

郭向萌, 张亚平, 付鹏程, 等. 枣树基因组中转座元件分析[J]. 江苏农业学报, 2017, 33(2): 412-416.  
doi:10.3969/j.issn.1000-4440.2017.02.027

## 枣树基因组中转座元件分析

郭向萌, 张亚平, 付鹏程, 押辉远  
(洛阳师范学院生命科学学院, 河南 洛阳 471022)

**摘要:** 转座子是基因组演化的重要因素。为了探明转座子在枣树(*Ziziphus jujuba* Mill.) 基因组演化中可能的作用, 采用生物信息学工具对枣树基因组中的转座元件进行分析。结果表明, 所有转座元件占枣树基因组的 23.38%, 其中 DNA 转座子占 8.60% (包括 0.34% 的 MITEs)、LTR 反转录转座子占 12.23%、Non-LTR 反转录转座子占 2.30%、内源性反转录病毒占 0.25%。系统进化树分析结果表明, MITEs 和 LTR 反转录转座子在演化过程中进行过复制和扩增, 这说明转座子的活动和枣树基因组的进化是相关的。本研究将有助于研究枣树基因组的遗传多样性及进行相关种质资源的鉴定。

**关键词:** 枣树(*Ziziphus jujuba* Mill.); 基因组; 转座元件

**中图分类号:** S665.1      **文献标识码:** A      **文章编号:** 1000-4440(2017)02-0412-05

## Analysis of transposable elements in jujube genome

GUO Xiang-meng, ZHANG Ya-ping, FU Peng-cheng, YA Hui-yuan  
(College of Life Science in Luoyang Normal University, Luoyang 471022, China)

**Abstract:** Transposable elements (TEs) play important roles in plant genome evolution. Bioinformatics tools were used to analyze the TEs in jujube genome in this study. The TEs accounted for 23.38% of jujube genome, including 8.60% DNA transposons, 12.23% LTR-retrotransposons, 2.30% Non-LTR retrotransposons and 0.25% endogenous retrovirus. The phylogenetic analysis revealed that these TEs had been amplified previously, suggesting that transposable elements activities are relevant to the evolution of jujube genome.

**Key words:** jujube(*Ziziphus jujuba* Mill.); genome; transposable element

枣树(*Ziziphus jujuba* Mill.) 为重要的药用植物和生态经济林树种, 果实干鲜兼用。在长期驯化过程中, 由于分离、突变、迁移、自然选择、人工选择等作用, 枣树群体产生了大量的遗传变异, 形成区域性品种和品种群<sup>[1]</sup>。枣树多为无性繁殖, 栽培品种繁

多, 导致同物异名、同名异物的现象比较严重<sup>[2]</sup>。因此, 为了探明中国枣品种群的形成和发展, 指导资源利用和育种实践, 枣树种质资源的遗传多样性、进化和枣品种群遗传结构研究成为枣研究的焦点<sup>[3]</sup>。枣树的遗传多样性分析多是基于 AFLP、SSR、ISSR、SRAP 等分子标记技术<sup>[4-7]</sup>, 但由于枣树长期采用扦插方式繁殖, 遗传背景较为狭窄, 这些分子标记技术的多态性少, 在研究枣种质资源的演化和遗传方面并不理想。

近年来, 基于转座元件(Transposable elements TEs)的分子标记越来越受到遗传多样性研究者的

收稿日期: 2016-11-25

基金项目: 国家青年科学基金项目(30800204); 河南省科技攻关项目(152102310346)

作者简介: 郭向萌(1980-), 男, 河南洛阳人, 硕士, 讲师, 主要从事枣树功能基因组研究。(Tel) 13837984812; (E-mail) 23511243@qq.com

青睐。转座元件可分为 DNA 转座子和反转录转座子。DNA 转座子是植物基因组中最早发现的转座元件,分为自主和非自主两类,自主型转座子具有转座酶,可以独立转座<sup>[8]</sup>。如果转座酶突变就变成非自主型 DNA 转座子。在植物基因组中还存在结构与 DNA 转座子类似,但却高度重复的 MITEs (微型反向重复转座元件)<sup>[9]</sup>,它可以在同源自主 DNA 转座子存在的情况下发生转座。反转录转座子主要包括 LTR 反转录转座子<sup>[10]</sup> (LTR retrotransposon)、Non-LTR 反转录转座子 (Non-LTR retrotransposon) 和内源性反转录病毒元件 (Endogenous retrovirus)。基于 MITEs (Miniature Inverted Repeat Transposable Element, 微型反向重复转座元件) 的 IMP (Inter MITE Polymorphisms)、基于反转录转座子的 IRAP (Inverse Retrotransposon Amplified Polymorphisms), 这些分子标记扩增条带多、反应稳定、重复性好,在基因组中分布广泛,适合植物遗传标记分析<sup>[11-13]</sup>。但是基于转座子分子标记的枣树遗传多样性研究还鲜见报道。2014 年,中国科研工作者完成了枣基因组的测序工作<sup>[14]</sup>,这为枣树种质资源遗传多样性研究提供了便利,为基于转座子的分子标记的使用提供可能,可以利用生物信息学的方法鉴定枣基因组转座子的成分,包括其类别、数量、序列等信息。本研究利用生物信息学工具,对枣基因组中存在的转座元件加以鉴定和识别,以期进一步了解枣基因组的构成、枣遗传多样性和进化。

## 1 材料与方法

### 1.1 枣基因组序列的获取

枣树的全基因组有 12 条染色体,从 NCBI (<http://www.ncbi.nlm.nih.gov/>) 下载枣树的全基因组 (编号: 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。

### 1.2 基于在线分析软件 Censor 分析枣树基因组中的转座元件

利用 Censor 软件网络版 (<http://www.girinst.org/censor/index.php>) 将枣基因组 12 条染色体序列分别输入,依照软件默认参数分析枣树的全基因组中的已有成分。Censor 不是从头分析,而是与重复

序列库中重复序列进行匹配,因此,枣树基因组中新的转座成分不能识别。

### 1.3 基于 LTR-finder 在线分析枣基因组中 LTR-反转录转座子

LTR-FINDER ([tlife.fudan.edu.cn/ltr\\_finder/](http://tlife.fudan.edu.cn/ltr_finder/)) 的参数按照默认值进行设置<sup>[15]</sup>。运用 giri (<http://www.girinst.org/censor/index.php>) 将枣树基因组 12 条染色体上的 LTR 反转录座子 (copia 和 gypsy) 进行分类<sup>[16]</sup>。

### 1.4 基于 MITE-Digger 鉴定枣树的全基因组 MITEs

MITE-Digger 的软件参数: MITE 全长 100~800 bp; TSD (Target site duplication, 靶位点重复) 碱基数范围为 >2; 侧翼序列不同碱基数的阈值为 5 bp<sup>[17]</sup>。

### 1.5 构建系统发育树

采用 MEGA 6.0 软件构建系统发育树,将碱基序列以 Fasta 格式输入 MEGA 6.0<sup>[18]</sup>,以最大简约法构建系统发育树,使用软件默认参数。

## 2 结果与分析

对枣树基因组中的转座元件进行分析,主要分为 DNA 转座子、MITEs 和反转录元件。结果表明,所有转座元件占枣树基因组的 23.38%,其中 DNA 转座子占 8.60% (包括 0.34% 的 MITEs)、LTR 反转录转座子占 12.23%、Non-LTR 反转录转座子占 2.30%、内源性反转录病毒占 0.25%。

### 2.1 枣树基因组中的 DNA 转座子和 MITEs

在 Censor 中,把 MITEs 归为 DNA 转座子,但越来越多的研究把 MITEs 单独分类,并开发了相应的生物信息学识别工具。Censor 分析结果表明,DNA 转座子在枣树基因组中约占 9% (表 1),其中第 11 号染色体所占比例最大,第 1 号染色体的数量最多。有些片段碱基数较小,有些片段较大。有的距离很近的短片段可能是 1 个 DNA 转座子的不同区域。因此,试验结果中的片段数应该比实际的片段数多,而且每个长度的平均碱基数在 115 bp 左右。这与植物基因组中 MITEs 的长度相似,有必要把这其中的 MITEs 单独分析,但是 Censor 软件不能区分 MITEs,因此利用生物信息学工具 MITE-digger 分析枣树基因组中 MITEs 的成分,结果如表 2 所示。MITEs 占枣基因组的 0.34%,每 10 kb 出现 1 次。每个 MITE 的平均长度为 342.0 bp。

表 1 Censor 识别鉴定枣基因组中 DNA 转座子  
Table 1 Censor identification of jujube genome DNA transposons

染色体		DNA 转座子			
编号	碱基数 (bp)	片段数	总碱基数 (bp)	平均长度 (bp)	占染色体碱基数的比例
1	42 390 363	30 336	3 458 319	114.0	0.081 6
2	27 986 743	20 614	2 378 913	115.4	0.085 0
3	26 737 297	20 408	2 333 521	114.3	0.087 3
4	30 445 767	23 039	2 661 433	115.5	0.087 4
5	31 365 312	23 446	2 834 866	120.9	0.090 4
6	25 259 912	18 692	2 148 736	115.0	0.085 1
7	27 644 224	21 271	2 485 809	116.9	0.089 9
8	23 350 829	17 221	1 955 698	113.6	0.083 8
9	25 348 792	19 015	2 155 592	113.4	0.085 0
10	20 983 256	15 897	1 834 908	115.4	0.087 4
11	20 703 932	15 853	1 889 702	119.2	0.091 3
12	19 346 091	14 445	1 575 660	109.1	0.081 4

表 2 各条染色体检测出的 MITEs  
Table 2 MITEs in each chromosome

染色体编号	条数	总碱基数 (bp)	频数	平均长度 (bp)	占染色体碱基数比例
1	424	144 178	99 977.3	340.0	0.003 4
2	304	106 881	92 061.7	351.6	0.003 8
3	276	89 482	96 874.3	324.2	0.003 3
4	35	17 171	869 879.1	490.6	0.000 6
5	309	111 210	101 505.9	359.9	0.003 5
6	284	100 994	88 943.4	355.6	0.004 0
7	407	137 957	67 921.9	339.0	0.005 0
8	210	66 660	111 194.4	317.4	0.002 9
9	302	99 504	83 936.4	329.5	0.003 9
10	234	76 199	89 672.0	325.6	0.003 6
11	179	56 794	115 664.4	317.3	0.002 7
12	266	97 649	72 729.7	367.1	0.005 0

为了解 MITEs 在枣基因组中的功能,采用系统进化软件 MEGA6.0 对 MITEs 进行系统发育分析。结果表明,枣树基因组中的 MITEs 是由共同祖先经过扩增、突变而来。特别是类群 II 的重复次数高达上千次,这种 DNA 片段的扩增和突变可能与枣树的多态性有密切关系。

2.2 枣树基因组中反转录元件

在植物基因组中,反转录元件重复序列主要包

括 LTR 反转录转座子 (LTR retrotransposon)、Non-LTR 反转录转座子 (Non-LTR retrotransposon) 和内源性反转录病毒元件 (Endogenous retrovirus)。Censor 分析结果 (表 3) 表明:枣树基因组中 LTR 反转录转座子的含量达到了 12.23%,而基于 LTR-finder 的分析结果显示 LTR 反转录转座子只占枣树基因组的 3.2%,这是因为 LTR 反转录转座子在基因组的演化过程中会发生缺失和突变,甚至丢掉其特有的

特征结构变成短的重复片段,这样 LTR-finder 就无法识别。用 MEGA 构建系统进化树,结果显示,部分 LTR 反转录转座子在长期进化中有明显的数量扩增和突变。

表 3 枣树基因组中 LTR 反转录转座子  
Table 3 LTR retro transposons in jujube genome

染色体 编号	片段数	总碱基数 (bp)	平均长度 (bp)	占染色体碱 基数的比例
1	23 784	5 001 417	210.3	0.118 0
2	16 000	3 411 138	213.2	0.121 9
3	16 062	3 273 662	203.8	0.122 4
4	18 539	3 939 767	212.5	0.129 4
5	19 631	4 043 196	205.0	0.128 9
6	14 564	3 177 525	218.2	0.125 8
7	17 127	3 553 936	207.5	0.128 6
8	13 265	2 686 877	202.6	0.115 1
9	14 627	2 901 943	198.4	0.114 5
10	12 338	2 562 189	207.7	0.122 1
11	12 211	2 570 270	210.5	0.124 1
12	11 645	2 256 764	193.8	0.116 7

而基于 Censor 枣基因组中 Non-LTR 反转录转座子的分析结果(表 4)表明,枣基因组中的 Non-LTR 反转录转座子只占0.022 5(2.25%),远远小于 LTR-反转录转座子成分。

表 4 枣树基因组中 Non-LTR 反转录转座子  
Table 4 Non-LTR retro transposons in jujube genome

染色体 总碱基数	片段数	总碱基数 (bp)	平均长度 (bp)	占染色体碱 基数的比例
1	10 846	918 507	84.7	0.021 7
2	7 578	639 741	84.5	0.022 9
3	7 138	608 988	85.4	0.022 8
4	8 104	687 964	84.9	0.022 6
5	8 131	687 449	84.6	0.021 9
6	6 736	571 340	84.8	0.022 6
7	7 119	607 289	85.3	0.022 0
8	6 203	518 992	83.7	0.022 2
9	6 909	581 918	84.2	0.023 0
10	5 623	481 461	85.6	0.023 0
11	5 470	465 277	85.1	0.022 5
12	5 341	441 198	82.6	0.022 8

基于 Censor 的分析发现,枣树基因组中还有内源性反转录病毒,这是一类对基因表达有调控作用的重复序列,一般在动物中含量较高<sup>[19]</sup>。在人类基因组中可高达 8.00%,而枣树基因组中有 0.25%的碱基与已发现的内源性反转录病毒具有较高的同源性,说明枣基因组中有这种成分的存在(表 5)。

表 5 枣树基因组中内源性反转录病毒元件  
Table 5 Endogenous retrovirus in jujube genome

染色体 编号	片段数	碱基数 (bp)	平均长度 (bp)	所染色体碱 基数的比例
1	1 481	104 292	70.4	0.002 5
2	1 066	76 032	71.3	0.002 7
3	919	66 462	72.3	0.002 5
4	1 082	79 221	73.2	0.002 6
5	1 060	77 491	73.1	0.002 5
6	862	62 940	73.0	0.002 5
7	929	66 426	71.5	0.002 4
8	846	58 913	69.6	0.002 5
9	871	63 087	72.4	0.002 5
10	735	53 025	72.1	0.002 5
11	752	52 142	69.3	0.002 5
12	726	53 139	73.2	0.002 8

3 讨论

枣树品种繁多,中国有 700 多个枣树品种。鉴定并了解这些资源的遗传结构对枣生产和育种具有重要意义。随着分子生物学技术的发展,分子标记技术为种质资源的研究提供了有力工具。枣树多样性的遗传基础是基因组的多样性,而基因组的多样性来自于基因组在长期进化中积累的各种突变、重组、重排、复制以及缺失<sup>[20]</sup>。这些现象和基因组中的转座子密切相关。本研究结果表明,枣树基因组中的转座成分达到了 23.38%,这相对于拟南芥的 10%来说是较高的,但与水稻(35%)<sup>[21]</sup>、玉米(60%)、小麦(90%)<sup>[22]</sup>、松树(90%)<sup>[23]</sup>相比,含量又明显偏低。

由于 Censor 是基于已知转座子序列的相似性来鉴别的,对于未发现过的转座子是不能识别鉴定的。因此枣树基因组中的转座子的比例可能要高于 23.4%。转座子的活动可以造成染色体断裂及重排、插入突变和特定的序列扩增等结构变异,且常和

基因活性变化相联系。

本研究结果表明, MITEs 和反转录转座子在长期进化过程中, 有明显的扩增现象, 这必将对枣树基因组的结构稳定性有重要影响, 并可能是枣基因组向多个方向歧化的重要因素, 因此从转座成分入手, 研究枣树的种质资源显得尤为重要。

枣树基因组碱基序列的公布, 为枣树的分子生物学研究开辟了新时代, 也为研究枣基因组中转座子提供了重要素材和资源。在本研究中, 获得了枣树基因组中可识别的转座子的 DNA 序列, 这些序列通过生物信息学工具的分析, 可以获得枣树基因组中转座元件的起源和演化。同时依据这些 DNA 序列的同源性分析可以设计基于转座子分子标记的引物, 比如: IMP、IRAP 等。

#### 参考文献:

- [1] 刘孟军, 王玖瑞, 刘平, 等. 中国枣生产与科研成就及前沿进展[J]. 园艺学报, 2015, 42(9): 1683-1698.
- [2] 侯思宇, 孙朝霞, 申洁, 等. 30 个枣树种质资源遗传多样性的 ISSR 分析[J]. 植物生理学报, 2011, 47(3): 275-280.
- [3] 孙海艳, 党江波, 何波, 等. 武隆猪腰枣遗传多样性的 AFLP 分析[J]. 西南大学学报(自然科学版), 2014, 36(12): 30-35.
- [4] 乔勇, 赵锦, 杨海旭, 等. 21 个枣品种(系)的 AFLP 指纹分析[J]. 植物遗传资源学报, 2009, 10(2): 205-210.
- [5] 麻丽颖, 孔德仓, 刘华波, 等. 36 份枣品种 SSR 指纹图谱的构建[J]. 园艺学报, 2012, 39(4): 647-654.
- [6] 齐靖, 董祯, 申连英, 等. 枣 ISSR 扩增体系的建立[J]. 华北农学报, 2008, 23(增刊): 209-212.
- [7] 李莉, 彭建营, 白瑞霞. 中国枣属植物亲缘关系的 SRAP 分析[J]. 中国农业科学, 2009, 42(5): 1713-1719.
- [8] 刘琳, 张美丽, 黄粤. DNA 转座子在小鼠基因功能研究中的应用[J]. 遗传, 2011, 33(5): 485-493.
- [9] 朱克明, 杨艳华. 植物 MITEs 转座子及其应用[J]. 湖北农业科学, 2015, 54(22): 5497-5500.
- [10] 蒋爽, 滕元文, 宗宇, 等. 植物 LTR 反转录转座子的研究进展[J]. 西北植物学报, 2013, 33(11): 2354-2360.
- [11] 王利英, 杜永臣, 张斌, 等. 茄子 IRAP 和 REMAP 分子标记的开发[J]. 园艺学报, 2008, 35(9): 1363-1367.
- [12] 李芳, 徐良, 魏美甜, 等. 萝卜 IRAP 技术体系建立与品种指纹图谱构建[J]. 江苏农业学报, 2015, 31(1): 143-148.
- [13] 李静婷, 王健胜, 杨风岭. 小麦 MITE 转座子对 *sHSP* 基因的表达调控研究[J]. 江苏农业科学, 2015, 43(12): 29-32.
- [14] LIU M J, ZHAO J, CAI Q L, et al. The complex jujube genome provides insights into fruit tree biology[J]. Nat Commun, 2014, 10: 5315.
- [15] 侯小改, 张曦, 郭大龙. 植物 LTR 类反转录转座子序列分析识别方法[J]. 遗传, 2012, 34(11): 1491-1500.
- [16] 李卫涛, 张焕丽, 押辉远. 基于 LTR-FINDER 分析葡萄糖基因组中 LTR 反转录转座子[J]. 湖北农业科学, 2013, 52(8): 4-7.
- [17] 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.
- [18] 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, 46(4): 37-42.
- [19] SPENCER T E, BLACK S G, ARNAUD F, et al. Endogenous retroviruses of sheep: a model system for understanding physiological adaptation to an evolving ruminant genome[J]. J Reprod Dev, 2010, 67: 95-104.
- [20] 吴志俊, 金玮. 拷贝数变异: 基因组多样性的新形式[J]. 遗传, 2009, 31(4): 339-347.
- [21] SAKAI H, TANAKA T, ITOH T. Birth and death of genes promoted by transposable elements in *Oryza sativa*[J]. Gene, 2007, 392: 59-63.
- [22] 马庆华, 续九如. 枣树分子标记研究进展[J]. 分子植物育种, 2006, 6(4): 133-136.
- [23] 田海霞. 转座子在基因组和基因进化方面的研究进展[J]. 安徽农业科学, 2011, 39(20): 12018-12020.

(责任编辑: 陈海霞)