



研究论文

Research Articles

番茄基因组逆转座子的鉴定与演化分析

许莹修[✉]

江苏省农业科学院经济作物研究所生物信息项目组, 农业部长江下游棉花和油菜重点实验室(南京), 南京, 210014

✉ 通讯作者: mailbox_e@163.com; ✉ 作者

分子植物育种, 2013 年, 第 11 卷, 第 25 篇 doi: 10.5376/mpb.cn.2013.11.0025

这是一篇采用 Creative Commons Attribution License 进行授权的开放取阅论文。只要对本原作有恰当的引用, 版权所有人允许并同意第三方无条件的使用与传播。

引用格式(中文):

许莹修, 2013, 番茄基因组逆转座子的鉴定与演化分析, 分子植物育种(online), 11(25): 1181-1189 (doi: 10.5376/mpb.cn.2013.11.0025)

引用格式(英文):

Xu Y.X., 2013, Genome-scale Identification and Evolutionary Analysis of LTR Retrotransposons in Tomato, Fenzi Zhiwu Yuzhong (online) (Molecular Plant Breeding), 11(25): 1181-1189 (doi: 10.5376/mpb.cn.2013.11.0025)

摘要 逆转座子是真核生物特别是高等植物基因组的重要组分, 对其演化起到关键作用。为了解析逆转座子在番茄基因组中的分布及其在基因组进化中的作用, 采用结构分析和同源比较方法从全基因组水平鉴定了栽培番茄的 LTR 逆转座子, 得到 15 036 个逆转座元件, 共 86.01 Mb, 占测序序列的 11.02%。将这些元件分为 330 个家族, 其中 *copia* 类的 182 个家族, *gypsy* 类 80 个。完整转座元件的转座时间在 2.5 MYA 时间处于高峰, 比大豆、玉米等逆转座元件整体古老。比较栽培番茄与野生醋栗番茄中完整转座子的时空分布发现栽培番茄(*Solanum lycopersicum*)与野生番茄(*Solanum pimpinellifolium*)在最近的时间段内(25 万年以内)产生了较明显的分化。有 930 个逆转座元件插入到基因序列附近或内部, 可能在物种演化中起到了重要作用。

关键词 逆转座子; 逆转座元件; 分布; 基因组; 演化

Genome-scale Identification and Evolutionary Analysis of LTR Retrotransposons in Tomato

Xu Yingxiu[✉]

Key Laboratory of Cotton and Rapeseed(Nanjing), Ministry of Agriculture, Institute of Industrial Crops, Jiangsu Academy of Agricultural Sciences, Nanjing, 210014, PR. China

✉ Corresponding author, mailbox_e@163.com; ✉ Authors

Abstract Retrotransposons are major components of eukaryote genomes, especially for higher plants. They play important roles in genomes evolution. In order to understand distribution and function of retrotransposons in tomato genome evolution, 15 036 LTR transposable elements were identified by structure-based and homology-based methods. They have a total size of 86.01 Mb, about 11.02% of the sequenced genome. They were clustered and classified into 330 families, 182 of which were *copia* type and 80 were *gypsy* type. Intact elements' transposition time peaked at 2.5 MYA, indicating tomato retro-elements are older than that of other plants, such as soybean and maize. Comparison of intact elements space-time distribution between tomato (*Solanum lycopersicum*) and its wild ancestor current tomato (*Solanum pimpinellifolium*) suggests they have a continual genetic interchange and differentiate at very recent time. There are 930 retro-elements localized close to or inside genes and might play important role during tomato speciation and evolution.

Keywords Retrotransposon; Retro-element; Distribution; Genome; Evolution

研究背景

DNA 重复序列是构成真核生物基因组的主要组分, 其中移动元件(mobile elements)占有很大比

例。逆转座子(retrotransposon)是一类通过“复制-粘贴”方式转座的移动元件, 在真核生物中特别是许多植物基因组中普遍存在(Finnegan, 1989; Kumar and Bennetzen, 1999; Sabot and Schulman, 2006; SanMiguel et al., 1996; Wicker et al., 2007)。其中 LTR 逆转座子两端有正向长末端重复序列(long terminal repeat, LTR), 中间是一个或几个开放阅读

收稿日期: 2013 年 06 月 24 日

接受日期: 2013 年 06 月 26 日

发表日期: 2013 年 08 月 08 日

基金项目: 本研究由江苏省农业科技自主创新基金(CX(12)5037)资助



框(ORF), 编码转座所需的各个蛋白组分。依据编码各组分的序列排列顺序的不同和序列同源性 LTR 逆转座子大体分为 Ty1-*copia* 和 Ty3-*gypsy* 两类 (Havecker et al., 2004)。LTR 逆转座子在真核生物中含有大量拷贝, 是构成基因组尤其是高等植物基因组序列的主要组分 (Baucom et al., 2009; Du et al., 2010; International Rice Genome Sequencing Project, 2005; Ma et al., 2004; Schnable et al., 2009)。碱基突变和序列重组是基因组变异的主要来源, 而逆转座子通过转座、删除机制可以产生大量的序列变化, 带来丰富的遗传信息, 因此分析研究逆转座子的分布和变化是了解基因组进化发展和生物演化的重要手段。

番茄是重要的经济作物和研究果实发育的模式植物, 许多研究工作都尝试在番茄基因组中进行逆转座子的分析鉴定 (Rogers and Pauls, 2000; Su and Brown, 1997; Yañez et al., 1998), 而且番茄中已开发出基于逆转座子序列的多态性较高的分子标记系统 (Tam et al., 2005)。逆转座子在栽培番茄基因组演化和物种形成中起到了重要的作用, 包括其复制和转座对功能基因的影响。例如逆转座子 Rider 介导的一个 SUN 基因的重复使得其表达水平升高, 从而改变了番茄果实的形态 (Xiao et al., 2008)。Rider 是一个近期活跃的家族, 部分成员已经被分离鉴定, 研究表明 Rider 可能是在 1~6 百万年 (MYA) 之间, 通过水平转移方式进入番茄基因组的 (Cheng

et al., 2009; Jiang et al., 2009)。再比如一个 *copia* 逆转座子插入到一个高活性的酯酶基因附近, 增加了它的表达量, 降低了其底物酯类的水平, 从而影响了番茄果实的味道 (Goulet et al., 2012)。番茄全基因组序列测定的完成, 为全面分析其 LTR 逆转座子提供了条件, 测序结果也表明番茄基因组中含有大量逆转座子序列 (The Tomato Genome Consortium, 2012)。

本研究基于番茄全基因组测序结果, 利用生物信息学方法从整个基因组水平鉴定逆转座子, 对其组成、分布特点和种类等进行分析, 对其在栽培和野生醋栗番茄之间的差异进行了比较, 并探讨了逆转座子在番茄物种演化和基因功能进化中的作用。

1 结果和分析

1.1 逆转座子及相关组分的分离鉴定和分类

用 LTR_STRUC 程序对整个基因组进行分析, 分离得到 1 825 个结构完整的候选转座子。人工检测确定 1 410 个结构完整清晰的元件作为进一步研究的基础。用 perl 编写程序, 在番茄基因组中分析上一步得到的转座子同源序列, 得到 676 个完整元件, 9 408 个非完整元件及 3 542 个 solo 元件, 共计 15 036 个 (86.08 Mb), 占基因组测序序列的 11.02% (表 1)。另外检测到 284.04 Mb 的同源 DNA 片段, 占基因组序列的 36.19%。完整元件共 2 086 个, 序列总长度 15.80 Mb, 单个长度介于 943 bp 和 23 668 bp 之间, 平均 7 605 bp。

表 1 番茄基因组中的 LTR 转座子序列

Table 1 LTR retrotransposon sequences in tomato genome

	完整元件 Intact element	不完整元件 Truncated element	Solo 元件 Solo element	片断 Fragment	总和 Total
结构方法	1 410	/	/	/	1 410
LTR_STRUC					
同源方法	676	9 408	3 542	/	13 626
Homologous					
总数	2 086	9 408	3 542	/	15 036
Total					
平均大小(bp)	7 605	7 026	1 171	/	/
Average size (bp)					
总碱基数(Mb)	15.83	66.10	4.15	284.04	370.12
Total size(Mb)					
占基因组比例(%)	2.03	8.46	0.53	36.19	47.21
Genome ratio (%)					

对 2 086 个人工确认的完整转座子的 LTR 序列进行聚类分析, 依据其遗传聚类和相似程度, 共得到 330 个家族, 其他元件根据其最初鉴定所依据的完整 LTR 序列划分到相应的家族, 其中 182 个 *Copia*

类, 80 个 *gypsy* 类, 其他 68 个未能鉴定(not determined, ND)(表 2)。有 29 个家族含有 100 个以上元件(表 3), 其中 3 个家族 SL_RT_F94、SL_RT_F324 和 SL_RT_F322 含有超过 1 000 个元件, 均是 *gypsy* 类型。

表 2 番茄中 LTR 逆转座子的家族分类

Table 2 Families of LTR retrotransposons in tomato

聚类元件数量	家族数量	<i>Copia</i> 家族数量	<i>Gypsy</i> 家族数量	未鉴定家族数量
Element num	Family num	<i>Copia</i> type	<i>Gypsy</i> type	ND family
2086	330	182	80	68

表 3 100 个以上元件家族列表

Table 3 Families with more than 100 elements

超家族名称	家族名称	完整 LTR 元件数	Trunc 元件数	Solo 元件数	总数
Super family	Family name	Intact element	Truncated element	Solo element	Sum
Gypsy	SL_RT_F94	197	1376	576	2149
Gypsy	SL_RT_F324	65	1084	260	1409
Gypsy	SL_RT_F322	72	999	308	1379
Gypsy	SL_RT_F218	21	380	178	579
Copia	SL_RT_F159	29	168	333	530
Copia	SL_RT_F113	63	258	85	406
Gypsy	SL_RT_F314	25	256	78	359
Copia	SL_RT_F108	90	162	82	334
Copia	SL_RT_F160	50	132	141	323
Gypsy	SL_RT_F166	6	259	20	285
Gypsy	SL_RT_F307	35	179	51	265
Gypsy	SL_RT_F209	2	26	221	249
Copia	SL_RT_F132	68	114	48	230
Copia	SL_RT_F149	18	139	72	229
Copia	SL_RT_F164	21	134	55	210
Copia	SL_RT_F123	12	149	39	200
Copia	SL_RT_F120	21	147	30	198
Copia	SL_RT_F124	11	123	34	168
Copia	SL_RT_F299	35	111	21	167
Gypsy	SL_RT_F139	10	135	22	167
Copia	SL_RT_F169	27	99	40	166
Copia	SL_RT_F118	4	76	55	135
Gypsy	SL_RT_F318	16	88	30	134
Gypsy	SL_RT_F325	13	94	27	134
Copia	SL_RT_F119	20	61	44	125
Gypsy	SL_RT_F203	25	69	25	119
Gypsy	SL_RT_F326	6	92	20	118
Gypsy	SL_RT_F225	13	91	10	114
Copia	SL_RT_F163	14	68	21	103
总数	/	989	7069	2926	10984
Sum					



1.2 转座子的转座时间和分布

逆转座子转座发生的时间是重要的信息，而且可以依据其两侧 LTR 序列的差异判断突变的比率从而确定转座发生至今的大概时间(SanMiguel et al., 1998)。比较分析 2 086 个完整转座子的 LTR 序列并计算突变率，确定了它们的转座年龄，结果表明在不同的时间段转座发生的频率是不同的。转座高峰时期在 2.5 MYA 时间左右，表明在这个时间段番茄基因组中逆转座子活性较强(图 1)。番茄逆转座子转座高峰时期与其他已知的高等植物有所不同，在接近现代时间段内转座事件逐渐减少，而如大豆、玉米等基因组中转座事件却随时间接近而逐渐增加(Baucom et al., 2009; Du et al., 2010)。

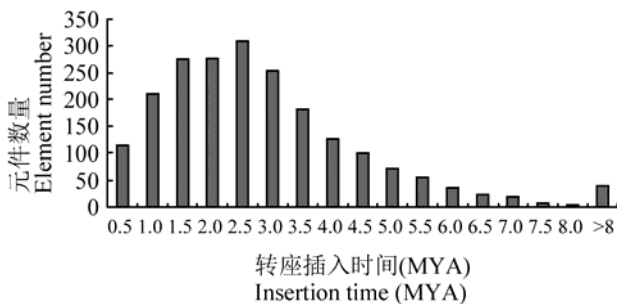


图 1 番茄中完整逆转座子的转座时间分布
Figure 1 Insertion time of tomato intact LTR retrotransposons

我们统计了不同染色体上逆转座元件的数量及其与各染色体测序长度的关系，结果表明二者基本成正比(图 2)，不同种类逆转座元件在各染色体上的分布也大体均衡，各染色体上都是非完整元件所占比例最大，完整元件较少(图 3)。

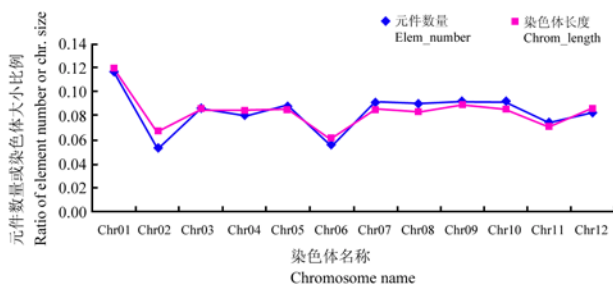


图 2 逆转座元件数量与相应染色体测序长度的关系(Y 轴显示标准化后的数值)
Figure 2 Relation between retro-element numbers and corresponding chromosomes (Y axis shows standardized values)

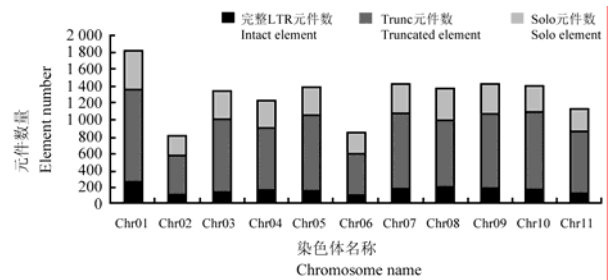


图 3 不同类型元件在各染色体上的数量和比例
Figure 3 Numbers of different type elements on chromosomes

1.3 完整逆转座子元件在栽培和野生番茄中的差异分析

除去单个碱基的自然突变之外，逆转座子还可以通过序列重组过程不断产生大量插入、缺失及异位重组等变异(Girard and Freeling, 1999; Morgante et al., 2007)。逆转座子的转座和消除机制以及突变等过程表明它们可以在较短时间内改变基因组序列组成并影响基因功能，而且这种快速的变化使得在不同物种之间逆转座子很少存在较高的序列同源性。但是在近缘物种中比较逆转座子的差异可以为研究物种分化、植物驯化及品种差异等提供丰富的信息。在番茄栽培种和野生种分化后发生的逆转座子转座事件必然在两个基因组中产生差异，依据发生转座的时间可以判断差异产生是在物种分化以前或以后，反之也可以推测物种分化的大体时间。我们分析了栽培番茄中完整转座子在野生番茄基因组中的相应分布，结果发现 1 341 个元件在两个基因组中是相同的(shared)，但是栽培番茄中另外鉴定到 733 个转座元件在野生番茄中并不存在，说明是在物种分化后形成或者是在野生番茄基因组进化过程中被删除(表 1)。依据不同时间段计算不同转座元件的比例的话，可以看出在最近的 0.25 MYA 时间段内 shared 元件的比例明显高于其它所有时间段，表明在最近的 25 万年里，栽培番茄基因组与野生番茄基因组发生了明显的分化(图 4)。

1.4 插入编码基因内部或附近的逆转座元件

逆转座子对基因组进化的影响是多方面的。从功能角度来讲基因组的演化包括基因表达变化(包括调控序列的碱基变化和表观修饰)、编码序列的改变和可变剪切，而逆转座子的活动对这几个方面都可以产生影响(Hua-Van et al., 2011)。逆转座子也可以通过插入或者删除影响基因组的大小以及染色

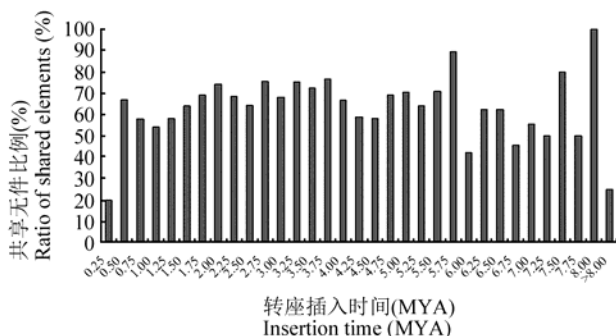


图4 共享元素在不同时间段的比例
Figure 4 Ratios of shared elements at time slices

体结构。插入基因序列附近甚至内部的逆转座子元件对于基因组的演化具有较直接的影响,可能在物种进化过程中起到了重要作用。我们比较了栽培番茄中所有逆转座元件与基因的相对位置,结果发现有 930 个元件位于基因序列附近(与基因序列 5'端距离小于 1 000 bp 或位于基因内部)(表 4)。

表 4 位于基因附近或内部的元件

Table 4 Elements close to or inside genes

距离	内部	≤35bp	36bp-100bp	101bp-500bp	501bp-1000bp	总数
Distance	Inside					Sum
数量	270	22	56	281	301	930
Number						

逆转座元件变异迅速,种类繁多,并没有统一的分类系统。依据 *pol* 基因编码产物的排列顺序和序列同源性 LTR 转座子可分为 *copia* 和 *gypsy* 两个主要类型。我们依据前人的一些标准(Wicker et al., 2007)用同源分类方法将番茄中逆转座子分为 330 个家族,其中鉴定到 *copia* 类(182 个)远多于 *gypsy* (80 个),但是从元件数量上来看 *copia* 类(5 314 个)却少于 *gypsy* (8 978 个),并且含有 1 000 个以上元件的三个家族都是 *gypsy* 类型。因此, *copia* 逆转座元件的多态性较高,而 *gypsy* 在进化中相对保守。

研究表明逆转座元件的转座插入位置序列并没有特异性,是随机的过程。但是逆转座子在染色体上的分布有一定的差异,通常在基因较多的区域,逆转座元件相对较少,而基因密度小的区域逆转座元件相对较多(Baucom et al., 2009)。这种不均匀可能是由于逆转座子在转座时对特定区域有一定选择性或者是由于逆转座子的消除机制更多地作用于常染色质区,也可能是不同区域所面对的选择压力差异的反映。番茄中逆转座元件在 12 条染

2 讨论

2.1 番茄基因组中逆转座子的数量和分布

我们用结构分析和同源比对方法在番茄基因组中鉴定到 15 036 个逆转座元件,总共 86.1 Mb,占测序序列总长度(759.9 Mb)的 11.3%。这些元件不包括零散的难于精确识别的逆转座子序列片段。其中完整元件有 2 068 个,只有总数的 13.7%,说明大多数元件的结构已经随着时间的流逝通过突变和各种消除机制被破坏。从转座时间分布来看也符合这种推测。番茄完整转座子的转座插入时间峰值在 2.5 MYA 左右,在这之后就迅速下降。而在大豆基因组中逆转座子随时间的接近数量不断增加,转座插入时间峰值在 0~0.5 MYA (Du et al., 2010),玉米中 LTR 逆转座子也表现出相似分布模式(Baucom et al., 2009),说明番茄中大多数逆转座元件相对古老,发生突变和结构变化的比例会更高。

染色体上分布大体均匀,数量基本与染色体序列长度成正比,表明对于不同染色体来讲,逆转座子的转座并没有选择性。

2.2 番茄中逆转座子的转座时间及种间差异比较

逆转座子的转座机制决定了它们在转座后两端的 LTR 序列是完全一致的,因此比较完整转座子两侧 LTR 序列就可以依据其碱基突变率计算其年龄。同时由于逆转座子的转座插入是随机的过程并且转座事件在物种进化的各个时期都在不断进行,因而可以为研究物种及其基因组的分化和演化提供丰富的信息。由地理隔离发展到生殖隔离,即种群间不存在基因交流是物种形成的条件和标志,许多作物栽培种群的形成是人为阻断基因交流并产生生殖隔离的过程。多位点突变率计算表明栽培番茄与醋栗番茄在 1.4 MYA 发生分化(Nesbitt and Tanksley, 2002),同时研究表也明栽培番茄的起源和驯化伴随着野生番茄基因的不断渗入,而樱桃番茄 *S. lycopersicum* var. *cerasiforme* (*S. l. cerasiforme*)被



认为是野生番茄向栽培番茄发展的一个过渡种(Nesbitt and Tanksley, 2002; Ranc et al., 2008)。逆转座子的copy-paste 转座模式决定了它的每一个转座事件都会在基因组进化中留下特异的标记,又由于可以借助两端LTR 的差异解析其自身存在的时间,因此含有大量的进化信息。我们分析栽培番茄和醋栗番茄之间LTR逆转座子的等位性分布发现在大部分时间段内都存在很大比例在两个基因组中分布相同(shared)的LTR逆转座元件,表明确实野生番茄和栽培番茄的基因组序列存在遗传物质交流。在物种分化后期基因组必然是独立演化的,因此这个时间以后发生的逆转座子的复制和删除应该是物种特异的,而在这个时间点之前发生的转座子复制在两个物种中应该大部分保持相同的分布。栽培和野生醋栗番茄中0.25 MYA 以上的时间段内相同分布的逆转座元件比例均在50% 以上,但是在0.25 MYA 以内双方共有的LTR 元件大大低于其他时间段(约20%),表明在近期栽培种与野生种产生了较明显的分化,是栽培番茄物种形成的阶段,明显晚于已知的分化时间。

2.3 逆转座子对功能基因进化的影响

逆转座子的插入位点序列是随机的,而且在真核生物特别是高等植物中存在大量转座事件,因此从理论上来说逆转座子可以插入基因附近影响其功能,甚至进入基因内部破坏其结构。很多研究已经表明逆转座子可以影响基因的表达,甚至通过自身含有的转座元件直接调控基因表达(Roman et al., 2008)。基因的启动子包括核心启动子区域和调控区域,总长度可以在2 000 bp以上。我们发现930个元件位于基因上游1 000 bp 以内或者位于基因内部,这些元件在转座过程中很可能会影响基因的表达调控及生物学功能。除直接插入到基因内部的元件以外,还有22个元件位于基因上游35 bp以内,可以肯定这些元件在转座过程中至少改变了原始基因的结构或表达水平。

3 材料和方法

3.1 材料

栽培番茄(*Solanum lycopersicum*)和野生醋栗番茄(*Solanum pimpinellifolium*)全基因组序列 *S_lycopersicum_chromosomes.2.40* 和 *S_lycopersicum_chromosomes.2.40.fa* 来源于 <ftp://ftp.sgn.cornell.edu/>。

3.2 方法

3.2.1 LTR-逆转座子的鉴定

用结构分析和同源比对的方法鉴定反转座子元件。首先用 LTR_STRUC 程序(Mccarthy and Mcdonald, 2003)寻找完整结构的逆转座子,然后利用其 LTR 序列进行同源比对,并应用 perl 语言编程从中寻找非完整元件(truncated elements)和 solo 元件(Ma et al., 2004)。

3.2.2 逆转座子的分类

逆转座子家族的分类方法依据前人的标准修改后进行(Wicker et al., 2007)。对完整逆转座子的 LTR 序列用 clustalW 程序进行聚类分析,计算各个元件的遗传距离,差异小于 0.2 的转座元件划分为同一家族。其它同源分析方法得到的转座元件依据其匹配的相应完整元件确定所属家族。转座子家族种类的确定(*copia* 或者 *gypsy*)利用最近时间段内发生转座插入的元件,用 NCBI 的 RPS-BLAST 程序分析是否存在与 *copia* 或 *gypsy* 转座子同源的反转录酶蛋白序列(Marchler-Bauer et al., 2009; Marchler-Bauer and Bryant, 2004; Marchler-Bauer et al., 2011; Xiong and Eickbush, 1990)。

3.2.3 转座插入时间鉴定

通过比较完整转座子两端的 LTR 序列来确定转座子的插入时间。首先用 MUSCLE 程序比对转座子两端的 LTR 序列(Edgar, 2004),计算比对碱基的总数和碱基突变的数目,计算得到碱基替换率(r)。遗传距离(K)用 Jukes-Cantor 方法(Kimura and Ohta, 1972)进行修正。以平均每年每个位点 1.3×10^{-8} 碱基替换数作为计算转座插入年龄的依据(Ma and Jackson, 2006)。最后转座子的插入时间(T)用公式 $T = K/2r$ 计算得到。

3.2.4 栽培番茄逆转座子在野生醋栗番茄的等位鉴定

依据已有的方法完成(Tian et al., 2009)。完整转座元件的 5'和 3'端侧翼序列(25 bp)结合转座子本身边界序列(25 bp)用 cross_match 在野生醋栗番茄中做序列比对,5'或者 3'结合序列在野生番茄基因组中有良好匹配结果的认为该元件在两个物种中是都存在的(shared),如果 5'和 3'序列均不存在匹配结果,则该元件在栽培番茄中是特异的(unshared)。

作者贡献

许莹修是本研究的构思者及负责人,实验设计,数据

分析, 论文写作与修改。

致谢

本研究得到江苏省农业科技自主创新基金(CX(12)5037)资助。

参考文献

- Baucom R.S., Estill J.C., Chaparro C., Upshaw N., Jogi A., Deragon J.M., Westerman R.P., Sanmiguel P.J., and Bennetzen J.L., 2009, Exceptional diversity, non-random distribution, and rapid evolution of retroelements in the B73 maize genome, *PLoS Genet*, 5(11): e1000732
<http://dx.doi.org/10.1371/journal.pgen.1000732>
 PMid:19936065 PMCID:PMC2774510
- Cheng X.D., Zhang D.F., Cheng Z.K., Keller B., and Ling H.Q., 2009, A new family of Ty1-copia-like retrotransposons originated in the tomato genome by a recent horizontal transfer event, *Genetics*, 181(4): 1183-1193
<http://dx.doi.org/10.1534/genetics.108.099150>
 PMid:19153256 PMCID:PMC2666490
- Du J.C., Tian Z.X., Hans C.S., Laten H.M., Cannon S.B., Jackson S.A., Shoemaker R.C., and Ma J.X., 2010, Evolutionary conservation, diversity and specificity of LTR-retrotransposons in flowering plants: insights from genome-wide analysis and multi-specific comparison, *Plant J.*, 63(4): 584-598
<http://dx.doi.org/10.1111/j.1365-313X.2010.04263.x>
 PMid:20525006
- Edgar R.C., 2004, MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Res.*, 32(5): 1792-1797
<http://dx.doi.org/10.1093/nar/gkh340>
 PMid:15034147 PMCID:PMC390337
- Finnegan D.J., 1989, Eukaryotic transposable elements and genome evolution, *Trends Genet.*, 5(4): 103-107
[http://dx.doi.org/10.1016/0168-9525\(89\)90039-5](http://dx.doi.org/10.1016/0168-9525(89)90039-5)
- Girard L., and Freeling M., 1999, Regulatory changes as a consequence of transposon insertion, *Dev. Genet.*, 25(4): 291-296
[http://dx.doi.org/10.1002/\(SICI\)1520-6408\(1999\)25:4<291::AID-DVG2>3.0.CO;2-5](http://dx.doi.org/10.1002/(SICI)1520-6408(1999)25:4<291::AID-DVG2>3.0.CO;2-5)
- Goulet C., Mageroy M.H., Lam N.B., Floystad A., Tieman D.M., and Klee H.J., 2012, Role of an esterase in flavor volatile variation within the tomato clade, *Proc. Natl. Acad. Sci., USA*, 109(46): 19009-19014
<http://dx.doi.org/10.1073/pnas.1216515109>
 PMid:23112200 PMCID:PMC3503167
- Havecker E.R., Gao X., and Voytas D.F., 2004, The diversity of LTR retrotransposons, *Genome Biol.*, 5(6): 225
<http://dx.doi.org/10.1186/gb-2004-5-6-225>
 PMid:15186483 PMCID:PMC463057
- Hua-Van A., Le Rouzic A., Boutin T.S., Filée J., and Capy P., 2011, The struggle for life of the genome's selfish architects, *Biol. Direct*, 6: 19
<http://dx.doi.org/10.1186/1745-6150-6-19>
 PMid:21414203 PMCID:PMC3072357
- International Rice Genome Sequencing Project, 2005, The map-based sequence of the rice genome, *Nature*, 436(7052): 793-800
<http://dx.doi.org/10.1038/nature03895>
 PMid:16100779
- Jiang N., Gao D., Xiao H., and van der Knaap E., 2009, Genome organization of the tomato sun locus and characterization of the unusual retrotransposon rider, *Plant J.*, 60(1): 181-193
<http://dx.doi.org/10.1111/j.1365-313X.2009.03946.x>
 PMid:19508380
- Kimura M., and Ohta T., 1972, On the stochastic model for estimation of mutational distance between homologous proteins, *J. Mol. Evol.*, 2(1): 87-90
<http://dx.doi.org/10.1007/BF01653945>
 PMid:4668865
- Kumar A., and Bennetzen J.L., 1999, Plant Retrotransposons, *Annu. Rev. Genet.*, 33: 479-532
<http://dx.doi.org/10.1146/annurev.genet.33.1.479>
 PMid:10690416
- Ma J., Devos K.M., and Bennetzen J.L., 2004, Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice, *Genome Res.*, 14(5): 860-869
<http://dx.doi.org/10.1101/gr.1466204>
 PMid:15078861 PMCID:PMC479113
- Ma J., and Jackson S.A., 2006, Retrotransposon accumulation and satellite amplification mediated by segmental duplication facilitate centromere expansion in rice, *Genome Res.*, 16(2): 251-259
<http://dx.doi.org/10.1101/gr.4583106>
 PMid:16354755 PMCID:PMC1361721



- Marchler-Bauer A., Anderson J.B., Chitsaz F., Derbyshire M.K., DeWeese-Scott C., Fong J.H., Geer L.Y., Geer R.C., Gonzales N.R., Gwadz M., He S., Hurwitz D.I., Jackson J.D., Ke Z., Lanczycki C.J., Liebert C.A., Liu C., Lu F., Lu S., Marchler G.H., Mullokandov M., Song J.S., Tasneem A., Thanki N., Yamashita R.A., Zhang D., Zhang N., and Bryant S.H., 2009, CDD: specific functional annotation with the conserved domain database, *Nucleic Acids Res.*, 37(Database issue): D205- D210
<http://dx.doi.org/10.1093/nar/gkn845>
PMid:18984618 PMCID:PMC2686570
- Marchler-Bauer A., and Bryant S.H., 2004, CD-Search: protein domain annotations on the fly, *Nucleic Acids Res.*, 32(Web Server issue): W327- W331
- Marchler-Bauer A., Lu S., Anderson J.B., Chitsaz F., Derbyshire M.K., DeWeese-Scott C., Fong J.H., Geer L.Y., Geer R.C., Gonzales N.R., Gwadz M., Hurwitz D.I., Jackson J.D., Ke Z., Lanczycki C.J., Lu F., Marchler G.H., Mullokandov M., Omelchenko M.V., Robertson C.L., Song J.S., Thanki N., Yamashita R.A., Zhang D., Zhang N., Zheng C., and Bryant S.H., 2011, CDD: a conserved domain database for the functional annotation of proteins, *Nucleic Acids Res.*, 39(Database issue): D225- D229
<http://dx.doi.org/10.1093/nar/gkq1189>
PMid:21109532 PMCID:PMC3013737
- Mccarthy E.M., and Mcdonald J.F., 2003, LTR_STRUC: a novel search and identification program for LTR retrotransposons, *Bioinformatics*, 19(3): 362-367
<http://dx.doi.org/10.1093/bioinformatics/btf878>
PMid:12584121
- Morgante M., De Paoli E., and Radovic S., 2007, Transposable elements and the plant pan-genomes, *Curr. Opin. Plant Biol.*, 10(2): 149-155
<http://dx.doi.org/10.1016/j.pbi.2007.02.001>
PMid:17300983
- Nesbitt T.C., and Tanksley S.D., 2002, Comparative sequencing in the genus *lycopersicon*: implications for the evolution of fruit size in the domestication of cultivated tomatoes, *Genetics*, 162(1): 365-379
PMid:12242247 PMCID:PMC1462239
- Ranc N., Mu-os S., Santoni S., and Causse M., 2008, A clarified position for *solanum lycopersicum* var. *cerasiforme* in the evolutionary history of tomatoes (solanaceae), *BMC Plant Biology*, 8: 130
<http://dx.doi.org/10.1186/1471-2229-8-130>
PMid:19099601 PMCID:PMC2657798
- Rogers S.A., and Pauls K.P., 2000, Ty1-copia-like retrotransposons of tomato (*Lycopersicon esculentum* Mill.), *Genome*, 43(5): 887-894
<http://dx.doi.org/10.1139/g00-056>
<http://dx.doi.org/10.1139/gen-43-5-887>
PMid:11081980
- Roman A.C., Benitez D.A., Carvajal-Gonzalez J.M., and Fernandez-Salguero P.M., 2008, Genome-wide B1 retrotransposon binds the transcription factors dioxin receptor and Slug and regulates gene expression in vivo, *Proc. Natl. Acad. Sci., USA*, 105(5): 1632-1637
<http://dx.doi.org/10.1073/pnas.0708366105>
PMid:18223155 PMCID:PMC2234196
- Sabot F., and Schulman A.H., 2006, Parasitism and the retrotransposon life cycle in plants: a hitchhiker's guide to the genome, *Heredity*, 97(6): 381-388
<http://dx.doi.org/10.1038/sj.hdy.6800903>
PMid:16985508
- SanMiguel P., Gaut B.S., Tikhonov A., Nakajima Y., and Bennetzen J.L., 1998, The paleontology of intergene retrotransposons of maize, *Nat. Genet.*, 20(1): 43-45
<http://dx.doi.org/10.1038/1695>
PMid:9731528
- SanMiguel P., Tikhonov A., Jin Y.K., Motchoulskaia N., Zakharov D., Melake-Berhan A., Springer P.S., Edwards K.J., Lee M., Avramova Z., and Bennetzen J.L., 1996, Nested retrotransposons in the intergenic regions of the maize genome, *Science*, 274(5288): 765-768
<http://dx.doi.org/10.1126/science.274.5288.765>
PMid:8864112
- Schnable P.S., Ware D., Fulton R.S., Stein J.C., Wei F.S., Pasternak S., Liang C.Z., Zhang J.W., Fulton L., Graves T.A., Minx P., Reily A.D., Courtney L., Kruchowski S.S., Tomlinson C., Strong C., Delehaunty K., Fronick C., Courtney B., Rock S.M., Belter E., Du F.Y., Kim K., Abbott R.M., Cotton M., Levy A., Marchetto P., Ochoa K., Jackson S.M., Gillam B., Chen W.Z., Yan L., Higginbotham J., Cardenas M., Waligorski J., Applebaum E., Phelps L., Falcone J., Kanchi K., Thane T., Scimone A., Thane N., Henke J., Wang T., Ruppert J., Shah N.,

- Rotter K., Hodges J., Ingenthron E., Cordes M., Kohlberg S., Sgro J., Delgado B., Mead K., Chinwalla A., Leonard S., Crouse K., Collura K., Kudrna D., Currie J., He R.F., Angelova A., Rajasekar S., Mueller T., Lomeli R., Scara G., Ko A., Delaney K., Wissotski M., Lopez G., Campos D., Braidotti M., Ashley E., Golser W., Kim H., Lee S., Lin J., Dujmic Z., Kim W., Talag J., Zuccolo A., Fan C.Z., Sebastian A., Kramer M., Spiegel L., Nascimento L., Zutavern T., Miller B., Ambroise C., Muller S., Spooner W., Narechania A., Ren L., Wei S., Kumari S., Faga B., Levy M.J., McMahan L., Van Buren P., Vaughn M.W., Ying K., Yeh C.T., Emrich S.J., Jia Y., Kalyanaraman A., Hsia A.P., Brad Barbazuk W., Baucom R.S., Brutnell T.P., Carpita N.C., Chaparro C., Chia J., Deragon J., Estill J.C., Fu Y., Jeddeloh J.A., Han Y.J., Lee H., Li P.H., Lisch D.R., Liu S.Z., Liu Z.J., Nagel D.H., McCann M.C., SanMiguel P., Myers A.M., Nettleton D., Nguyen J., Penning B.W., Ponnala L., Schneider K.L., Schwartz D.C., Sharma A., Soderlund C., Springer N.M., Sun Q., Wang H., Waterman M., Westerman R., Wolfgruber T.K., Yang L.X., Yu Y., Zhang L.F., Zhou S.G., Zhu Q.H., Bennetzen J.L., Kelly Dawe R., Jiang J.M., Jiang N., Presting G.G., Wessler S.R., Aluru S., Martienssen R.A., Clifton S.W., Richard McCombie W., Wing R.A., and Wilson R.K., 2009, The B73 maize genome: complexity, diversity, and dynamics, *Science*, 326(5956): 1112-1115
<http://dx.doi.org/10.1126/science.1178534>
 PMid:19965430
- Su P.Y., and Brown T.A., 1997, Ty3/gypsy-like retrotransposon sequences in tomato, *Plasmid*, 38(3): 148-157
<http://dx.doi.org/10.1006/plas.1997.1310>
 PMid:9435017
- Tam S.M., Mhiri C., Vogelaar A., Kerkveld M., Pearce S.R., and Grandbastien M.A., 2005, Comparative analyses of genetic diversities within tomato and pepper collections detected by retrotransposon-based SSAP, AFLP and SSR, *Theor. Appl. Genet.*, 110(5): 819-831
<http://dx.doi.org/10.1007/s00122-004-1837-z>
 PMid:15700147
- The Tomato Genome Consortium, 2012, The tomato genome sequence provides insights into fleshy fruit evolution, *Nature*, 485(7400): 635-641
<http://dx.doi.org/10.1038/nature11119>
 PMid:22660326 PMCID:PMC3378239
- Tian Z., Rizzon C., Du J., Zhu L., Bennetzen J.L., Jackson S.A., Gaut B.S., and Ma J., 2009, Do genetic recombination and gene density shape the pattern of DNA elimination in rice long terminal repeat retrotransposons? *Genome Res.*, 19(12): 2221-2230
<http://dx.doi.org/10.1101/gr.083899.108>
 PMid:19789376 PMCID:PMC2792168
- Wicker T., Sabot F., Hua-Van A., Bennetzen J.L., Capy P., Chalhou B., Flavell A., Leroy P., Morgante M., Panaud O., Paux E., SanMiguel P., and Schulman A.H., 2007, A unified classification system for eukaryotic transposable elements, *Nat. Rev. Genet.*, 8(12): 973-982
<http://dx.doi.org/10.1038/nrg2165>
 PMid:17984973
- Xiao H., Jiang N., Schaffner E., Stockinger E.J., and van der Knaap E., 2008, A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit, *Science*, 319(5869): 1527-1530
<http://dx.doi.org/10.1126/science.1153040>
 PMid:18339939
- Xiong Y., and Eickbush T.H., 1990, Origin and evolution of retroelements based upon their reverse transcriptase sequences, *EMBO J.*, 9(10): 3353-3362
 PMid:1698615 PMCID:PMC552073
- Ya-ez M., Verdugo I., Rodríguez M., Prat S., Ruiz-Lara S., 1998, Highly heterogeneous families of Ty1/copia retrotransposons in the *Lycopersicon chilense* genome, *Gene*, 222(2): 223-228
[http://dx.doi.org/10.1016/S0378-1119\(98\)00486-7](http://dx.doi.org/10.1016/S0378-1119(98)00486-7)