

毛竹长末端重复序列反转录转座子的全基因组特征及进化分析

陈娅欣, 周明兵

(浙江农林大学 省部共建亚热带森林培育国家重点实验室, 浙江 杭州 311300)

摘要: 【目的】研究毛竹 *Phyllostachys edulis* 基因组中的长末端重复序列反转录转座子 (long terminal repeat retrotransposons, LTR-REs) 的特征, 为今后利用 LTR 反转录转座子对毛竹基因组的功能和对竹种资源遗传多样性的研究奠定基础。【方法】通过生物信息学方法, 利用 LTRharvest 和 RepeatMakser 软件对第 2 版毛竹基因组中的 LTR 反转录转座子进行全面注释与分类, 并对得到的 LTR 反转录转座子的分布特征、进化特性和插入时间进行分析。【结果】在毛竹基因组中共注释得到 1 014 565 个 LTR 反转录转座子, 1 562 个家族, 占毛竹基因组的 54.97%。其中 solo LTR 反转录转座子与完整 LTR 反转录转座子 (S/F) 的比例较高 (约 1.77:1.00), 表明在毛竹 LTR 反转录转座子中可能发生了相对较高频率的非法重组和不平衡重组。毛竹 LTR 反转录转座子分为 Ty1-copia 和 Ty3-gypsy 超家族, Tork、Reftrofit、Sire、Oryco、Del、Reina、Crm、Tat、Galadriel、Athila 等 10 个谱系。毛竹 LTR 反转录转座子的 Ty1-copia 和 Ty3-gypsy 超家族对 PBS 位点的偏好性呈相反趋势, 较长的 LTR 反转录转座子具有更长的 LTR 序列, 结构也更加完整。毛竹 LTR 反转录转座子的插入时间主要集中在 0~2.0 Ma, 且还处于不断缓慢增长的状态。【结论】第 2 版毛竹基因组的高质量组装, 能更好地注释和分析毛竹基因组中的 LTR 反转录转座子。基于结构预测的 LTRharvest 法, 能更精准地预测毛竹 LTR 反转录转座子。不同谱系的毛竹 LTR 反转录转座子在进化过程中具有不同的分化和扩增活性。毛竹 LTR 反转录转座子总体上处于不断扩增状态, 这是导致毛竹基因组较大的主要原因之一。图 3 表 3 参 52

关键词: LTR 反转录转座子; 毛竹; 基因组; 进化

中图分类号: Q753; S795.7 文献标志码: A 文章编号: 2095-0756(2021)03-0455-09

Genome-wide characteristics and evolution analysis of long terminal repeat retrotransposons in *Phyllostachys edulis*

CHEN Yaxin, ZHOU Mingbing

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

Abstract: [Objective] This study aims to analyze the characteristics of long terminal repeat retrotransposons (LTR-REs) in moso bamboo genome, so as to promote the research on the function of LTR-REs in moso bamboo genome and the genetic diversity of bamboo resources. [Method] Based on bioinformatics methods, LTR retrotransposons in the second edition of moso bamboo genome were annotated and classified by LTRharvest and RepeatMakser software, and the distribution characteristics, evolution characteristics and insertion time of the obtained LTR retrotransposons were analyzed. [Result] A total of 1 014 565 LTR retrotransposons and 1 562 families were identified, accounting for 54.97% of moso bamboo genome. Among them, the ratio of solo LTR retrotransposons to intact LTR retrotransposons (S/F) was relatively high (about

收稿日期: 2020-07-15; 修回日期: 2021-01-04

基金项目: 国家自然科学基金资助项目 (31870656, 31470615); 浙江自然科学基金资助项目 (LZ19C160001)

作者简介: 陈娅欣 (ORCID: 0000-0003-3889-2502), 从事毛竹 LTR 反转录转座子在非生物胁迫下的研究。E-mail: 2849124318@qq.com。通信作者: 周明兵 (ORCID: 0000-0001-5674-4410), 教授, 博士, 从事竹子生长发育表观遗传学和竹子快速生长机制研究。E-mail: zhoumingbing@zafu.edu.cn

1.77:1.00), indicating that a higher frequency of illegitimate recombination and unbalanced recombination might have occurred in the LTR-REs of moso bamboo genome. LTR retrotransposons were divided into Ty1-copia and Ty3-gypsy superfamilies, and ten lineages included Tork, Refrofit, Sire, Oryco, Del, Reina, Crm, Tat, Galadriel, and Athila. The preference of Ty1-copia and Ty3-gypsy superfamilies for PBS sites showed an opposite tendency. The longer LTR retrotransposons had longer LTR sequences and more complete structures. The insertion time of LTR retrotransposon in moso bamboo was mainly concentrated in the 0–2.0 Ma region, and it was still in a state of slow growth. [Conclusion] The high-quality assembly of the second edition of moso bamboo genome can better annotate and analyze the LTR retrotransposons in moso bamboo genome. The LTR harvest method based on structure prediction can more accurately predict the LTR retrotransposons of moso bamboo. The LTR retrotransposons of different lineages have different differentiation and amplification activities during evolution. LTR retrotransposons are generally in a state of continuous amplification, which is one of the reasons for the large genome of moso bamboo. [Ch, 3 fig. 3 tab. 52 ref.]

Key words: LTR retrotransposons; *Phyllostachys edulis*; genome; evolution

转座子 (transposable elements, TEs) 是一种自私的基因组“寄生虫”，能够增加拷贝数并改变其在宿主基因组中的位置^[1]。转座子由于具有突变的可能，会对邻近基因表达产生潜在的危害，并导致染色体重排，对基因组的稳定性构成威胁^[2-4]。转座子根据其转座中间产物被分为 2 类，其中 I 类转座子包括复制黏贴式反转录转座子 (retrotransposons, REs)^[5]，II 类转座子包括剪切黏贴式转座子^[6]。I 类转座子根据其内部的结构可以分为 LTR 类 (long terminal repeat retrotransposons)^[6]、DIRS 类 (dictyostelium intermediate repeat sequence elements)^[7]、PLE 类 (penelope-like elements)^[8]、LINE 类 (long interspersed nuclear elements)^[9]、SINE 类 (short interspersed nuclear elements)^[10]。其中 LTR 反转录转座子是迄今为止研究最多的一类^[6]。LTR 反转录转座子具有 4 个结构特点。第一，在序列两端有 1 对靶位点重复序列 (target site repeats, TSD)，约 4~6 bp^[11]；第二，5'端和 3'端有 1 对长为几十到几千 bp 不等高度相似的长末端重复序列^[12]；第三，LTR 主要包括 GAG 衣壳蛋白编码区和 POL 多蛋白编码区 (polyprotein)，其中 POL 包括 RH 核糖核苷酸酶 (ribonuclease h, RNaseH)、RT 反转录酶 (reverse transcriptase)、INT 整合酶 (integrase) 和 AP 蛋白酶 (aspartic proteinase)。还有一些 Retrovirus 和 ENV 超家族 (superfamily) 的 LTR 含有 ENV 序列 (envelope protein, EN 或 ENV)^[13]；第四，在 5'端附近有 1 个引物结合位点 (primer binding site, PBS)，可调控其基因组 RNA 反转录所必需的 tRNA 引物，它以染色体外线性 DNA (extrachromosomal linear DNA, ecDNA) 的形式产生 LTR 反转录转座子生命周期中间体。在 3'端附近有 1 个富嘌呤位点 (poly purine trait, PPT)^[14]，协助反转录的完成。根据 LTR 反转录转座子开放阅读框 (open reading frames, ORF) 的完整性分为自主 LTR 反转录转座子和非自主 LTR 反转录转座子^[15]。自主 LTR 反转录转座子又可以根据 POL 中 RT、INT 和 RH 编码序列的排列方式，分为 Ty1-copia 超家族 (5'-INT-RT-RH-3') 和 Ty3-gypsy 超家族 (5'-RH-RH-INT-3')^[16]。根据每个超家族的序列同源性 (80-80-80 的分类规则^[17]) 可以分为不同的家族。根据 LTR 反转录转座子的同源性、蛋白结构、进化关系可以划分为不同的谱系，如梨 *Pyrus* 基因组中被划分为 Ale、Ivana、Bianca、Angela、Tar、Tat、Athila、Renia、Crm、Galadriel、Tekay 等 11 个谱系^[18]。LTR 反转录转座子的活性包括转录活性和转座活性。转录是转座的第 1 步，许多 LTR 反转录转座子在植物杂交、多倍体化或在环境挑战下发生去甲基化，被转录激活^[19]。转座活性不仅包括转录活性，还受转录后调控，如植物基因组为了抑制转座子的活性通过转录基因沉默 (transcriptional gene silencing, TGS) 机制抑制它们的能力^[20]。如果 TGS 得到缓解，则受 21~22 个核苷酸作用的转录后基因沉默机制 (post-transcriptional gene silencing, PTGS) 会将靶向转座子转录物进行降解^[21]，所以在实验条件下 LTR 反转录转座子很难被转座激活。在整个植物王国中 LTR 反转录转座子的进化特别成功，不断复制转座，导致基因组大小增加，基因组尺寸产生差异。在被子植物基因组之间由于几个 LTR 反转录转座子家族的扩增，产生一些巨大的基因组。例如 2 400 Mb 玉米 *Zea mays*^[22] 和 400 Mb 水稻 *Oryza sativa*^[23] 的基因组中 LTR 反转录转座子家族数相同，但是玉米基因组中 5 个谱系的 LTR 反转录转座子拷贝数较高。即使是亲缘关系很近的品种，LTR 反转录转座子也会促使它们的基因结构产生巨

大差异。如玉米与大刍草 *Zea mexicana* 是近亲, 但是大刍草的基因组比玉米大 1 倍^[24]。LTR 反转录转座子在植物基因组中处于动态变化的过程, 不仅会扩增, 也会丢失, 不平衡重组 (illegitimate recombination) 和非法重组 (unequal recombination) 活动就是丢失的主要原因^[25-26]。不平衡重组和非法重组的产物主要包括含有 TSD 位点的 solo LTR, 不含 TSD 位点的 Truncated LTR^[27]。预测 LTR 反转录转座子可以通过 4 种方式^[28]: 比较基因组法 (comparative genomic methods)^[29]、重复序列从头算法 (de novo repeat discovery)^[30]、同源比对法 (homology-based methods)^[31]、基于结构预测法 (structure-based methods)。基于结构预测法是通过 LTR 反转录转座子的序列结构和转座机制分析来捕获, 如 LTR_STRUC、LTR_FINDER、LTRharvest、LTR_par、LTR_Rho 等^[32]。毛竹 *Phyllostachys edulis* 具有较高的经济和生态价值, 其种植面积在中国的竹子总种植面积 (443 万 hm^2) 中占 73.76%^[33]。2018 年第 2 版毛竹基因组的公布^[34] 为深入分析毛竹基因组中 LTR 反转录转座子提供了良好的条件。本研究运用了 LTRharvest 的方法^[35], 对第 2 版毛竹基因组中的 LTR 反转录转座子进行预测, 并对 LTR 反转录转座子的结构、在基因组中的分布特征、插入时间等进行系统分析, 以期能了解毛竹 LTR 反转录转座子对基因组的影响。

1 材料与方法

1.1 毛竹 LTR 反转录转座子的鉴定与分类

第 2 版毛竹基因组序列、基因组注释文件来自毛竹基因组数据库 (<http://bamboo.bamboogdb.org/#/download>)^[34]。先利用 LTRharvest 软件^[35] 预测毛竹基因组 LTR 反转录转座子, 利用 cd-hit 根据 80-80-80 的分类规则^[17] 进行序列同源性聚类, 然后利用 LTRdigestion^[36] 注释各个 LTR 反转录转座子结构域。接着根据 LTR 反转录转座子的 POL 编码区 RT、INT 和 RH 的顺序将其分为 Ty1-Copia(INT-RT-RH) 和 Ty3-Gypsy (RT-RH-INT) 2 个超家族^[16]。根据序列中的 GAG 和 POL 编码区与 Gypsy Database 2.0 网站中 (http://gydb.org/index.php/Phylogeny:POL_LTR_retroelements) 植物典型的 Tork、Reftrofit、Sire、Oryco、Del、Reina、Crm、Tat、Galadriel 和 Athila 谱系对应编码区的同源性, 进一步将 2 个超家族分为 10 个谱系^[35]。最后利用 RepeatMakser 软件分析 LTR 反转录转座子在毛竹基因组中的含量^[31]。

1.2 毛竹 LTR 反转录转座子插入时间计算

使用 MA 等^[37] 对每条结构完整的毛竹 LTR 反转录转座子的插入时间进行计算: ①对每条结构完整的 LTR 反转录转座子的两端 LTR 序列利用 MUSCLE 软件^[38](使用默认参数) 进行比对; ②利用 JUCKES-CANTOR 的方法^[39] 计算碱基突变频率 (K); ③利用公式 $T=K/(2r)$ 计算插入时间, T 表示时间, r 表示生物钟, $r=1.3 \times 10^{-8} \text{ bp} \cdot \text{a}^{-1}$ ^[37]。

2 结果与分析

2.1 毛竹中 LTR 反转录转座子的鉴定和分类

由表 1 所示: 得到 1 014 565 条 LTR 反转录转座子, 占整个毛竹基因组的 54.97%。毛竹中 LTR 反转录转座子比例与其他基因组相比, 低于玉米基因组的 70.1%^[22], 相近于高粱 *Sorghum bicolor* 基因组的 55%^[40], 远高于水稻基因组的 26%^[23]。其中两端具有完整 LTR 序列, 编码结构域完整的 LTR 反转录转座子 (full-length LTR) 有 7 731 条, 两端具有完整 LTR 序列, 编码结构域不完整的 LTR 反转录转座子 (solo LTR) 有 13 656 条 (其余不含 TSD 位点的 LTR 反转录转座子忽略不计)。然后按照 WICKER 等^[16] 提出的真核生物转座子的分类方法, 将 blastn(all-vs-all) 的方法和 80-80-80 的规则相结合, 对 7 731 条完整的 LTR 反转录转座子进行分类, 共分为 1 562 个家族。

毛竹 LTR 反转录转座子分为 Ty1-copia 和 Ty3-gypsy 2 个超家族, 在 1 562 个 LTR 反转录转座子家族中有 819 个家族属于 Ty1-Copia 超家族, 共包括 433 137 条序列, 长度为 400 788 135 bp, 占毛竹基因组的 21.01%。743 个家族属于 Ty3-Gypsy 超家族, 共包括 581 429 条序列, 长度为 647 905 081 bp, 占毛竹基因组的 33.96%(表 2)。Ty3-gypsy 与 Ty1-copia 数量之比为 1.3:1.0, 低于大豆 *Glycine max*(1.4:1.0)^[19] 和玉米 (1.6:1.0)^[22], 远低于水稻 (4.9:1.0)^[41] 和高粱 (3.7:1.0)^[40], 但远高于苜蓿 *Medicago sativa*(0.3:1.0)^[42]。

根据 LTR 反转录转座子不同家族之间的进化关系和结构特征, Ty1-copia 超家族和 Ty3-gypsy 超家族可以被分为多个不同的谱系^[42-43]。根据 Gypsy Database2.0^[37] 中植物典型的谱系序列特征, 对毛竹

表1 毛竹 LTR 反转录转座子超家族分类

Table 1 Classification of LTR retrotransposons superfamily of moso bamboo genome

超家族	谱系	家族 ^a	结构	数量/个	全长/bp	百分比 ^b /%
Ty1-copia	Tork	236	GAG-PR-INT-RT-RH	145 708	124 219 995	6.51
	Retrofit	342	GAG-PR-INT-RT-RH	41 965	43 615 815	2.29
	Sire	136	GAG-PR-INT-RT-RH-ENV	223 386	210 097 734	11.01
	Oryco	105	GAG-PR-INT-RT-RH	22 078	22 854 591	1.20
	合计	819		433 137	400 788 135	21.01
Ty3-gypsy	Del	207	GAG-PR-RT-RH-INT-CHR	295 222	334 005 916	17.51
	Reina	249	GAG-PR-RT-RH-INT-CHR	27 803	39 235 939	2.06
	Crm	47	GAG-PR-RT-RH-INT	40 781	44 298 955	2.32
	Tat	238	GAG-PR-RT-RH-INT	217 288	230 055 053	12.06
	Galadriel	1	GAG-PR-RT-RH-INT-CHR	23	51 248	0.00
	Athila	1	GAG-PR-RT-RH-INT-ENV	311	257 970	0.01
	合计	743		581 428	647 905 081	33.96
总计		1 562		1 014 565	1 048 693 216	54.97

说明: a表示每个谱系的数量; b表示在毛竹基因组中LTR反转录转座子所占的比例

表2 毛竹 LTR 反转录转座子谱系特征

Table 2 Structure of LTR retrotransposon family of moso bamboo

谱系	家族 ^a	百分比 ^c /%	全长LTR ^d	Solo LTR ^e	全长LTR/Solo LTR	全长 LTR+Solo LTR	百分比 ^f /%
Tork	236	28.82	1 169	1 492	1.28	2 661	28.76
Retrofit	342	41.76	302	1 158	3.83	1 460	15.78
Sire	136	16.61	464	3 139	6.77	3 603	38.95
Oryco	105	12.82	521	1 006	1.93	1 527	16.51
Ty1-copia	819	100.00	2 456	6 795	2.77	9 251	100.00
Del	207	27.86	2 102	2 992	1.42	5 094	41.97
Reina	249	33.51	495	1 245	2.52	1 740	14.34
Crm	47	6.33	510	1 352	2.65	1 862	15.34
Tat	238	32.03	2 168	1 251	0.58	3 419	28.17
Galadriel	1	0.13	0	7	0	7	0.06
Athila	1	0.13	0	14	0	14	0.12
Ty3-gypsy	743	100.00	5 275	6 861	1.30	12 136	100.00
总计	1 562	100.00	7 731	13 656	1.77	21 387	100.00

说明: a表示每个谱系的数量; c表示每个谱系在超家族中所占的比例; d表示结构完整的LTR反转录转座子(full-length LTR), 包含两端LTR序列和完整的编码结构域^[44]; e表示仅含有两端LTR序列, 编码结构域有缺失的LTR反转录转座子(solo LTR)^[44]; f表示每个谱系中full-length LTR和solo LTR在超家族中所占的比例

LTR 反转录转座子进行分类, 将 Ty1-copia 超家族分为 4 个谱系, 分别为 Tork、Retrofit、Sire、Oryco; Ty3-gypsy 超家族分为 6 个谱系, 分别为 Del、Reina、Crm、Tat、Galadriel、Athila。其中 Tork 包含 236 个家族, Retrofit 包含 342 个家族, Sire 包含 136 个家族, Oryco 包含 105 个家族, Del 包含 207 个家族, Reina 包含 249 个家族, Crm 包含 47 个家族, Tat 包含 238 个家族, Galadriel 包含 1 个家族, Athila 包含 1 个家族。在 Ty1-copia 超家族的 4 个谱系中, Sire 的含量最高(达 11.01%), 紧随其后的是 Tork(6.51%)。在 Ty3-gypsy 超家族的 6 个谱系中, Del 的含量最高(达 17.51%), 紧随其后的是 Tat(12.06%)(表 1)。Tat 和 Del 在植物中普遍存在并且是植物所特有的。ENV 域在 Sire 中被识别, CHR 域在 Del 和 Reina 中被识别(表 1~2)。

2.2 毛竹 LTR 反转录转座子的 PBS 偏好性及 LTR 序列长度分析

在转座过程中, PBS 是 LTR 反转录转座子反转录开始的重要位点, 因为 LTR 反转录转座子开始反

转录时 tRNA 会结合到 RNA 的 PBS 处, 然后通过反转录酶合成 cDNA^[45]。不同超家族和谱系的 LTR 反转录转座子对 PBS 具有不同的偏好性。由表 3 所示: MetCAT24 是转座子反转过程中使用频率最高的 PBS 位点, 占 4.05%, 比其他的位点要高, 其次是 LysTTT 和 LysTTT10。表 3 中 Ty1-copia 和 Ty3-gypsy 超家族对 PBS 位点的偏好性呈相反趋势, MetCAT24 是 Ty1-copia 超家族中使用最多的 PBS 位点, LysTTT 是 Ty3-gypsy 超家族使用最多的 PBS 位点, 但在 Ty1-copia 超家族中频率很低, 仅有 1 个。LTR 序列是 LTR 反转录转座子中特有的, 它们位于 LTR 反转录转座子的 5'端和 3'端, 是一对高度相似的序列, 通常较长的 LTR 反转录转座子具有更长的 LTR 序列, 结构也更加完整。所以把 5'端 LTR 作为参照, 对 LTR 序列长度进行统计, 结果如图 1 显示。对 LTR 反转录转座子而言, LTR 序列长度与其全长序列的长度成正比。

表 3 LTR 反转录转座子 PBS 使用统计

Table 3 Usage status of PBS in LTR retrotransposons				
tRNA	数量/个	百分比/%	Ty1-copia 使用比例/%	Ty3-gypsy 使用比例/%
MetCAT24	1 383	4.05	1.70	0.83
LysTTT	486	1.42	0.00	1.10
LysTTT10	285	0.83	0.31	0.03
LeuAAG21	131	0.38	0.00	0.15
LysTTT3	111	0.32	0.16	0.01
LeuTAG9	66	0.19	0.07	0.01

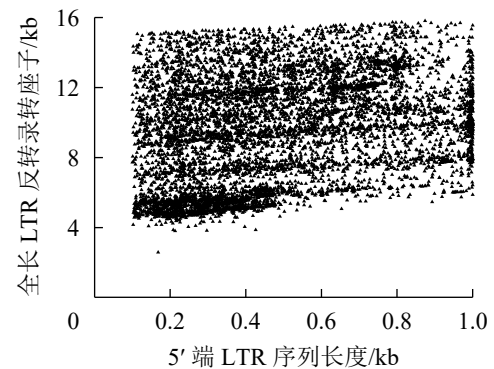


图 1 LTR 反转录转座子全长与 LTR 序列长度的相关性

Figure 1 Correlation of length between LTR and LTR retrotransposons

2.3 毛竹 LTR 反转录转座子的插入时间分布

对 21 387 个含有 TSD 位点的毛竹 LTR 反转录转座子的插入时间进行统计, 如图 2 和图 3 所示。毛竹 LTR 反转录转座子的插入时间集中于 0~2.0 Ma, 其中插入最旺盛的是在 1.0~1.5 Ma, 有 4 426 个, 占 21.06%, 插入较少的是在 3.0 Ma 之前, 有 891 个, 仅占 2.61%, 插入时间为 0 的有 508 个 (占 1.4%), 说明这部分 LTR 反转录转座子可能还具有转座潜力。Del 转座频率最高, 有 4 777 个拷贝, 占 22.62%, 且在 0.5~1.5 Ma 转座活动最为旺盛, 其次为 Sire 和 Tat。而 Retrofit、Oryco、Reina、Crm 转座频率都较低, Retrofit 最低, 仅有 1 527 个拷贝, 占 6.91%。以上数据说明毛竹基因组中 LTR 反转录转座子在 0~2.0 Ma 内大量复制增长, 且还处于不断增长的状态, 但增长趋势在减弱。

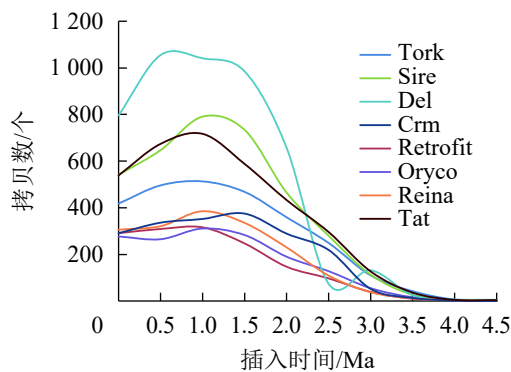


图 2 毛竹 LTR 反转录转座子各个谱系的插入时间

Figure 2 Insertion time distribution of different lineages of moso bamboo LTR retrotransposons

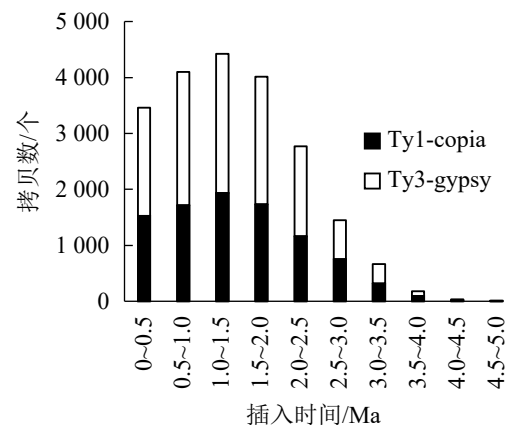


图 3 毛竹 LTR 反转录转座子超家族的插入时间

Figure 3 Insertion times of superfamily of moso bamboo LTR retrotransposons

3 讨论

3.1 第2版毛竹基因组的公布提高有助于LTR反转录转座子准确鉴定

HU等^[46]利用第1版毛竹基因组数据分析了LTR反转录转座子的分布结构和进化模式。但由于第1版毛竹基因组数据的碎片化严重,限制了对LTR反转录转座子的完整预测。毛竹第2版基因数据的覆盖范围、精准度都有所提高,完整性达95.2%^[34],相比玉米(92.2%)^[22]要高,与水稻(95.6%)^[23]接近,所以本研究的结果准确性较高。然而,LTR反转录转座子是一种重复序列^[47],软件无法准确逐条识别,产生假阴性。在不同的基因组中LTR反转录转座子所占的比例不同,如在酵母基因组中占3%^[48],在玉米基因组中占70.1%^[22],分布特点也有所差异,且基因组中还存在其他重复序列,所以也很容易产生假阳性。因此,利用较完整的毛竹第2版基因组,能在一定程度上避免误差。

在本研究中,通过对第2版毛竹基因组的注释共得到1 014 565条LTR反转录转座子(表1),占基因组的54.7%,较对第1版毛竹基因组注释的结果(1 954 616, 39.83%,不包括未知LTR)数量有所下降,比例有所上升。对7 731条结构完整的LTR反转录转座子进行家族分类,共划分为1 562个家族,较第1版(959个)也有所上升。通过对LTR反转录转座子的结构完整性判断,solo-LTR较完整LTR反转录转座子比例更高(S/F为1.77),说明可能毛竹LTR反转录转座子不平衡重组和非法重组的活动频率较高,这可能是由于毛竹基因组在不断进化的过程中对转座子产生抑制,碎片化严重。但在第1版毛竹基因组中完整的LTR反转录转座子的占比更高(S/F为0.28),这是由于第1版基因组不完整所以信息显示不全面。总的来说,第2版毛竹基因组在转座子的注释、鉴定、家族和超家族的分类等方面都相比第1版毛竹基因组更加准确,但是依然存在数万条短的组装碎片,因此本研究对毛竹LTR反转录转座子的注释仍是保守值,随着毛竹基因组的更加完善会有更多的LTR反转录转座子被发现。

3.2 不同谱系的毛竹LTR反转录转座子在进化过程中具有不同的分化和扩增活性

根据Gypsy Database 2.0网站中植物典型的谱系序列特征,毛竹LTR反转录转座子共分为Sire、Oryco、Retrofit、Tork、Crm、Del、Reina、Tatol、Galadriel、Athila等10个谱系。其中Retrofit、Reina和Tat是数量最多的3个谱系。Galadriel和Athila在植物界中虽然广泛存在^[15],但在毛竹基因组发现较少。一方面可能是Athila的大小通常为8.5~12.0 kb,并且具有相对较长的LTR序列(1.5~2.5 kb),很难被LTRhvest程序识别^[35];另一方面,毛竹基因组的不完整和碎片化,可能会遗漏Galadriel和Athila。

将毛竹LTR反转录转座子的谱系与水稻^[47]、拟南芥*Arabidopsis thaliana*^[49]进行比较,发现毛竹的10个谱系包含了双子叶植物和单子叶植物共同的进化特征,这是它们分裂后分化的结果^[50],其中Tork是例外,它在毛竹、水稻、拟南芥中并无明显差异,这表明Tork相对其他谱系更加保守。并且毛竹与水稻亚科在LTR反转录转座子都为多拷贝和多进化谱系并存,而拟南芥的谱系数则相对较少^[49],这可能是由于不同进化速率导致的结果。

在不同谱系之间,LTR反转录转座子增殖差异较大。在毛竹Ty1-copia超家族中,Retrofit的LTR反转录转座子家族数量最多(342个),但拷贝数仅占15.78%,相比之下,Sire只有136个家族,拷贝数却达11.1%,占比最高。这种情况在Ty3-gypsy超家族更为显著,Reina包含249个家族,拷贝数却仅占2.06%,Crm只有47个家族,拷贝数所占比例比Reina还高一些。各谱系中LTR反转录转座子的数量反映了它们最近的扩增情况,而各谱系中家族的数量则代表了历史上不同的分化情况。因此,不同谱系的LTR反转录转座子在进化过程中具有不同的分化和扩增活性。

3.3 毛竹LTR反转录转座子插入时间主要集中在0~2.0 Ma

LTR反转录转座子两端的LTR序列是相同的,但在不断转座过程中,LTR序列会发生突变并分化,根据剪辑替换速率,可以得出LTR反转录转座子的插入时间^[31]。毛竹LTR反转录转座子的插入时间在0~1.5 Ma呈直线式递增,但在大于1.5 Ma呈指数衰减,总体上呈现抛物线形式,与小麦*Triticum aestivum*^[51]、桑树*Morus notabilis*^[52]类似,这说明毛竹LTR反转录转座子的插入时间主要集中于0~2.0 Ma,在1.5 Ma处于转座活动爆发期,但其增长趋势处于回缩的状态。

毛竹LTR反转录转座子在大于5.0 Ma区域缺失,可能原因:第一,5.0 Ma之前的老LTR反转录转座子与年轻的LTR反转录转座子发生重组,所以无法识别^[42];第二,根据PENG等^[33]的研究,在

7.0~12.0 Ma 中, 毛竹基因组发生了四倍体事件, 之后又不断进化为二倍体, 在这个过程中毛竹基因组经历了较大的选择压力, 所以 5.0 Ma 之前的毛竹 LTR 反转录转座子在基因组中被删除或严重破坏, 无法通过结构预测和同源性比对来鉴定。

4 参考文献

- [1] 殷豪. 梨基因组 LTR 反转座子注释及进化分析研究[D]. 南京: 南京农业大学, 2014.
YIN Hao. *Genome-wide Annotation and Evolutionary Analysis of Long Terminal Repeat Retrotransposons in Pear (Pyrus bretschneideri Rehd.)*[D]. Nanjing: Nanjing Agricultural University, 2014.
- [2] NIE Qiong, QIAO Guang, PENG Lei, *et al.* Transcriptional activation of long terminal repeat retrotransposon sequences in the genome of pitaya under abiotic stress [J]. *Plant Physiol Biochem*, 2019, **135**: 460 – 468.
- [3] 蒋爽. 基于反转录转座子标记的梨属植物亲缘关系研究[D]. 杭州: 浙江大学, 2015.
JIANG Shuang. *Studies on Genetic Relationships of Pyrus Species and Cultivars based on Retrotransposons Markers*[D]. Hangzhou: Zhejiang University, 2015.
- [4] 汪浩. 植物基因组 LTR 反转录转座子注释和比较研究[D]. 上海: 复旦大学, 2008.
WANG Hao. *Annotation and Comparative Study of LTR Retrotransposons in Plant Genomes*[D]. Shanghai: Fudan University, 2008.
- [5] KOBAYASHI S, GOTO-YAMAMOTO N, HIROCHIKA H. Retrotransposon-induced mutations in grape skin color[J]. *Science*, 2004, **304**(5673): 982. doi: 10.1126/science.1095011.
- [6] ZHOU Mingbing, LIANG Linlin, HANNINEN H. A transposition-active *Phyllostachys edulis* long terminal repeat (LTR) retrotransposon [J]. *J Plant Res*, 2018, **131**(2): 203 – 210.
- [7] JIANG Shuang, TENG Yuanwen, ZONG Yu, *et al.* Review of LTR retrotransposons in plants [J]. *Acta Bot Boreali-Occident Sin*, 2013, **33**(11): 2354 – 2360.
- [8] 张赞一. 毛竹 LTR 反转录转座子-PHRE6 的克隆与转座活性鉴定以及转座监测系统的构建[D]. 杭州: 浙江农林大学, 2018.
ZHANG Zanyi. *Phyllostachys edulis LTR Transposon-cloning and Transposition Activity Identification of PHRE6 and Construction of Transposition Monitoring System*[D]. Hangzhou: Zhejiang A&F University, 2018.
- [9] 吴骏澜. 长末端重复序列反转录转座子分析流程构建及应用[D]. 合肥: 安徽农业大学, 2017.
WU Julian. *Construction and Application of Identification and Analysis Process of Full-length LTR-retrotransposons*[D]. Hefei: Anhui Agricultural University, 2017.
- [10] ROY N S, CHOI J Y, LEE S I, *et al.* Marker utility of transposable elements for plant genetics, breeding, and ecology: a review [J]. *Genes Genomics*, 2015, **37**(2): 141 – 151.
- [11] 周鹏. 梨 Ty1-copia 反转录转座子的分子特性研究[D]. 杨凌: 西北农林科技大学, 2013.
ZHOU PENG. *Molecular Character of Novel Ty1-copia Retrotransposons in Pear*[D]. Yangling: Northwest A&F University, 2013.
- [12] 马焱. 桑树全基因组转座子的鉴定及特征分析[D]. 重庆: 西南大学, 2014.
MA Bi. *Genome-wide Identification and Characterization of Transposable Elements in Mulberry (Morus notabilis)*[D]. Chongqing: Southwest University, 2014.
- [13] 侯菲. 蔷薇目 7 个物种间 LTR 反转录转座子水平转移的鉴定以及转座活性分析[D]. 重庆: 西南大学, 2018.
HOU Fei. *Horizontal Transfers and Activity Analysis of LTR Retrotransposons in Seven Rosales Species*[D]. Chongqing: Southwest University, 2018.
- [14] FINATTO T, de OLIVEIRA A C, CHAPARRO C, *et al.* Abiotic stress and genome dynamics: specific genes and transposable elements response to iron excess in rice [J]. *Rice*, 2015, **8**: 13. doi: 10.1186/s12284-015-0045-6.
- [15] GALINDO-GONZALEZ L, MHIRI C, DEYHOLOS M K, *et al.* LTR-retrotransposons in plants: engines of evolution [J]. *Gene*, 2017, **626**: 14 – 25.
- [16] 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.
- [17] SHINGOTE P R, MITHRA S V A, SHARMA P, *et al.* LTR retrotransposons and highly informative ISSRs in combination

- are potential markers for genetic fidelity testing of tissue culture-raised plants in sugarcane [J]. *Mol Breed*, 2019, **39**(2): 25.
- [18] SAZE H, KAKUTANI T. Differentiation of epigenetic modifications between transposons and genes [J]. *Curr Opin Plant Biol*, 2011, **14**(1): 81 – 87.
- [19] DU Jianchang, TIAN Zhixi, BOWEN N J, *et al.* Bifurcation and enhancement of autonomous-nonautonomous retrotransposon partnership through LTR swapping in soybean [J]. *Plant Cell*, 2010, **22**(1): 48 – 61.
- [20] LLORENS C, MUNOZ-POMER A, BERNAD L, *et al.* Network dynamics of eukaryotic LTR retroelements beyond phylogenetic trees [J]. *Biol Dir*, 2009, **4**(12): 41 – 72.
- [21] 虞洪杰. 植物 LTR 反转录转座子的预测和注释及邻聚法构建系统进化树研究[D]. 杭州: 浙江大学, 2011.
YU Hongjie. *Prediction and Annotation of LTR Retrotransposons in Plant and a New Method to Construct Phylogenetic Trees*[D]. Hangzhou: Zhejiang University, 2011.
- [22] XU Ling, ZHANG Yue, SU Yuan, *et al.* Structure and evolution of full-length LTR retrotransposons in rice genome [J]. *Plant Syst Evol*, 2010, **287**(1/2): 19 – 28.
- [23] WANG Qinghua, DOONER H K. Dynamic evolution of bz orthologous regions in the *Andropogoneae* and other grasses [J]. *Plant J*, 2012, **72**(2): 212 – 221.
- [24] LAVERGNE S, MUENKE N J, MOLOFSKY J. Genome size reduction can trigger rapid phenotypic evolution in invasive plants [J]. *Ann Bot*, 2010, **105**(1): 109 – 116.
- [25] ELLINGHAUS D, KURTZ S, WILLHOEFT U. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons[J]. *BMC Bioinf*, 2008, **9**(1). doi: 10.1186/1471-2105-9-18.
- [26] WANG Hao, LIU Jinsong. LTR retrotransposon landscape in *Medicago truncatula*: more rapid removal than in rice [J]. *BMC Genomics*, 2008, **9**(1). doi: 10.1186/1471-2164-9-382.
- [27] LERAT E. Identifying repeats and transposable elements in sequenced genomes: how to find your way through the dense forest of programs [J]. *Heredity*, 2010, **104**(6): 520 – 533.
- [28] SU Shuai, CUI Ning, SUN Aijun, *et al.* Sequence analysis of the whole genome of a recombinant Marek's disease virus strain, GX0101, with a reticuloendotheliosis virus LTR insert [J]. *Arch Virol*, 2013, **158**(9): 2007 – 2014.
- [29] LIAN Shuaibin, CHEN Xinwu, WANG Peng, *et al.* A complete and accurate Ab initio repeat finding algorithm [J]. *Interdisciplinary Sci Comput Life Sci*, 2016, **8**(1): 75 – 83.
- [30] OU Shujun, JIANG Ning. LTR_FINDER_parallel: parallelization of LTR_FINDER enabling rapid identification of long terminal repeat retrotransposons[J]. *Mobile DNA*, 2019, **10**(6403). doi: 10.1101/1722736.
- [31] BEDELL J A, KORF I, GISH W, *et al.* MaskerAid: a performance enhancement to RepeatMaskerf [J]. *Bioinformatics*, 2000, **16**(11): 1040 – 1041.
- [32] 周敏. 竹子 LINES, Ty3-gypsy 类转座子的克隆、鉴定及特性分析[D]. 杭州: 浙江农林大学, 2014.
ZHOU Min. *Cloning, Identification and Analysis Characteristics of LINES and Ty3-gypsy Retrotransposons from Bamboo*[D]. Hangzhou: Zhejiang A&F University, 2014.
- [33] PENG Zhenhua, LU Yuying, LI Lubin, *et al.* The draft genome of the fast-growing non-timber forest species moso bamboo (*Phyllostachys heterocycla*) [J]. *Nat Genet*, 2013, **45**(4): 456 – 461.
- [34] ZHAO Hansheng, GAO Zhimin, WANG Le, *et al.* Chromosome-level reference genome and alternative splicing atlas of moso bamboo (*Phyllostachys edulis*)[J]. *GigaScience*, 2018, **7**(10): giy115. doi: 10.1093/gigascience/giy115.
- [35] MONAT C, TANDON, TRANCHANT-DUBREUIL C, *et al.* LTRclassifier: a website for fast structural LTR retrotransposons classification in plants[J]. *Mobile Genet Elem*, 2016, **6**(6). doi: 10.1080/2159256X.2016.1241050.
- [36] BERNARD H R, WUTICH A, RYAN G W. *Analyzing Qualitative Data: Systematic Approaches*[M]. New York: SAGE Publications, 2016.
- [37] MA Jianxin, BENNETZEN J L. Rapid recent growth and divergence of rice nuclear genomes [J]. *Proc Nat Acad Sci*, 2004, **101**(34): 12404 – 12410.
- [38] EDGAR R C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity [J]. *BMC Bioinf*, 2004, **5**(1): 113. doi: 10.1186/1471-2105-5-113.
- [39] KIMURA M, OHTA T. On the stochastic model for estimation of mutational distance between homologous proteins [J]. *J*

- Mol Evol*, 1972, **2**(1): 87 – 90.
- [40] PATERSON A H, BOWERS J E, BRUGGMANN R, *et al.* The Sorghum bicolor genome and the diversification of grasses [J]. *Nature*, 2009, **457**(7229): 551 – 556.
- [41] WANG Hao, XU Zhao, YU Hongjie. LTR retrotransposons reveal recent extensive inter-subspecies nonreciprocal recombination in Asian cultivated rice [J]. *BMC Genomics*, 2008, **9**(1): 1 – 13.
- [42] HAVECKER E R, GAO Xiang, VOYTAS D F. The Sireviruses, a plant-specific lineage of the Ty1/copia retrotransposons, interact with a family of proteins related to dynein light chain 8 [J]. *Plant Physiol*, 2005, **139**(2): 857 – 868.
- [43] CHADHA S, SHARMA M. Genetic differentiation and phylogenetic potential of Ty3/Gypsy LTR retrotransposon markers in soil and plant pathogenic fungi [J]. *J Basic Microbiol*, 2020, **60**(6): 508 – 516.
- [44] BENNETZEN J L. Transposable element contributions to plant gene and genome evolution [J]. *Plant Mol Biol*, 2000, **42**(1): 251 – 269.
- [45] PICAULT N, CHAPARRO C, PIEGU B, *et al.* Identification of an active LTR retrotransposon in rice [J]. *Plant J*, 2009, **58**(5): 754 – 765.
- [46] HU Bingjie, ZHOU Mingbing, ZHU Yihang. Genome-wide characterization and evolution analysis of long terminal repeat retroelements in moso bamboo (*Phyllostachys edulis*) [J]. *Tree Genet Genomes*, 2017, **13**(2): 43. doi: 10.1007/s11295-017-1114-3.
- [47] PENG Yu, ZHANG Yingying, GUI Yijie, *et al.* Elimination of a retrotransposon for quenching genome instability in modern rice [J]. *Mol Plant*, 2019, **12**(10): 1395 – 1407.
- [48] WAGNER A. Distribution of transcription factor binding sites in the yeast genome suggests abundance of coordinately regulated genes [J]. *Genomics*, 1998, **50**(2): 293 – 295.
- [49] LOCKTON S, GAUT B S. The contribution of transposable elements to expressed coding sequence in *Arabidopsis thaliana* [J]. *J Mol Evol*, 2009, **68**(1): 80 – 89.
- [50] FESCHOTTE C, JIANG N, WESSLER S R. Plant transposable elements: where genetics meets genomics [J]. *Nat Rev Genet*, 2002, **3**(5): 329 – 341.
- [51] KASHKUSH K, FELDMAN M, LEVY A A. Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat [J]. *Nat Genet*, 2003, **33**(1): 102 – 106.
- [52] HE Ningjia, ZHANG Chi, QI Xiwu, *et al.* Draft genome sequence of the mulberry tree *Morus notabilis* [J]. *Nat Commun*, 2013, **4**(1): 1 – 9.