

## 瓠瓜 LINE 逆转座子 RT 序列的克隆与特征分析

赵芹<sup>①,②</sup>, 谢大森<sup>①</sup>, 何晓明, 罗少波, 彭庆务, 陈俊秋

(广东省农业科学院蔬菜研究所, 广东 广州 510640)

**摘要:** 以瓠瓜 [*Lagenaria siceraria* (Molina) Standl.] 品种‘大籽葫芦’(‘Dazihulu’)和‘杂交瓠瓜’(‘Hybrid bottle gourd’)为供试材料,对其基因组总 DNA 中的 LINE 逆转座子 RT 序列进行扩增,并对 30 个克隆产物的核苷酸序列及其编码的氨基酸序列进行比对、同源性分析和系统进化分析。结果表明:瓠瓜品种‘大籽葫芦’与‘杂交瓠瓜’的基因组总 DNA 均包含长度约 580 bp 的 RT 序列片段。从 2 个瓠瓜品种中获得的 30 条 LINE 逆转座子 RT 核苷酸序列(编号为 LsRT1 至 LsRT30)长度为 564~599 bp, 碱基 A、T、G 和 C 的数量分别为 143~193、157~205、104~139 和 83~134, AT/GC 比为 1.29~1.76, 表现出高度异质性。缺失突变和终止密码子突变是造成瓠瓜 LINE 逆转座子 RT 核苷酸序列长度差异的主要因素。30 条瓠瓜 LINE 逆转座子 RT 核苷酸序列的相似性为 47.1%~99.5%, 其编码的氨基酸序列相似性为 26.7%~100.0%。根据核苷酸替代值, 30 条瓠瓜 LINE 逆转座子 RT 核苷酸序列可分为 4 个家族(family), 分别包含 14、8、1 和 7 条序列。氨基酸序列分析结果显示:瓠瓜 LINE 逆转座子 RT 氨基酸序列包含 20 个保守氨基酸残基和多个半保守氨基酸残基; 有 14 条氨基酸序列具有终止密码子突变。Family 1、Family 2 和 Family 4 是可能具有转座活性的逆转座子家族, 分别包含 8、3 和 5 条无终止密码子的 RT 氨基酸序列。根据瓠瓜与其他 15 种植物的 LINE 逆转座子 RT 氨基酸序列构建的系统进化树, 瓠瓜与葡萄(*Vitis vinifera* Linn.) 和黄瓜(*Cucumis sativus* Linn.) 等种类的 LINE 逆转座子 RT 氨基酸序列有较高同源性。研究结果表明:瓠瓜 LINE 逆转座子是一类较古老元件, LINE 逆转座子可在瓠瓜与其他种类的基因组间横向传递。

**关键词:** 瓠瓜; LINE 逆转座子; RT 序列; 序列分析; 相似性; 系统进化

中图分类号: Q785; S642.9 文献标志码: A 文章编号: 1674-7895(2015)01-0001-11

DOI: 10.3969/j.issn.1674-7895.2015.01.01

**Cloning and characteristic analysis on RT sequence of LINE retrotransposon from *Lagenaria siceraria*** ZHAO Qin<sup>①,②</sup>, XIE Dasen<sup>①</sup>, HE Xiaoming, LUO Shaobo, PENG Qingwu, CHEN Junqiu (Vegetable Research Institute, Guangdong Academy of Agricultural Sciences, Guangzhou 510640, China), *J. Plant Resour. & Environ.*, 2015, 24(1): 1-11

**Abstract:** Taking cultivars ‘Dazihulu’ and ‘Hybrid bottle gourd’ of *Lagenaria siceraria* (Molina) Standl. as tested materials, RT sequence of LINE retrotransposon in their genomic total DNA was amplified, and alignment, homology analysis and phylogeny analysis of nucleotide sequence of 30 cloning products and their amino acid sequences encoded were carried out. The results show that genomic total DNA of both ‘Dazihulu’ and ‘Hybrid bottle gourd’ includes RT sequence fragment with length about 580 bp. Length of 30 RT nucleotide sequences (Nos. from LsRT1 to LsRT30) of LINE retrotransposon obtained from the two cultivars is 564-599 bp, base number of A, T, G and C is 143-193, 157-205, 104-139 and 83-134, respectively and AT/GC ratio is 1.29-1.76, which shows high heterogeneity. The main factors inducing difference in length of RT nucleotide sequence of LINE retrotransposon from *L. siceraria* are deletion mutation and stop codon mutation. Similarity of 30 RT nucleotide sequences of LINE

收稿日期: 2014-06-24

基金项目: 国家自然科学基金资助项目(31311643); 广东省自然科学基金博士科研启动基金资助项目(10451064001006063); 广州市珠江科技新星专项资助项目(2013086)

作者简介: 赵芹(1982—), 女, 山东泰安人, 博士, 副研究员, 主要从事瓜类遗传育种与分子生物技术方面的研究。

谢大森(1970—), 男, 江西遂川人, 博士, 研究员, 主要从事瓜类抗病育种与生物技术方面的研究。

①并列第一作者

②通信作者 E-mail: zhaqin0802@126.com

retrotransposon from *L. siceraria* is 47.1%–99.5%, and that of amino acid sequence encoded is 26.7%–100.0%. Based on nucleotide substitution, 30 RT nucleotide sequences of LINE retrotransposon from *L. siceraria* can be divided into four families, which contains 14, 8, 1 and 7 sequences, respectively. The analysis result on amino acid sequence shows that RT amino acid sequence of LINE retrotransposon from *L. siceraria* contains 20 conservative amino acid residues and many semi-conservative amino acid residues, and 14 amino acid sequences possess stop codon mutation. Family 1, Family 2 and Family 4 may be the retrotransposon families with transposition activity, and include 8, 3 and 5 RT amino acid sequences without stop codon, respectively. According to phylogenetic tree constructed on RT amino acid sequence of LINE retrotransposon from *L. siceraria* and other 15 species, there is higher homology in RT amino acid sequence of LINE retrotransposon from *L. siceraria* with those from *Vitis vinifera* Linn., *Cucumis sativus* Linn. and other species. It is suggested that LINE retrotransposon from *L. siceraria* is a more ancient element, and can be horizontally transferred among genome of *L. siceraria* and other species.

**Key words:** *Lagenaria siceraria* (Molina) Standl.; LINE retrotransposon; RT sequence; sequence analysis; similarity; phylogeny

瓠瓜(*Lagenaria siceraria* (Molina) Standl.)为葫芦科(Cucurbitaceae)1年生攀缘草本植物,主要分布于中国长江流域及长江以南地区,品种十分丰富。瓠瓜易栽培且产量高,嫩瓜柔软多汁且风味可口,是华南地区蔬菜淡季的重要蔬菜<sup>[1-2]</sup>。根据形态和基因,非洲和亚洲的瓠瓜被分为2个亚种<sup>[3-4]</sup>。目前对瓠瓜种质资源研究较少,仅见利用分子标记初步分析部分瓠瓜种质遗传多样性的研究报告<sup>[5-8]</sup>,但对瓠瓜不同种质的形成及进化关系仍不甚明了。另外,瓠瓜花色、花型、瓜型及抗性等性状差异很大,但这些性状的形成机制尚不明确。

逆转座子在植物中分布广泛,种类多、拷贝数量高、插入位点专一;逆转座子以RNA为中间产物,反转录为染色体DNA,在寄主基因组内不断转座增殖,影响寄主基因组的大小、结构、功能和进化<sup>[9]</sup>。逆转座子可用于基因组的组成、表达调控和功能进化分析,以及遗传连锁的分子标记和生物多样性评价及体细胞无性系变异评估等方面<sup>[10-12]</sup>,是当前研究的热点之一。外界因素易激活逆转座子,进而导致逆转座子发生基因突变、基因组扩增及重排等,并最终引起植物的遗传变异<sup>[13-16]</sup>,有关葡萄(*Vitis vinifera* Linn.)果皮颜色及菜豆(*Phaseolus vulgaris* Linn.)花色形成的研究<sup>[17-18]</sup>即证实了这一点。因此,研究瓠瓜的特异生物学性状及遗传进化与逆转座子的关系具有十分重要的意义。

逆转座子包括长末端重复序列(long terminal repeat, LTR)逆转座子和非长末端重复序列(non-long terminal repeat, non-LTR)逆转座子,其中LINEs(long

interspersed nuclear elements)逆转座子为非长末端重复序列逆转座子之一<sup>[19-20]</sup>。迄今non-LTR类逆转座子已被广泛应用于多种植物的多个研究领域<sup>[11,21-23]</sup>,但目前尚未见瓠瓜LINE逆转座子RT序列分离及转录活性研究的报道。本文对瓠瓜LINE逆转座子RT序列及其相互关系进行研究,以为瓠瓜种质资源的分子鉴定、种质评价与遗传进化研究奠定基础。

## 1 材料和方法

### 1.1 材料

供试材料为瓠瓜品种‘大籽葫芦’(‘Dazihulu’)和‘杂交瓠瓜’(‘Hybrid bottle gourd’),均由广东省农业科学院蔬菜研究所瓜类研究一室保存。将2个瓠瓜品种的种子播种于营养盘内并置于温室中培养,待幼苗长至2片真叶时采集叶片供试。

### 1.2 方法

1.2.1 DNA提取及逆转录酶序列PCR扩增 采用改良CTAB法<sup>[8,24]</sup>提取瓠瓜基因组总DNA,用质量体积分数1.0%琼脂糖凝胶电泳及GENEQUANT纳米紫外分光光度计(德国Eppendorf公司)检测DNA浓度及纯度。

参照Hill等<sup>[21]</sup>的方法合成扩增LINE逆转座子RT序列所需的引物。上游引物LINE-F的序列为5'-GGGATCCNGCNGCCNGAYGGNWT-3',下游引物LINE-R的序列为5'-SWNARNGGRTCNCCTG-3',其中N为A/C/G/T, Y为C/T, W为A/T, S为C/G, R为A/G。PCR反应体系包含50 ng DNA、2.5 μL

10×*Ex* buffer(含  $MgCl_2$ )、 $1.0 \mu mol \cdot L^{-1}$  上游和下游引物、 $0.20 mmol \cdot L^{-1}$  dNTPs、1 U *Ex Taq* DNA 聚合酶,用灭菌双蒸水补足至  $25 \mu L$ 。扩增程序为:95 °C 预变性 3 min;95 °C 变性 1 min,50 °C 退火 1 min,72 °C 延伸 1 min,35 个循环;最后于 72 °C 延伸 10 min。用 GeneAmp PCR System 9700 型扩增仪(美国 ABI 公司)进行 PCR 扩增。扩增产物用质量体积分数 1.2% 琼脂糖凝胶电泳检测,并用 GeneGeniusBio Imaging System 凝胶成像系统(美国 Bio-Rad 公司)观察和拍照。

1.2.2 PCR 产物克隆及序列测定 回收纯化 PCR 产物,与 pMD19-T Vector 连接,转化大肠杆菌 DH5 $\alpha$  感受态细胞并培养过夜;用 Amp 抗生素与 IPTG 及底物 X-gal 进行蓝白斑筛选;挑取白斑,利用 M13 通用引物进行 PCR 扩增并鉴定出阳性克隆,送至广州英骏生物技术有限公司测序。

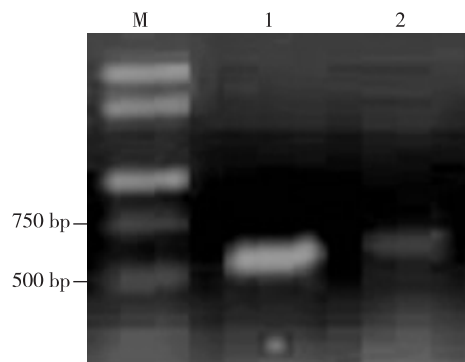
1.2.3 逆转录酶序列比对及系统进化树分析 利用 NCBI VecScreen 程序去除序列的载体部分,并用 NCBI BLASTx 程序同源比对推导获得 RT 片段序列及氨基酸序列;应用 DNASTar 软件进行序列长度、碱基组成、AT/GC 比、氨基酸序列和多重比对等分析;结合 GenBank 报道的其他植物的 RT 序列构建系统进化树。

## 2 结果和分析

### 2.1 瓠瓜 LINE 逆转座子 RT 核苷酸序列的扩增与序列分析

利用简并引物从瓠瓜品种‘大籽葫芦’与‘杂交瓠瓜’的基因组总 DNA 中均获得长度约 580 bp 的特异片段(图 1),预期的片段长度与甜菜(*Beta vulgaris* Linn.)<sup>[25]</sup> 和牡丹(*Paeonia suffruticosa* Andr.)<sup>[26]</sup> 的 LINE 逆转座子 RT 序列基本一致,说明该类逆转座子在瓠瓜中普遍存在。

将 PCR 产物回收克隆,随机挑取 32 个阳性克隆进行测序,去除 1 条重复序列。用 NCBI BLASTx 程序进行同源搜索,结果显示:有 30 条序列与已报道的其他种类的 non-LTR 逆转座子 RT 核苷酸序列具有较高的同源性,且均含有 RT 保守结构域,其中 11 条来自品种‘大籽葫芦’、19 条来自品种‘杂交瓠瓜’。说明这些来源于瓠瓜的克隆均为 LINE 逆转座子 RT 序列,依次命名为 LsRT1 至 LsRT30。



M: DL2000 DNA marker; 1: 品种‘大籽葫芦’ Cultivar ‘Dazihulu’; 2: 品种‘杂交瓠瓜’ Cultivar ‘Hybrid bottle gourd’.

图 1 瓠瓜 LINE 逆转座子 RT 核苷酸序列的 PCR 扩增图谱  
Fig. 1 PCR amplification pattern of RT nucleotide sequence of LINE retrotransposon from *Lagenaria siceraria* (Molina) Standl.

### 2.2 瓠瓜 LINE 逆转座子 RT 核苷酸序列的相似性分析

DNASTar 软件分析结果表明(表 1):30 条瓠瓜 LINE 逆转座子 RT 核苷酸序列的长度并不完全一致,变化范围为 564 ~ 599 bp,其中, LsRT10 序列最长、LsRT2 序列最短,其余序列长度为 577 ~ 591 bp。碱基分析结果显示:30 条 RT 核苷酸序列均富含碱基 A 和 T,AT/GC 比为 1.29 ~ 1.76,A、T、G 和 C 的数量分别为 143 ~ 193、157 ~ 205、104 ~ 139 和 83 ~ 134。

30 条瓠瓜 LINE 逆转座子 RT 核苷酸序列的相似性为 47.1% ~ 99.5%(表 2),其中 LsRT3 与 LsRT4 的核苷酸序列相似性最高、LsRT10 与 LsRT26 的核苷酸序列相似性最低(表 2,图 2)。由此可见,同一引物扩增所得的 LINE 逆转座子 RT 核苷酸序列并不一致,在序列长度、碱基组成及序列相似性等方面存在较大差异,反映出瓠瓜同一类群逆转座子的高度异质性。甜菜和牡丹等其他植物也有类似特点<sup>[25-26]</sup>。

### 2.3 瓠瓜 LINE 逆转座子 RT 核苷酸序列的系统进化分析

为阐明瓠瓜 LINE 逆转座子 RT 核苷酸序列间的相互关系,利用 MegAlign 程序对克隆获得的瓠瓜 LINE 逆转座子 RT 核苷酸序列构建系统进化树(图 3)。根据核苷酸替代值,30 条 RT 核苷酸序列可分为 4 个家族(Family 1、Family 2、Family 3 和 Family 4),其中 LsRT25 核苷酸序列单独列为 1 个家族(Family 3),与其他家族的遗传距离较远;Family 1、Family 2 和 Family 4 分别包含 14、8 和 7 条序列,分别占序列总数

表1 瓠瓜 LINE 逆转座子 RT 核苷酸序列的长度和组成比较

Table 1 Comparison on length and composition of RT nucleotide sequence of LINE retrotransposon from *Lagenaria siceraria* (Molina) Standl.

序列 Sequence	长度/bp Length	AT/GC 比 AT/GC ratio	碱基数量 Base number				序列 Sequence	长度/bp Length	AT/GC 比 AT/GC ratio	碱基数量 Base number			
			A	T	G	C				A	T	G	C
LsRT1	582	1.51	193	157	111	121	LsRT16	589	1.43	170	177	138	104
LsRT2	564	1.47	143	193	125	103	LsRT17	590	1.43	169	178	137	106
LsRT3	589	1.43	168	179	139	103	LsRT18	584	1.60	180	179	111	114
LsRT4	590	1.44	169	179	138	104	LsRT19	580	1.60	187	170	136	87
LsRT5	580	1.57	186	168	137	89	LsRT20	580	1.57	187	167	138	88
LsRT6	581	1.45	146	198	131	106	LsRT21	580	1.43	147	194	129	110
LsRT7	580	1.36	176	158	112	134	LsRT22	580	1.42	147	193	131	109
LsRT8	590	1.41	168	177	139	106	LsRT23	580	1.61	190	168	134	88
LsRT9	580	1.76	192	178	104	106	LsRT24	580	1.60	189	168	136	87
LsRT10	599	1.43	148	205	135	111	LsRT25	577	1.37	172	162	136	107
LsRT11	588	1.42	166	179	139	104	LsRT26	591	1.47	170	182	139	100
LsRT12	580	1.42	147	193	133	107	LsRT27	581	1.57	189	166	137	89
LsRT13	580	1.44	149	193	130	108	LsRT28	586	1.76	186	188	129	83
LsRT14	580	1.46	174	170	112	124	LsRT29	580	1.59	189	167	136	88
LsRT15	580	1.29	146	181	127	126	LsRT30	580	1.45	147	196	131	106

表2 瓠瓜 LINE 逆转座子 RT 核苷酸序列及氨基酸序列的相似性<sup>1)</sup>Table 2 Similarity of RT nucleotide sequence and amino acid sequence of LINE retrotransposon from *Lagenaria siceraria* (Molina) Standl.<sup>1)</sup>

序列 Sequence	各序列间的相似性/% Similarity among different sequences														
	LsRT1	LsRT2	LsRT3	LsRT4	LsRT5	LsRT6	LsRT7	LsRT8	LsRT9	LsRT10	LsRT11	LsRT12	LsRT13	LsRT14	LsRT15
LsRT1	—	54.5	41.4	41.1	55.2	54.4	76.6	40.6	82.3	48.2	34.6	53.1	55.2	69.8	55.7
LsRT2	58.3	—	38.7	38.7	52.4	95.7	53.5	38.7	54.5	89.3	30.1	94.1	96.3	51.9	81.3
LsRT3	51.9	50.4	—	100.0	41.9	39.8	41.9	98.5	41.9	35.1	76.9	39.3	40.3	39.3	40.8
LsRT4	52.0	50.4	99.5	—	41.9	39.2	41.9	98.0	41.9	34.9	76.9	39.3	40.3	39.3	40.8
LsRT5	58.1	57.4	49.5	49.1	—	53.2	54.7	41.9	53.1	46.9	37.7	51.6	53.6	53.6	53.1
LsRT6	58.9	97.5	51.7	51.6	57.9	—	54.7	39.1	55.7	87.6	32.9	94.8	97.4	52.6	81.8
LsRT7	69.0	59.0	52.2	52.2	55.5	60.2	—	41.4	75.0	49.5	34.0	53.1	55.2	67.7	57.3
LsRT8	51.1	49.8	98.6	98.5	49.0	51.5	51.4	—	41.4	34.9	77.4	38.7	39.8	38.7	40.3
LsRT9	75.0	60.6	52.4	52.2	58.6	61.9	70.5	51.7	—	49.5	33.5	54.7	56.2	70.3	57.8
LsRT10	56.8	92.9	47.9	47.7	54.5	91.4	56.2	47.5	58.1	—	26.7	91.1	88.5	45.8	74.0
LsRT11	51.8	49.7	99.1	99.0	49.4	51.1	52.3	98.5	52.2	47.5	—	30.4	31.4	30.9	33.0
LsRT12	59.0	96.8	51.2	51.0	57.4	97.1	59.8	50.0	61.4	93.3	50.6	—	95.3	51.6	79.7
LsRT13	58.3	97.3	51.7	51.4	58.1	98.1	59.7	51.0	61.6	91.4	51.6	97.1	—	54.2	81.8
LsRT14	70.5	59.6	52.1	51.9	59.1	60.9	69.5	51.2	72.2	56.7	52.2	60.7	61.4	—	55.7
LsRT15	57.4	75.4	49.3	49.0	56.6	76.2	59.1	48.3	57.8	71.4	48.4	75.3	75.7	59.1	—
LsRT16	51.9	49.6	99.2	99.2	49.0	51.2	52.4	98.5	52.2	47.6	99.0	50.7	51.4	51.9	48.4
LsRT17	51.6	50.0	99.0	99.2	48.8	51.1	51.6	98.5	52.1	47.5	98.6	50.3	51.0	51.2	48.4
LsRT18	70.4	58.3	49.3	48.7	56.0	58.5	70.3	48.5	73.3	55.7	48.9	58.4	58.1	68.3	55.2
LsRT19	57.9	57.3	49.3	49.1	97.6	57.9	55.2	49.0	58.8	54.8	49.1	57.8	58.6	59.5	56.4
LsRT20	58.1	57.4	50.3	50.0	97.9	58.3	55.7	49.1	59.0	54.5	49.6	58.4	58.8	59.8	57.1
LsRT21	58.8	97.5	51.7	51.4	58.4	97.8	60.2	50.9	61.2	91.6	51.8	96.9	99.3	61.2	75.7
LsRT22	59.1	96.8	51.2	50.7	57.9	97.4	60.2	50.3	61.6	93.6	50.9	99.3	97.2	60.7	75.3
LsRT23	58.1	57.4	49.8	49.3	98.6	58.3	55.9	49.0	58.8	54.7	49.2	57.9	58.6	59.7	56.7
LsRT24	58.4	57.8	50.0	50.0	97.9	58.3	56.2	48.8	59.0	54.8	49.4	58.6	58.8	60.0	57.1
LsRT25	58.5	56.8	52.8	52.4	56.6	57.5	57.8	52.4	57.8	54.0	52.5	56.8	57.8	56.8	58.9
LsRT26	51.6	48.9	95.1	94.9	49.0	50.8	51.7	94.1	51.7	47.1	94.6	50.5	50.5	51.2	48.1
LsRT27	57.7	57.4	49.0	49.1	98.1	57.5	55.2	48.4	58.1	54.4	48.7	57.4	57.8	59.0	56.0
LsRT28	49.7	51.2	60.9	60.6	49.2	52.3	50.7	59.6	52.0	48.4	60.5	52.2	52.9	49.0	47.7
LsRT29	58.4	58.0	50.0	49.5	98.3	58.3	55.5	49.0	58.8	55.0	49.1	58.3	58.6	59.7	57.1
LsRT30	59.5	97.7	51.7	51.6	57.9	99.1	60.5	50.7	61.6	91.7	51.3	97.1	97.6	60.7	76.0

续表 2 Table 2 (Continued)

序列 Sequence	各序列间的相似性/% Similarity among different sequences														
	LsRT16	LsRT17	LsRT18	LsRT19	LsRT20	LsRT21	LsRT22	LsRT23	LsRT24	LsRT25	LsRT26	LsRT27	LsRT28	LsRT29	LsRT30
LsRT1	41.1	41.1	74.1	54.7	55.7	55.2	53.6	56.2	55.7	43.5	40.6	54.4	30.4	55.7	54.7
LsRT2	38.7	38.7	55.6	51.9	52.4	96.3	94.1	52.9	52.4	45.5	37.1	52.4	32.4	52.4	92.7
LsRT3	99.0	99.5	39.8	41.9	42.4	40.3	39.3	42.4	42.4	32.6	90.8	40.8	45.9	42.4	39.8
LsRT4	99.0	99.5	39.6	41.9	42.4	40.3	39.3	42.4	42.4	32.6	90.3	40.8	45.6	42.4	39.8
LsRT5	41.9	41.9	52.6	99.0	98.4	53.6	51.6	98.4	99.0	44.0	41.4	97.9	30.5	99.0	53.1
LsRT6	39.8	40.1	54.9	52.6	53.6	97.4	95.3	53.1	53.6	44.5	39.6	51.8	31.9	53.1	100.0
LsRT7	41.9	41.9	71.9	54.2	55.2	55.2	53.6	54.7	55.2	44.0	41.4	53.6	31.1	55.2	54.7
LsRT8	98.5	97.4	39.6	41.9	41.9	39.8	38.7	41.9	41.9	32.6	89.8	40.6	46.7	41.9	39.3
LsRT9	41.9	41.9	74.5	52.6	53.6	56.2	54.7	53.1	53.6	44.0	40.8	52.1	28.9	53.6	55.7
LsRT10	35.1	34.9	50.5	46.4	46.9	88.5	91.7	46.4	46.9	43.5	32.8	47.2	29.8	46.9	88.0
LsRT11	76.4	76.4	33.5	37.7	37.2	31.4	30.4	37.2	37.2	28.9	69.7	36.1	40.2	37.2	30.9
LsRT12	39.3	39.3	54.2	51.0	52.1	95.3	99.5	51.6	52.1	43.5	37.2	50.5	31.6	52.1	94.8
LsRT13	40.3	40.3	55.7	53.1	54.2	100.0	95.8	53.6	54.2	45.0	39.3	52.6	32.1	54.2	97.4
LsRT14	39.3	39.3	66.1	53.1	54.2	54.2	51.6	53.6	54.2	42.4	38.2	52.6	29.5	54.2	52.6
LsRT15	40.8	40.8	55.7	52.6	53.6	81.8	80.2	53.1	53.6	47.1	39.8	52.1	28.9	53.6	81.8
LsRT16	—	99.5	39.8	41.9	42.4	40.3	39.3	42.4	42.4	32.1	90.8	40.8	45.4	42.4	39.8
LsRT17	98.8	—	39.6	41.9	42.4	40.3	39.3	42.4	42.4	32.1	89.8	40.6	45.1	42.4	39.8
LsRT18	48.8	48.4	—	52.1	52.6	55.7	54.2	52.1	52.6	44.0	37.5	52.8	28.3	52.6	55.2
LsRT19	49.5	49.1	55.9	—	98.4	53.1	51.0	98.4	99.0	44.0	41.4	97.9	30.5	99.0	52.6
LsRT20	49.7	49.5	56.2	98.3	—	54.2	52.1	99.0	99.5	44.0	41.9	97.9	30.5	99.5	53.6
LsRT21	51.6	50.9	58.1	58.4	58.6	—	95.8	53.6	54.2	45.0	39.3	52.6	32.1	54.2	97.4
LsRT22	50.7	50.2	58.4	57.8	58.1	97.6	—	51.6	52.1	43.5	37.2	50.5	31.6	52.1	95.3
LsRT23	49.3	48.8	56.6	98.1	99.0	58.6	58.1	—	99.5	43.5	41.9	97.9	30.5	99.5	53.1
LsRT24	49.8	49.3	55.9	98.4	99.1	59.0	58.1	99.0	—	44.0	41.9	98.4	30.5	100.0	53.6
LsRT25	52.6	52.6	56.6	56.4	56.4	57.8	56.9	56.8	56.6	—	32.6	44.0	29.1	44.0	44.5
LsRT26	94.9	94.2	47.9	48.8	49.5	50.3	50.2	49.1	49.7	50.5	—	40.6	44.6	41.9	39.8
LsRT27	48.6	48.5	55.9	97.6	98.3	57.6	57.1	98.4	98.4	55.9	48.9	—	29.8	98.4	52.1
LsRT28	60.4	60.4	51.0	49.0	49.9	52.5	51.8	49.7	49.7	51.7	58.5	48.7	—	30.5	32.1
LsRT29	49.1	49.3	56.4	98.1	99.0	58.6	58.1	99.3	99.3	57.1	49.3	98.3	49.9	—	53.6
LsRT30	51.2	50.9	58.4	57.9	58.3	98.1	97.6	58.1	58.4	57.3	50.7	57.2	52.3	58.3	—

<sup>1)</sup> 横线下方的数值为核苷酸序列间的相似性, 横线上方的数值为氨基酸序列间的相似性 Datums below the diagonal are similarity among nucleotide sequences and those above the diagonal are similarity among amino acid sequences.

的 46.7%、26.7% 和 23.3%。Family 1 包含的 14 条序列的相似性为 55.2% ~ 99.3%，其中仅 LsRT2 与 LsRT10 核苷酸序列长度差异较大, 其余序列的长度均较一致, 但存在由不同程度碱基替换、点突变与缺失突变导致的少许差异性(图 2); Family 1 又可分为 2 个亚家族, 分别包含 9 和 5 条 RT 核苷酸序列, 其中 LsRT12 与 LsRT22、LsRT13 与 LsRT21 核苷酸序列的相似性最高, 推测可能是在进化过程中由同一类逆转座子发生突变造成的。Family 2 包含 8 个序列, 相似性为 58.5% ~ 99.5%, 该家族又可分为 2 个亚家族, 分别包含 7 和 1 条 RT 核苷酸序列, 其中 LsRT3 与 LsRT4 的核苷酸序列相似性最高, 该家族序列间同样存在缺失突变、碱基替换与点突变(图 2)。Family 4

包含 7 条 RT 核苷酸序列, 相似性为 97.6% ~ 99.3%，其中 LsRT29 与 LsRT24 的核苷酸序列相似性最高。由此可见, 碱基替换、点突变或缺失插入突变可能是同一家族逆转座子异质性的原因之一。此外, 不同家族包含的序列数量不同, 也反映了其转座过程的差异性。

#### 2.4 瓠瓜 LINE 逆转座子 RT 氨基酸序列分析和相似性分析

利用 NCBI BLASTx 程序进行同源搜索, 依据与瓠瓜 LINE 逆转座子 RT 核苷酸序列相似性较高的其他物种的 RT 氨基酸序列, 对 30 条瓠瓜 LINE 逆转座子 RT 核苷酸序列进行推导翻译, 其中 LsRT11 和 LsRT28 核苷酸序列从第 2 个核苷酸开始翻译, 其他



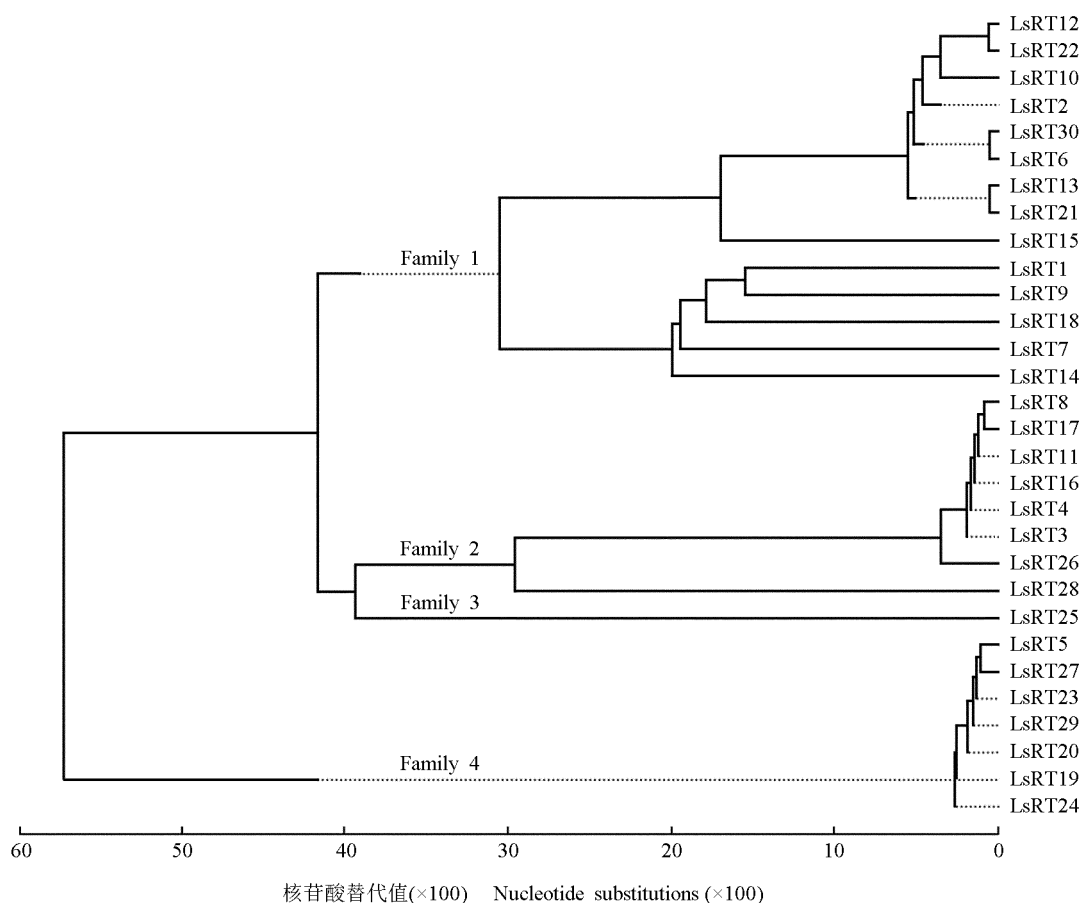


图 3 瓠瓜 LINE 逆转座子 RT 核苷酸序列的系统进化树  
 Fig. 3 Phylogenetic tree of RT nucleotide sequence of LINE retrotransposon from *Lagenaria siceraria* (Molina) Standl.

序列则从第 3 个核苷酸开始翻译。分析结果表明(图 4):瓠瓜 LINE 逆转座子 RT 氨基酸序列包含 20 个保守氨基酸残基,分别为第 12(Phe)、第 51(Lys)、第 62(Arg)、第 64(Ile)、第 65(Ser)、第 66(Leu)、第 71(Tyr)、第 72(Lys)、第 82(Leu)、第 94(Gln)、第 96(Ala)、第 97(Phe)、第 101(Arg)、第 103(Ile)、第 105(Asp)、第 112(Glu)、第 124(Gly)、第 132(Lys)、第 133(Leu)和第 147(Leu)位,推测这些位点可能是瓠瓜 LINE 逆转座子 RT 序列中非常保守的位点;另外还存在多个半保守氨基酸残基,如第 78(Leu)、第 81(Arg)、第 83(Lys)、第 134(Asp)和第 137(Lys)位等。有 14 条序列存在终止密码子突变,其中 LsRT25 氨基酸序列存在 3 个终止密码子突变(第 8、第 93 和第 192 位氨基酸);LsRT11 和 LsRT28 氨基酸序列存在 2 个终止密码子突变,前者为第 18 和第 25 位氨基酸,后者为第 169 和第 171 位氨基酸;存在 1 个终止密码子突变的氨基酸序列包括 LsRT1(第 158 位氨基酸)、

LsRT2(第 158 位氨基酸)、LsRT5(第 24 位氨基酸)、LsRT10(第 139 位氨基酸)、LsRT12(第 137 位氨基酸)、LsRT16(第 137 位氨基酸)、LsRT17(第 137 位氨基酸)、LsRT18(第 16 位氨基酸)、LsRT22(第 139 位氨基酸)、LsRT23(第 158 位氨基酸)和 LsRT26(第 147 位氨基酸);有 16 条氨基酸序列无终止密码子,但可能具有转录活性,包括 LsRT3、LsRT4、LsRT6 至 LsRT9、LsRT13 至 LsRT15、LsRT19 至 LsRT21、LsRT24、LsRT27、LsRT29 和 LsRT30,占瓠瓜 LINE 逆转座子 RT 氨基酸序列数量的 53.3%。因此,终止密码子突变可能是逆转座子发生突变与停止表达、进而导致逆转座子异质性的主要原因。

由表 2 还可以看出:经 DNASTar 软件 MegAlign 程序同源比对,30 条瓠瓜 LINE 逆转座子 RT 氨基酸序列的相似性为 26.7% ~ 100.0%;其中,LsRT11 与 LsRT10 氨基酸序列相似性最低,而二者核苷酸序列的相似性却达到 47.5%;LsRT3 与 LsRT4、LsRT6 与

```

LsRT1 -DGGDGGTIEFFKTFNNIIRKPDLMKQPHDFEHSKIVNRAVNAITYALIPKLSKQKVTIYRPIISLTIISYKILAKVLAERLKKVLPETISAHQSAPVKRQITDPILLIANEVVDHNCN
LsRT2 -DGGDGGTAEFFKTFNWTIKSDIMKLIQDEYTSGVVAVKLNIEYICLIPKQVDSKIVSDYRPIISLIPCAKIVAVLNRKLIIPSTIAANQLAFVTRQILDASIMANELIDWSSN
LsRT3 -DGGDGLKQFFQKYWEELGTTIVDLCINILNGEGKIEDLNFSTVLIPEKSNKFRAMNDFRPSILCNVYKIIANTLANRKSVDHIIISPPQSAPFIPSRLLITDIAIGFECIHSLN-N
LsRT4 -DGGDGLKQFFQKYWEELGTTIVDLCINILNGEGKIEDLNFSTVLIPEKSNKFRAMNDFRPSILCNVYKIIANTLANRKSVDHIIISPPQSAPFIPSRLLITDIAIGFECIHSLN-N
LsRT5 -DGGDGGTAEFFKTFNWTIKSDIMKLIQDEYTSGVVAVKLNIEYICLIPKQVDSKIVSDYRPIISLIPCAKIVAVLNRKLIIPSTIAANQLAFVTRQILDASIMANELIDWSSN
LsRT6 -DGGDGGTAEFFKTFNWTIKSDIMKLIQDEYTSGVVAVKLNIEYICLIPKQVDSKIVSDYRPIISLIPCAKIVAVLNRKLIIPSTIAANQLAFVTRQILDASIMANELIDWSSN
LsRT7 -DGGDGGTIEFFKTFNWTIKSDIMKLIQDEYTSGVVAVKLNIEYICLIPKQVDSKIVSDYRPIISLIPCAKIVAVLNRKLIIPSTIAANQLAFVTRQILDASIMANELIDWSSN
LsRT8 -DGGDGGTAEFFKTFNWTIKSDIMKLIQDEYTSGVVAVKLNIEYICLIPKQVDSKIVSDYRPIISLIPCAKIVAVLNRKLIIPSTIAANQLAFVTRQILDASIMANELIDWSSN
LsRT9 -DGGDGGTIEFFKTFNWTIKSDIMKLIQDEYTSGVVAVKLNIEYICLIPKQVDSKIVSDYRPIISLIPCAKIVAVLNRKLIIPSTIAANQLAFVTRQILDASIMANELIDWSSN
LsRT10 -DGGDGGTAEFFKTFNWTIKSDIMKLIQDEYTSGVVAVKLNIEYICLIPKQVDSKIVSDYRPIISLIPCAKIVAVLNRKLIIPSTIAANQLAFVTRQILDASIMANELIDWSSN
LsRT11 -GSSWGRVGFLLPKILA-VGYYS-SVPHFEWGRKDRGFKLHIYCPHSKSNKFRAMNDFRPSILCNVYKIIANTLANRKSVDHIIISPPQSAPFIPSRLLITDIAIGFECIHSLN-N
LsRT12 -DGGDGGTAEFFKTFNWTIKSDIMKLIQDEYTSGVVAVKLNIEYICLIPKQVDSKIVSDYRPIISLIPCAKIVAVLNRKLIIPSTIAANQLAFVTRQILDASIMANELIDWSSN
LsRT13 -DGGDGGTAEFFKTFNWTIKSDIMKLIQDEYTSGVVAVKLNIEYICLIPKQVDSKIVSDYRPIISLIPCAKIVAVLNRKLIIPSTIAANQLAFVTRQILDASIMANELIDWSSN
LsRT14 -DGGDGGTIEFFKTFNWTIKSDIMKLIQDEYTSGVVAVKLNIEYICLIPKQVDSKIVSDYRPIISLIPCAKIVAVLNRKLIIPSTIAANQLAFVTRQILDASIMANELIDWSSN
LsRT15 -DGGDGGTAEFFKTFNWTIKSDIMKLIQDEYTSGVVAVKLNIEYICLIPKQVDSKIVSDYRPIISLIPCAKIVAVLNRKLIIPSTIAANQLAFVTRQILDASIMANELIDWSSN
LsRT16 -DGGDGGTAEFFKTFNWTIKSDIMKLIQDEYTSGVVAVKLNIEYICLIPKQVDSKIVSDYRPIISLIPCAKIVAVLNRKLIIPSTIAANQLAFVTRQILDASIMANELIDWSSN
LsRT17 -DGGDGLKQFFQKYWEELGTTIVDLCINILNGEGKIEDLNFSTVLIPEKSNKFRAMNDFRPSILCNVYKIIANTLANRKSVDHIIISPPQSAPFIPSRLLITDIAIGFECIHSLN-N
LsRT18 -DGGDGGTAEFFKTFNWTIKSDIMKLIQDEYTSGVVAVKLNIEYICLIPKQVDSKIVSDYRPIISLIPCAKIVAVLNRKLIIPSTIAANQLAFVTRQILDASIMANELIDWSSN
LsRT19 -DGGDGGTAEFFKTFNWTIKSDIMKLIQDEYTSGVVAVKLNIEYICLIPKQVDSKIVSDYRPIISLIPCAKIVAVLNRKLIIPSTIAANQLAFVTRQILDASIMANELIDWSSN
LsRT20 -DGGDGGTAEFFKTFNWTIKSDIMKLIQDEYTSGVVAVKLNIEYICLIPKQVDSKIVSDYRPIISLIPCAKIVAVLNRKLIIPSTIAANQLAFVTRQILDASIMANELIDWSSN
LsRT21 -DGGDGGTAEFFKTFNWTIKSDIMKLIQDEYTSGVVAVKLNIEYICLIPKQVDSKIVSDYRPIISLIPCAKIVAVLNRKLIIPSTIAANQLAFVTRQILDASIMANELIDWSSN
LsRT22 -DGGDGGTAEFFKTFNWTIKSDIMKLIQDEYTSGVVAVKLNIEYICLIPKQVDSKIVSDYRPIISLIPCAKIVAVLNRKLIIPSTIAANQLAFVTRQILDASIMANELIDWSSN
LsRT23 -DGGDGGTAEFFKTFNWTIKSDIMKLIQDEYTSGVVAVKLNIEYICLIPKQVDSKIVSDYRPIISLIPCAKIVAVLNRKLIIPSTIAANQLAFVTRQILDASIMANELIDWSSN
LsRT24 -DGGDGGTAEFFKTFNWTIKSDIMKLIQDEYTSGVVAVKLNIEYICLIPKQVDSKIVSDYRPIISLIPCAKIVAVLNRKLIIPSTIAANQLAFVTRQILDASIMANELIDWSSN
LsRT25 -GIRVLMV-QVSFLIKSNMILKSDLVGVVQES-VGIIINRCHETIYICLIPKQVDSKIVSDYRPIISLIPCAKIVAVLNRKLIIPSTIAANQLAFVTRQILDASIMANELIDWSSN
LsRT26 -DGGDGGTAEFFKTFNWTIKSDIMKLIQDEYTSGVVAVKLNIEYICLIPKQVDSKIVSDYRPIISLIPCAKIVAVLNRKLIIPSTIAANQLAFVTRQILDASIMANELIDWSSN
LsRT27 -DGGDGGTAEFFKTFNWTIKSDIMKLIQDEYTSGVVAVKLNIEYICLIPKQVDSKIVSDYRPIISLIPCAKIVAVLNRKLIIPSTIAANQLAFVTRQILDASIMANELIDWSSN
LsRT28 -GSSWGRVGFLLPKILA-VGYYS-SVPHFEWGRKDRGFKLHIYCPHSKSNKFRAMNDFRPSILCNVYKIIANTLANRKSVDHIIISPPQSAPFIPSRLLITDIAIGFECIHSLN-N
LsRT29 -DGGDGGTAEFFKTFNWTIKSDIMKLIQDEYTSGVVAVKLNIEYICLIPKQVDSKIVSDYRPIISLIPCAKIVAVLNRKLIIPSTIAANQLAFVTRQILDASIMANELIDWSSN
LsRT30 -DGGDGGTAEFFKTFNWTIKSDIMKLIQDEYTSGVVAVKLNIEYICLIPKQVDSKIVSDYRPIISLIPCAKIVAVLNRKLIIPSTIAANQLAFVTRQILDASIMANELIDWSSN
LsRT1 RTKG---FIIKLDIEKAFQKINWFLSMLKRGFHDKWDN-VIACISTVYVSIIMNGCPRGYIQAISGIRQGDPLN
LsRT2 KRKG---VVKLDIEKAFQKINWFLSMLKRGFHDKWDN-VIACISTVYVSIIMNGCPRGYIQAISGIRQGDPLN
LsRT3 KRVGKDGVAALKDMSKAYDRVWSYLRQVHTKRGFCSRWIQLIMNCVETVTSVLFNGNQRITFRSRLRQGDPL
LsRT4 KRVGKDGVAALKDMSKAYDRVWSYLRQVHTKRGFCSRWIQLIMNCVETVTSVLFNGNQRITFRSRLRQGDPL
LsRT5 KRKG---WILKLDIEKAFQKINWFLSMLKRGFHDKWDN-VIACISTVYVSIIMNGCPRGYIQAISGIRQGDPL
LsRT6 KRKG---FIIKLDIEKAFQKINWFLSMLKRGFHDKWDN-VIACISTVYVSIIMNGCPRGYIQAISGIRQGDPL
LsRT7 KRKG---FIIKLDIEKAFQKINWFLSMLKRGFHDKWDN-VIACISTVYVSIIMNGCPRGYIQAISGIRQGDPL
LsRT8 KRVGKDGVAALKDMSKAYDRVWSYLRQVHTKRGFCSRWIQLIMNCVETVTSVLFNGNQRITFRSRLRQGDPL
LsRT9 KRVGKDGVAALKDMSKAYDRVWSYLRQVHTKRGFCSRWIQLIMNCVETVTSVLFNGNQRITFRSRLRQGDPL
LsRT10 KRVG---VVKLDIEKAFQKINWFLSMLKRGFHDKWDN-VIACISTVYVSIIMNGCPRGYIQAISGIRQGDPL
LsRT11 KRVGKDGVAALKDMSKAYDRVWSYLRQVHTKRGFCSRWIQLIMNCVETVTSVLFNGNQRITFRSRLRQGDPL
LsRT12 KRVG---VVKLDIEKAFQKINWFLSMLKRGFHDKWDN-VIACISTVYVSIIMNGCPRGYIQAISGIRQGDPL
LsRT13 KRKG---IVKLDIEKAFQKINWFLSMLKRGFHDKWDN-VIACISTVYVSIIMNGCPRGYIQAISGIRQGDPL
LsRT14 KRKG---IMKLDIEKAFQKINWFLSMLKRGFHDKWDN-VIACISTVYVSIIMNGCPRGYIQAISGIRQGDPL
LsRT15 KRKG---VVKLDIEKAFQKINWFLSMLKRGFHDKWDN-VIACISTVYVSIIMNGCPRGYIQAISGIRQGDPL
LsRT16 KRVGKDGVAALKDMSKAYDRVWSYLRQVHTKRGFCSRWIQLIMNCVETVTSVLFNGNQRITFRSRLRQGDPL
LsRT17 KRVGKDGVAALKDMSKAYDRVWSYLRQVHTKRGFCSRWIQLIMNCVETVTSVLFNGNQRITFRSRLRQGDPL
LsRT18 KRKG---FIIKLDIEKAFQKINWFLSMLKRGFHDKWDN-VIACISTVYVSIIMNGCPRGYIQAISGIRQGDPL
LsRT19 KRKG---WILKLDIEKAFQKINWFLSMLKRGFHDKWDN-VIACISTVYVSIIMNGCPRGYIQAISGIRQGDPL
LsRT20 KRKG---WILKLDIEKAFQKINWFLSMLKRGFHDKWDN-VIACISTVYVSIIMNGCPRGYIQAISGIRQGDPL
LsRT21 KRKG---IVKLDIEKAFQKINWFLSMLKRGFHDKWDN-VIACISTVYVSIIMNGCPRGYIQAISGIRQGDPL
LsRT22 KRVG---VVKLDIEKAFQKINWFLSMLKRGFHDKWDN-VIACISTVYVSIIMNGCPRGYIQAISGIRQGDPL
LsRT23 KRKG---WILKLDIEKAFQKINWFLSMLKRGFHDKWDN-VIACISTVYVSIIMNGCPRGYIQAISGIRQGDPL
LsRT24 KRKG---WILKLDIEKAFQKINWFLSMLKRGFHDKWDN-VIACISTVYVSIIMNGCPRGYIQAISGIRQGDPL
LsRT25 KRKG---ALKLDIEKAFQKINWFLSMLKRGFHDKWDN-VIACISTVYVSIIMNGCPRGYIQAISGIRQGDPL
LsRT26 KRVGKDGVAALKDMSKAYDRVWSYLRQVHTKRGFCSRWIQLIMNCVETVTSVLFNGNQRITFRSRLRQGDPL
LsRT27 KRKG---WILKLDIEKAFQKINWFLSMLKRGFHDKWDN-VIACISTVYVSIIMNGCPRGYIQAISGIRQGDPL
LsRT28 KRIGKDGVAALKDMSKAYDRVWSYLRQVHTKRGFCSRWIQLIMNCVETVTSVLFNGNQRITFRSRLRQGDPL
LsRT29 KRKG---WILKLDIEKAFQKINWFLSMLKRGFHDKWDN-VIACISTVYVSIIMNGCPRGYIQAISGIRQGDPL
LsRT30 KRKG---VVKLDIEKAFQKINWFLSMLKRGFHDKWDN-VIACISTVYVSIIMNGCPRGYIQAISGIRQGDPL

```

\* : 终止密码子 Stop codon; -: 用于优化联配的缺口 Gaps used for optimal alignment.

图 4 瓠瓜 LINE 反转座子 RT 氨基酸序列比对

Fig. 4 Alignment of RT amino acid sequence of LINE retrotransposon from *Lagenaria siceraria* (Molina) Standl.

LsRT30、LsRT13 与 LsRT21、LsRT24 与 LsRT29 的氨基酸序列相似性均为 100.0%，其核苷酸序列相似性分别为 99.5%、99.1%、99.3% 和 99.3%。结合图 3 可见，Family 1 中 LsRT12 与 LsRT22 核苷酸序列的遗传距离最近、其相似性为 99.3%，氨基酸序列的相似性为 99.5%；Family 2 中 LsRT8 与 LsRT17 核苷酸序列的遗传距离最近、其相似性为 98.5%，氨基酸序列的相似性为 97.4%；Family 4 中 LsRT5 与 LsRT27 核苷酸序列的遗传距离最近、其相似性为 98.1%，氨基酸序列的相似性为 97.9%。瓠瓜 LINE 反转座子 RT 序

列各家族所包含的序列数目越多、相似性越高，表明发生转座的时间越近，具转座活性的可能性也越大。Family1、Family 2 和 Family 4 分别包含 8、3 和 5 条无终止密码子的 RT 氨基酸序列，因此可能是具有转座活性的家族。

### 2.5 瓠瓜 LINE 反转座子 RT 氨基酸序列的系统进化分析

将 30 条瓠瓜 LINE 反转座子 RT 氨基酸序列与 GenBank 上登录的其他 15 种植物的 LINE 反转座子 RT 氨基酸序列进行同源比对并构建系统进化树，结



果见图 5。瓠瓜 LINE 逆转座子 RT 氨基酸序列与葡萄(CAN76892)、拟南芥[*Arabidopsis thaliana* (Linn.) Heynh., AAB48348]、黄瓜(*Cucumis sativus* Linn., XP\_004142331)等种类的 LINE 逆转座子 RT 氨基酸序列具有较高相似性,相似性为 41% ~ 59%。根据系统进化树,可将涉及的 RT 氨基酸序列聚为 4 组: LsRT1、

LsRT2、LsRT5 至 LsRT7、LsRT9、LsRT10、LsRT12 至 LsRT15、LsRT18 至 LsRT25、LsRT27、LsRT29 和 LsRT30 等 22 条氨基酸序列与黄瓜 LINE 逆转座子 RT 氨基酸序列遗传距离较近,聚为第 1 组(Group I),该组包含的瓠瓜 LINE 逆转座子 RT 氨基酸序列数约占瓠瓜 RT 氨基酸序列总数的 73.33%;第 3 组

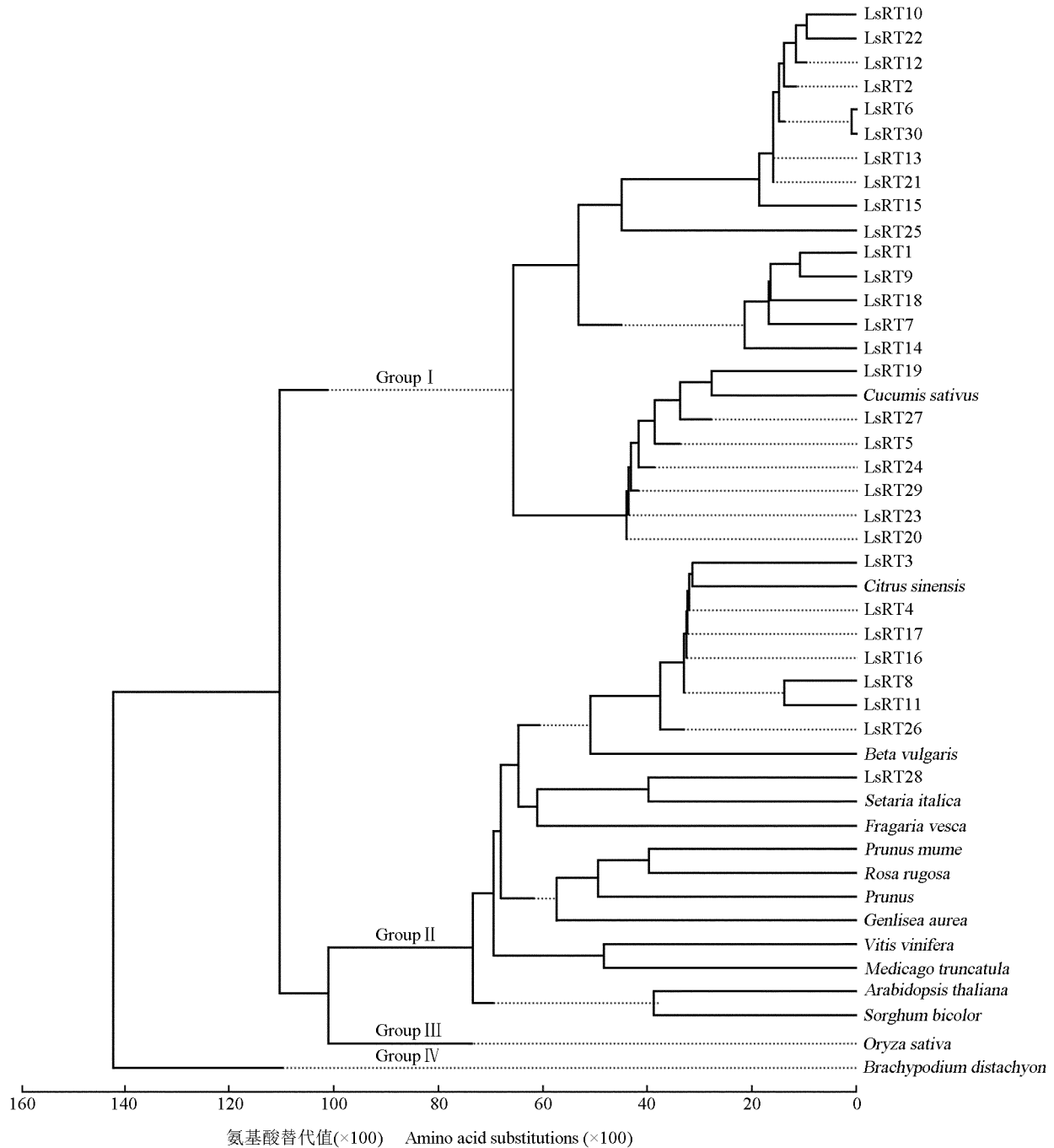


图 5 瓠瓜与其他植物 LINE 逆转座子 RT 氨基酸序列的系统进化树  
 Fig. 5 Phylogenetic tree of RT amino acid sequence of LINE retrotransposon from *Lagenaria siceraria* (Molina) Standl. and other species

(Group III) 仅包含 1 条水稻 (*Oryza sativa* Linn., EEE59033) LINE 逆转座子 RT 序列; 第 4 组 (Group IV) 仅包含二穗短柄草 [*Brachypodium distachyon* (Linn.) P. Beauv., XP\_003579203] LINE 逆转座子 RT 氨基酸序列, 与其他组的遗传距离最远; LsRT3、LsRT4、LsRT8、LsRT11、LsRT16、LsRT17、LsRT26 和 LsRT28 等 8 条氨基酸序列与其他 12 种植物的 LINE 逆转座子 RT 氨基酸序列聚为第 2 组 (Group II), 它们的进化关系相对较近。30 条瓠瓜 LINE 逆转座子 RT 氨基酸序列主要集中在 Group I 与 Group II 中, 说明瓠瓜 LINE 逆转座子 RT 序列兼具一定的保守性与较高的异质性。上述分析结果表明: 不同植物种类间存在起源相同或相近的逆转座子; 由于 30 条瓠瓜 LINE 逆转座子 RT 氨基酸序列相似性为 26.7% ~ 100.0%, 因此不同种属间逆转座子的相似性可能比种属内更高。由此可见, 瓠瓜与上述植物在进化过程中可能存在 LINE 逆转座子的横向传递现象。

### 3 讨 论

逆转座子广泛散布于植物基因组中, 其具有的高异质性、高拷贝数及专一插入位点等特性对植物基因组结构大小与功能进化具有显著影响, 对研究植物遗传多样性、遗传进化及基因克隆也具有重要意义<sup>[27]</sup>。瓠瓜种质资源的鉴定、评价及系统进化研究是加速品种改良和杂种优势利用的前提。作者利用 LINE 逆转座子 RT 序列简并引物从瓠瓜种质中扩增得到目的片段, 扩增获得的 30 条瓠瓜 LINE 逆转座子 RT 核苷酸序列长度为 564 ~ 599 bp, 长度差为 35 bp, 其碱基组成与 AT/GC 比也存在较大差异。系统进化分析结果显示: 30 条瓠瓜 LINE 逆转座子 RT 核苷酸序列可划分为 4 个家族 (family), 但 4 个家族的相似性变异范围较大, 最小为 1.7%, 表现出很高的异质性。此外, 终止密码子突变也普遍发生在瓠瓜 LINE 逆转座子 RT 核苷酸序列中, 推测这是瓠瓜逆转座子高度异质性的主要原因之一, 与牡丹和甜菜等作物逆转座子的丰富多态性及形成原因相似<sup>[26-27]</sup>。由于 30 条瓠瓜 LINE 逆转座子 RT 核苷酸序列为随机挑选克隆所得, 目前尚未在 30 条 RT 序列中发现移码突变, 而移码突变在许多物种的逆转座子 RT 序列普遍存在, 因此, 还需进一步研究瓠瓜 RT 序列中移码突变发生的可能性。以上分析结果表明: 瓠瓜基因组的 LINE 逆转座

子是一类较古老元件, 其异质性是后来伴随寄主基因组的进化在转座与世代传递过程中逐渐积累突变的。

30 条瓠瓜 LINE 逆转座子 RT 核苷酸序列分布于 4 个家族中, 4 个家族包含的 RT 核苷酸序列数量不同, 表明其转录过程存在差异以及家族历史久远程度不同。进一步的研究结果显示: 其中 14 条序列存在终止密码子突变, 而另外 16 条可能具有转录活性的序列则分布于 3 个家族中, 表明这些家族可能新近发生过转座整合事件, 目前还具有转座活性<sup>[28]</sup>, 可进一步用于瓠瓜种下变异的研究。Family 3 仅包含 1 条 RT 核苷酸序列, 与其他家族亲缘关系较远, 表明该家族成员可能由其他家族成员突变而来或经横向传递进入瓠瓜基因组, 因此相对较“年轻”; 另外, 鉴于克隆数目有限, 也存在尚未分离其他成员的可能。植物体内的逆转座子易受到生物和非生物胁迫而激活其转录活性<sup>[29-33]</sup>, 因此明确瓠瓜性状变异与逆转座子转录活性之间的关系, 对于挖掘瓠瓜种质的优良基因具有重要意义。

逆转座子的纵向传递与横向传递是推测逆转座子起源、物种形成及其与物种相互关系的重要工具<sup>[34-35]</sup>。Voytas 等<sup>[36]</sup>认为, 植物逆转座子序列的种内异质性差异可能大于种属间。瓠瓜 LINE 逆转座子 RT 氨基酸序列的相似性为 26.7% ~ 100.0%, 家族内部某些序列间的相似性极高, 可能是寄主基因组与逆转座子间互惠关系的体现; 而与葡萄和拟南芥等种类 LINE 逆转座子 RT 氨基酸序列的相似性为 41% ~ 59%, 说明瓠瓜 LINE 逆转座子虽较保守, 但种内异质性差异高于种属间, 可能是在进化历史上瓠瓜 LINE 逆转座子通过横向传递进入其他种类的基因组内, 由此也可说明生物物种源于共同祖先。

#### 参考文献:

- [1] 彭庆务, 何晓明, 谢大森, 等. 瓠瓜的特征特性及育种对策[J]. 广东农业科学, 2003, 30(1): 18-19.
- [2] 陈孟强, 牛玉, 林鉴荣, 等. 瓠瓜部分种质资源主要性状的初步评价[J]. 广东农业科学, 2012, 39(23): 22-24.
- [3] DECKER-WALTERS D S. Cucurbits, sanskrit and the Indo-Aryans [J]. Economic Botany, 1999, 53: 98-112.
- [4] DECKER-WALTERS D S, WILKINS-ELLER M, CHUNG S M, et al. Discovery and genetic assessment of wild bottle gourd (*Lagenaria siceraria* (Mol.) Standley; Cucurbitaceae) from Zimbabwe [J]. Economic Botany, 2004, 58: 501-508.
- [5] 高山, 许端祥, 林碧英, 等. 38 份瓠瓜种质资源遗传多样性的 ISSR 分析[J]. 植物遗传资源学报, 2007, 8(4): 396-400.

- [6] 高山,许端祥,林碧英,等.瓠瓜种质资源遗传多样性的RAPD分析[J].分子植物育种,2007,5(4):502-506.
- [7] 周先治,陈阳,陈晟,等.基于5.8S rDNA和ITS序列探讨亚洲瓠瓜的地理分化[J].中国蔬菜,2011(6):49-53.
- [8] 赵芹,谢大森,彭庆务,等.不同产地瓠瓜品种ITS序列的遗传多样性分析[J].植物资源与环境学报,2014,23(3):24-35.
- [9] FLAVELL A J, SMITH D B, KUMAR A. Extreme heterogeneity of Ty1-copia group retrotransposons in plants [J]. Molecular and General Genetics, 1992, 231: 233-242.
- [10] BISWAS M K, BAIG M N R, CHENG Y J, et al. Retrotransposon-based genetic similarity within the genus *Citrus* and its relatives [J]. Genetic Resources and Crop Evolution, 2010, 57: 963-972.
- [11] KALENDAR R, FLAVELL A J, ELLIS T H N, et al. Analysis of plant diversity with retrotransposon-based molecular markers [J]. Heredity, 2011, 106: 520-530.
- [12] MA Y, HE P, SUN H Y, et al. Isolation and characterization of transcriptionally active Ty1-copia retrotransposons in *Fragaria* × *ananassa* [J]. Agricultural Sciences in China, 2010, 9: 337-345.
- [13] McCLINTOCK B. The significance of responses of the genome to challenge [J]. Science, 1984, 226: 792-801.
- [14] GRANDBASTIEN M A, LUEAS H, MOREL J B, et al. The expression of the tobacco Tnt1 retrotransposon is linked to plant defense responses [J]. Genetica, 1997, 100: 241-252.
- [15] MHIRI C, De WILT P J, GRANDBASTIEN M A. Activation of the promoter of the Tnt1 retrotransposon in tomato after inoculation with the fungal pathogen *Cladosporium fulvum* [J]. Molecular Plant-Microbe Interactions, 1999, 12: 592-603.
- [16] BEGUIRISTAIN T, GRANDBASTIEN M A, PUIGDOMÈNECH P, et al. Three Tnt1 subfamilies show different stress-associated patterns of expression in tobacco. Consequences for retrotransposon control and evolution in plants [J]. Plant Physiology, 2001, 127: 212-221.
- [17] CARRIER G, Le CUNFF L, DEREPPER A, et al. Transposable elements are a major cause of somatic polymorphism in *Vitis vinifera* L. [J]. PLoS One, 2012, 7: e32973.
- [18] ERDMANN P M, LEE R K, BASSETT M J, et al. A molecular marker tightly linked to P, a gene required for flower and seedcoat color in common bean (*Phaseolus vulgaris* L.), contains the Ty3-gypsy retrotransposon Tpv3g [J]. Genome, 2002, 45: 728-736.
- [19] PAGAN H J, SMITH J D, HUBLEY R M, et al. PiggyBac-ing on a primate genome: novel elements, recent activity and horizontal transfer [J]. Genome Biology and Evolution, 2010, 2: 293-303.
- [20] PRITHAM E J, FESCHOTTE C. Massive amplification of rolling-circle transposons in the lineage of the bat *Myotis lucifugus* [J]. Proceedings of the National Academy of Sciences of the United States of America, 2007, 104: 1895-1900.
- [21] HILL P, BURFORD D, MARTIN D M, et al. Retrotransposon populations of *Vicia* species with varying genome size [J]. Molecular Genetics and Genomics, 2005, 273: 371-381.
- [22] MARTIN S L. Nucleic acid chaperone properties of ORF1p from the non-LTR retrotransposon, LINE-1 [J]. RNA Biology, 2010, 7: 706-711.
- [23] SEIBT K M, WENKE T, WOLLRAB C, et al. Development and application of SINE-based markers for genotyping of potato varieties [J]. Theoretical and Applied Genetics, 2012, 125: 185-196.
- [24] DOYLE J J, DOYLE J L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue [J]. Phytochemical Bulletin, 1987, 19: 11-15.
- [25] KUBIS S E, HESLOP-HARRISON J S, DESEL C, et al. The genomic organization of non-LTR retrotransposons (LINEs) from three *Beta* species and five other angiosperms [J]. Plant Molecular Biology, 1998, 36: 821-831.
- [26] 宋程威,郭大龙,张曦,等.牡丹LINE类反转录转座子RT序列的克隆及分析[J].园艺学报,2014,41(1):157-164.
- [27] KUMAR A, BENNETZEN J L. Plant retrotransposons [J]. Annual Review of Genetics, 1990, 33: 479-532.
- [28] GALINDO L M, GAITÁN-SOLÍS E, BACCAM P, et al. Isolation and characterization of RNase LTR sequences of Ty1-copia retrotransposons in common bean (*Phaseolus vulgaris* L.) [J]. Genome, 2004, 47: 84-95.
- [29] KALENDAR R, TANSKANEN J, IMMONEN S, et al. Genome evolution of wild barley (*Hordeum spontaneum*) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence [J]. Proceedings of the National Academy of Sciences of the United States of America, 2000, 97: 6603-6607.
- [30] ELLIS T H, POYSER S J, KNOX M R, et al. Polymorphism of insertion sites of Ty1-copia class retrotransposons and its use for linkage and diversity analysis in pea [J]. Molecular and General Genetics, 1998, 260: 9-19.
- [31] TAKEDA S, SUGIMOTO K, OTSUKI H, et al. Transcriptional activation of the tobacco retrotransposon Tto1 by wounding and methyl jasmonate [J]. Plant Molecular Biology, 1998, 36: 365-376.
- [32] PEARCE S R, HARRISON G, LI D, et al. The Ty1-copia group retrotransposons in *Vicia* species: copy number, sequence heterogeneity and chromosomal localisation [J]. Molecular and General Genetics, 1996, 250: 305-315.
- [33] 胡志昂,姜国强,邓馨,等.野大豆种群转座子和转录因子的多样性和分子适应[J].植物生态学报,2007,31(5):952-959.
- [34] KUMAR A. The evolution of plant retroviruses: moving to green pastures [J]. Trends in Plant Science, 1998, 3: 371-374.
- [35] GRANDBASTIEN M A. Retroelements in higher plants [J]. Trends in Genetics, 1992, 8: 100-108.
- [36] VOYTAS D F, CUMMINGS M P, KONICZNY A, et al. Copia-like retrotransposons are ubiquitous among plants [J]. Proceedings of the National Academy of Sciences of the United States of America, 1992, 89: 7124-7128.