

DOI: 10.3724/SP.J.1005.2013.01014

# 白菜和甘蓝基因组转座子表达及其对基因调控的潜在影响

赵美霞<sup>1,2</sup>, 张彪<sup>2</sup>, 刘胜毅<sup>1</sup>, 马渐新<sup>2</sup>

1. 中国农业科学院油料作物研究所, 农业部油料作物生物学重点实验室, 武汉 430062;  
2. 美国普渡大学农学系, 西拉法叶 47907

**摘要:** 转座子或转座元件是大多数真核生物基因组的主要组成成分。甘蓝(*Brassica oleracea*)基因组比白菜(*B. rapa*)大主要是转座子的扩增差异造成的。然而, 这两个芸薹属近缘物种转座子表达水平以及对基因的调控和功能的影响目前还不清楚。文章对白菜和甘蓝叶、根、茎 3 个器官的转录组数据进行了初步分析。结果显示, 转座子的表达量很低, 转录组 reads 中有 1% 来自转座子的转录本; 转座子的表达存在器官差异, 且不同类别和家族的转座子表达量相差很大, 相同类别和同一家族的转座子在白菜和甘蓝基因组中的表达活性也不相同。进一步鉴定到转录读出的 LTR 反转座子, 其与下游基因距离小于 2 kb 的有 41 个, 小于 100 bp 的有 9 个, 这些 LTR 的转录读出很可能通过正义或反义的转录本激活或干扰下游基因的表达。同时, 具有转录读出的 intact LTR 比 solo LTR 具有更强的读出活性。通过深入分析转座子的插入位点发现, 白菜基因组中转座子插入基因内部的频率比甘蓝基因组中的高; 与反转座子相比, DNA 转座子更偏向于插入或保留在基因的内含子当中。这些结果为认识转座子对其他蛋白编码基因的影响提供了基础。

**关键词:** 白菜; 甘蓝; 转座子; 转录组; 转录读出; 转座子插入位点

## Transposon expression and potential effects on gene regulation of *Brassica rapa* and *B. oleracea* genomes

ZHAO Mei-Xia<sup>1,2</sup>, ZHANG Biao<sup>2</sup>, LIU Sheng-Yi<sup>1</sup>, MA Jian-Xin<sup>2</sup>

1. Key Laboratory of Oil Crops Biology and Genetic Breeding, Oil Crops Research Institute, Chinese Academy of Agricultural Sciences, Wuhan 430062, China;  
2. Department of Agronomy, Purdue University, West Lafayette, IN 47907, USA

**Abstract:** Transposons or transposable elements (TEs) are ubiquitous and most abundant DNA components in higher eukaryotes. Recent sequencing of the *Brassica rapa* and *B. oleracea* genomes revealed that the amplification of TEs is one

收稿日期: 2013-03-11; 修回日期: 2013-04-16

基金项目: 国家自然科学基金项目(编号: 30671119)资助

作者简介: 赵美霞, 博士研究生, 研究方向: 功能和比较基因组学。E-mail: xiami1119@163.com

通讯作者: 刘胜毅, 研究员, 研究方向: 油菜功能基因组学和植物病理学。E-mail: liusy@oilcrops.cn

马渐新, 副教授, 研究方向: 植物遗传学与基因组学。E-mail: maj@purdue.edu

网络出版时间: 2013-5-9 10:58:45

URL: <http://www.cnki.net/kcms/detail/11.1913.R.20130509.1058.002.html>

of the main factors inducing the difference in genome size. However, the expressions of TEs and the TE effects on gene regulation and functions of these two *Brassica* diploid species were unclear. Here, we analyzed the RNA sequencing data of leaves, roots, and stems from *B. rapa* and *B. oleracea*. Our data showed that overall TEs in either genome expressed at very low levels, and the expression levels of different TE categories and families varied among different organs. Moreover, even for the same TE category or family, the expression activities were distinct between the two *Brassica* diploids. Forty-one and nine LTR retrotransposons with the transcripts that read into their adjacent sequences have the distances shorter than 2 kb and 100 bp compared to the downstream genes. These LTR retrotransposon readout transcriptions may produce sense or antisense transcripts of nearby genes, with the effects on activating or silencing corresponding genes. Meanwhile, intact LTRs were detected at stronger readout activities than solo LTRs. Of the TEs inserted into genes, the frequencies were observed at a higher level in *B. rapa* than in *B. oleracea*. In addition, DNA transposons were prone to insert or retain in the intronic regions of genes in either *Brassica* genomes. These results revealed that the TEs may have potential effects on regulating protein coding genes.

**Keywords:** *Brassica rapa*; *B. oleracea*; transposon; transcriptome; transcription readout; TE insertion

芸薹属(*Brassica*)属于双子叶植物纲, 十字花科 (Brassicaceae 或 Cruciferae), 包含我国主要的油料和蔬菜作物。其中的 6 个代表作物以及相互之间的关系可以很好地用禹式三角来阐述<sup>[1]</sup>。3 个基本种白菜 (*B. rapa*, AA, n=10)、甘蓝 (*B. oleracea*, CC, n=9) 以及黑芥 (*B. nigra*, BB, n=8) 通过两两的自然或人工杂交形成各自的异源四倍体甘蓝型油菜 (*B. napus*, AACC, n=19)、芥菜型油菜 (*B. juncea*, AABB, n=18) 以及埃塞俄比亚芥 (*B. carinata*, BBCC, n=17)<sup>[2,3]</sup>。

转座元件 (Transposable element, TE) 或转座子 (Transposon) 是一类可移动的 DNA 序列, 它能够通过自我转座从基因组的一个位点移动到另一个位点。根据转座子转座的机制以及转座的中间体, 可以将其分成两类<sup>[4,5]</sup>: 第一类被称为“复制-黏贴”型的反转录转座子 (Retrotransposon) 或称 RNA 转座子 (RNA transposon)。这一类又根据是否含有长末端重复序列 (Long terminal repeat, LTR) 分成两个亚类: LTR 反转座子 (LTR retrotransposon) 和非 LTR 反转座子 (Non-LTR retrotransposon), 后者中典型的有长散在核元件 (Long interspersed nuclear element, LINE) 和短散在核元件 (Short interspersed nuclear element, SINE)。一般 LTR 反转座子在植物中大量存在, LINE 和 SINE 在哺乳动物中分布较多。第二类被称为“剪切-黏贴”型的 DNA 转座子 (DNA transposon)。根据剪切过程中被剪切的 DNA 的链数, DNA 转座子可以分成两个亚类: 第 1 亚类的转座酶能够识别转座子

末端反向重复序列 (Terminal inverted repeat, TIR), 主要包括 *Tc1-Mariner*、*hAT*、*Mutator*、*PIF-Harbinger*、*Pong*、MITE (Miniature inverted repeat transposable element) 以及 CACTA。在植物中第 2 亚类主要指的是 *Helitron*, 它不含有 TIR, 复制机制也与第 1 亚类不同, 它主要通过滚环来进行复制转座<sup>[5,6]</sup>。除了多倍体化外, 转座子的大规模扩增是造成植物基因组快速膨胀的主要因素。

LTR 反转座子的转录起始于 5'LTR 的启动子, 终止于 3'LTR。由于 5'LTR 与 3'LTR 高度同源性, 使得有些 LTR 反转座子从 3'端启动读出, 从而带动下游序列的转录表达。在山羊草 (*Aegilops sharonensis*) 和一粒小麦 (*Triticum monococcum* ssp. *aegilopoides*) 人工新合成的异源四倍体当中, 找到一些反转座子 Wis 2-1A, 它们产生大量的转录本, 且 Wis 2-1A 3'端的读出使邻近基因正义或反义转录, 从而激活或干扰邻近基因表达<sup>[7,8]</sup>。

转座子一度被认为是“自私的”或“寄生的”DNA 片段, 它能够通过扩增增加自身的拷贝数, 但它插入功能编码基因, 引起基因突变或提前终止, 甚至会引起染色体断裂<sup>[9-12]</sup>。转座子能引起插入突变, 因而被作为基因沉默的有效工具<sup>[13]</sup>。有些转座子尽管没有插入到基因的编码区, 但同样也引起了性状的变化。大豆的一个 CACTA 超家族转座子 *Tgmt\** 插入到类黄酮 3'羟化酶 (Flavonoid 3'-hydroxylase, F3'H) 基因的内含子部位, 引起毛状体颜色发生改变<sup>[14]</sup>。

该转座子被认为目前还具有转座活性,这一家族的另一个成员 *Tgm9* 插入到黄烷酮醇-4-还原酶 2(Dihydroflavonol-4-reductase 2, DFR2)基因的第二个内含子中,影响了花和下胚轴色素的合成<sup>[15]</sup>。另有报道 *Spm* 插入一个基因位点附近<sup>[16]</sup>, *Mu* 插入 *hcf106* 基因附近<sup>[17]</sup>, *foldback-like* 插入果蝇发育相关基因 *CG13617* 附近<sup>[18]</sup>等,影响了基因的表达水平。转座子被认为并不仅仅是简单的跳跃 DNA 片段,它可能参与了基因的协同进化。

白菜和甘蓝约在 4 百万年前从它们的共同祖先种分化而来,经过长期的独立进化形成现在约 485 Mb 和 630 Mb 大小的基因组<sup>[19-21]</sup>。随着白菜和甘蓝基因组相继被测序,本文利用已建立的白菜和甘蓝转座子数据库<sup>[21, 22]</sup>,结合叶、根和茎 3 个器官的转录组数据,在全基因组水平上分析了各类转座子的表达差异、LTR 反转座子的转录读出以及含有转座子插入位点的基因的功能,初步探讨了转座子对基因表达调控的影响,在一定程度上阐述了芸薹属两个近缘物种进化上的差异。

## 1 材料和方法

### 1.1 转录组序列中 TE 含量的估计

利用 RepeatMasker(<http://www.repeatmasker.org>)软件,将 TE 数据库与白菜和甘蓝基因组叶、根、茎 3 个器官的转录组 reads 比对分析,鉴定来自转座子的转录本,并进行人工检验。TE 在叶、根、茎 3 个器官中的表达量按各器官转录组数据中匹配各类或各家族 TE 的长度占该器官总的转录组 reads 长度的百分比来计算。标准化的 TE 的表达量被定义为转录组每百万 reads 中来自于某类或某家族 TE 每千碱基长度的 reads 数(Reads per kilo bases of TE category per million RNA-Seq reads, RPKM)。一系列 Perl 程序的编写贯穿全文。

### 1.2 LTR 反转座子转录读出正义和反义转录本的确定

这里正义和反义指的是 LTR 反转座子转录的方向与下游邻近基因转录的方向相同,反义指相反。LTR 反转座子的方向主要依赖于内部的反转录酶基因的方向来判定。不含任何反转录酶基因或其保守区域片段的 LTR 反转座子(如非自主的 intact LTR 和

solo LTR 等)将与同一家族其他已知方向的成员比较来确定。如果一个家族所有成员都不含有反转录酶基因或其保守区域片段,它们将继续与邻近家族比较,最终确定其转录的方向。

### 1.3 含有 TE 插入位点的基因功能分析

利用 GO(gene ontology)数据库分别从细胞组分(Cellular component)、分子功能(Molecular function)和生物学过程(Biological process)3 个方面对含有 TE 插入位点的基因功能注释。统计每个 GO term 在这些基因中的拷贝数,并用 R 语言的热图包(Heatmap package)进行展示。

### 1.4 统计分析方法

利用 SAS 9.2 软件中卡平方适合性检验方法对白菜和甘蓝基因组中具有转录读出活性的 solo LTR 和 intact LTR 的比例、TE 插入基因内的频率以及 TE 插入或保留在基因内含子与外显子的比例进行了统计差异性测验。

## 2 结果与分析

### 2.1 TE 的低水平表达及器官间差异

白菜和甘蓝约在 4 百万年前从它们的共同祖先种分化而来,经过长期的独立进化形成现在约 485 Mb 和 630 Mb 大小的基因组<sup>[19-21]</sup>。随着白菜和甘蓝基因组相继被测序,本文利用已建立的白菜和甘蓝的转座子数据库<sup>[21, 22]</sup>,结合叶、根和茎 3 个器官的转录组数据,从全基因组水平上分析了各类转座子的转录本(白菜:叶 1.02%、根 1.22%、茎 1.23%;甘蓝:叶 1.37%、根 1.34%、茎 0.88%),甘蓝茎中的 TE 表达量最低,而叶中的最高(图 1:A、B)。LTR 反转座子以及 DNA 转座子中的 CACTA 和 *Pong* 超家族在 TE 转录本中占了较大比例。不同类型的 TE 在不同的器官中表达量差异明显。白菜的 SINE 反转座子在根中的表达量达到了叶中的 7 倍,甘蓝的 *Helitron* DNA 转座子在根中的表达量约为叶中的 3.5 倍。

为避免 TE 的长度对表达量检测的影响,本研究对各类 TE 以及 LTR、CACTA 家族的表达量进行了标准化处理(详见材料与方法)。标准化后,在白菜和甘蓝叶、根、茎 3 个器官中, DNA 转座子都呈现出较高的表达量,约达到了相应器官的反转座子表

达量的 2.1 ~ 5.7 倍(DNA 转座子相比于反转座子: 白菜叶 1.47 vs. 0.27, 根 3.58 vs. 1.80, 茎 5.81 vs. 1.77; 甘蓝叶 1.12 vs. 0.34, 根 0.69 vs. 0.32, 茎 0.64 vs. 0.24)(图 1 C、D), 特别是 *PIF/Harbinger* 和 *Pong*。标准化后 TE 的表达量也同样呈现出器官间差异。

LTR 反转座子和 CACTA 转座子是影响白菜和甘蓝基因组大小差异的主要因素<sup>[21,22]</sup>。通过对白菜和甘蓝 936 个 LTR 反转座子家族和 231 个 CACTA 转座子家族的表达量分析表明, 自白菜与甘蓝分化后, 大多数 LTR 反转座子和 CACTA 转座子的活性发生了变异; 同一物种内不同家族转座子的表达量

相差很大, 同一家族也具有很强的器官差异性(图 2)。如 LTR 反转座子 *F193* 家族在白菜叶和甘蓝茎中的表达量约为其他器官的 10 ~ 27 倍, CACTA 转座子 *Bc90* 家族在白菜叶、根和茎 3 个器官中都有很高的表达量, 而在甘蓝的各器官中却不表达。

## 2.2 LTR 反转座子的转录读出影响了邻近序列的表达

利用先前已分别检测到的具有 3'LTR 转录读出活性的 37 个白菜 LTR 和 70 个甘蓝 LTR<sup>[22]</sup>, 进一步分析发现, 在这两个芸薹属近缘物种中, 尽管未达

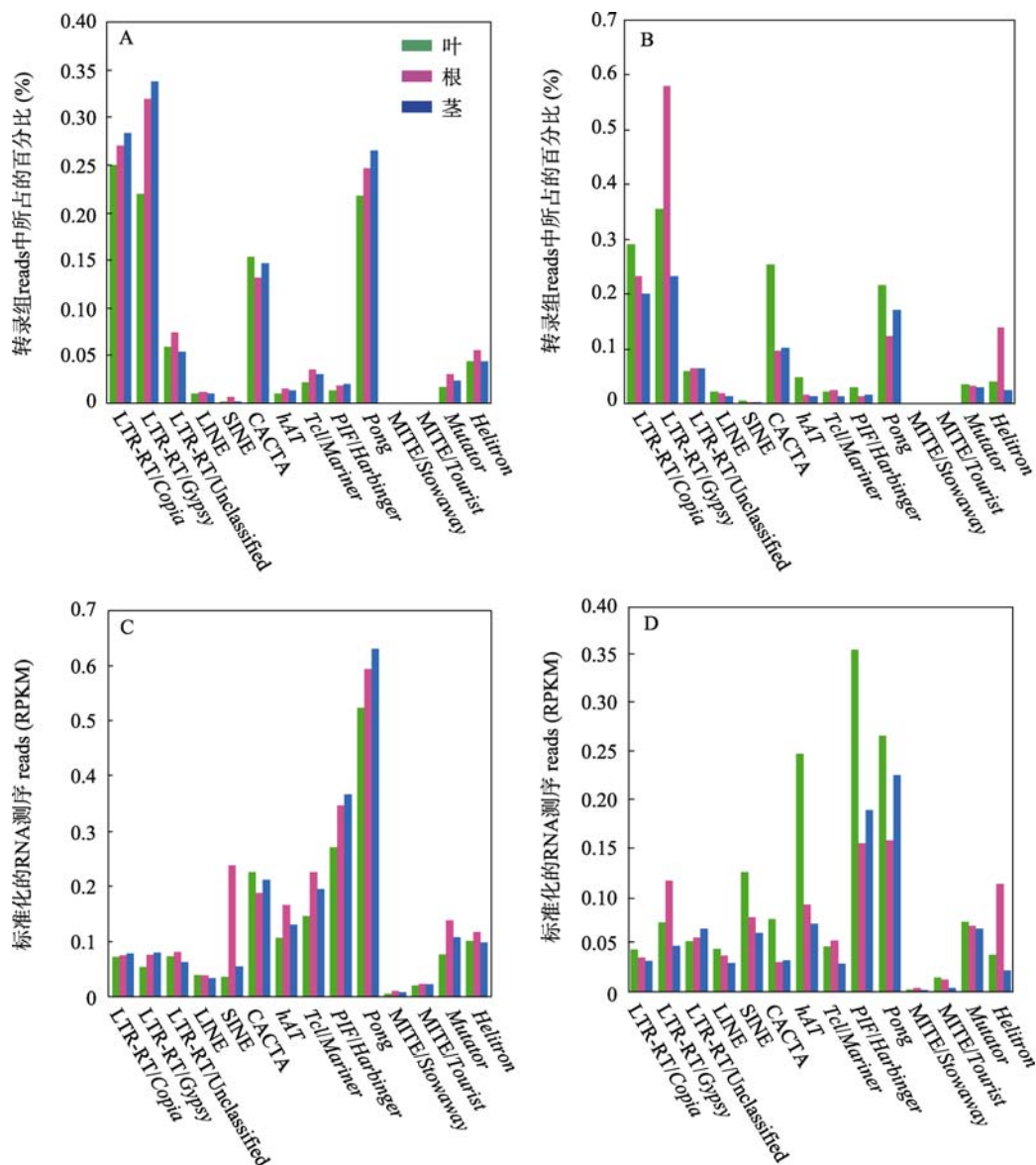


图 1 白菜和甘蓝叶、根、茎 3 个器官转录组数据中转座子的表达  
A 和 C: 白菜; B 和 D: 甘蓝。



到显著水平, 但 intact LTR(具有完整两端长末端重复序列)相比于 solo LTR(只具有单个重复序列, 一般是通过 intact LTR 两端长末端重复序列的不等同源重组而产生<sup>[25]</sup>)具有更高的转录读出活性, 其中白菜基因组具有转录读出活性的 solo LTR 与 intact LTR 的比例与全基因组水平上这两者的比例差异更大(表 1)。

通过对这 37 和 70 个 LTR 反转座子下游侧翼序列的分析发现, 其中分别有 14 和 27 个与下游邻近

基因的距离小于 2 kb(表 2)。进一步分析这 41 个 LTR 反转座子与下游基因转录的方向发现, 在白菜和甘蓝中分别存在 6 和 20 个的转录与下游邻近基因的方向相反, 因此它们可能产生该基因反义的转录本, 从而干扰该基因的表达。而一些正义的转录本则可能使下游基因过量表达。LTR 反转座子与下游基因的距离越小, 影响下游基因表达的可能性越大, 在白菜和甘蓝基因组中分别检测到 4 和 5 个转录读出的 LTR 反转座子与下游基因的距离小于 100 bp。

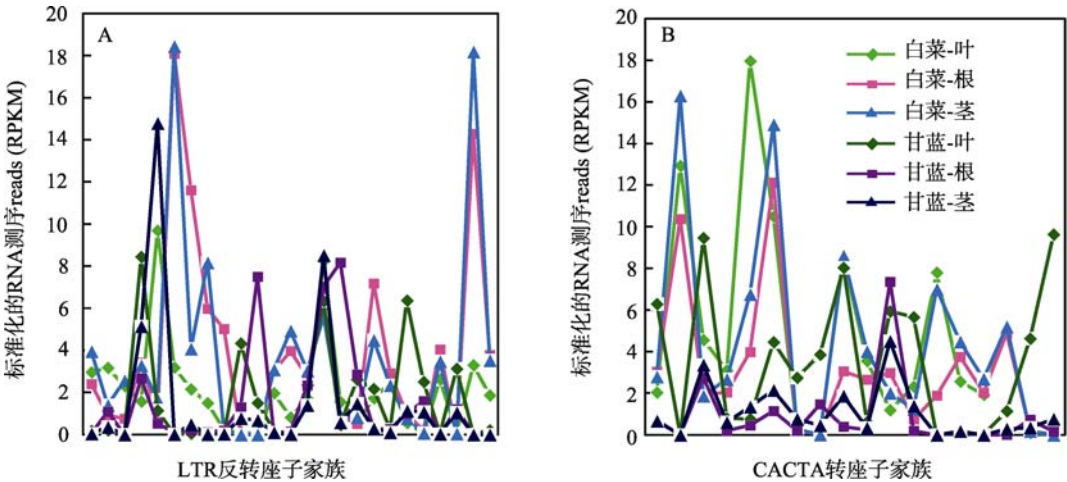


图 2 白菜和甘蓝基因组中 LTR 反转座子和 CACTA 转座子家族的表达量差异  
A: LTR 反转座子家族; B: CACTA 转座子家族。RPKM 在所有器官中都小于 2 的家族未展示。

表 1 白菜和甘蓝基因组中具有转录读出活性的 LTR 反转座子的特征

LTR 反转座子	转录读出的 LTR 反转座子			全基因组的 LTR 反转座子	<i>P</i> 值
	Solo LTR	Intact LTR	S/I 之比 <sup>a</sup>		
白菜( <i>B. rapa</i> )	7	30	0.23	0.50	0.0634
甘蓝( <i>B. oleracea</i> )	12	58	0.21	0.34	0.1110

注: <sup>a</sup>Solo LTR 与 Intact LTR 反转座子的数目之比。

表 2 白菜和甘蓝基因组 LTR 反转座子的转录读出

与下游邻近基 因的距离	白菜( <i>B. rapa</i> )			甘蓝( <i>B. oleracea</i> )		
	正义	反义	合计	正义	反义	合计
<100 bp	2	2	4	0	5	5
<500 bp	4	3	7	3	13	16
<1 kb	7	4	11	6	17	23
<2 kb	8	6	14	7	20	27
转录读出的 LTR 反转座子	37			70		

注: 正义指 LTR 反转座子与下游基因的方向相同, 反义指相反。

2.3 不同类型 TE 插入位点的比较及含有 TE 插入位点的基因功能的分析

先前的研究已鉴定 337 个白菜和甘蓝基因含有转座子插入位点<sup>[22]</sup>。本研究通过进一步比较分析发现, 白菜基因组中 TE 插入基因中的频率显著高于甘蓝基因组(142/4602 vs. 195/13382,  $P<0.01$ ,  $\chi^2$  检验), 且这些 TE 偏向于插入或更多地保留在基因的内含子当中(表 3), 如 MITE/Stowaway 家族在白菜和甘蓝中分别鉴定到 25 个和 50 个成员插入基因内部, 且全部都位于内含子当中。深入分析发现, 无论在白

菜还是甘蓝基因组中, DNA 转座子插入或保留在基因内含子与外显子的比例显著高于反转座子的这两者比例(白菜: 3.78 vs. 0.87,  $P < 0.01$ ; 甘蓝: 2.29 vs. 1.21,  $P < 0.01$ ;  $\chi^2$  检验)。DNA 转座子更偏向于插入基因的内含子可能与转座子在染色体上的分布以及重组率相关。

利用 GO 数据库, 从细胞组分、分子功能和生物学过程 3 个方面分析了 337 个基因的功能。这些基因分布很广, 主要集中在核中、叶绿体以及质膜上, 很多作为序列特异的转录因子、DNA 结合蛋白, 参与了对其他基因的转录调节、新陈代谢等过程(图 3)。

表 3 白菜和甘蓝基因组中不同类型 TE 插入基因内部的特征和比较分析

	白菜( <i>B. rapa</i> )		甘蓝( <i>B. oleracea</i> )	
	内含子	外显子	内含子	外显子
反转座子 <sup>[22]</sup>	26	30	37	33
LTR/Copia	16	20	23	18
LTR/Gypsy	5	10	7	13
LTR/Unclassified	3	0	0	1
LINE	0	0	1	0
SINE	2	0	6	1
DNA 转座子	68	18	87	38
CACTA	7	10	11	23
hAT	7	3	5	11
Tc1/Mariner	2	0	1	0
PIF/Harbinger	4	0	1	2
Pong	5	5	3	1
MITE/Stowaway	25	0	50	0
MITE/Tourist	8	0	3	1
Mutator	2	0	3	0
Helitron	8	0	10	0

### 3 讨论

TE 的扩增是造成白菜和甘蓝这两个芸薹属近缘物种基因组差异的主要因素之一<sup>[21, 22]</sup>。本研究从转录组水平, 结合第一个相对完整的芸薹属转座子数据库<sup>[21, 22]</sup>, 全面分析和比较了 TE 的表达以及对功能基因的影响。研究结果揭示 TE 以非常低的水平表达, 且表现出器官差异性。同时, 不同类别以及不同家族的 TE 表达量相差很大, 同一家族的 TE 自白菜和甘蓝基因组分化后活性也不相同(图 1, 图 2)。TE 的低水平表达以及不同器官、家族的差异可能与

DNA 甲基化状态以及 siRNA 的调控有关。拟南芥基因组包含 24%甲基化的 CG 位点, 6.7%甲基化的 CHG 位点(H = A、C、T), 1.7%甲基化的 CHH 位点<sup>[23]</sup>。拟南芥基因组的 TE 几乎都处于甲基化状态, 甲基化抑制了它们的活性和表达。另一个调控 TE 表达活性的机制即为 siRNA 介导的 TE 的沉默<sup>[24]</sup>。目前这些 TE 位点的甲基化状态尚不清楚, 下一步我们将会进一步分析和实验, 同时结合大规模的小 RNA 测序序列, 揭示这两个芸薹属近缘物种 TE 与甲基化以及 siRNA 之间的调控机制。

一个完整 LTR(intact LTR)通过元件内 5'LTR 和 3'LTR 发生不平等同源重组, 丢失了内部的成分, 形成了只有单个 LTR 末端的 solo LTR<sup>[25]</sup>。Solo LTR 的形成被认为是对抗基因组由于 TE 扩增造成扩张的主要机制之一<sup>[26]</sup>。本研究利用已鉴定到的一批可能具有转录读出活性的 LTR 反转座子<sup>[22]</sup>, 进行了更深入的探讨。研究揭示, 在白菜和甘蓝可能具有转录读出活性的 LTR 反转座子中, solo LTR 与 intact LTR 的比例在两个基因组中都小于相应的全基因组水平上 solo LTR 与 intact LTR 的比例, 白菜基因组几乎达到显著水平(表 1,  $P = 0.0634$ )。这说明 intact LTR 比 solo LTR 可能具有更高的转录读出活性。Solo LTR 作为 intact LTR 不平等同源重组的产物, 一般比较古老, 且不具有转录及复制所需要的各种酶类。此外, 基因的表达量还与转座子和基因之间的距离紧密相关<sup>[24]</sup>。LTR 反转座子的转录读出导致产生下游基因正义或反义的转录本, 从而激活或干扰下游基因的表达(表 2)。

TE 插入基因内部会破坏基因的结构, 严重的会引起基因功能的丧失。本研究发现, 相比于甘蓝, 白菜 TE 具有显著高的插入到基因内部的频率。这可能与 TE 的活性有关, 还可能与白菜基因组较高的基因密度有关(白菜: 145 基因/Mb, 甘蓝: 85 基因/Mb<sup>[19-22]</sup>)。共线性区间的比较分析揭示, TE 在甘蓝基因组基因富集区发生了扩增积累, 致使甘蓝的共线性区间得到了扩张<sup>[21, 22]</sup>。相比于外显子, TE 更偏向于插入内含子当中(表 3)。这一方面可能是由于植物基因组基因的内含子一般长于外显子使其能够获得更高的插入频率, 另一方面也可能是植物体的一种自我保护机制, 避免基因结构的破坏, 引起功能的丧失。TE 的分布与染色体结构显著相关, LTR 反

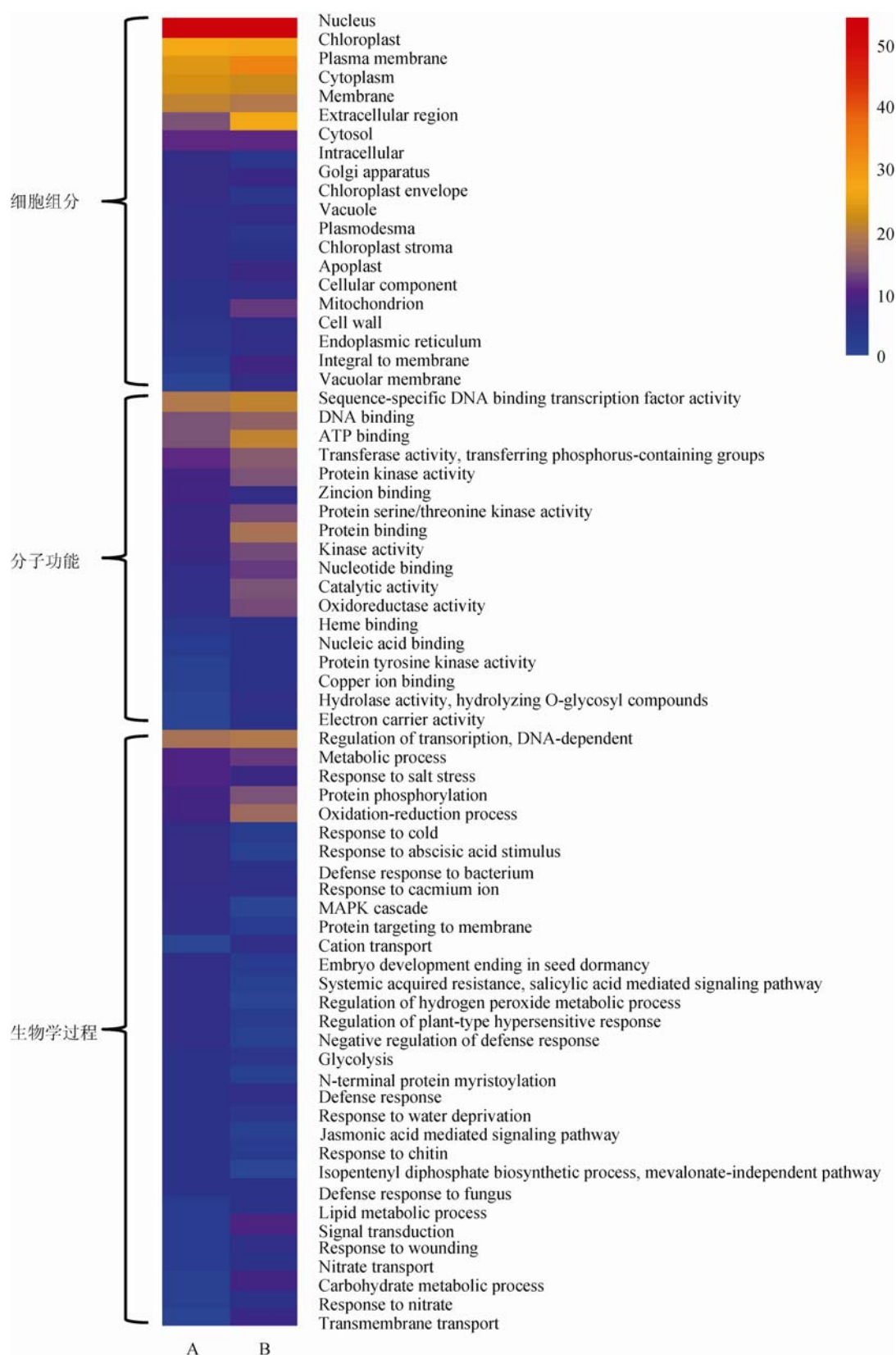


图 3 白菜和甘蓝基因组中含有 TE 插入的基因 GO 注释

A: 白菜; B: 甘蓝。白菜和甘蓝的基因 GO term 拷贝数都小于 5 的未列出。

转座子主要富集于重组抑制的异染色质区, 如着丝粒或近端着丝粒区, 这个区间基因密度很低<sup>[27]</sup>。DNA 转座子则普遍分布于常染色质区, 如臂端, 这个区间基因密度很高。这在一定程度上解释了白菜和甘蓝 DNA 转座子比反转座子更偏向插入或更多保留在基因内含子中的原因。

转座子引起的遗传多态性是选择功能变异的一个很重要的来源。另外, 具有活性的 TE 还可用来创制作为功能基因组学研究平台的转座子插入突变体, 进一步通过正向或反向遗传学的途径来研究基因的功能<sup>[13]</sup>。本研究从全基因组水平上初步分析了在不同器官中 TE 的表达, LTR 反转座子的转录读出对下游基因的潜在影响以及含有 TE 插入位点的基因的功能, 为进一步探讨造成这种现象的机制奠定了基础, 同时在一定程度上阐述了芸薹属这两个近缘物种的差异。

#### 参考文献(References):

- [1] Nagaharu U. Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. *Jap J Bot*, 1935, 7: 389–452. [DOI](#)
- [2] 刘后利. 几种芸薹属油菜的起源和进化. 作物学报, 1984, 10(1): 9–17. [DOI](#)
- [3] Beilstein MA, Al-Shehbaz IA, Kellogg EA. Brassicaceae phylogeny and trichome evolution. *Am J Bot*, 2006, 93(4): 607–619. [DOI](#)
- [4] Finnegan DJ. Eukaryotic transposable elements and genome evolution. *Trends Genet*, 1989, 5(4): 103–107. [DOI](#)
- [5] Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P, Schulman AH. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet*, 2007, 8(12): 973–982. [DOI](#)
- [6] Kapitonov VV, Jurka J. Rolling-circle transposons in eukaryotes. *Proc Natl Acad Sci USA*, 2001, 98(15): 8714–8719. [DOI](#)
- [7] Kashkush K, Feldman M, Levy AA. Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nat Genet*, 2003, 33(1): 102–106. [DOI](#)
- [8] Kashkush K, Khasdan V. Large-scale survey of cytosine methylation of retrotransposons and the impact of readout transcription from long terminal repeats on expression of adjacent rice genes. *Genetics*, 2007, 177(4): 1975–1985. [DOI](#)
- [9] White SE, Habera LF, Wessler SR. Retrotransposons in the flanking regions of normal plant genes: a role for *copia*-like elements in the evolution of gene structure and expression. *Proc Natl Acad Sci USA*, 1994, 91(25): 11792–11796. [DOI](#)
- [10] Bejerano G, Lowe CB, Ahituv N, King B, Siepel A, Salama SR, Rubin EM, Kent WJ, Haussler D. A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature*, 2006, 441(7089): 87–90. [DOI](#)
- [11] Xiao H, Jiang N, Schaffner E, Stockinger EJ, van der Knaap E. A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science*, 2008, 319(5869): 1527–1530. [DOI](#)
- [12] Naito K, Zhang F, Tsukiyama T, Saito H, Hancock CN, Richardson AO, Okumoto Y, Tanisaka T, Wessler SR. Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. *Nature*, 2009, 461(7267): 1130–1134. [DOI](#)
- [13] 廖鸣娟, 董爱华, 王正栋, 朱睦元. 植物转座子及其在功能基因组学中的应用. 遗传, 2000, 22(5): 345–348. [DOI](#)
- [14] Zabala G, Vodkin L. A putative autonomous 20.5 kb-CACTA transposon insertion in an *F3'H* allele identifies a new CACTA transposon subfamily in *Glycine max*. *BMC Plant Biol*, 2008, 8: 124. [DOI](#)
- [15] Xu M, Brar HK, Grosic S, Palmer RG, Bhattacharyya MK. Excision of an active CACTA-Like transposable element from *DFR2* causes variegated flowers in soybean [*Glycine max* (L.) Merr.]. *Genetics*, 2010, 184(1): 53–63. [DOI](#)
- [16] Masson P, Surosky R, Kingsbury JA, Fedoroff NV. Genetic and molecular analysis of the *Spm*-dependent *a-m2* alleles of the maize *a* locus. *Genetics*, 1987, 177(1): 117–137. [DOI](#)
- [17] Martienssen R, Barkan A, Taylor WC, Freeling M. Somatically heritable switches in the DNA modification of *Mu* transposable elements monitored with a suppressible mutant in maize. *Genes Dev*, 1990, 4(3): 331–343. [DOI](#)
- [18] Puig M, Cáceres M, Ruiz A. Silencing of a gene adjacent to the breakpoint of a widespread *Drosophila* inversion by a transposon-induced antisense RNA. *Proc Natl Acad Sci USA*, 2004, 101(24): 9013–9018. [DOI](#)
- [19] The *Brassica rapa* Genome Sequencing Project Consortium, Wang XW, Wang HZ, Wang J, Sun RF, Wu J, Liu SY, Bai YQ, Mun JH, Bancroft I, Cheng F, Huang SW, Li XX, Hua W, Wang JY, Wang XY, Freeling M, Pires JC, Paterson AH, Chalhoub B, Wang B, Hayward A, Sharpe AG, Park BS, Weissshaar B, Liu B, Li B, Liu B, Tong CB, Song C, Duran CF, Peng CF, Geng CY, Koh C, Lin CY, Edwards D, Mu DS, Shen D, Soumpourou E, Li F, Fraser F, Conant G, Lassalle G, King GJ, Bonnema G, Tang HB, Wang HP, Belcram H, Zhou HL, Hirakawa H, Abe H, Guo H, Wang



- H, Jin HZ, Parkin IA, Batley J, Kim JS, Just J, Li JW, Xu JH, Deng J, Kim JA, Li JP, Yu JY, Meng JL, Wang JP, Min JM, Poulain J, Wang J, Hatakeyama K, Wu K, Wang L, Fang L, Trick M, Links MG, Zhao MX, Jin MN, Ramchiary N, Drou N, Berkman PJ, Cai QL, Huang QF, Li RQ, Tabata S, Cheng SF, Zhang S, Zhang SJ, Huang SM, Sato S, Sun SL, Kwon SJ, Choi SR, Lee TH, Fan W, Zhao X, Tan X, Xu X, Wang Y, Qiu Y, Yin Y, Li YR, Du YC, Liao YC, Lim Y, Narusaka Y, Wang YP, Wang ZY, Li ZY, Wang ZW, Xiong ZY, Zhang ZH. The genome of the mesopolyploid crop species *Brassica rapa*. *Nat Genet*, 2011, 43(10): 1035–1039. [DOI](#)
- [20] Cheng F, Liu SY, Wu J, Fang L, Sun SL, Liu B, Li PX, Hua W, Wang XW. BRAD, the genetics and genomics database for Brassica plants. *BMC Plant Biol*, 2011, 11: 136. [DOI](#)
- [21] [Http://ocri-genomics.org/BrassicaTEdb](http://ocri-genomics.org/BrassicaTEdb). [DOI](#)
- [22] Zhao M, Du J, Lin F, Tong C, Yu J, Huang S, Wang X, Liu S, Ma J. Shifts in evolutionary rate and intensity of purifying selection between two *Brassica* genomes revealed by analyses of orthologous transposons and relics of a whole genome triplication. *Plant J*, 2013.
- [23] Cokus SJ, Feng SH, Zhang XY, Chen ZG, Merriman B, Haudenschild CD, Pradhan S, Nelson SF, Pellegrini M, Jacobsen SE. Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature*, 2008, 452(7184): 215–219. [DOI](#)
- [24] Hollister JD, Smith LM, Guo YL, Ott F, Weigel D, Gaut BS. Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and *Arabidopsis lyrata*. *Proc Natl Acad Sci USA*, 2011, 108(6): 2322–2327. [DOI](#)
- [25] Devos KM, Brown JKM, Bennetzen JL. Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome Res*, 2002, 12(7): 1075–1079. [DOI](#)
- [26] Ma JX, Devos KM, Bennetzen JL. Analyses of LTR-Retrotransposon structures reveal recent and rapid genomic DNA loss in rice. *Genome Res*, 2004, 14(5): 860–869. [DOI](#)
- [27] Tian ZX, Rizzon C, Du JC, Zhu LC, Bennetzen JL, Jackson SA, Gaut BS, Ma JX. Do genetic recombination and gene density shape the pattern of DNA elimination in rice long terminal repeat retrotransposons? *Genome Res*, 2009, 19(12): 2221–2230. [DOI](#)