

李 阳, 袁 娜, 刘大亮, 等. 绿豆 Copia 类反转座子全基因组注释及进化分析[J]. 江苏农业学报, 2020, 36(4): 858-867.
doi:10.3969/j.issn.1000-4440.2020.04.008

绿豆 Copia 类反转座子全基因组注释及进化分析

李 阳^{1,2}, 袁 娜², 刘大亮^{1,2}, 翟小杰^{1,2}, 徐照龙², 程 静^{1,2}, 杜建厂^{1,2}

(1. 南京农业大学园艺学院/作物遗传与种质创新国家重点实验室, 江苏 南京 210095; 2. 江苏省农业科学院种质资源与生物技术研究所/江苏省农业生物学重点实验室, 江苏 南京 210014)

摘要: 利用基于结构从头寻找和同源比对的方法, 从绿豆基因组中鉴定出插入位置明确的 1 198 个完整的 Copia 类反转座子和 1 038 个 solo LTR 转座子元件。这些元件不均匀分布在绿豆的 22 条染色体上, 并和功能基因的分布呈现显著的负相关关系。绝大部分 Copia 类反转座子 (91.8%) 在最近 5.0 MYA 内插入到寄主基因组中, 并在 1.0~2.0 MYA 呈现 1 个明显的活跃峰值。此外, 每个家族的 solo LTR 转座元件数量与完整转座元件数量的比值与进化时间无关, 而与 LTR 的长度呈现显著的正相关关系。尽管这些 Copia 类反转座子可划分为 Angela、Ale、Bianca、Ivana、Maximus 和 TAR 等 6 个谱系, 但各谱系内所包含的家族数和元件数差异较大。通过染色体上的位置比较, 发现 67 个 Copia 类反转座子家族的 713 个元件插入到 629 个功能基因的内部或附近区域 (<1 kb), 提示这些元件可能对基因的结构、表达和功能产生一定的影响。

关键词: Copia 类反转座子; 插入时间; 进化树; 绿豆

中图分类号: Q754 **文献标识码:** A **文章编号:** 1000-4440(2020)04-0858-10

Annotation and evolutionary analysis of the Copia retrotransposons in mung bean

LI Yang^{1,2}, YUAN Na², LIU Da-liang^{1,2}, ZHAI Xiao-jie^{1,2}, XU Zhao-long², CHENG Jing^{1,2},
DU Jian-chang^{1,2}

(1. College of Horticulture, Nanjing Agricultural University/State Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing 210095, China;
2. Institute of Germplasm Resources and Biotechnology, Jiangsu Academy of Agricultural Sciences/Jiangsu Provincial Key Laboratory of Agrobiotechnology, Nanjing 210014, China)

Abstract: Using structure-based and homology search approaches, 1 198 intact Copia retrotransposons and 1 038 solo LTR retrotransposons with well-defined insertion positions were identified from the mung bean genome. These elements were unevenly distributed along 22 chromosomes of mung bean, and showed a significant negative correlation with the distribution of functional genes. The vast majority of Copia retrotransposons (91.8%) were inserted into the host genome in the last 5.0 MYA, and showed a significant active peak during 1.0~2.0 MYA. In addition, the ratio of the number of solo LTRs

to the number of intact elements in each family had no correlation with the evolutionary time, but had a significant positive correlation with the size of the LTR. Although these Copia retrotransposons could be divided into six lineages, such as Angela, Ale, Bianca, Ivana, Maximus and TAR, the number of families and elements in each lineage

收稿日期: 2020-01-04

基金项目: 国家自然科学基金项目 (31370266); 江苏省农业科技自主创新基金项目 (CX17-3004)

作者简介: 李 阳 (1993-), 男, 河南焦作人, 硕士研究生, 主要从事生物信息学研究。(E-mail) yliroot@163.com

通讯作者: 杜建厂, (E-mail) dujianchang@hotmail.com

varied greatly. By comparing the positions on the chromosomes, it was found that 713 elements of 67 Copia retrotransposon families were inserted into or close to the 629 functional genes (<1 kb), suggesting that these elements may affect the structure, expression and function of genes.

Key words: Copia retrotransposons; insertion time; phylogenetic tree; mung bean

转座元件 (Transposable elements), 又叫转座子, 是基因组中可以移动的 DNA 片段。根据转座方式的不同, 植物中的转座子一般可分为 2 种类型, 即以 RNA 为媒介的反转座子和以 DNA 为媒介的 DNA 转座子。根据结构的不同, 反转座子又可以分为 LTR (Long terminal repeat) 类反转座子和非 LTR 类反转座子^[1]。研究表明, LTR 类反转座子是植物基因组的主要组成部分。在玉米中, LTR 类反转座子的 DNA 含量可以达到 75%^[2]。根据序列相似性和转座酶的先后顺序, LTR 类反转座子又可以分为 Copia 和 Gypsy 2 个超家族。两者在结构上的区别主要在于整合酶 *INT* 的位置不同。Copia 类反转座子具有 *INT-RT-RH* 结构, 而 Gypsy 型反转座子则具有 *RT-RH-INT* 结构^[1]。

一个完整的 LTR-反转座子 (Intact element) 通常具有一些明显的结构特征, 如包含 2 个正向重复的 LTR 序列, 通常以 TG 开头, CA 结尾, 往往在其插入位点附近形成 4~6 bp 的正向重复序列 (Target site duplications, TSDs), 含有与转录起始和终止有关的 tRNA 结合位点 (Primer binding site, PBS) 和多聚嘌呤序列 (Polypurine tract, PPT)。除了完整的 LTR-反转座子 (Intact element) 外, 基因组中也存在大量的单个 LTR 序列 (solo LTR) 以及部分缺失的 LTR 元件 (Truncated element)^[3-4]。一般认为, 这 2 种类型的 LTR 转座子是由于各种类型的重组事件而产生的^[5-6]。

由于同一个 LTR-反转座子的 2 个 LTR 序列来自于同一个 mRNA 分子, 因此, 一个新的 LTR-反转座子在形成时具有 2 个序列完全一致的 LTR 序列^[4]。随着进化时间的推移, 同样的 2 个 LTR 序列逐渐产生和积累变异。因此, 根据 2 个 LTR 序列之间的核苷酸序列差异, 以及 LTR-反转座子的大致进化速率, 可以估算 LTR-反转座子的插入时间。比如, 在大豆基因组中, >90% 的鉴定完整的 LTR-反转座子是在最近 3.0 MYA 产生的, 并且 >3 000 个 LTR-反转座子是在最近 0.5 MYA 内产生^[7]。

作为植物中主要的重复序列类型之一, Copia

类反转座子是构成植物基因组的重要组成部分。已有研究表明, 4% 的水稻基因组 DNA^[8]、12% 的大豆基因组 DNA^[9] 以及 24% 的玉米基因组 DNA^[10] 由各种类型的 Copia 类反转座子组成。尽管大多数转座元件含有插入/缺失、移码突变、终止密码子等, 在基因组中不再活跃, 一些 Copia 类反转座子在特殊生长条件下仍然具有转座活性。如在组织培养条件下, 水稻中的 *Tos17* 转座子被大量激活, 并偏向性的插入到基因富集的区域^[8]。利用 *Tos17* 这种特性, 科研人员创造了大量的水稻突变体库, 从而更有助于水稻基因功能的研究^[11]。另外, 一些 Copia 类反转座子还直接介导了植物器官的产生和性状的形成。已有研究表明, Copia 类反转座子 *Rider* 参与了番茄 *SUN* 基因的复制, 进而影响了番茄果实的形状和大小^[12]。橙子中 Copia 类反转座子 *Tcs1* 插入到 *Ruby* 基因的上游, 促进了花色素苷的合成, 使得果肉颜色呈现红色^[13]。这些结果表明, Copia 类反转座子对于植物基因组的构成, 突变体库的建立, 以及植物表型的产生都可能具有重要作用。

绿豆 (*Vigna radiata*) 是一种重要的经济作物。由于其具有生长期短、适应性广、营养价值高等优点, 被广泛作为倒茬轮作、间作套种、减灾救灾等作物, 以及用于豆芽、粉丝、糕点等食品的制作^[14]。绿豆基因组具有 22 条染色体, 基因组大小约 579 Mb^[15]。尽管绿豆基因组序列已组装完成, 但是由于其基因组进化时间较长, 成分较多, 很多重复序列种类多、结构复杂。因此, 绿豆基因组的注释还有待于进一步完善, 这对于后续功能基因的研究以及绿豆分子育种工作都可能提供更好的便利和数据来源。

鉴于 Copia 类反转座子在植物基因和基因组进化过程中的重要作用, 本研究利用公开发表的二倍体绿豆基因组序列 (*V. radiata* var. *radiata* VC1973A), 从基因组水平上系统注释了绿豆 Copia 类反转座子, 初步明确了绿豆 Copia 类反转座子的数量、类型、分类、分布、亲缘关系, 以及与功能基因

之间的相互关系,为后续深入研究 *Copia* 类转座子在绿豆基因和基因组进化过程中所起的作用提供依据,也为后续进一步开展绿豆功能基因的研究,活性转座子的鉴定和筛选可利用的转座子分子标记提供数据来源。

1 材料与方法

1.1 绿豆基因组数据来源及 *Copia* 类反转座子的注释

本研究中所用的绿豆 (*Vigna radiata*) 基因组数据来自 GenBank 数据库 (<https://www.ncbi.nlm.nih.gov/>)。基于结构从头寻找和同源比对相结合的方法,利用 LTR_STRUC 软件搜索绿豆全基因组序列,获得完整的 LTR-反转座子元件^[16]。然后,以这些完整的 LTR-反转座子元件为目标序列,利用 Cross_match 软件(默认参数)对绿豆基因组序列进行同源比对搜索。根据比对结果,利用 perl 脚本进一步分析处理,再经人工检查确定所有 LTR-反转座子的结构和插入位置。根据 LTR-反转座子的内部结构、同源性和 80-80-80 规则,鉴定出所有的 *Copia* 类 LTR-反转座子,并将它们分为不同的家族^[1]。

1.2 绿豆 *Copia* 类反转座子基因组分布特征分析

本研究使用以下方法对绿豆 LTR-反转座子进行随机性分布检验。首先,将绿豆基因组按连续无重叠 1 Mb 的窗口进行划分,并对每个窗口内鉴定的 LTR-反转座子数量进行统计。然后,借助 perl 脚本,将鉴定的所有 LTR-反转座子分别随机分配到各个窗口中,并重复 10 000 次,同时计算每次每个窗口内的 LTR-反转座子数量。最后,根据公式 $P = (m + 1) / (10\ 000 + 1)$ 计算每个窗口中的 P 值。其中, m 指在 10 000 次模拟结果中有 m 次结果等于或小于实际统计数值,当 $0.025 < P < 0.975$ 时符合随机分布标准^[17]。

1.3 绿豆 *Copia* 类反转座子插入时间的估算

提取含有 TSD 位点的完整 LTR-反转座子两端的 LTR 序列,利用 MUSCLE 软件(默认参数)进行序列比对^[18]。根据公式 $T = K / 2r$ 计算插入时间。其中进化速率 $r = 1.3 \times 10^{-8}$,核苷酸差异 K 用 Jukes-Cantor 方法进行计算。

1.4 绿豆 *Copia* 类反转座子的系统进化树分析

提取绿豆 *Copia* 类反转座子中反转座酶 RT 基

因的保守序列,并从每个家族中选取年轻且 RT 基因结构相对完整的序列作为参照序列。利用 MUSCLE 软件(默认参数),对 *Copia* 类家族的 RT 基因序列进行序列比对。用 MEGA7 软件中的 P -distance 模块构建邻近进化树,重复 500 次。

1.5 绿豆 *Copia* 类反转座子与功能基因之间的关系

根据绿豆 *Copia* 类反转座子与基因间的位置关系,提取内部和上下游 1 kb 范围内存在转座元件的功能基因,利用 Blast2GO 5 软件对这些功能基因进行注释,并使用 WEGO2.0 软件(<http://wego.genomics.org.cn/>)进行富集分析^[19-20]。

2 结果与分析

2.1 绿豆 *Copia* 类反转座子的鉴定及注释

本研究基于结构从头寻找和同源比对的方法,从绿豆基因组中共鉴定出 2 236 个插入位置明确的 *Copia* 类反转座子,其中,1 198 个为完整转座子元件,1 038 个为 solo LTR。从数量上来看,两者的比值 (S/I) 为 0.87 : 1.00。在 1 198 个完整转座子元件中,长度在 20 kb 以上的元件有 11 个,15~20 kb 的元件有 9 个,10~15 kb 的元件有 33 个,5~10 kb 的元件有 212 个,1~5 kb 的元件有 622 个,1 kb 以下的元件有 311 个。根据统一的转座子的分类标准,这些元件可分为 126 个不同的家族,拷贝数从 1 到 589 不等。其中,共有 111 个 *Copia* 类反转座子家族含有 RT 转座酶序列,而剩余的 15 个家族不含有 RT 转座酶序列,提示它们可能属于非自主型转座子。拷贝数最大的家族为 Vrar_C1,含有 333 个完整转座子元件和 256 个 solo LTR;元件两端含有 2 个 120 bp 的 LTR 序列;元件全长为 6 134 bp(表 1)。另外,其他高拷贝(>100)的家族有 5 个,分别是 Vrar_C2(317 个),Vrar_C3(277 个),Vrar_C4(157 个),Vrar_C5(112 个),和 Vrar_C6(109 个)。元件长度最长的为 Vrar_C2,达到 16 029 bp,含有 44 个完整转座子元件 273 个 solo LTR(表 1)。元件长度最短的为 Vrar_C30,全长仅有 909 bp,LTR 序列为 173 bp,仅含有 6 个拷贝(表 1)。这些结果说明,尽管同属于 *Copia* 类反转座子,但不同家族具有各自的遗传特征,可能在绿豆基因和基因组的进化过程中扮演着不同的角色。

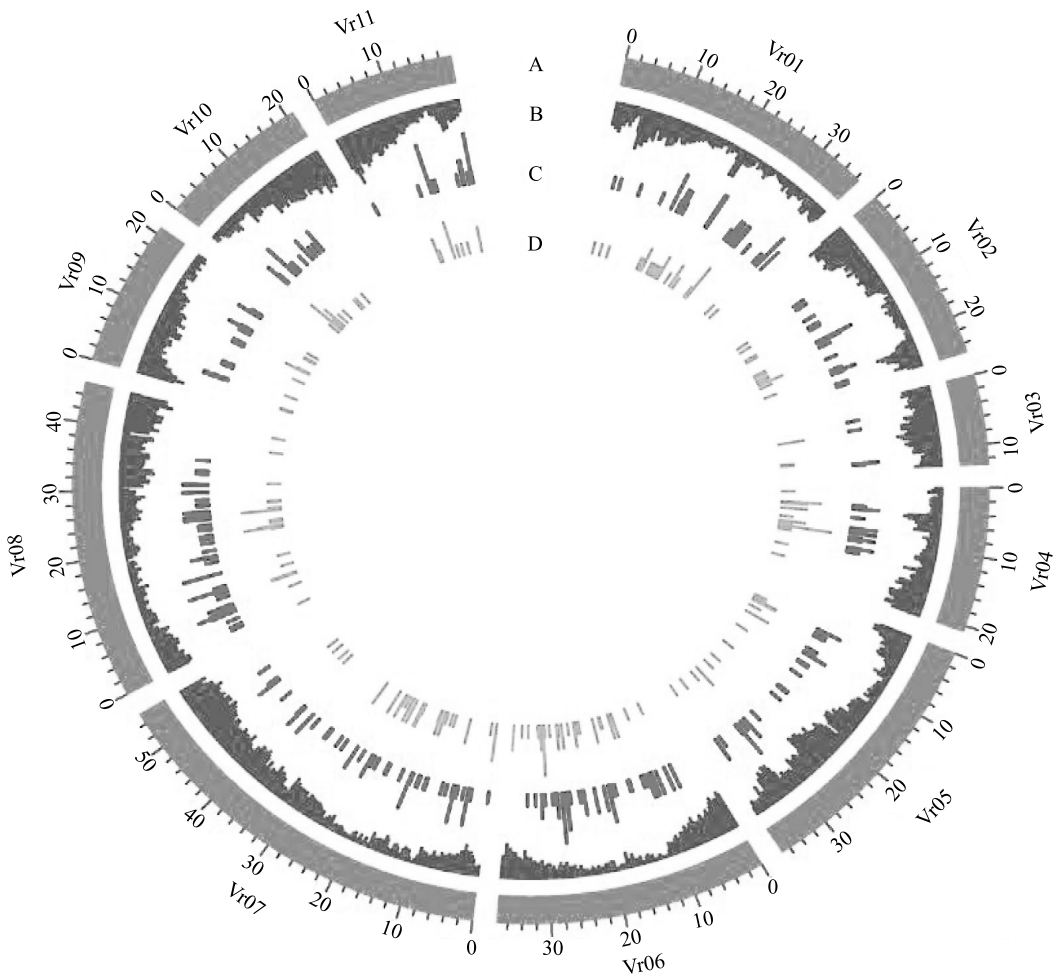
表 1 绿豆 *Copia* 类反转座子部分家族信息汇总表Table 1 Summary of some *Copia* families in mung bean

家族	自主型转座子 (RT)	LTR 大小 (bp)	完整元件大小 (bp)	完整元件 数量(个)	solo LTR 数量 (个)	元件总数 (个)	插入时间 (MYA)	<i>S/I</i>	谱系
Vrar_C1	是	120	6134	333	256	589	3.28	0.77	Ale
Vrar_C2	是	1455	16 029	44	273	317	2.42	6.2	Maximus
Vrar_C3	是	191	14 473	198	79	277	2.21	0.4	Ale
Vrar_C4	是	996	10 447	36	121	157	1.74	3.36	Maximus
Vrar_C5	是	1 065	8 520	10	102	112	1.78	10.2	Maximus
Vrar_C6	是	193	9 645	104	5	109	1.52	0.05	Ale
Vrar_C7	是	206	7 350	49	32	81	1.76	0.65	Ale
Vrar_C8	是	220	5 022	36	7	43	2.22	0.19	Ale
Vrar_C9	否	1 394	11 092	6	33	39	3.01	5.5	/
Vrar_C10	是	677	7 872	21	17	38	2.54	0.81	Bianca
Vrar_C11	是	189	4 568	25	6	31	1.66	0.24	Ale
Vrar_C12	否	318	4 085	18	7	25	2.49	0.39	/
Vrar_C13	是	150	4 088	12	11	23	1.92	0.92	Ale
Vrar_C14	是	205	5 384	20	3	23	1.29	0.15	Ale
Vrar_C15	否	271	2 844	13	9	22	0.98	0.69	/
Vrar_C16	否	235	3 229	16	5	21	2.44	0.31	/
Vrar_C17	是	454	9 053	14	5	19	2.57	0.36	Angela
Vrar_C18	是	439	5 528	10	9	19	1.65	0.9	TAR
Vrar_C19	是	207	5 935	10	5	15	1.72	0.5	Ale
Vrar_C20	是	506	6 510	5	8	13	1.5	1.6	Angela
Vrar_C21	是	409	4 889	8	4	12	1.2	0.5	Ivana
Vrar_C22	是	163	4 813	3	9	12	0.74	3	Ale
Vrar_C23	是	200	4 480	9	2	11	1.63	0.22	Ivana
Vrar_C24	是	250	4 395	8	1	9	1.37	0.13	Ale
Vrar_C25	是	173	4 900	6	2	8	1	0.33	Ale
Vrar_C26	是	282	5 038	4	4	8	0.73	1	Ale
Vrar_C27	是	211	7 784	6	1	7	2.58	0.17	Ale
Vrar_C28	否	253	2 695	4	3	7	2.17	0.75	/
Vrar_C29	是	213	4 459	5	2	7	1.81	0.4	Bianca
Vrar_C30	否	173	909	2	4	6	3.16	2	/

2.2 绿豆 *Copia* 类反转座子在基因组中的分布

为了探讨 *Copia* 类反转座子在绿豆基因组中的分布规律及特征,我们对 463 个 1 Mb 窗口中 LTR-反转座子的分布进行了随机性检验。其中,有 61 个 (13%)窗口中 *Copia* 类反转座子的模拟值与理论值之间存在显著差异,说明这些 LTR-反转座子具有偏

向分布的特点(图 1)。其发生的频率远低于大豆的 79%,略高于梨的 10%^[7,21],提示在不同的基因组中,LTR-反转座子在染色体上的分布存在较大差异。另外,我们的研究结果也显示,solo LTR 和完整转座子元件具有相似的分布特点,均富集在染色体的异染色质区域(图 1)。



A:染色体;B:基因;C:完整 LTR 元件;D:单个 LTR 元件。
图 1 绿豆 Copia 类反转座子在染色体上的分布
Fig.1 Distribution of Copia retrotransposons along the chromosomes in mung bean

为了便于比较,我们用同样的方法,对绿豆基因组中功能基因的分布也进行了随机性检验。研究结果显示,在绿豆中,336 个(73%)窗口中的基因具有偏向性分布的特点。这一比例略低于大豆基因的 80%,略高于梨基因的 68%^[7,21]。从染色体上的分布来看,绿豆中的基因更多分布在染色体两端(图 1)。

为了明确 Copia 类反转座子的分布是否与基因的分布呈一定的相关性,我们对两者进行了相关性分析。结果显示,在绿豆基因组中,Copia 类反转座子元件密度与基因密度之间呈现显著的负相关关系(图 2)。这一结果提示,基因的密度可能是影响 Copia 类反转座子分布的一个因素。

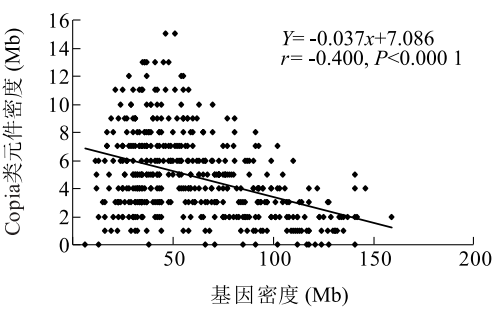
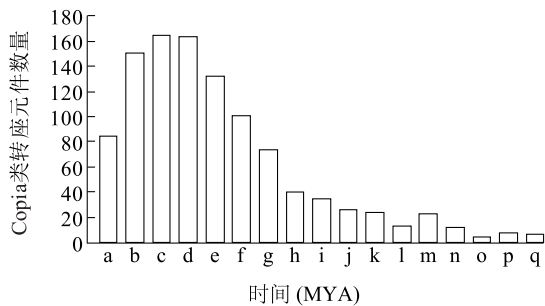


图 2 Copia 类元件与基因的相关性
Fig.2 The association between Copia element density and gene density

2.3 绿豆 Copia 类反转座子的插入时间
为估算绿豆 Copia 类反转座子的大致插入时

间,我们对含有 *TSD* 位点的完整转座元件进行了进一步分析。结果显示,在绿豆中,绝大多数元件(91.8%)在最近 5.0 MYA 插入到基因组中,并在 1.0~2.0 MYA 左右具有最高活性(图 3)。有 84 个元件在 0.5 MYA 内插入到寄主基因组中(图 3)。需要特别指出的是,23 个元件的插入时间为 0,提示它们是新近插入到绿豆基因组中的,可能仍然具有转录和转座活性(图 3)。

从单个家族来看,在 126 个 *Copia* 类反转座子家族中,有 106 个家族(84.1%)中元件的平均插入时间<3.0 MYA;47 个家族(37.3%)中元件的平均插入时间在 1.0~2.0 MYA;28 个家族(22.2%)中元件的平均插入时间<1.0 MYA;有 6 个家族中元件的平均插入时间为 0,提示该家族可能仍然具有活性。



a:0~0.5 MYA;b:0.5~1.0 MYA;c:1.0~1.5 MYA;d:1.5~2.0 MYA;e:2.0~2.5 MYA;f:2.5~3.0 MYA;g:3.0~3.5 MYA;h:3.5~4.0 MYA;i:4.0~4.5 MYA;j:4.5~5.0 MYA;k:5.0~5.5 MYA;l:5.5~6.0 MYA;m:6.0~6.5 MYA;n:6.5~7.0 MYA;o:7.0~7.5 MYA;p:7.5~8.0 MYA;q:>8.0 MYA。

图 3 绿豆 *Copia* 类反转座子插入时间分布

Fig.3 Insertion time of *Copia* retrotransposons in mung bean

2.4 绿豆 *Copia* 类反转座子 *S/I* 与插入时间和 LTR 长度之间的相关性

我们前期在大豆上的研究结果表明,solo LTR 与完整转座子元件拷贝数之间的比例(*S/I*)与 LTR 长度存在显著的相关性,而与平均插入时间不存在相关性^[7]。为了验证这一结论在绿豆中是否仍然成立,我们统计分析了拷贝数最多的前 30 个家族。研究结果显示,绿豆中 *S/I* 值与 LTR 长度之间具有显著的正相关关系(图 4)。这说明,随着 LTR 长度的增加,可能更有利于同一个 *Copia* 类反转座子 2 个 LTR 序列之间重组事件的发生,

从而更有利于 solo LTR 序列的形成。研究结果也显示,绿豆中 *S/I* 值与 *Copia* 类反转座子家族的平均插入时间之间不存在相关关系(图 4),说明 solo LTR 的形成可能主要与重组率有关,而与进化时间关系不大。这一结论可能不是物种特异性的,而可能适用于不同的豆类植物基因组中。

2.5 绿豆 *Copia* 类反转座子的系统进化树分析

已有的研究结果表明,植物中 *Copia* 类反转座子可大致分为 6 种谱系,包括 *Angela*、*Ale*、*Bianca*、*Ivana*、*Maximus* 和 *TAR*^[6,22-23]。为了进一步明确绿豆 *Copia* 类反转座子在系统进化树上的相互关系和发生的频率,我们对保守的逆转录酶 *RT* 进行多序列比对。结果显示,绿豆 *Copia* 类反转座子具有植物普遍存在的 6 种谱系,但是家族的数量和每个家族的拷贝数差异很大(图 5、表 2)。比如,在 6 种谱系中,*Ivana* 谱系含有 57 个家族,占总家族数的 51.4%,但是其对应的元件数量仅为 118 个,占总拷贝数的 5.6%;*Ale* 谱系含有 37 个家族,占总家族数的 33.3%,其拷贝数则为 1 277 个,占总拷贝数的 60.8%;*Bianca* 谱系家族和拷贝数均较少,分别为 6 个和 54 个,分别占总家族数和总拷贝数的 5.4%和 2.6%;*Maximus* 谱系家族数很少,但是拷贝数却很多;*TAR* 和 *Angela* 谱系含有的家族数和拷贝数均很低。这些结果表明,在不同的谱系中,家族数反映的是转座元件遗传多样性的大小,而拷贝数反映的是转座子元件转座活性的高低和活性持续的时间长短。因此,家族数和拷贝数没有必然的正相关关系。

2.6 绿豆 *Copia* 类反转座子与功能基因之间的关系

为了明确绿豆中 *Copia* 类反转座子与功能基因之间的关系,我们统计了绿豆中 *Copia* 类反转座子插入基因及基因两端临近区域(<1 kb)的元件数量(表 3)。统计结果显示,共有 43 个家族的 563 个 *Copia* 类反转座子插入到基因的内部,48 个家族的 157 个 *Copia* 类反转座子插入到基因的附近区域(<1 kb)。功能富集分析结果显示,这些基因的功能主要为细胞组分、分子功能和生物进程。进一步分析显示,细胞组分方面主要集中在细胞,细胞组分及细胞器等;分子功能方面主要集中在催化活动及拼接;生物进程方面主要集中在细胞进程和代谢进程(图 6)。

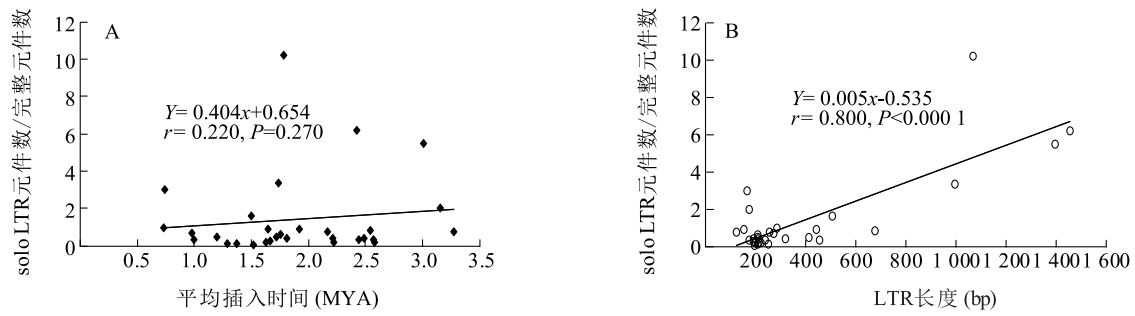


图 4 S/I 与平均插入时间 (A) 和 LTR 长度 (B) 之间的相关性

Fig.4 Association of S/I with average insertion time (A) and LTR size (B)

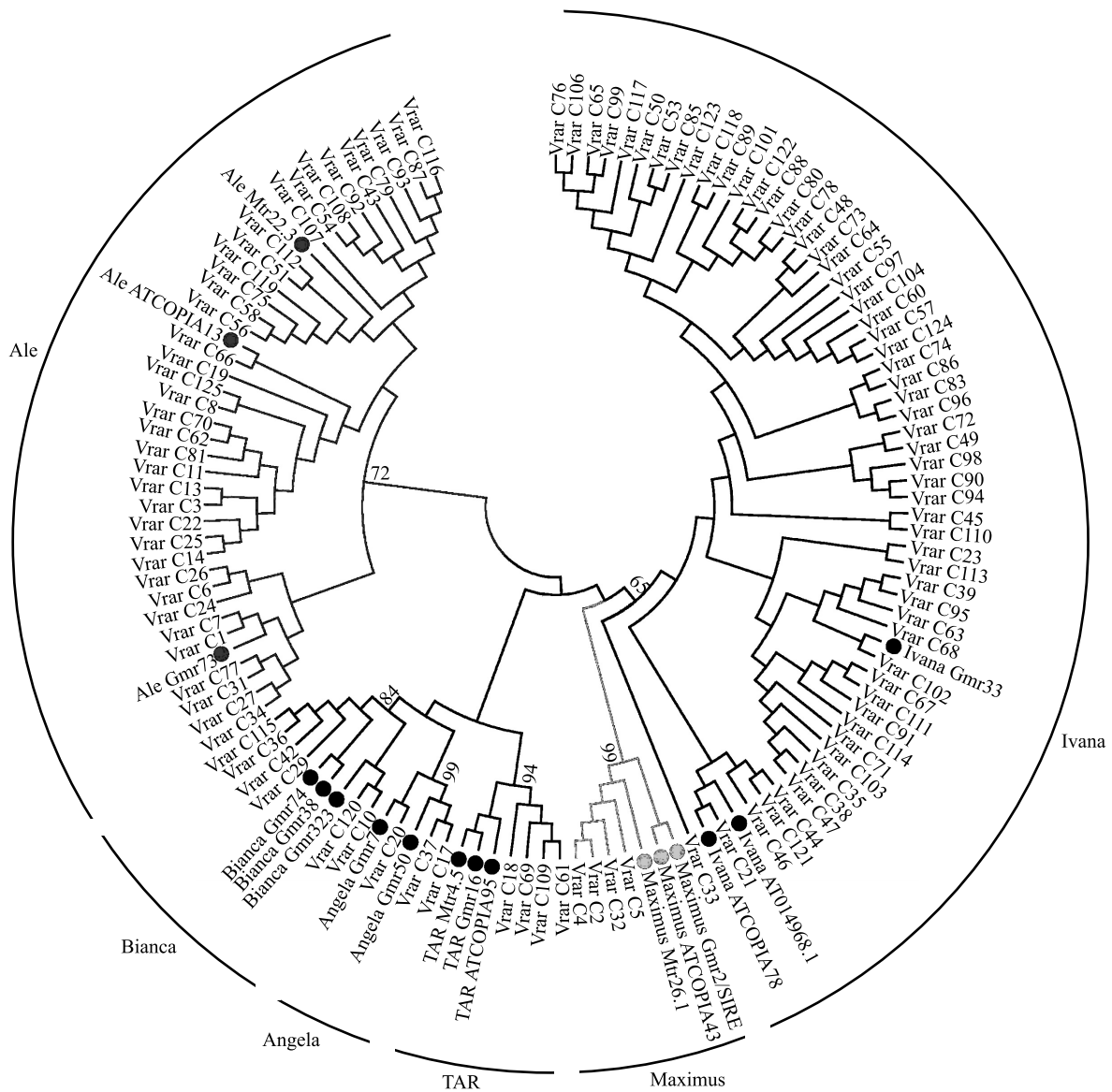


图 5 绿豆 Copia 类反转座子谱系进化树

Fig.5 Phylogenetic relationships of Copia retrotransposon lineages in mung bean

表 2 绿豆 Copia 类反转座子谱系信息汇总表

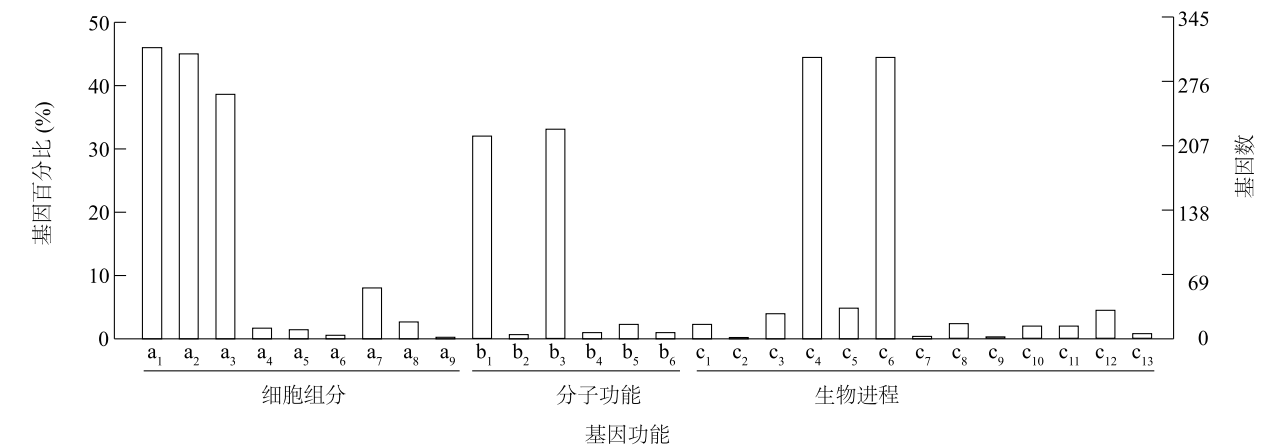
Table 2 Summary of Copia retrotransposon lineages in mung bean

谱系	家族数量 (个)	家族数量占比 (%)	元件数量 (个)	元件数量占比 (%)
Angela	3	2.7	36	1.7
Ale	37	33.3	1 277	60.8
Bianca	6	5.4	54	2.6
Ivana	57	51.4	118	5.6
Maximus	4	3.6	591	28.1
TAR	4	3.6	24	1.1

表 3 绿豆 Copia 类反转座子与功能基因信息汇总表

Table 3 Summary of Copia retrotransposons and functional genes in mung bean

插入位置	基因数量 (个)	家族数量 (个)	转座元件数量 (个)
基因内部	476	43	563
基因附近区域	158	48	157
总数	629	67	713



a₁:细胞;a₂:细胞组分;a₃:细胞器;a₄:细胞器组分;a₅:腔上包膜;a₆:胞外区;a₇:含蛋白质复合物;a₈:膜;a₉:胞外区组分。b₁:催化活性;b₂:结构分子活性;b₃:拼接;b₄:转录调节活性;b₅:转运活性;b₆:分子功能调节剂。c₁:发展过程;c₂:多细胞生物过程;c₃:细胞组分组织或合成;c₄:细胞进程;c₅:定位;c₆:代谢进程;c₇:生长;c₈:生物调节;c₉:细胞增殖;c₁₀:生物过程调节;c₁₁:信号;c₁₂:刺激反应;c₁₃:再生。

图 6 绿豆 Copia 类反转座子相关基因的功能聚类分析
Fig.6 Functional cluster analysis of genes related to Copia retrotransposons in mung bean

3 讨论

尽管绿豆基因组序列已于 2014 年对外公开发布,但是原有的研究主要是估算了转座元件的 DNA 在整个基因组中所占的比例^[15]。由于转座元件种类多、结构复杂,在基因组中又经常以“巢式”的方式存在,通常的软件(如 RepeatMasker)仅能检测转座元件的某些片段,不能完整反映转座元件在基因组中的准确位置和进化特征。具有 TSD 完整的 LTR-反转座子和 solo LTR 转座子由于具有明确的插入位置,在许多完成测序的植物基因组中都先后被注释出来^[2,7,24]。本研究中,联合运用基于结构从头寻找和同源比对的方法,我们从绿豆基因组中系统鉴定出 1 198 个完整转座子元件和 1 038 个 solo LTR 的 Copia 类反转座子。这些转座元件的注释为

深入研究 Copia 类反转座子在绿豆基因和基因组进化中的作用提供了数据来源。

从插入时间来看,绝大多数 Copia 类反转座子是在最近 5.0 MYA 插入到绿豆基因组中的,这与之之前水稻的研究结果是基本一致的^[4]。而从爆发的时期来看,与大豆基因组中的 LTR-反转座子不同,绿豆 Copia 类反转座子不是在最近(<0.5 MYA)具有最高的拷贝数,而是在 1.0~2.0 MYA 具有最高的拷贝数,提示它们在该时期具有一个活性升高的爆发时期。这一现象也在我们之前番茄的研究结果中得到了印证^[25]。对番茄 LTR-反转座子插入时间的统计分析结果表明,其最活跃的时期是 2.0~3.0 MYA^[25]。而对于梨的 Copia 类反转座子来说,除了最近(<0.5 MYA)具有最高的拷贝数以外,在 2.5~3.0 MYA 和 5.0~5.5 MYA 也具有较多的拷贝数^[21]。这些结果表

明,对于不同的植物基因组而言,尽管转座元件的拷贝数呈现出随进化时间延长逐渐减少的指数分布,但是爆发的集中时间和爆发所持续的时间不相同,也说明不同的基因组可能具有不同的进化历史。

前人对拟南芥、水稻和麦类作物的 599 个 *Copia* 类反转座子进行了系统进化树构建和比较研究^[6]。结果表明,*Copia* 类反转座子可划分为 6 个非常保守的谱系,即 *Angela*、*Ale*、*Bianca*、*Ivana*、*Maximus* 和 *TAR*,并在单子叶和双子叶植物分化之前就已经产生^[6]。我们在大豆上的研究结果也表明,除 *Bianca* 谱系外,大豆 *Copia* 类反转座子也具有其他 5 个谱系^[7]。本研究的结果表明,绿豆 *Copia* 类反转座子也可以聚类到上述 6 个谱系中。这进一步说明,植物 *Copia* 类反转座子的谱系具有古老性和保守性。需要指出的是,对于不同的基因组,同一谱系内家族的数量和拷贝数可能不同。比如,在大豆中 *Maximus* 谱系和 *Ivana* 谱系具有较高的元件数量和家族数量^[7],水稻中 *Maximus* 谱系和 *Ale* 谱系元件数量和家族数量较高^[24],而在绿豆中,*Ale* 谱系具有较高的元件数量,*Ivana* 谱系具有较高的家族数量。这些结果也进一步表明,对于不同的基因组而言,*Copia* 类反转座子家族爆发的程度和持续的时间可能很不相同,这也可能是造成不同植物基因组大小具有明显差异的原因之一。

总之,本研究以公开发表的绿豆基因组为研究对象,在对 *Copia* 类反转座子进行精细注释的基础上,系统研究了其家族分类、染色体分布、插入时间、系统进化树,以及与功能基因的关系,为后续进一步鉴定活性转座子,开发可利用的转座子分子标记,筛选相关的绿豆突变体,加快相关基因功能的研究提供了数据来源和理论依据。

参考文献:

- [1] 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.
- [2] BAUCOM R S, ESTILL J C, CHAPARRO C, et al. Exceptional diversity, non-random distribution, and rapid evolution of retroelements in the B73 maize genome[J]. *PLoS Genet*, 2009, 5(11): e1000732.
- [3] DEVOS K M, BROWN J K, BENNETZEN J L. Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis* [J]. *Genome Res*, 2002, 12(7): 1075-1079.
- [4] MA J, DEVOS K M, BENNETZEN J L. Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice[J]. *Genome Res*, 2004, 14(5): 860-869.
- [5] NYSTEDT B, STREET N R, WETTERBOM A, et al. The Norway spruce genome sequence and conifer genome evolution [J]. *Nature*, 2013, 497(7451): 579-584.
- [6] WICKER T, KELLER B. Genome-wide comparative analysis of *Copia* retrotransposons in *Triticeae*, rice, and *Arabidopsis* reveals conserved ancient evolutionary lineages and distinct dynamics of individual *Copia* families [J]. *Genome Res*, 2007, 17(7): 1072-1081.
- [7] DU J, TIAN Z, HANS C S, et al. Evolutionary conservation, diversity and specificity of LTR-retrotransposons in flowering plants: insights from genome-wide analysis and multi-specific comparison [J]. *Plant J*, 2010, 63(4): 584-598.
- [8] INTERNATIONAL RICE GENOME SEQUENCING PROJECT. The map-based sequence of the rice genome [J]. *Nature*, 2005, 436(7052): 793-800.
- [9] SCHMUTZ J, CANNON S B, SCHLUETER J, et al. Genome sequence of the palaeopolyploid soybean [J]. *Nature*, 2010, 463(7278): 178-183.
- [10] SCHNABLE P S, WARE D, FULTON R S, et al. The B73 maize genome: complexity, diversity, and dynamics [J]. *Science*, 2009, 326(5956): 1112-1115.
- [11] MIYAO A, TANAKA K, MURATA K, et al. Target site specificity of the *Tos17* retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome [J]. *Plant Cell*, 2003, 15(8): 1771-1780.
- [12] XIAO H, JIANG N, SCHAFFNER E, et al. A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit [J]. *Science*, 2008, 319(5869): 1527-1530.
- [13] BUTELLI E, LICCIARDELLO C, ZHANG Y, et al. Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges [J]. *Plant Cell*, 2012, 24(3): 1242-1255.
- [14] 赵丹, 程须珍, 王丽侠, 等. 绿豆基因组研究进展 [J]. *植物遗传资源学报*, 2010, 11(5): 583-588.
- [15] KANG Y J, KIM S K, KIM M Y, et al. Genome sequence of mungbean and insights into evolution within *Vigna* species [J]. *Nat Commun*, 2014, 5: 5443.
- [16] MCCARTHY E M, MCDONALD J F. LTR_STRUC: a novel search and identification program for LTR retrotransposons [J]. *Bioinformatics*, 2003, 19(3): 362-367.
- [17] YIN H, DU J, LI L, et al. Comparative genomic analysis reveals multiple long terminal repeats, lineage-specific amplification, and frequent interelement recombination for *Cassandra* retrotransposon in pear (*Pyrus bretschneideri* Rehd.) [J]. *Genome Biol Evol*, 2014, 6(6): 1423-1436.
- [18] EDGAR R C. MUSCLE: multiple sequence alignment with high accuracy and high throughput [J]. *Nucleic Acids Res*, 2004, 32(5): 1792-1797.

- [19] CONESA A, GOTZ S, GARCIA-GOMEZ J M, et al. Blast2GO; a universal tool for annotation, visualization and analysis in functional genomics research[J]. *Bioinformatics*, 2005, 21(18): 3674-3676.
- [20] YE J, ZHANG Y, CUI H, et al. WEGO 2.0: a web tool for analyzing and plotting GO annotations, 2018 update[J]. *Nucleic Acids Res*, 2018, 46(W1):71-75.
- [21] YIN H, DU J, WU J, et al. Genome-wide annotation and comparative analysis of long terminal repeat retrotransposons between pear species of *P. bretschneideri* and *P. Communis*[J]. *Sci Rep*, 2015, 5: 17644.
- [22] WANG H, LIU J S. LTR retrotransposon landscape in *Medicago truncatula*: more rapid removal than in rice[J]. *BMC Genomics*, 2008, 9: 382.
- [23] MASCAGNI F, GIORDANI T, CECCARELLI M, et al. Genome-wide analysis of LTR-retrotransposon diversity and its impact on the evolution of the genus *Helianthus* (L.) [J]. *BMC Genomics*, 2017, 18(1): 634.
- [24] TIAN Z, RIZZON C, DU J, et al. Do genetic recombination and gene density shape the pattern of DNA elimination in rice long terminal repeat retrotransposons? [J]. *Genome Res*, 2009, 19(12): 2221-2230.
- [25] XU Y, DU J. Young but not relatively old retrotransposons are preferentially located in gene-rich euchromatic regions in tomato (*Solanum lycopersicum*) plants[J]. *Plant J*, 2014, 80(4): 582-591.

(责任编辑:陈海霞)