

枣树全基因组 MITEs 成分分析与系统进化研究

李勇慧, 冯爱青, 押辉远, 程彦伟, 逯海朋

(洛阳师范学院 生命科学学院 河南 洛阳 471022)

摘要: 微型反向重复转座基因 (minialure inverted repeat transposable elements, MITEs) 的演化可能是造成枣树基因组及品种多样性的重要原因. 为研究 MITEs 在枣树全基因组中的分布、种类及演化, 使用 MITE 预测软件 (MITE-Digger) 识别枣树全基因组中的 MITEs 序列, 在植物 MITEs 数据库中进行分类注释, 并用 MEGA 6.0 构建枣树基因组中 MITEs 的系统进化树. 结果显示, MITE-Digger 检测出 MITEs 总共 3 230 条, 碱基数为 1 104 679 bp, 占基因组比例为 0.34%; 其中共有 948 条在植物 MITEs 数据库中比对到相似序列, 这些 MITEs 可以归类到 5 个家族: hAT, PIF/Harbinger, CACTA, Mutator 和 Tc1/Mariner, MITEs 系统进化树显示: 枣树基因组中 MITEs 来源于少数几个共同祖先, 在长期的进化中有明显的扩增和突变, 这可能与枣树品种多样性有关. 研究结果将为枣树的遗传资源及育种研究提供重要参考.

关键词: MITEs; 枣树基因组; 系统进化; IMP 分子标记技术

中图分类号: Q949.4

文献标志码: A

文章编号: 1671-6841(2015)04-0103-05

DOI: 10.3969/j.issn.1671-6841.2015.04.020

0 引言

转座因子 (transposable element) 是一类可移动的 DNA 片段, 在真核生物的基因组中分布非常广泛. 研究显示转座因子的活性在功能、结构及进化等方面对生物基因组有一定的影响, 也是形成物种多样性的主要原因之一^[1-2]. 依据转座机理的不同, 把转座因子分为两种类型: I 型转座元件, 也称反转录转座子 (retrotransposon), II 型转座元件, 也叫 DNA 转座子 (DNA transposon). 按转座活性的有无来划分, DNA 转座子又分为自主型转座子、非自主型转座子和微型颠倒重复序列 (MITEs). MITEs 长度一般为 100 ~ 800 bp, 比较短, 包括末端倒转重复 (terminal inverted repeats, TIRs) 及靶位点的重复序列 (target site duplication, TSD), 不能编码转座酶. MITEs 与反转座子类似的地方是拷贝数较高, 研究发现它们和植物中很多基因相连, 主要是由于它们有很强的插入低拷贝区域的偏好, 能够插入和跳出基因的某一片段^[3]. 大部分的 MITEs 很可能来源于不同的 DNA 转座子亚家族 Tc1/mariner, PIF/IS5, CACATA, hAT 或 Mutator. 通过对水稻基因组测序进行分析显示^[4-5], MITEs 主要分布于常染色质区, 能够形成巢式结构, 与反转录转座子不同. 研究表明^[6], 大多数 MITEs 都存在于靠近基因附近的非编码区. 转座子进行转座会对基因组的大小、结构排列以及染色质的组成成分产生一定的影响, 从而进一步影响到基因组的结构、功能及进化.

枣树 (*Zizyphus jujube Mill.*) 果实是中国北方的主要水果之一. 目前, 有数百个枣树品种, 不同地区果实形状、果肉味道和肉质会有所不同, 这可能与枣树基因组的多样性有关. 枣树的这种多样性可能与枣树基因组内的转座子有关. MITEs 是一类与基因组进化密切相关的转座子, MITEs 分子标记技术 IMP (inter-MITE polymorphisms) 在许多植物中都已经成功应用. 最近几年有人开发了检测 MITEs 的新软件: MITE-Digger^[6], 对分析从头合成比较有效, 尤其对基因组比较大的序列更有效. 而且 MITE-Digger 分析的正确率高, 错误率低, 且分析速度比较快, 在 Windows 系统下就可以使用, 所以本文采用 MITE-Digger 这个软件来分析枣树的基因

收稿日期: 2015-07-10

基金项目: NSFC-河南人才培养联合基金资助项目, 编号 U1204307; 河南省高等学校重点科研项目, 编号 15A2100410; 洛阳师范学院省级培育项目, 编号 2010-PYJJ-007.

作者简介: 李勇慧 (1977-), 女, 河南孟津人, 副教授, 硕士, 主要从事植物学生物技术研究, E-mail: huiyongli8209@126.com; 通讯作者: 押辉远 (1977-), 河南许昌人, 副教授, 博士, 主要从事植物辐照研究, E-mail: yahuiyuan@lynu.edu.cn.

引用本文: 李勇慧, 冯爱青, 押辉远, 等. 枣树全基因组 MITEs 成分分析与系统进化研究 [J]. 郑州大学学报: 理学版, 2015, 47 (4): 103-107.

组. 鉴于目前枣树基因组测序刚刚完成, MITEs 在枣树全基因组中的分布、种类、所占比例、枣树基因组 MITEs 演化的研究还没有报道, 本研究试图通过分析枣树基因组中 MITEs 转座子来研究枣树基因组多样性的原因, 为枣树多样性研究和遗传资源鉴别提供有力工具.

1 数据来源

枣树的全基因组有 12 条染色体, 从 NCBI(<http://www.ncbi.nlm.nih.gov/>) 数据库中下载枣树的全基因组(12 条染色体, 序列的 NCBI 号分别为 CM003114.1, CM003115, CM003116.1, CM003117.1, CM003118.1, CM003119.1, CM003120.1, CM003121.1, CM003122.1, CM003123.1, CM003124.1, CM003125.1), 序列保存为 FASTA 格式.

2 生物信息学方法

2.1 基于 MITE-Digger 鉴定枣树的全基因组 MITEs

利用 MITE-Digger 软件分析枣树的全基因组. MITE-Digger 在分析枣树基因组序列时识别枣树基因组中 MITEs 的运行参数: MITE 全长 100 ~ 800 bp(也就是大于 100 bp 小于 800 bp 的碱基序列鉴定为 MITEs); TSD 碱基数范围为 >2; 侧翼序列不同碱基数的阈值为 5 bp^[7].

2.2 MITEs 的注释及分类 BLAST

植物 MITEs 数据库(plant MITE database, P-MITE)在对基因和基因组序列的注释有很大的帮助和益处, 它经常被用于研究 MITEs 的起源和扩增^[8]. 我们通过 P-MITE 的 BLAST(<http://pmite.hzau.edu.cn/MITE-blast/blast.htm>) 功能来注释所得 MITEs 的归类, 从而找到相应的 MITEs 所对应的家族.

2.3 MEGA 6.0 软件构建系统发育树

将能够注释的 MITEs 序列以 FASTA 格式输入 MEGA 6.0, 以最大简约法构建系统发育树, 使用软件默认参数.

3 结果与分析

3.1 枣树 MITEs 的分布

MITE-Digger 检测出枣树基因组中的 MITEs 总共 3 230 条, 总碱基数为 1 104 679 bp, 占全基因组的比例为 0.34%. 共有 948 条在植物 MITEs 数据库(P-MITE)中比对到相似序列, 碱基数为 323 856 bp, 平均每 339 201.0 bp 出现一个注释的 MITE, 占基因组比例为 0.10%. 各条染色体检测出的 MITE 条数及占各染色体的比例如表 1.

表 1 各条染色体检测出的 MITEs

Tab. 1 Detected MITEs of each chromosome

染色体编号	染色体总碱基数/bp	MITE 条数	MITE 总碱基数/bp	MITE 频数	MITE 平均长度/bp	MITE 占染色体比例
1	42 390 363	424	144 178	99 977.3	340.0	0.003 4
2	27 986 743	304	106 881	92 061.7	351.6	0.003 8
3	26 737 297	276	89 482	96 874.3	324.2	0.003 3
4	30 445 767	35	17 171	869 879.1	490.6	0.000 6
5	31 365 312	309	111 210	101 505.9	359.9	0.003 5
6	25 259 912	284	100 994	88 943.4	355.6	0.004 0
7	27 644 224	407	137 957	67 921.9	339.0	0.005 0
8	23 350 829	210	66 660	111 194.4	317.4	0.002 9
9	25 348 792	302	99 504	83 936.4	329.5	0.003 9
10	20 983 256	234	76 199	89 672.0	325.6	0.003 6
11	20 703 932	179	56 794	115 664.4	317.3	0.002 7
12	19 346 091	266	97 649	72 729.7	367.1	0.005 0
基因组	321 562 518	3 230	1 104 679	99 555.0	342.0	0.003 4

从表 1 中可以看出第 4 条染色体检测出来的 MITEs 是最少的;第 1 条染色体检测出来的 MITEs 是最多的;其他的几条染色体检测出来的 MITEs 相差不太大. 平均每 99 555.0 bp 出现一个 MITE,每个 MITE 的平均长度为 342.0 bp. 还可以看出,每条染色体检测出来的 MITEs 所占各自染色体的比例都很小,与水稻相比还是有一定差别的^[9].

3.2 枣基因组 MITEs 的注释及分类

通过序列比对,共有 948 条在 P-MITE 中找到相似序列,碱基数为 323 856 bp,占基因组比例为 0.10%. 发现这些 MITEs 来源于 5 个 DNA 转座子家族:hAT, PIF/Harbinger, CACTA, Mutator 和 Tc1/Mariner;还有一部分 MITE 没有和任何 DNA 转座子家族匹配上,这些 MITE 可能是枣基因组特有的. 这 5 个家族的分布见表 2. 结果显示主要分布在 PIF, Mutator, hAT 家族.

3.3 枣基因组中 MITEs 系统发育

MITE-Digger 检测出确认的 MITEs 总共 3 230 条,其中共有 948 条在 P-MITE 中比对到相似序列,对注释上的 MITEs 利用 MEGA 6.0 构建枣树基因组 MITEs 的系统发育树,如图 1 所示. 从图中可以看出,此进化树分为 4 个类群 I, II, III, IV. 其中第 I 类群和 ZRSiTEMTO5900030 (来源于玉米基因组中的一个 MITE 序列) 在一个分枝上,类群 II 和 ORSgTEMTO0100126 (来源于水稻基因组中的一个 MITE 序列), ORSgTEMTO1601525 (来源于水稻基因组中的一个 MITE 序列), ZRSiTEMTOOT00002 (来源于玉米基因组中的一个 MITE 序列) 在一个分枝上,类群 III 和 ORSgTEMTO0502514 (来源于水稻基因组中的一个 MITE 序列), ZRSiTEMTO5600001 (来源于玉米基因组中的一个 MITE 序列) 在一个分枝上,证明它们和相应的水稻、玉米基因组内的 MITEs 具有一定的亲缘关系,可能来源于一个共同的祖先. 而类群 IV 没有水稻、玉米基因组 MITEs, 说明是枣树特有的类群.

表 2 MITEs 在 5 个家族中的分布

Tab.2 The distribution of MITEs in five families

MITEs 家族名称	MITE 条数
hAT	214
PIF/Harbinger	355
CACTA	18
Mutator	296
Tc1/Mariner	39
micro-like sequence	26

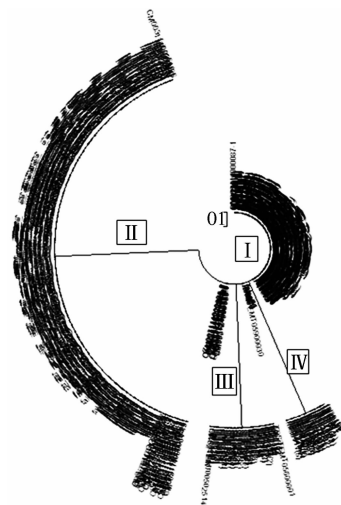


图 1 匹配上的 MITEs 的系统发育树(MP Tree)

Fig.1 Matched MITEs on the phylogenetic tree (MP Tree)

从图 1 还可以得出,这 4 个分支源于少数几个共同的祖先,经过长期进化中,祖先的基础上有明显的扩增和突变,说明枣树基因组 MITEs 的活动处于不稳定状态. 分支 II, III, IV 上各自的亲缘关系都很近,这说明这 3 个分支上的 MITEs 在发育过程中进行了一系列的扩增,从而引起了枣树基因组的多样性. 在这 3 个分支中,分支 II 上 MITEs 最多.

4 讨论

由表 1 知枣树中的 MITEs 总共 3 230 条,总碱基数为 1 104 679 bp,占基因组比例为 0.34%,其中有 948 条在 P-MITE 中比对到相似序列,碱基数为 323 856 bp,占基因组比例为 0.10%. 无论是检测出来的总的 MITEs,还是匹配上的 MITEs,在枣树基因组中所占的比例都很小. 6 种植物中 MITEs 所占的比例如表 3 所示^[10]. 从表 3 中可以看出水稻中的 MITEs 所占的比例是最大的,高粱中的 MITEs 所占的比例是最小的,枣树

基因组中的 MITEs 所占的比例相对这几种植物来说还不算太低,而且枣树中 MITEs 的全长为 323 856 bp,平均长度为 341.26 bp,仅次于高粱,所以应用 MITEs 转座子来研究枣树的基因组多样性是比较可行的.

表3 6种植物中的 MITEs

Tab.3 MITEs in six plants

植物	总碱基数/kb	MITEs 全长/bp	MITEs 平均长度/bp	所占比例/%
枣树	321 563	323 856	341.26	0.34
毛竹	2 034 000	920 000	926	0.045
玉米	2 580 000	26 600	320.9	0.001
水稻	363 000	14 590 877	198.6	4.02
高粱	697 000	3 200	406.2	0.000 5
苜蓿	215 000	13 800	234.6	0.006 4

由图1枣树基因组 MITEs 的系统发育树知,分支 II, III, IV 上各自的亲缘关系都很近,这说明这3个分支上的 MITEs 在发育过程中进行了一系列的扩增,从而引起了枣树基因组的多态性.因此可以利用这3个分支上的 MITEs 的保守序列分别设计引物,来进一步扩增枣树基因组内的 MITEs,以研究枣树基因组的结构和功能.同时还可以根据枣树基因组内 MITEs 的相对稳定性来开发枣树体内的 MITEs 的分子标记,以期研究枣树基因组内的 MITEs 所引起的基因多态性(IMP).应用 IMP 研究不同遗传背景的枣树基因组的多态性,也就是说可以开发枣树基因组的 IMP 引物.

人们已经开发了许多基于转座元件的分子标记技术.现在经常用的分子标记技术有 RAPD, AFLP, RFLP, ISSR, SSR 等^[11-12]. AFLP, RFLP 和其他技术相比操作相对复杂, RAPD 存在重复性和稳定性较差的问题,而 SSR 要求知道重复序列的两端的序列情况,并且费用比较高, ISSR 标记技术在 PCR 扩增时需要找到其最佳反应条件,不能区分显性纯合基因型和杂合基因型. IMP 分子标记是以 PCR 为基础的技术,有可重复性的结果,设计引物是根据 MITE 的 TIR 序列进行设计,用来检测两个相邻的 MITE 位点间的多态性,属于共显性的分子标记技术^[13].文献[14]研究显示 MITE 和它的侧翼所在的部位是基因组中多态性比较高的区域,而且富含大量的 AT 重复序列,因而可以根据这些特点开发一些新的高效多态遗传标记 IMP(inter-MITE polymorphism).而大量研究也表明 IMP 分子标记技术已有效用于小麦、大麦、燕麦、玉米、甘蔗等许多禾本科植物的遗传图谱的构建及遗传分析研究^[13,15-16].由表3知枣树基因组内的 MITEs 含量为 0.34%,玉米基因组内的 MITEs 含量仅仅为 0.001%,可以看出枣树体内的 MITEs 含量比玉米的要高,既然 IMP 分子标记在玉米基因中可行,那么 IMP 在枣树基因组中顺利开发和实施相对来说还是可行的.虽然目前 MITEs 研究的进展非常迅速,但已鉴定的 MITEs 仍然非常有限.本文对枣树基因组中大量的 MITEs 的分离和鉴定进行了研究,为从表观遗传方面来研究枣树在栽培和移植过程中出现的一些变化的分子机制奠定了一定的理论基础.

总而言之,转座元件是植物基因组中非常重要的组成部分,对于植物基因组的进化以及基因表达都起着非常重要的调节作用,所以 MITEs 对枣树基因的表达和基因组的进化调控等方面都有非常重要的意义.如果我们不了解转座子的功能,就不可能对植物基因组的结构、功能和进化有一个明确的概念^[17].因此对于转座因子的深入研究将了解枣树基因组的组成和进化秘密提供一把钥匙,也会促进枣树相关遗传育种理论与技术的进一步发展.

参考文献:

- [1] 温小杰,张学勇,郝晨阳,等. MITE 转座元件在植物中的研究进展[J]. 中国农业科学, 2008, 41(8): 2219-2226.
- [2] 李勇慧,押辉远,程彦伟,等. 水稻转座子对低能离子束辐照反应研究[J]. 郑州大学学报:理学版, 2015, 47(1): 107-111.
- [3] 时羽,韩康顺,孟令君,等. 稻/菰远缘杂交后代变异材料 MITE 类转座子 mPing 的激活[J]. 北方水稻, 2013, 6(11): 7-12.
- [4] Chen Jiongjiang, Lu Chen, Zhang Yu, et al. Miniature inverted-repeat transposable elements (MITEs) in rice were originated and amplified predominantly after the divergence of *Oryza and Brachypodium* and contributed considerable diversity to the species[J]. Mobile

- Genetic Elements, 2012,5(5):23-25.
- [5] 代红艳,张志宏. 植物转座元件及其分子标记的研究进展[J]. 农业生物技术学报,2006,3(12):434-439.
- [6] 刘铁,许金山,李田,等. 从金凤蝶分离的一株微孢子虫的全基因组 MITEs 转座子鉴定与系统进化分析[J]. 蚕业科学, 2014, 9(1): 52-58.
- [7] Yang Guojun. MITE Digger, an efficient and accurate algorithm for genome wide discovery of miniature inverted repeat transposable elements[J]. BMC Bioinformatics,2013, 14(1): 14-16.
- [8] Chen Jiongjiang, Hu Qun, Zhang Yu, et al. P-MITE: a database for plant miniature inverted repeat transposable elements[J]. Nucleic Acids Research,2013,16(1):42-45.
- [9] 卢辰. 微小反向重复转座子(MITE)对水稻基因表达和基因组分化的作用及番茄特有 miRNA 基因的演化[D]. 武汉:华中农业大学, 2013.
- [10] 钟浩,周明兵,白有煌,等. 毛竹 Stowaway-like MITEs 转座子的分离与分析[J]. 林业科学, 2010, 34(4): 37-42.
- [11] 高玉龙,桂毅杰,肖炳光,等. 烟草 MITE 位点间多态性(IMP)标记开发及其遗传作图应用[J]. 浙江大学学报:农业与生命科学版,2012, 21(6):655-661.
- [12] 谢永波,冯殿齐,赵进红,等. 基于 AFLP 技术的枣树主栽品种亲缘关系的研究[J]. 中国农学通报, 2014, 13(18):137-141.
- [13] Chang Y, O'donoughuels, Bureau T E. Inter-MITE polymorphisms (IMP): a high throughput transposon based genome mapping and fingerprinting approach[J]. Theor Appl Genet,2001,102(5): 773-781.
- [14] Komori T, Nitta N. High frequency of sequence polymorphism in rice MITEs and application to efficient development of PCR based markers[J]. Breeding Sci, 2003,53(2), 85-92.
- [15] Lyons M, Cardle L, Rostoks N, et al. Isolation, analysis and marker utility of novel miniature inverted repeat transposable elements from the barley genome[J]. Mol Genet Genom, 2008,280(4): 275-285.
- [16] Nakayama S. Inter-MITE polymorphisms of a newly identified MITE show relationships among sugarcane (*Saccharum*) species [J]. Genet Resour Crop Evol,2012, 59(7): 1389-1396.
- [17] Bennetezn J L. Transposable element contributions to plant gene and genome evolution[J]. Plant Molecular Biology, 2000, 42(1),251-269.

Analysis and System Evolution of MITEs in the Genome of *Ziziphus Jujuba Mill*

LI Yonghui, FENG Aiqing, YA Huiyuan, CHENG Yanwei, LU Haipeng
(Life Science College, Luoyang Normal University, Luoyang 471022, China)

Abstract: The evolution of the MITEs might be the important cause of Chinese jujube genome and species diversity. In order to explore the distribution, types and evolution of MITEs in jujube genome, MITE prediction software (MITE-Digger) was used to identify the MITEs sequence of jujube genome, annotated in plant MITEs classified database. The phylogenetic tree of MITEs in jujube genome was built up with MEGA 6.0. The results indicated that the MITE-Digger detected 3 230 pieces of MITE, with 1 104 679 bp bases which account for 0.34% of the whole genome. The similar sequence compared had 948 pieces of MITEs in the plant MITEs database (P-MITE). These MITEs could be categorized into five families: hAT, PIF/Harbinger, CACTA, Mutator and Tc1/Mariner. MITEs phylogenetic tree showed that MITEs of jujube genome came from a few common ancestor, and had no obvious amplification and mutation in the long-term evolution, which might be related to the diversity of jujube varieties. The findings could provide important reference for genetic resources and breeding of *Ziziphus Jujuba*.

Key words: MITEs; jujuba genome; system evolution; IMP

(责任编辑:王浩毅)