19
<i>PIF/Harbinger</i>	5 025	6 267	80.18
<i>Mutator</i>	33 421	40 931	81.65
UN	450	581	77.45
合计 Total	11 175	145 091	76.62

海棠中各 MITEs 超家族的 AT 含量较高, 稳定性相对较低, 导致转座活动较为活跃。

盾叶秋海棠 4 个超家族包含 18 个家族, 依据系统发育可以分为 5 个支系(图 2), 其中 *hAT* 超家族包含 9 个家族分布在 5 个支系; *Tc1/Mariner* 分为 4 个家族分属 3 个支系; *Mutator* 超家族包含 3 个家族同属于 1 个支系; *PIF/Harbinger* 超家族仅有一个家族成员。除 *PIF/Harbinger* 外, 其他超家族均不构成单系。虽然可以根据 TIR 和 TSD 序列特征将不同超家族成员区分开来, 但是在系统发育树上它们并没有构成单系类群, 这可能与 MITEs 富含 AT, 组成上不太稳定, 容易发生突变或不同超家族成员之间的重组交换, 原本的系统发育信息丢失, 因而在系统发育树上呈现出同一超家族成员多系分布的样式。

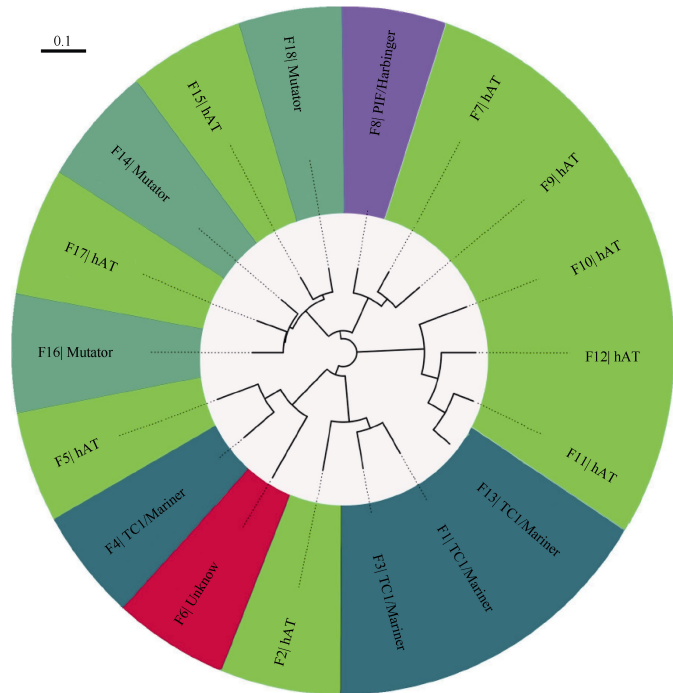


图 2 盾叶秋海棠 MITEs 超家族的系统发育树。不同颜色表示不同的超家族, Unknown 为不具有 MITEs 转座子完整结构的分类群。
Fig. 2 Phylogenetic tree based on superfamily of MITEs in the genome of *Begonia peltatifolia*. Different colored represent different superfamily, while Unknown represents a taxonomic group without a complete structure of MITEs transposons.

2.3 MITEs 分布和插入偏好分析

对盾叶秋海棠基因组上 925 个 MITEs 元件拷贝的插入位点分析,显示共有 549 个拷贝插入到基因或基因侧翼(上下游各 5 kb 范围),占总拷贝数的 59.35%。其他部分可能插入到端粒、着丝粒或其它基因稀疏区域。549 个 MITEs 拷贝中,有 285 和 143 拷贝分别位于基因 5'侧翼区和 3'侧翼区,占 78.0%;在基因的转录非翻译区(UTRs)和内含子区域则均有 121 个 MITEs 拷贝;在基因外显子上没有发现 MITEs 插入。绝大部分的 MITEs 元件插入到基因侧翼,部分插入到基因上的则主要分布在非编码区(UTRs 和内含子),在外显子上没有观察到 MITEs 的插入。位于基因附近的插入位置分布分析表明,基因的 5'端比 3'端插入频率高,且基因侧翼 1 000~3 000 bp 是

插入热点区域(图 3)。基因侧翼 1 001~1 500 bp 区域插入频度最高,共有 84 个插入事件。在距离基因 4 000~5 000 bp 和 0~1 000 bp 的区域,插入数量相对较低。

盾叶秋海棠蛋白质编码基因在 15 条染色体上分布存在较大差异,比如 11 号染色体上基因分布密度整体较高,3 号染色体基因分布密度较低;在染色体内部,染色体两臂外端基因密度较高,而中间靠近着丝粒区域密度较低(图 4: A);对于 MITEs 转座元件,不同染色体间和染色体内部分布相对均匀(图 4: B)。插入位置大多位于基因密度分布较高的区域,这表明 MITEs 转座子在插入位点的选择上有一定的偏好,倾向于插入到基因密度较高的区域,这可能会对盾叶秋海棠基因组结构和功能产生影响。

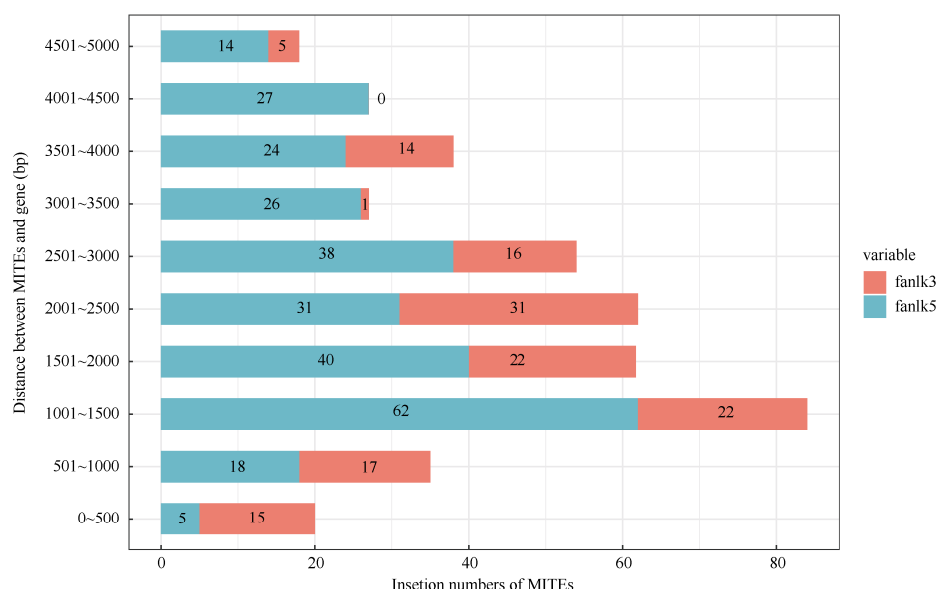


图 3 MITEs 转座子在基因附近的插入分布

Fig. 3 Insertion distribution of MITEs transposons flanking genes

2.4 MITEs 转座子关联基因的功能富集分析

对内部含有 MITEs 元件的基因功能富集分析表明,共有 54 个 GO 条目显著富集($P < 0.05$, FDR 校正没有显著富集的条目,这里列出的是没有校正的 P -value,下同),包括内质网出口位点(GO: 0070971, $P = 0.00$)、非同源末端连接的双链断裂修复(GO: 0006303, $P = 0.00$)、DNA 重组(GO: 0006310, $P = 0.00$)、激酶活性(GO: 0016301)等(图 5: A)。代谢通路 KEGG 富集分析显示,仅有 3 个条目显著富集($P < 0.05$),分别是非同源末端连接(KOG03450, $P = 0.00$)、脂肪酸生物合成(KOG00061, $P = 0.00$)、蛋白质输出(KOG03060,

$P = 0.03$) (图 5: B)。这些富集的功能与蛋白质合成和遗传信息的表达有关,具有重要的生物学功能。MITEs 插入位点位于 UTRs 或者内含子,可以通过影响转录因子结合效率或选择性剪切方式,在基因转录水平进行调控。

对 5'端侧翼有 MITEs 插入元件的基因富集分析表明,共有 47 个 GO 条目显著富集($P < 0.05$),包括碳水化合物:质子转运体活性(GO0005351, $P = 0.00$)、质子转运 v 型 ATP 酶(GO0033180, $P = 0.00$)、叶绿体淀粉粒(GO009569, $P = 0.01$)、蛋白激酶调控活性(GO0045859, $P = 0.01$)、蛋白复合物结合(GO0044877,

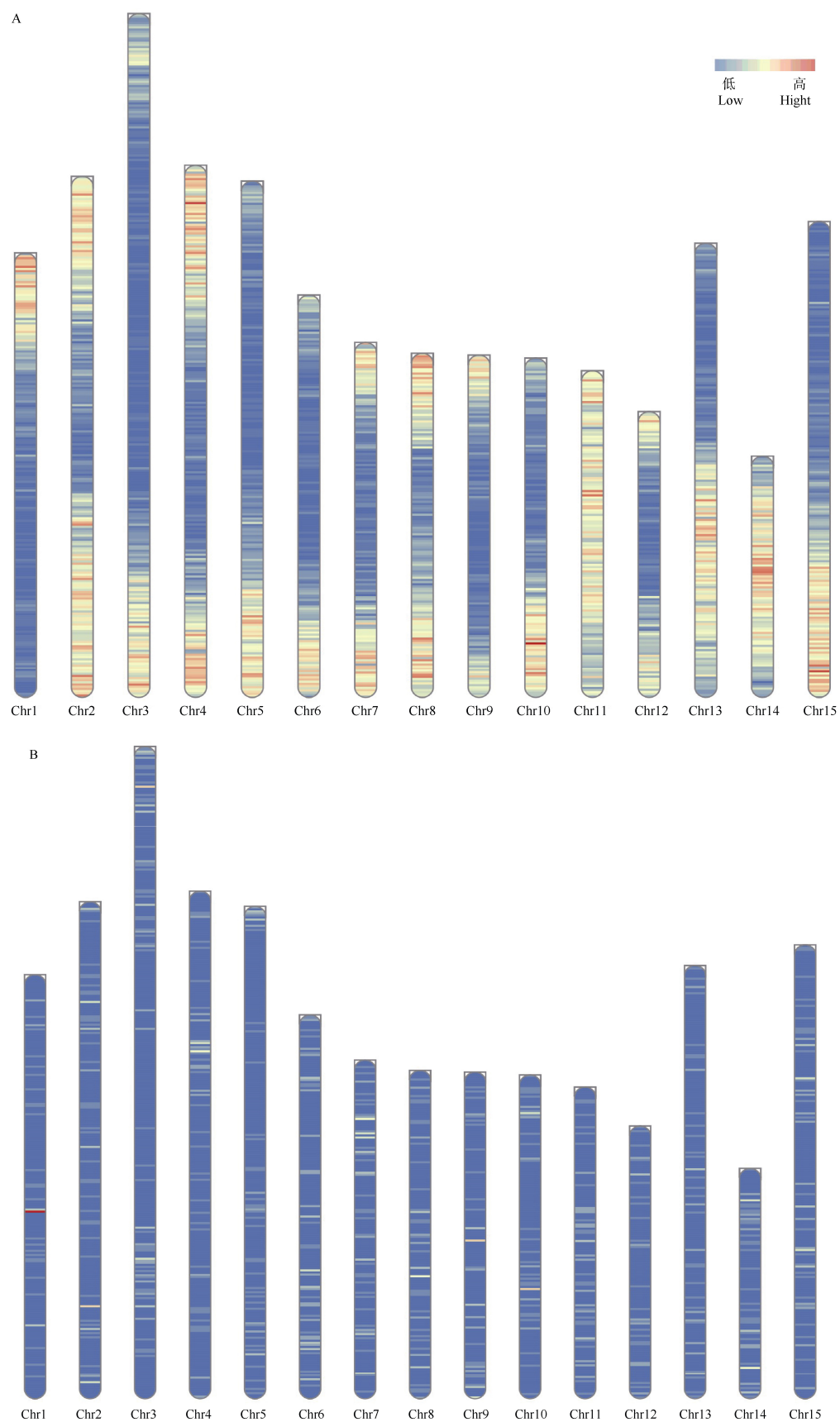


图 4 盾叶秋海棠基因组蛋白质编码基因(A)和 MITEs 元件染色体密度(B)分布热图。Chr1~Chr15 为染色体编号。
Fig. 4 Heatmap of protein encoding genes (A) and MITEs across the chromosomes (B) of *Begonia peltatifolia*. Chr1~Chr15 are chromosome No.

$P=0.01$) (图 5: C); 代谢通路中仅有 α -亚麻酸代谢。可以看到, 5'端侧翼有 MITEs 元件的基因多与次生代谢、能量合成和转运等功能有关。

2.5 不同组织 MITEs 共转录本和同源异构体的表达比较

在盾叶秋海棠 5 个不同组织(花梗、花、根、茎、叶各 2 个生物学重复) 10 个转录组数据共鉴定到 83 次 MITEs 转录表达事件, 其中 68 个 MITEs 与 30 个基因(68 个转录本异构体)发生共转录(表 3, 附录)。4 个超家族中, *hAT* 超家族发生转录的频率最高, 占总数的 81.9%。比较 MITEs 共表达转录本与无

MITEs 共表达异构体的表达水平 TPM (transcripts per million, 每百万个转录本的片段数)的比值, 定义比值 ≥ 1.5 为 MITEs 异构体转录表达量提高, 比值 ≤ 0.5 为 MITEs 异构体转录表达量下降, 0.5~1.5 为表达量无显著差异。结果表明, 42.6%的 MITEs 共表达转录本与无 MITEs 异构转录本表达没有显著差别; 38.2%的上述转录本对之间表达量显著下降, 其中发生在根的最多, 有 13 个 MITEs 共表达转录本与其异构体表达量下降, 在花和茎分别有 6 个。由于与 MITEs 共表达, 导致其转录水平相对于无 MITEs 共表达异构体上调的共有 13 个, 分别在花、根中有

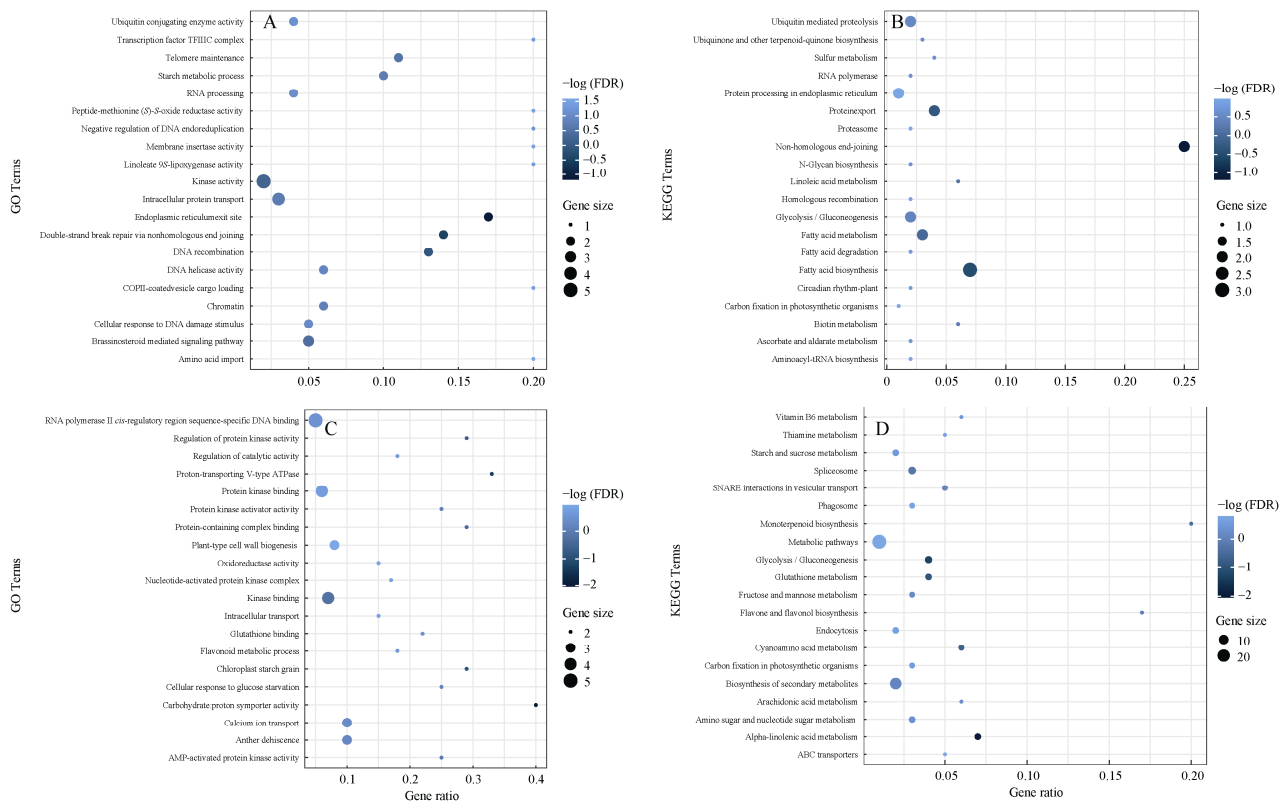


图 5 秋海棠基因组内部(A, B)和侧翼(5 kb 以内)(C, D)插入 MITEs 转座元件的编码基因功能(显著水平前 20)富集分析

Fig. 5 Functional enrichment analysis for encodings genes with internal (A, B) or flanking (C, D) MITEs insertions in the genome of *Begonia peltatifolia*

表 3 与 MITE 共转录基因异构体表达水平情况

Table 3 Expression levels of isoforms of co-transcribed genes with MITEs

组织 Tissue	高表达 High expression	无显著差异 No significant difference	低表达 Low expression	总数 Total
花梗 Scape	1	1	1	3
花 Flower	6	6	6	18
茎 Stem	0	10	6	16
根 Root	6	11	13	30
叶 Leaf	0	1	0	1
总数 Total	13	29	26	68
%	19.1	42.6	38.2	100

6个, 1个发生在花柄。选取根、茎、花3个组织都发生 MITEs 共转录的基因 *Bpe011924.1* (Ser/Thr protein kinase CK2)、*Bpe019454.3* (small auxin-up RNA (SAUR)和 *Bpe001757.1* (WD40 repeat-like protein) 进行比较(图 6)。其中 *Bpe011924.1* 基因和 MITEs 在花梗、花、根共转录, 表达量都高于无 MITEs 共转录异构体, 在多组织中保持着一致性。与此相反, 和 MITEs 共转录的 *Bpe019454.3* 都低于没有 MITEs 共转录的异构体。而与 MITEs 共转录的 *Bpe001757.1* 仅在花和茎中有表达, 且表达水平与无共表达 MITEs 异构体无显著差别。

3 讨论和结论

虽然半个多世纪前, 人们就开始认识到转座子元件作为‘控制元件’, 是植物表型多样性产生的重要来源, 然而, 利用早期测序技术准确获取转座子十分困难, 限制了进一步认识其功能与进化规律。随着高通量测序技术的变革, 特别是三代长读长测序技术如美国太平洋生物的 HiFi 测序技术和英国牛津纳米孔公司的超读长 UltraLong 测序技术, 可以进行端粒到端粒‘T2T’的高质量基因组从头组装, 转座子的研究迎来前所未有的机遇。

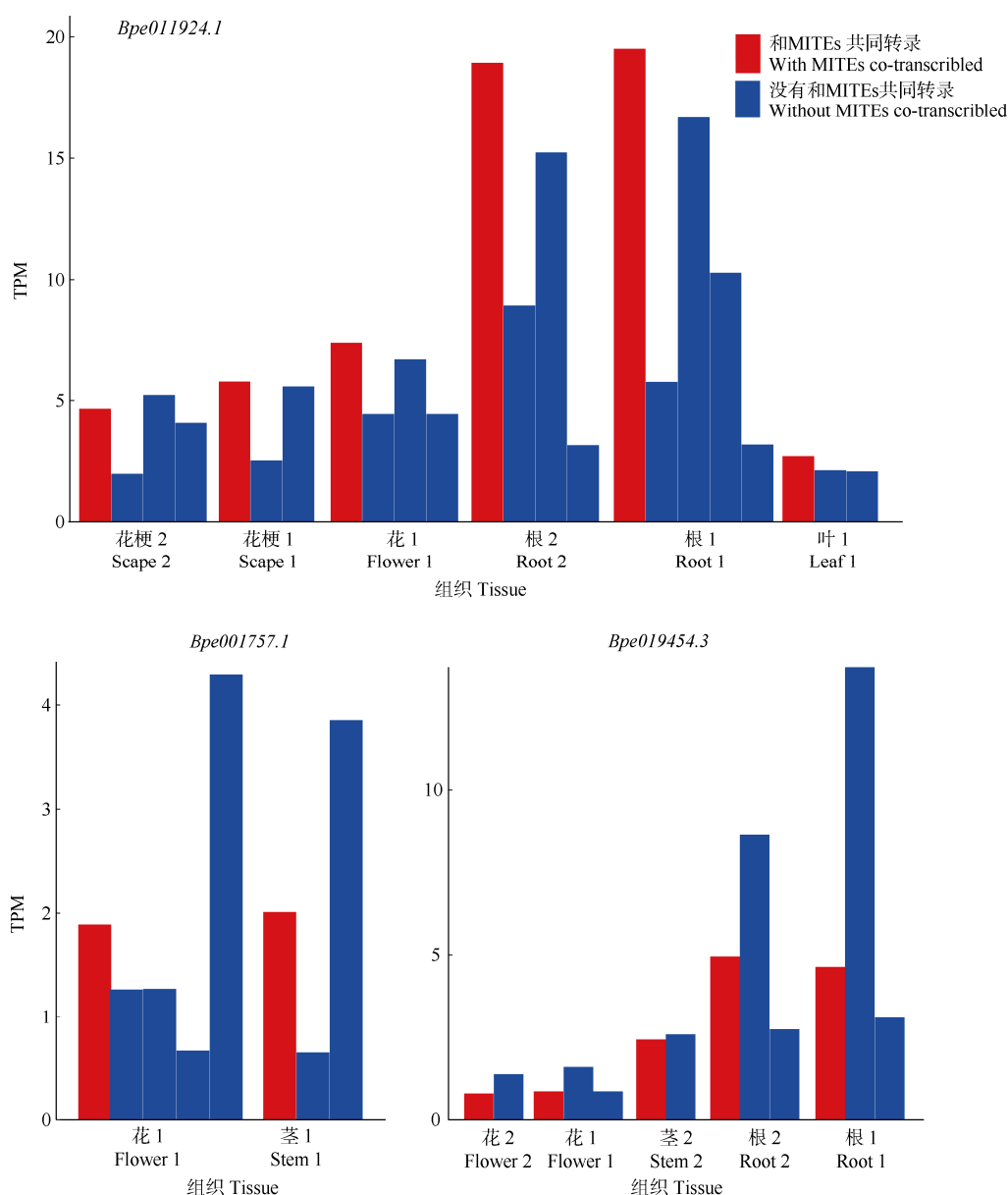


图 6 盾叶秋海棠不同组织 MITEs 转座子对转录本及其异构体表达量的影响。TPM: 每百万个转录本的片段数。

Fig. 6 Influences of MITEs on expression level between transcripts and isoforms in different tissues of *Begonia peltatifolia*. TPM: Transcripts per million.

在绝大部分已报道的植物基因组中, DNA 转座子的相对含量通常会比逆转座子低, 但是植物 LTR 逆转座子通常倾向于分布在基因密度较低的端粒和着丝粒附近, 这些区域相对稳定, 不容易对宿主基因组产生不利影响, 而 DNA 转座子没有如此强烈的偏好, 在基因组上的分布更为均匀, 有研究表明 DNA 转座子比 LTR 反转座子更倾向于插入到基因丰富的区域, 或是染色体的臂部^[30-31]。DNA 转座子的这种分布可能与它们通过“剪切与粘贴”转座的方式有关, 这使得 DNA 转座子对宿主基因组的影响更加直接^[32-33]。因此, 本研究结果表明, 有 59.5% 的 MITEs 元件位于基因侧翼(上下游各 5 kb)或基因内, 其他约 40% 的 MITEs 则位于端粒、着丝粒或其他基因稀疏区域; 且位于基因 5'侧翼的 MITEs 拷贝大约是 3'侧翼的 MITEs 的 2 倍, 这可能是 5'侧翼上游主要是启动子, MITEs 插入到启动子影响到基因的调控, 这些调控上的变化如果对宿主是中性甚至有适合度优势, 在随机遗传漂变或自然选择的作用下就会被保留下来^[34-35]。盾叶秋海棠 MITEs 不论 5'侧翼还是 3'侧翼, 插入热点在 1 000~3 000 bp 碱基距离内, 可能与这个区域有较多的基因调控元件如启动子、增强子等有关。MITEs 插入到这些区域可能促进了调控元件的多样性, 促进了邻近基因的表达模式多样化, 具有适合度优势, 因而通过自然选择被固定下来的机会较大。同样, 盾叶秋海棠中 MITEs 在基因内插入全部发生在 UTR 和内含子, 外显子没有观察到。

对内部含有 MITEs 元件和 5'端侧翼有 MITEs 的基因分别做功能富集分析, 可以看到, 由于这 2 个区域对突变的选择压力不一样, 导致富集的功能也有不同。前者与蛋白质合成和遗传信息的表达有关, 具有重要的生物学功能。这些 MITEs 插入位点位于 UTRs 或者内含子, 可以通过影响转录因子结合效率或选择性剪切方式, 在基因转录水平进行调控^[36]; 后者与次生代谢, 能量合成与转运这些功能有关, MITEs 插入到这些基因的上游如启动子、增强子结合区域, 极大丰富了这些基因表达调控机制的多样性^[33,37], 可能在盾叶秋海棠生长发育过程和应对生物/非生物胁迫中发挥重要功能^[36]。

本研究盾叶秋海棠转录组数据来自正常生长非胁迫状态下, 所以检测到表达的 MITEs 数目不多, 且不同组织表达情况差别显著。在营养器官花梗和叶中极少表达, 在生殖器官花和活跃分生组织的根

中表达较多, 这与转座子表达主要发生在生长发育过程和胁迫状态下相一致。因为在上述状态下, 植物染色质结构、组蛋白乙酰化、DNA 甲基化模式和小分子 RNA (siRNAs 和 piRNAs) 活性都会发生变化, 从而导致转座子的激活^[37]。且含有 MITEs 的转录本, 80.8% 和其异构体相比, 表达量降低或差别不显著, 仅有 19.1% 的表达量增加。含有 MITEs 的转录本与不含 MITEs 的同源异构体之间的表达水平差异可以归因于多种因素, 包括① MITEs 作为调节元件: MITEs 可能作为影响基因表达的调节序列, 在基因附近或内部可以改变启动子活性, 增强或抑制增强子活性, 或影响剪接调节因子, 导致基因表达的改变^[38]; ② 表观遗传效应: MITEs 可以影响周围基因组区域的表观遗传状态, 它们存在于基因内部或附近可能导致 DNA 甲基化或组蛋白修饰模式的改变, 从而影响基因表达^[39]; ③ 选择性剪接: MITEs 可以引入新的剪接位点或破坏现有的剪接位点, 这可以导致包含或不包含 MITE 序列的转录本, 造成不同表达水平的异构体^[40]; ④ 影响转录本稳定性: MITEs 可以影响 mRNA 的稳定性和降解率。含有 MITE 的转录本可能具有改变的二级结构或被 RNA 结合蛋白和 RNA 降解机制识别的基序, 导致转录本丰度的差异^[41]; ⑤ 选择压力: 如果含有 MITE 的转录本倾向于产生非功能性或有害蛋白, 可能会存在针对高表达 MITE 转录本的选择压力, 导致相较于不含 MITE 插入的同源异构体总体表达水平较低^[42]。本研究中, 80.8% 含有 MITEs 的转录本与其不含 MITEs 同源异构体相比, 表现出较低或没有显著变化, 这可能意味着 MITE 插入通常与抑制基因表达有关。相反, 在表达增加的情况下(本研究中为 19.1%), MITE 插入可能是通过插入到提升转录的调控元素附近或内部。分析每个 MITE 插入的基因组背景和每个基因的调控网络, 以确定表达变化的具体原因非常重要。此外, 诸如报告基因测定、染色质免疫沉淀和突变体研究等实验验证, 可能有助于明确 MITEs 在转录本表达改变中的确切角色。

盾叶秋海棠是生长于我国海南石灰岩干燥环境的特有植物, 是选育耐旱秋海棠新品种的优异种质资源。本研究在盾叶秋海棠基因组上鉴定了 925 个高置信度的 MITEs 元件, 分属于 *hAT*、*Mutator*、*PIF/Harbinger* 和 *Tc1/Mariner* 等 4 个超家族, 其中 *hAT* 家族最多。MITEs 主要插入在非编码区, 外显子没有观察到插入, 影响的基因功能主要是生物功

能、蛋白质合成和信息表达。MITEs 共转录在不同组织中差异表达, 在生殖器官和活跃分生组织中表达较多。本研究为从转座子角度了解其耐旱性机制提供了新的视角, 为新品种选育提供了新的基因资源。

参考文献

- [1] WICKER T, GUNDLACH H, SPANNAGL M, et al. Impact of transposable elements on genome structure and evolution in bread wheat [J]. *Genome Biol*, 2018, 19(1): 103. doi: 10.1186/s13059-018-1479-0.
- [2] FINNEGAN D J. Eukaryotic transposable elements and genome evolution [J]. *Trends Genet*, 1989, 5: 103–107. doi: 10.1016/0168-9525(89)90039-5.
- [3] WICKER T, SABOT F, HUA-VAN A, et al. A unified classification system for eukaryotic transposable elements [J]. *Nat Rev Genet*, 2007, 8(12): 973–982. doi: 10.1038/nrg2165.
- [4] CAI X, LIN R M, LIANG J L, et al. Transposable element insertion: A hidden major source of domesticated phenotypic variation in *Brassica rapa* [J]. *Plant Biotechnol J*, 2022, 20(7): 1298–1310. doi: 10.1111/pbi.13807.
- [5] MAKAREVITCH I, WATERS A J, WEST P T, et al. Transposable elements contribute to activation of maize genes in response to abiotic stress [J]. *PLoS Genet*, 2015, 11(10): e1004915. doi: 10.1371/journal.pgen.1004915.
- [6] KLEIN S P, ANDERSON S N. The evolution and function of transposons in epigenetic regulation in response to the environment [J]. *Curr Opin Plant Biol*, 2022, 69: 102277. doi: 10.1016/j.pbi.2022.102277.
- [7] WANG M J, LI J Y, QI Z Y, et al. Genomic innovation and regulatory rewiring during evolution of the cotton genus *Gossypium* [J]. *Nat Genet*, 2022, 54(12): 1959–1971. doi: 10.1038/s41588-022-01237-2.
- [8] FESCHOTTE C, JIANG N, WESSLER S R. Plant transposable elements: Where genetics meets genomics [J]. *Nat Rev Genet*, 2002, 3(5): 329–341. doi: 10.1038/nrg793.
- [9] HU T T, PATTYN P, BAKKER E G, et al. The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change [J]. *Nat Genet*, 2011, 43(5): 476–481. doi: 10.1038/ng.807.
- [10] OHMORI Y, ABIKO M, HORIBATA A, et al. A transposon, Ping, is integrated into intron 4 of the DROOPING LEAF gene of rice, weakly reducing its expression and causing a mild drooping leaf phenotype [J]. *Plant Cell Physiol*, 2008, 49(8): 1176–1184. doi: 10.1093/pcp/pcn093.
- [11] KUANG H H, PADMANABHAN C, LI F, et al. Identification of miniature inverted-repeat transposable elements (MITEs) and biogenesis of their siRNAs in the Solanaceae: New functional implications for MITEs [J]. *Genome Res*, 2009, 19(1): 42–56. doi: 10.1101/gr.078196.108.
- [12] LU C, CHEN J, ZHANG Y, et al. Miniature inverted-repeat transposable elements (MITEs) have been accumulated through amplification bursts and play important roles in gene expression and species diversity in *Oryza sativa* [J]. *Mol Biol Evol*, 2012, 29(3): 1005–1017. doi: 10.1093/molbev/msr282.
- [13] CHEN J J, HU Q, ZHANG Y, et al. P-MITE: A database for plant miniature inverted-repeat transposable elements [J]. *Nucl Acids Res*, 2012, 42(D1): D1176–D1181. doi: 10.1093/nar/gkt1000.
- [14] HAN M J, SHEN Y H, GAO Y H, et al. Burst expansion, distribution and diversification of MITEs in the silkworm genome [J]. *BMC Genom*, 2010, 11(1): 520. doi: 10.1186/1471-2164-11-520.
- [15] YE C T, JI G L, LIANG C. detectMITE: A novel approach to detect miniature inverted repeat transposable elements in genomes [J]. *Sci Rep*, 2016, 6(1): 19688. doi: 10.1038/srep19688.
- [16] CRESCENTE J M, ZAVALLO D, HELGUERA M, et al. MITE Tracker: An accurate approach to identify miniature inverted-repeat transposable elements in large genomes [J]. *BMC Bioinform*, 2018, 19(1): 348. doi: 10.1186/s12859-018-2376-y.
- [17] GU C Z, PENG C I, TURLAND N J. Begoniaceae [M]// WU Z Y, RAVEN P H, HONG D Y. *Flora of China*, Vol. 13. Beijing: Science Press & St. Louis: Missouri Botanical Garden, 1998: 153–207.
- [18] LI L F, CHEN X L, FANG D M, et al. Genomes shed light on the evolution of *Begonia*, a mega-diverse genus [J]. *New Phytol*, 2022, 234(1): 295–310. doi: 10.1111/nph.17949.
- [19] MADEIRA F, PARK Y M, LEE J, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019 [J]. *Nucl Acids Res*, 2019, 47(W1): W636–W641. doi: 10.1093/nar/gkz268.
- [20] HALL T A. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT [J]. *Nucl Acids Symp Ser*, 1999, 41: 95–98.
- [21] MA Y S, ZHANG W P, MING R. G. Identification and evolutionary analysis of MITE transposons in rambutan genome [J]. *J Sichuan Agric Univ*, 2022, 40(5): 746–757. [马誉生, 张文萍, 明瑞光. 红毛丹基因组中 MITEs 转座子的鉴定与演化分析 [J]. *四川农业大学学报*, 2022, 40(5): 746–757. doi: 10.16036/j.issn.1000-2650.202205115.]
- [22] YAN H D, BOMBARELY A, LI S. DeepTE: A computational method for *de novo* classification of transposons with convolutional neural network [J]. *Bioinformatics*, 2020, 36(15): 4269–4275. doi: 10.1093/bioinformatics/btaa519.
- [23] TAMURA K, STECHER G, KUMAR S. MEGA11: Molecular evolutionary genetics analysis version 11 [J]. *Mol Biol Evol*, 2021, 38(7): 1–7.

- 3022–3027. doi: 10.1093/molbev/msab120.
- [24] HAO Z D, LV D K, GE Y, et al. *Rldeogram*: Drawing SVG graphics to visualize and map genome-wide data on the idiograms [J]. *PeerJ Comput Sci*, 2020, 6: e251. doi: 10.7717/peerj-cs.251.
- [25] WICKHAM H. *ggplot2: Elegant Graphics for Data Analysis* [M]. Cham: Springer, 2016.
- [26] BU D C, LUO H T, HUO P P, et al. KOBAS-I: Intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis [J]. *Nucl Acids Res*, 2021, 49(W1): W317–W325. doi: 10.1093/nar/gkab447.
- [27] CHEN S F, ZHOU Y Q, CHEN Y R, et al. Fastp: An ultra-fast all-in-one FASTQ preprocessor [J]. *Bioinformatics*, 2018, 34(17): i884–i890. doi: 10.1093/bioinformatics/bty560.
- [28] DOBIN A, DAVIS C A, SCHLESINGER F, et al. STAR: Ultrafast universal RNA-seq aligner [J]. *Bioinformatics*, 2013, 29(1): 15–21. doi: 10.1093/bioinformatics/bts635.
- [29] XU L, YUAN K, YUAN M, et al. Regulation of rice tillering by RNA-directed DNA methylation at miniature inverted-repeat transposable elements [J]. *Mol Plant*, 2020, 13(6): 851–863. doi: 10.1016/j.molp.2020.02.009.
- [30] HUA-VAN A, LE ROUZIC A, MAISONHAUTE C, et al. Abundance, distribution and dynamics of retrotransposable elements and transposons: Similarities and differences [J]. *Cytogen Genome Res*, 2005, 110(1/2/3/4): 426–440. doi: 10.1159/000084975.
- [31] ZHAO D Y, FERGUSON A A, JIANG N. What makes up plant genomes: The vanishing line between transposable elements and genes [J]. *Biochim Biophys Acta Gene Regul Mech*, 2016, 1859(2): 366–380. doi: 10.1016/j.bbagr.2015.12.005.
- [32] MACH J. On the habits of transposons: Dissociation mapping in maize and megabase sequencing in wheat reveal site preferences distribution, and evolutionary history [J]. *Plant Cell*, 2010, 22(6): 1650–1652. doi: 10.1105/tpc.110.077396.
- [33] COSBY R L, JUDD J, ZHANG R L, et al. Recurrent evolution of vertebrate transcription factors by transposase capture [J]. *Science*, 2021, 371(6531): eabc6405. doi: 10.1126/science.abc6405.
- [34] AMRANI A, MARIE L, AINOUCHE A, et al. Genome-wide distribution and potential regulatory functions of *AtATE*, a novel family of miniature inverted-repeat transposable elements in *Arabidopsis thaliana* [J]. *Mol Genet Genom*, 2002, 267(4): 459–471. doi: 10.1007/s00438-002-0675-4.
- [35] WACHTER S, RAGHAVAN R, WACHTER J, et al. Identification of novel MITEs (miniature inverted-repeat transposable elements) in *Coxiella burnetii*: Implications for protein and small RNA evolution [J]. *BMC Genom*, 2018, 19(1): 247. doi: 10.1186/s12864-018-4608-y.
- [36] GEBRIE A. Transposable elements as essential elements in the control of gene expression [J]. *Mob DNA*, 2023, 14(1): 9. doi: 10.1186/s13100-023-00297-3.
- [37] HORVÁTH V, MERENCIANO M, GONZÁLEZ J. Revisiting the relationship between transposable elements and the eukaryotic stress response [J]. *Trends Genet*, 2017, 33(11): 832–841. doi: 10.1016/j.tig.2017.08.007.
- [38] XU L, WANG L, LIU T, et al. *Triton*, a novel family of miniature inverted-repeat transposable elements (MITEs) in *Trichosanthes kirilowii* Maximowicz and its effect on gene regulation [J]. *Biochem Biophys Res Co*, 2007, 364(3): 668–674. doi: 10.1016/j.bbrc.2007.10.069.
- [39] XIN Y C, MA B, ZENG Q W, et al. Dynamic changes in transposable element and gene methylation in mulberry (*Morus notabilis*) in response to *Botrytis cinerea* [J]. *Hort Res*, 2021, 8(1): 154. doi: 10.1038/s41438-021-00588-x.
- [40] XIN Y C, MA B, XIANG Z H, et al. Amplification of miniature inverted-repeat transposable elements and the associated impact on gene regulation and alternative splicing in mulberry (*Morus notabilis*) [J]. *Mob DNA*, 2019, 10(1): 27. doi: 10.1186/s13100-019-0169-0.
- [41] KAN Q X, LI Q. Post-transcriptional and translational regulation of plant gene expression by transposons [J]. *Curr Opin Plant Biol*, 2023, 75: 102438. doi: 10.1016/j.pbi.2023.102438.
- [42] MACKO-PODGORNI A, MACHAJ G, GRZEBELUS D. A global landscape of miniature inverted-repeat transposable elements in the carrot genome [J]. *Genes*, 2021, 12(6): 859. doi: 10.3390/genes12060859.

附录 与基因共转录的 MITEs 信息

http://jtsb.ijournals.cn/ajax/common/download_attache_file.aspx?seq_id=20240728212934001&file_no