

研究论文

毛竹微型颠倒重复序列的鉴定及分子标记开发

虞莹玉 汤定钦 周明兵*

(浙江农林大学亚热带森林培育国家重点实验室, 杭州 311300)

摘要 MITEs(miniature inverted-repeat transposable elements)又称颠倒重复序列,是缺少转座酶序列的非自主型转座子,在真核生物基因组含量丰富,是基因组多态性形成的重要驱动力之一。该研究利用MITE Tracker软件,在毛竹(*Phyllostachys edulis*)新版基因组中鉴定到1 579个MITEs家族,共包括18 373个全长MITEs,占毛竹基因组的0.34%,被归入到7个超家族。超家族的插入时间跨度为0.5~16.5 mya,其中有3个超家族在2~4 mya经历过一次扩增事件;2个超家族分别在1~2 mya和3~5 mya经历两次扩增事件;2个超家族在0.5~16.5 mya经历一次长期扩增。MITEs偏好插入基因或基因附近,且Micron-like超家族偏好插入ATT与ATA之间。该研究开发了3个分子标记,可从12份雷竹变种变型材料中鉴定出4份。综上所述,该文主要分析了毛竹基因组中MITEs的分布、进化及插入情况,并获得了3对可以区分雷竹变种变型的分子标记,为下一步验证MITEs功能奠定基础。

关键词 毛竹;微型反向重复转座元件;基因组;转座子

Identification of Miniature Inverted-Repeat Transposable Elements and Development of Molecular Markers in *Phyllostachys edulis*

YU Yingyu, TANG Dingqin, ZHOU Mingbing*

(State Key Laboratory of Subtropical Silviculture, Zhejiang A&F University, Hangzhou 311300, China)

Abstract MITEs (miniature inverted-repeat transposable elements) are non-autonomous transposons lacking transposable enzyme sequences. It is abundant in eukaryotes genomes and is one of the most important driving forces for the formation of genomic polymorphism. In the present study, MITE Tracker, a novel software program, was used to identify MITEs in large genomes. A total of 1 579 MITEs families (18 373 full-length MITEs) in the latest genome version of *Phyllostachys edulis* (version 2018) were identified, accounting for 0.34% of the total genome of *Ph. edulis*. These MITEs families were clustered into seven super-families. The range of insertion time of the seven MITEs super-families was 0.5-16.5 mya. Three super-families might experience an expansion event during 2-4 mya. Two super-families might undergo two expansion events during 1-2 mya and 3-5 mya, respectively. Two super-families might experience a long expansion period during 0.5-16.5 mya. The MITEs in *Ph. edulis* genome preferred to insert into or near genes, and the Micron-like super-family preferred to

收稿日期: 2020-08-20 接受日期: 2020-11-09

浙江自然科学基金重点项目(批准号: LZ19C160001)和国家自然科学基金(批准号: 31870656、31470615)资助的课题

*通讯作者。Tel: 0571-63732738, E-mail: zhoutingbing@zafu.edu.cn

Received: August 20, 2020 Accepted: November 9, 2020

This work was supported by the Zhejiang Provincial Natural Science Foundation of China (Grant No.LZ19C160001) and the National Natural Science Foundation of China (Grant No.31870656, 31470615)

*Corresponding author. Tel: +86-571-63732738, E-mail: zhoutingbing@zafu.edu.cn

URL: <http://www.cjcb.org/arts.asp?id=5451>

insert between ATT and ATA regions. Furthermore, three pairs of primers were developed for the classification of the varieties and forms of *Ph. violascens* 'Prevernalis'. In conclusion, the distribution, evolution and insertion of MITEs in *P. edulis* genome were analyzed and three molecular markers were obtained, which laid the foundation for further verification of MITEs function.

Keywords *Phyllostachys edulis*; miniature inverted-repeat transposable elements; genome; transposon

转座子(Transposon)又称跳动DNA,是一种特殊的DNA元件,通过将新产生的拷贝插入到染色体的其他位置,实现在宿主基因组中移动,此过程通常会导致转座子数量扩增^[1-3]。转座子曾经一度被认为是垃圾DNA,但随着研究的深入,发现转座子不仅是基因组的重要组成部分,还驱动着基因组的进化、扩增和可塑性^[4-5]。同时,研究表明转座子在调控基因表达和表观遗传修饰上具有重要的作用^[6]。根据转座子是否具有RNA转座中间体,将其分为两大类:RNA类转座子(class I或反转录转座子)和DNA类转座子(class II)^[7]。

MITEs(miniature inverted-repeat transposable elements)属于DNA类转座子(class II),是缺少编码转座酶能力的非自主型转座子,是自主型转座子内部缺失的衍生物^[8-10]。MITEs最早在突变的玉米(*Zea mays*)等位基因中被发现^[11],随后的研究表明,MITEs在植物和动物界分布广泛且含量很高,在稻科植物类^[11]、真菌^[12]、昆虫^[13]和人类^[14]等生物中已经鉴定到大量的MITEs,如MITEs占水稻(*Oryza sativa*)基因组的10%,小麦(*Triticum aestivum*)的16.46%,拟南芥(*Arabidopsis thaliana*)的3.90%,蒺藜苜蓿(*Medicago truncatula*)的8%^[15-16]。与一般非自主型转座子相比,MITEs长度要小得多,为50~800 bp,(A+T)%丰富,两端具有高度保守的反向重复序列(terminal inverted repeats, TIRs, ≥ 10 nt)和2个正向重复短翼序列,被称为靶位点重复序列(target site duplications, TSDs, 2~10 nt)^[17]。MITEs偏好插入基因或近基因区,如在水稻中,MITEs倾向于分布在近基因区^[18]。MITEs的插入还会影响相邻基因表达调控,例如,水稻HTD1(high-tillering dwarf1)基因外显子中插入1个246 bp的MITE,导致水稻表现出高分蘖矮化^[19]。MITEs在真核生物基因组中占比高,具有高拷贝性、高插入多态性以及和基因相关联的特点,因此常被用于开发分子标记进行数量性状位点(quantitative trait locus, QTL)分析^[20]。此外,MITEs衍生的microRNAs或者siRNAs在转录和转录后水平调控特定靶基因,诱导表观遗传沉默。例如,水稻中MITEs衍生的

siRNAs靶向沉默赤霉素和油菜素甾体稳态相关基因,从而影响植物高度和叶角^[21]。最新研究发现,小RNA介导的MITEs甲基化,可通过抑制OsMIR156d/j(microRNA156d/j)和促进D14(DWARF14)基因的表达,调控水稻分蘖^[22]。

竹子是生长最快的植物,具有较强的固碳能力和较高的柔韧性,是世界上最重要的非木材森林资源之一^[23]。据报道,竹子在中国有近500个物种、48个属^[24]。其中毛竹(*Phyllostachys edulis*)分布范围广、栽培面积大,占竹子总生长面积的70%^[25],在所有竹子中具有最高的生态、经济和文化价值。

毛竹为二倍体植株,含有24对染色体(2N=48),基因组大小为1.91 Gb^[26]。转座子在毛竹基因组中含量高,且有证据表明毛竹转座子参与了基因组多态性形成^[27]和基因表达调控^[28-29],而MITEs倾向于插入基因区,可能是毛竹基因表达调控网的重要组成部分^[30],因此,系统研究毛竹基因组中MITEs的分布和进化模式,是全面研究MITEs功能的前提和基础。

我们课题组之前使用MITE Digger工具在毛竹中鉴定到362个MITEs家族,共489 592个MITEs相关序列,占毛竹基因组的4.74%^[31]。但是第一版的毛竹基因组组装碎片化明显,大大影响了MITEs鉴定的精确性。2018年高质量的新版毛竹基因组数据被公布,近几年MITEs鉴定工具检测也更加准确,假阳性率得到显著降低^[16],因此,非常有必要对毛竹基因组MITEs的分布情况和进化模式进行更系统的研究。本研究借助最新公布的毛竹基因组数据,使用最新MITEs鉴定工具——MITE Tracker^[16]对毛竹MITEs进行鉴定,分析了MITEs的分布、进化模式及插入情况,为进一步研究毛竹中MITEs的功能以及MITEs对基因表达和基因组进化的影响奠定基础。

1 材料与方法

1.1 毛竹MITEs的鉴定

从毛竹基因组数据库(<http://bamboo.bamboogdb.org/#/download>)中下载毛竹基因组序列和

基因注释信息。采用基于结构搜索的MITE Tracker软件(<https://github.com/INTABiotechMJ/MITE-Tracker>)^[16], 鉴定毛竹基因组中的MITEs[MITEs全长为40~800 bp, TIRs \geq 10 nt, TSDs为2~12 nt, 末端序列(G+C)%为15%~95%, 片段局部组成复杂性分数(local composition complexity, LCC) \geq 1]。使用VSEARCH工具(<https://github.com/torognes/vsearch?files=1>)^[32]对候选MITEs序列进行相似性比对, 再将各家族中MITEs的左、右侧翼序列(TSDs侧翼序列, 默认长度为40 nt)分别进行局部比对。具有相同侧翼序列的MITEs被视为同一个体, 只保留具有唯一侧翼序列的MITEs, 且每个家族中MITEs个数要 \geq 3。获取每个MITEs家族的一致性序列(consensus sequence), 将其作为问询序列(query sequence)分别在RepBase(http://www.girinst.org/Repbase_Update.html)^[33]、P-MITE(<http://pmite.hzau.edu.cn/django/mite>)^[15]和Genbank(<https://www.ncbi.nlm.nih.gov/genbank>) 3个数据库中搜索, 从而将家族分为已知保守性家族和Unknown超家族, 再依据已知保守性家族的TSD、TIR特征, 将其分到不同超家族中。将最终鉴定出的MITEs家族按照统一格式, 命名为PhXY#, 其中Ph代表毛竹(*Ph. edulis*), XY代表超家族(取前2个字母); #代表序号。

1.2 MITEs插入时间分析

使用DAMBE软件(<http://dambe.bio.uottawa.ca/DAMBE/dambe.aspx>)^[34]中的MUSCLE工具对每个家族中的MITEs序列进行多重比对, 获取每个家族的一致性序列。使用MEGA7软件(<http://www.megasoftware.net>)计算家族中每个MITE与对应一致性序列之间的核苷酸取代数K, 采用Kimura双参数距离法^[35]、公式“ $T=K/2r$ ”(假定 $r=1.30\times 10^{-8}$)^[36]预测MITEs插入时间。

1.3 MITEs家族的插入偏好性分析

编写Perl脚本, 从毛竹基因组中提取MITEs上下游10 bp的侧翼序列, 将同一超家族的侧翼序列比后用TBtools(<https://github.com/CJ-Chen/TBtools/releases>)绘制logo图^[37](默认参数)。

1.4 毛竹基因组中MITEs的分布分析

编写一个Perl脚本计算毛竹基因组上的基因和MITEs位置间的距离。本文将MITEs与基因位置关系分为八类, 即5'flank、3'flank、5'overlap、3'overlap、include、In、All、not_find。5'flank、3'flank分别表示MITEs插入于基因上游、基因下游;

include表示MITEs中包含基因; In表示MITEs插入基因内部; All表示MITEs与基因完全重叠; 5'overlap、3'overlap分别表示MITEs与基因的5'端、3'端有重叠。随机从毛竹基因组中抽取大小为10 Kb的片段1 000个, 假定中央位点为插入位点, 计算中央位点与毛竹基因的位置关系, 将其作为对照组^[38-39]。

1.5 MITEs在毛竹、雷竹变种变型中的插入多态性检测

2020年6月, 从浙江省杭州市临安区太湖源观赏竹种园采集毛竹和12个毛竹变种变型。2020年6月, 从浙江农林大学翠竹园采集雷竹和11个雷竹变种变型(表1)。用改进的十六烷基三甲基溴化铵(hexadecyltrimethylammonium bromide, CTAB)法^[40]制备基因组总DNA。我们随机选择了20个长度为300~800 bp的MITEs, 取其上下游100 bp序列设计引物(表2), 通过PCR进行插入多态性检测, 并测序验证。PCR反应程序: 94 °C预变性5 min; 94 °C变性30 s, 57 °C退火30 s, 72 °C延伸1 min, 30个循环; 72 °C完全反应7 min, 4 °C保存。采用1%琼脂糖凝胶进行电泳, 电压为120 V, 电泳30 min, 拍照成像。

2 结果与分析

2.1 MITEs的鉴定与特征分析

MITE Tracker是一个能在大基因组中精确识别MITEs的最新开源程序^[16], 通过识别有效TIRs和TSDs序列鉴定MITEs。本研究鉴定到1 579个MITEs家族, 共18 373个全长MITEs, 占毛竹基因组的0.34%, 根据MITEs两端的TIR和TSD特性, 将1 579个家族划分到7个超家族中, 每个超家族包含的家族数量、MITEs数量以及结构特征如表3所示。其中PIF/Harbinger-like超家族所包含的家族数量最多, 共有480个家族, 5 713个MITEs; Mutator-like超家族包含的MITEs最多, 共有460个家族、5 847个MITEs。候选MITEs的长度为40~800 bp, TSD长度为2~10 nt, TIR长度 \geq 10 nt。每个家族的(A+T)%差异很大, 但每个超家族的(A+T)%都 $>$ 50%, 其中Tc1/Mariner-like超家族(A+T)%最高, 达70.35%。

2.2 MITEs插入时间

MITEs的插入时间跨度为0~16.5 mya(million years ago), 其中CACTA-like、hAT-like和PIF/Harbinger-like 3个超家族在2~4 mya期间可能经历一次扩增事件, 随后CACTA-like和hAT-like又在0~1 mya经历一次扩

表1 毛竹和雷竹变种变型

Table 1 The varieties and forms of *Ph. edulis* and *Ph. violascens* 'Prevernalis'

种质编号 Accession No.	拉丁名 Latin name	种质编号 Accession No.	拉丁名 Latin name
M1	<i>Ph. edulis</i> (Carr.) H. de Lehaie f. holochrysa	L1	<i>Ph. violascens</i> cv. atrovaginis
M2	<i>Ph. edulis</i> (Carr.) H. de Lehaie f. edulis	L2	/
M3	<i>Ph. edulis</i> (Carr.) H. de Lehaie f. tubaeformis	L3	<i>Ph. violascens</i> cv. linansis
M4	<i>Ph. edulis</i> (Carr.) H. de Lehaie f. gracilis	L4	/
M5	<i>Ph. edulis</i> (Carr.) H. de Lehaie f. pachyloen	L5	/
M6	<i>Ph. edulis</i> (Carr.) H. de Lehaie f. huamozhu	L6	<i>Ph. violascens</i> cv. anhuiensis
M7	<i>Ph. edulis</i> (Carr.) H. de Lehaie f. exaurita	L7	/
M8	<i>Ph. edulis</i> (Carr.) H. de Lehaie 'Mira'	L8	/
M9	<i>Ph. edulis</i> (Carr.) H. de Lehaie f. purpureosulcata	L9	<i>Ph. violascens</i> cv. notata.
M10	<i>Ph. edulis</i> (Carr.) H. de Lehaie 'Kikko-chiku'	L10	<i>Ph. violascens</i> 'Prevernalis'
M11	<i>Ph. edulis</i> (Carr.) H. de Lehaie f. nabeshimana	L11	/
M12	<i>Ph. edulis</i> (Carr.) H. de Lehaie f. porphyrosticta	L12	/
M13	<i>Ph. edulis</i> (Carr.) H. de Lehaie f. epruinosa	/	/

“/”: 没有拉丁名。

“/”: latin name is not available.

表2 引物序列

Table 2 The sequences of primers

引物 Primer	序列(5'→3') Sequence (5'→3')	引物 Primer	序列(5'→3') Sequence (5'→3')
MITE 1	F: ATG CTC CCA ACA ACC AAC CA R: TTT GCT TGC TGC TGG TTC AC	MITE 11	F: GGT GCA CTG AAA ACT AGA TGC C R: GGA ACA GGA GGT GAA TGC CA
MITE 2	F: GGA ATC GAC CCC TCC AAT GT R: TCT CTC AAA AAC ACT CTC TCT ATG T	MITE 12	F: TAA CAA CAC CGC TCC CCA AG R: GCC GTT GAT TGG CTG GTT AC
MITE 3	F: GAA CAA GGA GAA AGG AAG AAG TGT R: ATT ATT CGG TGC GTG GGG A	MITE 13	F: AGC ATG AAG AAA GTG GAA AAA GGG R: TGA CAG TAT AGG CCG CAG GT
MITE 4	F: GGA AGG TAG CGT CAC CGT AG R: CCA CCC TTC TTC CCA TCG TC	MITE 14	F: GCA AGC CCG ACA GTA TCA AC R: TCA CAT TCT CTC CTC ATC TTC TCT
MITE 5	F: GAT TCG GTG TTC TTT TCC TCT CT R: CTC CAG TCT ACC AAC TTT ACA TCC T	MITE 15	F: GTT AAG CGA TTC TCA CCC GC R: ATG GGG ACG AGG CAT ATT CG
MITE 6	F: CTA TTC ATT GCG AGA AGC CCT R: CTC TTT TGC TAT GCT ACT TCA TGG T	MITE 16	F: CTT CCG TGC TTG CTA AAC AAA T R: TCC GTA ACA AAC ATG GAG ACC T
MITE 7	F: AGA AGG CAA CCA GTC TTG AGT R: CTC CTA ATC GTG TTTG CTT GGA	MITE 17	F: CAC AAT CCA CAT AGA CAT ACA CAC AG R: GCC ATC ATT TTA CCA AGT ATC CAG T
MITE 8	F: TAG GTC CCC CAA GTG AGA TGT R: ACG GTC AAG TAG GGG GAA GG	MITE 18	F: CGA GTT GGT CGT AGA CTA GTG AAC A R: ATT TGA TGC ACG ATT AGG TCC T
MITE 9	F: AAT ATC CTT GTT CCC TGG ACA GC R: TAA AGG GTT CAT CAC CCA AAG CC	MITE 19	F: TGC AAA ATC AAA CCC TCC TAA CC R: AGG GCT CAA CCT TTG CAA CT
MITE 10	F: AAC TAC ACA AAT CCG CCA GG R: AAT GGA ATA CTG GTG CAG CG	MITE 20	F: CAG TCG GTT CCC TAG AAA CTG T R: CAG ACA CTT TCA ACC ACA AAA CG

增。Micron-like和Unknown超家族可能分别在1~2 mya和3~5 mya经历了两次主要的扩增事件。Mutator-like和PIF/Harbinger-like在0.5~16.5 mya经历了一次长期扩增,是含有MITEs数量最多的2个超家族。Micron-like与Unknown 2个超家族的MITEs扩增趋势相似(图1)。

2.3 MITEs在基因组中的分布、与基因关系及各超家族插入偏好性分析

我们进一步研究了MITEs插入位点与基因的位置关系,主要包括5'flank、3'flank、5'overlap、3'overlap、include、In、All、not_find,然后分别将5'flank、3'flank

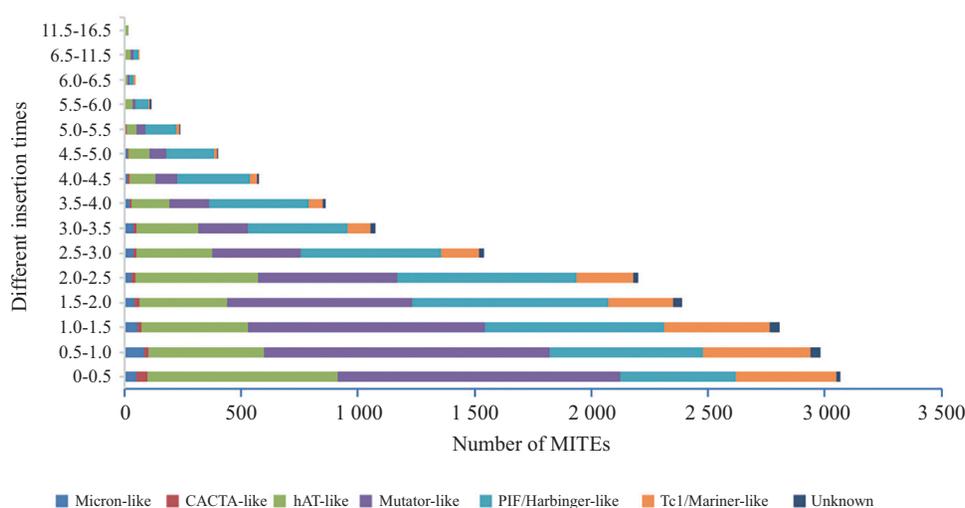
表3 毛竹全基因组MITEs的分类与特征分析

Table 3 Classification and characteristics analysis of MITEs in *Ph. edulis* genome

超家族 Super-family	家族数量 The number of family	全长MITEs的数量 The number of full-length MITEs ^a	MITEs长度范围及其总长度/bp The range of MITEs size and total length ^b /bp	超家族平均(A+T)含量/% The average (A+T) content in super-family /%
Micron-like	27	406	107-778 (140 692)	68.08
CACTA-like	22	180	48-736 (62 215)	50.48
hAT-like	325	3 738	40-800 (1 779 924)	55.03
Mutator-like	460	5 847	48-800 (2 011 703)	62.35
PIF/Harbinger-like	480	5 713	48-799 (1 960 224)	62.26
Tc1/Mariner-like	230	2 257	60-628 (535 159)	70.35
Unknown	35	232	99-791 (71 907)	61.64
Total	1 579	18 373	40-800 (6 561 824)	50.48-70.35

a: 两端TIRs完整的MITEs数量。b: 各超家族中MITEs的总长度。

a: the number of MITEs with complete TIRs in both ends. b: the length of all MITEs in each super-family.



X轴上显示的是各超家族的MITEs数量, Y轴上显示的是不同的插入日期(mya)。

The number of MITEs in each super-family is shown on the X-axis. The insertion time (mya) is shown on the Y-axis.

图1 毛竹基因组中MITEs超家族的插入时间分布

Fig.1 Distribution of insertion times of MITEs super-families in *Ph. edulis* genome

按照距离再分三类: $\text{flank} \leq 2 \text{ Kb}$ 、 $2 \text{ Kb} < \text{flank} \leq 5 \text{ Kb}$ 、 $\text{flank} > 5 \text{ Kb}$ 。结果表明, 共有2 895个MITEs与基因位置有交集(插入、包含和重叠); 6 768个MITEs插入位点位于近基因区($0 \text{ Kb} \leq \text{flank} \leq 5 \text{ Kb}$); 7 776个MITEs插入在的远基因区($> 5 \text{ Kb}$)(表4)。其中, Mutator-like超家族插入在毛竹基因中的MITEs数量最多, 共有876个, 占有所有基因内MITEs的35.36%; PIF/Harbinger-like超家族在近基因区分布的MITEs最多, 共2 316个。此外, 发现各超家族分布在基因上游(5' flank)的MITEs数量都大于基因下游(3' flank)。通过卡方检验, 将基因区、近基因区及远基因区的插入位点与随机抽取的对照组比较, 发现MITEs在基因附近的实际分布与对照组有极显著性差异(χ^2 检验, $P < 0.01$)。

为了进一步研究各超家族MITEs的插入偏好性, 我们选取MITEs前后10 bp, 按超家族归类对比后, 发现Micron-like超家族偏好插入ATT与ATA之间, 且插入位点附近(A+T)%非常高; Mutator-like、Tc1/Mariner-like、PIF/Harbinger-like和CACTA-like超家族分别偏好插入ATA/CTA、T/GTA、TA/CT和CAA/CA之间, hAT-like超家族无明显插入偏好性(图2), 这都与前人研究结果相一致^[15]。

2.4 MITEs在毛竹、雷竹变种变型中的插入多态性

我们用20对引物(表2), 分别对毛竹变种变型(13份)和雷竹变种变型(12份)进行插入位点的PCR验证并测序, 从而评估MITE Tracker预测结果的可靠性和MITEs作为分子标记的多态性。结果表明, 20对引物

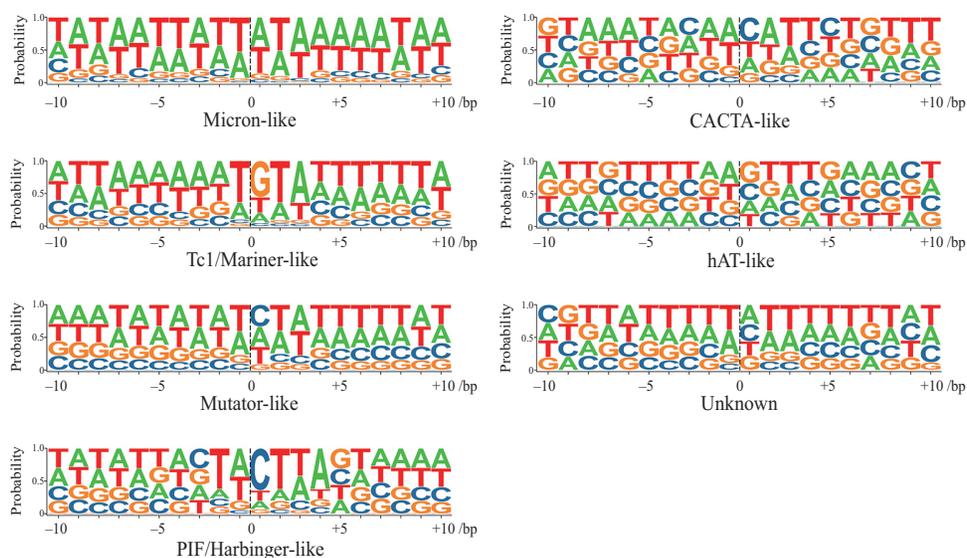
表4 毛竹基因组中MITEs在基因附近的分布

Table 4 The distribution of MITEs near the closest genes in *Ph. edulis* genome

超家族 Super-family	5'overlap ^a	3'overlap ^b	Include ^c	In ^d	All ^e	Flank ^f ≤2 Kb	2 Kb<flank ^f ≤5 Kb	Flank ^f >5 Kb
Micron-like	2	0	48	0	0	84	57	191
CACTA-like	0	1	25	0	0	45	16	80
hAT-like	11	12	716	3	0	623	493	1 705
Mutator-like	32	19	823	1	1	1 530	733	2 433
PIF/Harbinger-like	26	29	581	0	0	1 483	833	2 454
Tc1/Mariner-like	12	8	498	1	0	496	295	817
Unknown	2	0	44	0	0	49	31	96
Total	85	69	2 735	5	1	4 310	2 458	7 776

a: 与基因5'端重叠的MITEs数量。b: 与基因3'端重叠的MITEs数量。c: 包含有基因的MITEs数量。d: 插入基因内的MITEs数量。e: 与基因完全重叠的MITEs数量。f: 在基因侧翼区的MITEs数量。

a: the number of MITEs overlapped with 5' end of gene. b: the number of MITEs overlapped with 3' end of gene. c: the number of MITEs containing genes. d: the number of MITEs inserted into genes. e: the number of MITEs completely overlapped with genes. f: the number of MITEs in 5' and 3' flank regions of genes.



虚线处是MITEs的插入位点, 虚线左边为MITEs左侧10 bp侧翼序列, 右边为MITEs右侧10 bp侧翼序列。

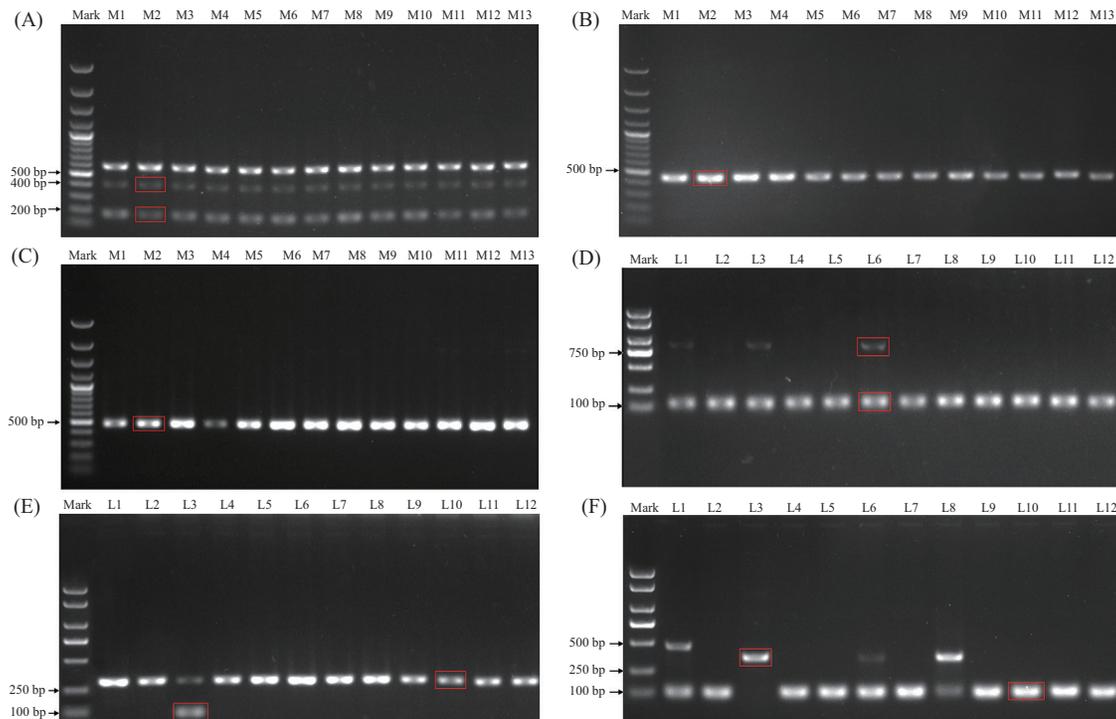
The position of the dotted line is the insertion site of MITEs. The left-flanking sequence of MITEs is on the left side of the dotted line, and right-flanking sequence of MITEs is on the right side of the dotted line.

图2 MITEs超家族的插入偏好分析

Fig.2 The insertion preference analysis of MITEs super-family

在毛竹及变种变型材料上都能扩增出预期大小的片段, 但未表现出多态性, 13份材料上扩增出的条带都是一致的(图3)。将前3对引物在毛竹(M2)上扩增出来的条带进行测序, 结果显示扩增产物与MITE Track预测的MITEs序列一致, 且匹配度都在91.47%以上(表2和图4)。这20对引物在雷竹及变种变型材料上也能扩出预期大小的片段, 且MITE 4、MITE 5、MITE 6在12份材料上呈现出多态性(图3): 在12份材料中, MITE 4在3份材料(L1、L3、L6)上表现为在1条染色体上存

在, 而在其同源染色体上等位位点处缺失, 在其余9份材料上表现为在1条染色体上都缺失; MITE 5在11份材料上表现为在1条染色体上都存在, 在1份材料(L3)上表现为在1条染色体上存在, 而在其同源染色体上等位位点处缺失; MITE 6在1份材料(L3)上表现为在1条染色体上都存在, 在3份材料(L1、L6、L8)上表现为在1条染色体上存在, 而在其同源染色体上等位位点处缺失, 在剩余8份材料上表现为在1条染色体上都缺失。分别将MITEs存在与缺失的条带回测



A: 毛竹变种型中MITE 1的PCR扩增结果。PCR产物长度在400 bp的表示MITE 1在该位置存在, 产物长度在180 bp左右的表示MITE 1在该位置缺失; B: 毛竹变种型中MITE 2的PCR扩增结果。产物长度在450 bp左右的表示 MITE 2在该位置存在; C: 毛竹变种型中MITE 3的PCR扩增结果。产物长度在500 bp左右的表示MITE 3在该位置存在; D: 雷竹变种型中MITE 4的PCR扩增结果。产物长度在750 bp左右的表示MITE 4在该位置存在, 产物长度在100 bp左右的表示MITE 4在该位置缺失; E: 雷竹变种型中MITE 5的PCR扩增结果。产物长度在300 bp左右的表示MITE 5在该位置存在, 产物长度在100 bp左右的表示MITE 5在该位置缺失; F: 雷竹变种型中MITE6的PCR扩增结果。产物长度在350 bp左右的表示MITE 6在该位置存在, 产物长度在100 bp左右的表示MITE 6在该位置缺失。红色框为测序条带, M1~M13、L1~L12代表的竹种见表1。

A: the PCR amplification result of MITE 1 in genomes of the varieties and forms of *Ph. edulis*. The PCR products about 180 bp reveal the absence of MITE 1, and the PCR products about 400 bp reveal the presence of MITE 1 at the locus; B: the PCR result of MITE 2 in genomes of the varieties and forms of *Ph. edulis*. The PCR products about 450 bp reveal the presence of MITE 2 at the locus; C: the PCR result of MITE 3 in genomes of the varieties and forms of *Ph. edulis*. The PCR products about 500 bp reveal the presence of MITE 3 at the locus; D: the PCR result of MITE 4 in genomes of the varieties and forms of *Ph. violascens* 'Prevernalis'. The PCR products about 100 bp reveal the absence of MITE 4 at the locus, and the PCR products about 750 bp reveal the presence of MITE 4 at the locus; E: the PCR result of MITE 5 in genomes of the varieties and forms of *Ph. violascens* 'Prevernalis'. The PCR products about 100 bp reveal the absence of MITE 5 at the locus, and the PCR products about 300 bp reveal that the presence of MITE 5 at the locus; F: the PCR result of MITE 6 in genomes of the varieties and forms of *Ph. violascens* 'Prevernalis'. The PCR products about 100 bp reveal the absence of MITE 6 at the locus, and the PCR products about 350 bp reveal the presence of MITE 6 at the locus. Sequencing bands are shown in red boxes. Bamboo species represented by M1-M13 and L1-L12 are shown in Table 1.

图3 毛竹变种型MITE 1、MITE 2、MITE 3及雷竹变种型MITE 4、MITE 5、MITE 6扩增电泳图

Fig.3 Amplification electrophoretograms of MITE 1, MITE 2, MITE 3 in genomes of the varieties and forms of *Ph. edulis*, and MITE 4, MITE 5, MITE 6 in genomes of the varieties and forms of *Ph. violascens* 'Prevernalis'

序, 测序结果与MITE Tracker预测的MITEs序列相匹配(表2和图4)。将缺失位点进行测序比对, 可发现MITE 1(hAT超家族)、MITE 4(hAT超家族)、MITE 6(Tc1超家族)切割不精确, 未将MITEs全部序列切除, 而MITE 5(Mutator超家族)切走了全部MITEs序列及部分侧翼序列(图5)。

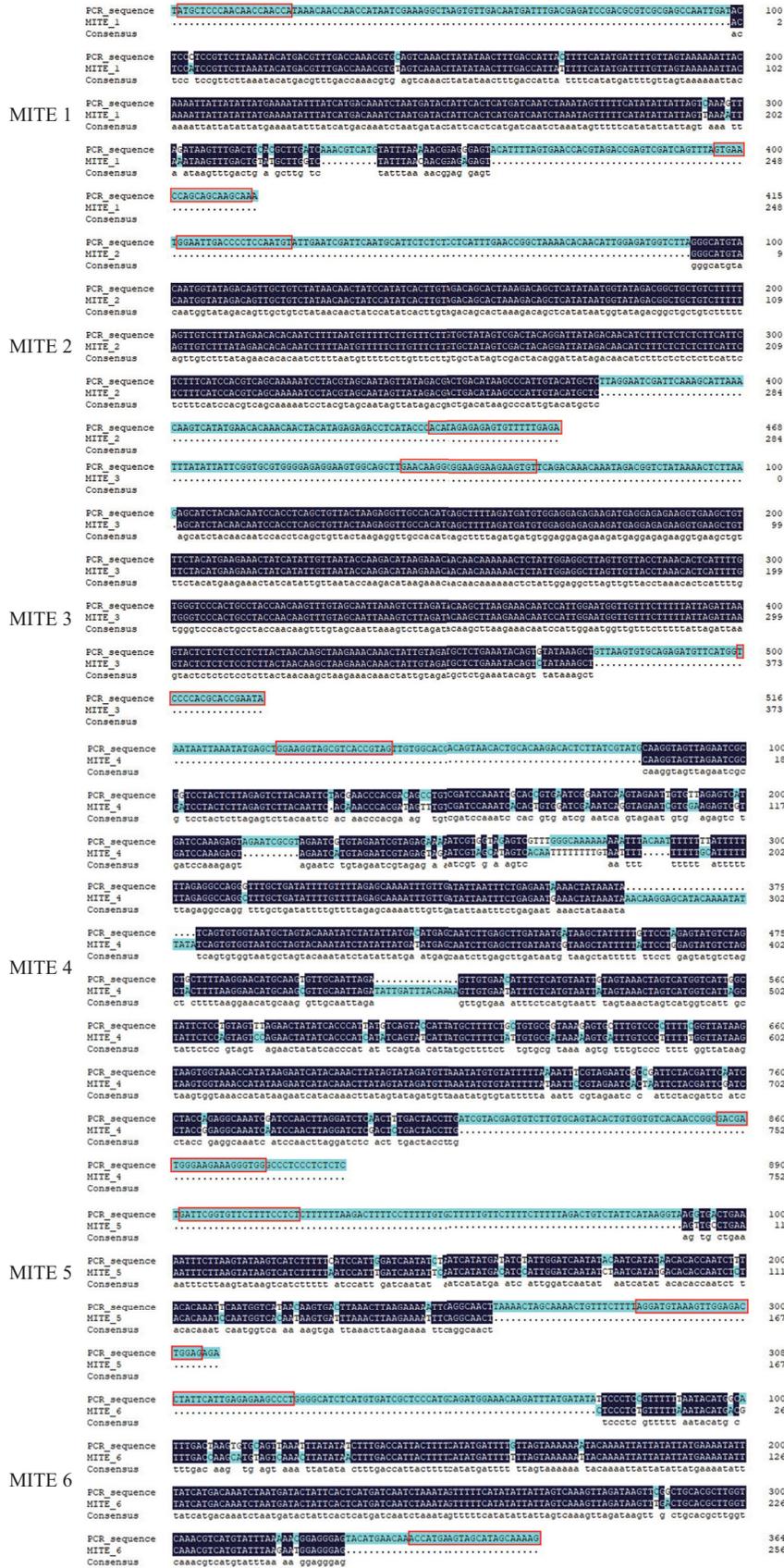
3 讨论

3.1 毛竹中MITEs的鉴定与特征

本研究中, 我们首次利用MITE Tracker对毛竹

MITEs进行全基因组分析, 鉴定到1 579个MITEs家族, 归属于7个超家族。虽然各家族在TSD、TIR序列和全长序列上存在巨大差异, 但大多数家族都具有MITEs高拷贝数和高(A+T)%的特点。此外, 通过PCR和测序进一步验证了竹子中MITEs的存在, 这些都验证了MITE Tracker在鉴定毛竹MITEs中的可靠性。

毛竹的基因组大小为1.91 Gb, 我们共鉴定到1 579个MITEs家族共18 373个MITEs相关序列。已有研究表明, 基因组中MITEs数量与基因

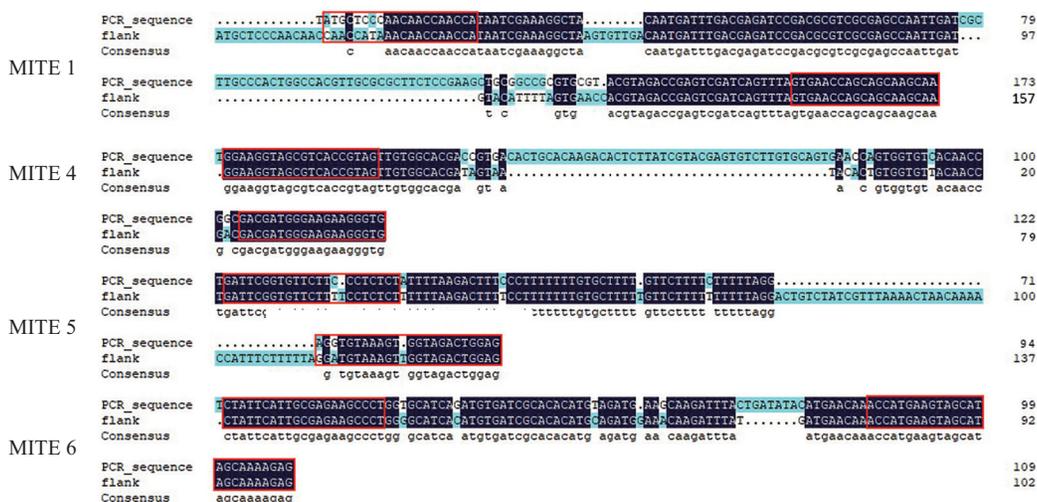


红色框为上下游引物。

The red boxes marked upstream and downstream primers.

图4 毛竹的MITE 1、MITE 2、MITE 3及雷竹变种变型的MITE 4、MITE 5、MITE 6扩增产物测序比对结果

Fig.4 Sequence alignment results of the amplified products of MITE 1, MITE 2, MITE 3 in genomes of the varieties and forms of *Ph. edulis*, and MITE 4, MITE 5, MITE 6 in genomes of the varieties and forms of *Ph. violascens* 'Prevernalis'



红色框为上下游引物。

The red boxes marked upstream and downstream primers.

图5 毛竹的MITE 1及雷竹变种变型中MITE 4、MITE 5和MITE 6转座足迹

Fig.5 Transposition footprints of MITE 1 in genomes of the varieties and forms of *Ph. edulis*, and MITE 4, MITE 5, MITE 6 in genomes of the varieties and forms of *Ph. violascens* 'Prevernalis'

组大小显著相关, 如水稻 (*O. sativa*) 基因组大小为 389 Mbp, 采用 MITE Tracker 鉴定出 17 651 个全长 MITEs, 而小麦 (*T. aestivum*) 基因组大小 (14 Gb) 是水稻的 35 倍, 包含全长 MITEs 序列 128 453 个^[16]。因此, 本研究在毛竹基因组中鉴定到的 MITEs 数量是合理的。与课题组之前的结果相比较^[31], MITE Tracker 在毛竹基因组中检测到的全长 MITEs 数量减少, 这可能与软件巧妙的计算策略有关, 其能够过滤重复计算的 MITEs 并且有效克服嵌套问题^[16]。此外, 本文使用的毛竹基因组是最近更新的组装版, 与原版相比组装版质量显著改善, 这些都在很大程度上降低了 MITEs 鉴定的假阳性。同时, Unknown 家族占比大大减少, 也从侧面印证了 MITE Tracker 鉴定的准确性的提高。然而, 目前毛竹基因组的组装还是较为分散, 完整度还有待提高^[26]。即使我们用了结构与同源性相结合的方法来鉴定 MITEs, 毛竹基因组的断裂仍然可能导致一些 MITEs 遗漏。因此, 这里确定的 18 373 个 MITEs 相关序列可能是一个保守的数字, 在未来还会发现更多。

3.2 毛竹中 MITEs 的扩增与多样性模式

超家族间 MITEs 的插入时间范围是 0~16.5 mya, 与课题组之前的研究^[31]相比, 缩短了五分之三, 并且两次研究得到的各超家族的扩增趋势也有差异, 而造成差异的原因可能是 MITEs 鉴定软件的准确性提

高, 大多假阳性结果被过滤, 同时新版毛竹基因组的准确性和连续性提高, 也提高了软件鉴定的准确性。本研究中各家族的扩增主要集中于 0~5 mya 和 6~11.5 mya。最初家族内只有少数几个拷贝, 但经历多次连续的扩增后家族内 MITEs 拷贝数呈爆炸式增长^[41], 通常毛竹基因组中 MITEs 的扩增可能在大部分时间段是不活跃的, 但“基因组休克”或同源转座酶的触发能引起 MITEs 的激活^[42-43]。如通过冷胁迫、盐胁迫、辐照、细胞培养等方式能激活水稻中的 mping 转座子^[10,44-45]。

3.3 MITEs 在毛竹基因组中的分布与基因关系

MITEs 在基因附近的实际分布与随机插入对照组相比有极显著差异, MITEs 更偏好插入基因或近基因区, 这与课题组之前的研究^[31]相一致, 但本研究中基因附近 MITEs 的百分含量明显增高, 且基因区与近基因区 MITEs 的数量和甚至多于远基因区 (表 4), 尤其是基因编码区上下游 2 Kb 范围内有大量的 MITEs 插入, 这些 MITEs 大多属于 Mutator-like 和 PIF/Harbinger-like 这 2 个一直处于扩增状态的超家族, 说明 MITEs 可能对基因调控具有重要作用^[19,41]。此外, 我们还发现 Micron-like 超家族偏好插入 (TA)_n 个重复序列中, 只有少数插入了 (CA)_n/(GT)_n 个重复序列中, 这与前人研究结果一致^[15]。

3.4 MITEs 在雷竹变种和栽培种中的插入多态性在毛竹变种变型中能够扩增并通过测序得到

相应的MITEs序列,证明MITE Tracker预测的数据是可靠的。这些MITEs是从毛竹基因组中获得的,但在雷竹变种变型中也能扩增出相应片段,说明毛竹和雷竹中的MITEs具有保守性。

另外,缺失位点的测序表明,有些MITEs家族在发生跳跃前的切割是不精准的,会丢失部分MITEs序列或切走部分侧翼序列^[19],其中PIF/Harbinger-like超家族为精准切割,其余家族为非精准切割。

植物中MITEs分布丰富、多态性高,并且MITEs的插入影响毛竹基因功能,因此MITEs常用于设计分子标记。MITE 4、MITE 5、MITE 6引物在雷竹变种变型上有扩增多态性,可作为分子标记用于区分雷竹竹种,在12份雷竹变种变型材料中可鉴别出雷山乌、弯杆雷竹、安徽雷竹和阔叶青头雷竹。本文开发的分子标记的多态性来源主要是MITEs内部的变异,因此多态性低,但我们可以利用这些内部有变异的MITEs开发其他多态性丰富的标记,如SSAP(sequence-specific amplification polymorphism)分子标记,从而将所有变种变型都区分开来。

4 总结

本文用最新MITEs鉴定工具MITE Tracker^[16]对最新毛竹基因组中MITEs进行鉴定,获得1 579个MITEs家族,共18 373个拷贝,根据MITEs两端的TIR和TSD特性,将1 579个家族归入6个已知超家族和1个Unknown超家族,家族间的比较进一步揭示了MITEs的进化模式。我们发现毛竹中MITEs偏好插入基因或近基因区,Micron-like超家族偏好插入ATT与ATA之间,且插入位点附近AT含量非常丰富。此外,我们还开发了MITEs分子标记,获得了能够区分雷竹变种和变型的3个分子标记。本研究为进一步理解毛竹中MITEs的功能以及MITEs对基因表达的影响等奠定了基础。

参考文献 (References)

- [1] CHENAIS B, CARUSO A, HIARD S, et al. The impact of transposable elements on eukaryotic genomes: from genome size increase to genetic adaptation to stressful environments [J]. *Gene*, 2012, 509(1): 7-15.
- [2] CASACUBERTA E, GONZALEZ J. The impact of transposable elements in environmental adaptation [J]. *Mol Ecol*, 2013, 22(6): 1503-17.
- [3] GRZEBELUS D. The functional impact of transposable elements on the diversity of plant genomes [J]. *Diversity*, 2018, 10(2): 18.
- [4] DOOLITTLE W, SAPIENZA C. Selfish genes, the phenotype paradigm and genome evolution [J]. *Nature*, 1980, 284(5757): 601-3.
- [5] ORGEL L, CRICK F. Selfish DNA: the ultimate parasite [J]. *Nature*, 1980, 284(5757): 604-7.
- [6] DUBIN M J, SCHEID O M, BECKER C. Transposons: a blessing curse [J]. *Curr Opin Plant Biol*, 2018, 42: 23-9.
- [7] WICKER T, SABOT F, HUAVAN A, et al. A unified classification system for eukaryotic transposable elements [J]. *Nat Rev Genet*, 2007, 8(12): 973.
- [8] BUREAU T, RONALD P, WESSLER S, et al. A computer-based systematic survey reveals the predominance of small inverted-repeat elements in wild-type rice genes [J]. *Proc Natl Acad Sci USA*, 1996, 93(16): 8524-9.
- [9] FESCHOTTE C, MOUCHES C. Evidence that a family of miniature inverted-repeat transposable elements (MITEs) from the *Arabidopsis thaliana* genome has arisen from a pogo-like DNA transposon [J]. *Mol Biol Evol*, 2000, 17(5): 730-7.
- [10] JIANG N, BAO Z R, ZHANG X Y, et al. An active DNA transposon family in rice [J]. *Nature*, 2003, 421(6919): 163-7.
- [11] BUREAU T E, WESSLER S R. Tourist: a large family of small inverted repeat elements frequently associated with maize genes [J]. *Plant Cell*, 1992, 4(10): 1283-94.
- [12] YEADON P J, CATCHESIDE D E. Guest: a 98 bp inverted repeat transposable element in *Neurospora crassa* [J]. *Mol Genet*, 1995, 247(1): 105-9.
- [13] TU Z. Three novel families of miniature inverted-repeat transposable elements are associated with genes of the yellow fever mosquito, *Aedes aegypti* [J]. *Proc Natl Acad Sci USA*, 1997, 94(14): 7475-80.
- [14] SMIT A F. The origin of interspersed repeats in the human genome [J]. *Curr Opin Genet Dev*, 1996, 6(6): 743-8.
- [15] CHEN J J, HU Q H, ZHANG Y, et al. P-MITE: a database for plant miniature inverted-repeat transposable elements [J]. *Nucleic Acids Res*, 2014, 42: D1176-81.
- [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 Bioinformatics*, 2018, 19(1): 348.
- [17] GUO C, SPINELLI M, YE C, et al. Genome-wide comparative analysis of miniature inverted repeat transposable elements in 19 *Arabidopsis thaliana* ecotype accessions [J]. *Sci Rep*, 2017, 7(1): 2634.
- [18] LU C, CHEN J 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*, 2011, 29(3): 1005-17.
- [19] TANG Y, MA X, ZHAO S, et al. Identification of an active miniature inverted-repeat transposable element mJing in rice [J]. *Plant J*, 2019, 98(4): 639-53.
- [20] MONDEN Y, NAITO K, OKUMOTO Y, et al. High potential of a transposon mPing as a marker system in *japonica* × *japonica* cross in rice [J]. *DNA Res*, 2009, 16(2): 131-40.
- [21] SONG X W, CAO X F. Transposon-mediated epigenetic regulation contributes to phenotypic diversity and environmental adaptation in rice [J]. *Curr Opin Plant Biol*, 2017, 36: 111-8.

- [22] 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): 851-63.
- [23] ZHAO H S, YANG L, PENG Z H, et al. Developing genome-wide microsatellite markers of bamboo and their applications on molecular marker assisted taxonomy for accessions in the genus *Phyllostachys* [J]. *Sci Rep*, 2015, 5: 8018.
- [24] GUI Y J, WANG S, QUAN L Y, et al. Genome size and sequence composition of moso bamboo: a comparative study [J]. *Sci China C Life Sci*, 2007, 50(5): 700-5.
- [25] LI L, CHENG Z C, MA Y J, et al. The association of hormone signalling genes, transcription and changes in shoot anatomy during moso bamboo growth [J]. *Plant Biotechnol J*, 2018, 16(1): 72-85.
- [26] ZHAO H S, GAO Z M, WANG L, et al. Chromosome-level reference genome and alternative splicing atlas of moso bamboo (*Phyllostachys edulis*) [J]. *GigaScience*, 2018, 7(10): 115.
- [27] LI S T, RAMAKRISHNAN M, VINOD K K, et al. Development and deployment of high-throughput retrotransposon-based markers reveal genetic diversity and population structure of Asian bamboo [J]. *Forests*, 2020, 11(1): 31.
- [28] ZHOU M B, CHEN A, ZHOU Q Q, et al. A moso bamboo (*Phyllostachys edulis*) miniature inverted-repeat transposable element (MITE): the possible role of a suppressor [J]. *Tree Genet Genomes*, 2017, 13: 129.
- [29] 胡慧, 周明兵, 杨萍, 等. 毛竹微型颠倒重复序列转座子 PhTourist1 的克隆与分析 [J]. *林业科学* (HU H, ZHOU M B, YANG P, et al. Cloning and analysis of miniature inverted repeat transposable elements PhTourist1 from *Phyllostachys edulis* [J]. *Sci Silv Sin*), 2015, 51(5): 127-34.
- [30] 周倩倩. 转座子在毛竹基因组启动子区域的分布及功能分析 [D]. 杭州: 浙江农林大学, 2016.
- [31] ZHOU M B, TAO G Y, PI P Y, et al. Genome-wide characterization and evolution analysis of miniature inverted-repeat transposable elements (MITEs) in moso bamboo (*Phyllostachys heterocycla*) [J]. *Planta*, 2016, 244(4): 775-87.
- [32] ROGNES T, FLOURI T, NICHOLS B, et al. Vsearch: a versatile open source tool for metagenomics [J]. *PeerJ*, 2016, 4: 2584.
- [33] JURKA J, KAPITONOV V, PAVLICEK A, et al. Repbase update, a database of eukaryotic repetitive elements [J]. *Cytogenet Genome Res*, 2005, 110(1/4): 462-7.
- [34] XIA X, XIE Z. DAMBE: software package for data analysis in molecular biology and evolution [J]. *J Hered*, 2001, 92(4): 371-3.
- [35] KIMURA M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences [J]. *J Mol Evol*, 1980, 16(2): 111-20.
- [36] YANG Y, KANG L, WU R, et al. Genome-wide identification and characterization of UDP-glucose dehydrogenase family genes in moso bamboo and functional analysis of *PeUGDH4* in hemicellulose synthesis [J]. *Sci Rep*, 2020, 10(1): 10124.
- [37] CHEN C J, CHEN H, ZHANG Y, et al. TBtools: An integrative toolkit developed for interactive analyses of big biological data [J]. *Mol Plant*, 2020, 13(8): 1194-202.
- [38] NAITO K, CHO E, YANG G J, et al. Dramatic amplification of a rice transposable element during recent domestication [J]. *Proc Natl Acad Sci USA*, 2006, 103(47): 17620-5.
- [39] HAN M J, SHEN Y H, GAO Y H, et al. Burst expansion, distribution and diversification of MITEs in the silkworm genome [J]. *BMC Genomics*, 2010, 11: 520.
- [40] DOYLE J J, DOYLE J L. A rapid isolation procedure for small quantities of fresh leaf materials [J]. *Phytochem Bull*, 1987, 19: 11-5.
- [41] ZERJAL T, JOETS J, ALIX K, et al. Contrasting evolutionary patterns and target specificities among three Tourist-like MITE families in the maize genome [J]. *Plant Mol Biol*, 2009, 71(1/2): 99-114.
- [42] MCCLINTOCK B. The significance of responses of the genome to challenge [J]. *Science*, 1984, 226(4676): 792-801.
- [43] YANG G J, ZHANG F, HANCOCK C, et al. Transposition of the rice miniature inverted repeat transposable element mPing in *Arabidopsis thaliana* [J]. *Proc Natl Acad Sci USA*, 2007, 104(26): 10962-7.
- [44] NAITO K, ZHANG F, TSUKIYAMA T, et al. Unexpected consequences of a sudden and massive transposon amplification on rice gene expression [J]. *Nature*, 2009, 461(7267): 1130-4.
- [45] NAKAZAKI T, OKUMOTO Y, HORIBATA A, et al. Mobilization of a transposon in the rice genome [J]. *Nature*, 2003, 421(6919): 170-2.