

Queensland University of Technology Brisbane Australia

This is the author's version of a work that was submitted/accepted for publication in the following source:

Okpul, Tom, Harding, Robert M., Dieters, Mark J., & Godwin, Ian (2011) Occurrence of LINE, gypsy-like, and copia-like retrotransposons in the clonally propagated sweet potato (Ipomoea batatas L.). *Genome*, *54*(7), pp. 603-609.

This file was downloaded from: http://eprints.qut.edu.au/45469/

© Copyright 2011 National Research Council Canada

Notice: Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source:

http://dx.doi.org/10.1139/g11-027

1	Occurrence of LINE, gypsy- and copia-like retrotransposons in
2	the clonally-propagated sweetpotato (Ipomoea batatas L.)
3 4	Tom Okpul ¹ , Robert M. Harding ² , Mark J. Dieters ¹ and Ian D. Godwin ¹
5	
6	¹ The University of Queensland, School of Agriculture and Food Sciences, St. Lucia 4072,
7	QLD, Australia.
8	² Centre for Tropical Crops and Biocommodities, Queensland University of Technology,
9	Brisbane, QLD, Australia.
10	
11	Correspondence
12	T. Okpul, The University of Queensland, School of Agriculture and Food Sciences, St. Lucia
13	4072, Brisbane, QLD, Australia.
14	Email: <u>tokpul@uq.edu.au</u>
15	

16 Abstract: Retrotransposons are a class of transposable elements that represent a major fraction of the repetitive DNA of most eukaryotes. Their abundance stems from their 17 18 expansive replication strategies. We screened and isolated sequence fragments of long 19 terminal repeat (LTR), reverse transcriptase (RT) and *Gypsy*-like RT and envelop (ENV)-like 20 domains of LTR retrotransposons, and two partial sequences of non-LTR long interspersed 21 element (LINE) in the clonally-propagated allohexaploid sweetpotato (*Ipomoea batatas*) 22 genome. Using dot blot hybridisation, these elements were found to be present in the \sim 1597 23 Mb haploid sweetpotato genome with copy numbers ranging from ~ 50 to ~ 4100 as observed 24 in the partial LTR (IbLtr-1) and LINE (IbLi-1) sequences, respectively. The continuous 25 clonal propagation of sweetpotato may have contributed to such a multitude of copies of 26 some of these genomic elements. Interestingly, the isolated Gypsy-like ENV and RT 27 sequence fragments, IbGy-1 (~2100 copies) and IbGy-2 (~540 copies) respectively, were 28 found to be homologous to the Bagy-2 cDNA sequences of barley (Hordeum valgare). 29 Although the isolated partial sequences were found to be homologous to other 30 transcriptionally active elements, future studies are required to determine whether they 31 represent elements that are transcriptionally active under normal and/or stressful conditions.

32

33 *Keywords:* Sweetpotato, clonal propagation, *env*-class retroelement, retrotransposons.

35 Introduction

Retrotransposons are a class of transposable elements (TEs) that are ubiquitous in almost all eukaryotic genomes and are believed to be a major force shaping their evolution (Hawkins et al. 2008). When they are not down-regulated, retrotransposons proliferate through cascades of genome-wide retrotransposition by the "copy-and-paste" action of reverse transcriptase. In so doing, they can augment the genome and generate insertional mutations as demonstrated in rice (Sakai et al. 2007), and cause chromosomal rearrangements by serving as sites for ectopic recombination (Kejnovsky et al. 2009).

Retrotransposons are divided into two groups depending on whether or not they are
flanked by long terminal repeat (LTR) sequences (i.e. LTR and non-LTR retrotransposons).
The LTR retrotransposons have been further divided into *Ty1/copia*, *Ty3/gypsy* and *Bel*subgroups based on sequence heterogeneity, especially in the organisation of their proteincoding domains (Eickbush and Jamburuthugoda 2008). Non-LTR retrotransposons comprise
both long and short interspersed elements, abbreviated as LINEs and SINEs respectively.

49 The proliferation of TEs can be detrimental to the survival of their host especially in 50 asexual species where the mechanisms regulating TE activity (such as through mechanisms 51 of DNA repair) are lax. Polyploids generally have been shown to undergo rapid and extensive 52 genomic changes (Chen 2007). Therefore, it is not surprising that evidence of genomic 53 instability in a clonally-propagated polyploid such as sweetpotato (2n=6x=90), Ipomoea 54 batatas (L.) Lam. (Convolvulaceae) is accumulating. These genomic instabilities have been 55 associated with TEs, particularly retrotransposons (Table 1; La Bonte 2001). Apart from 56 LINEs (e.g. Yamashita and Tahara 2006) and SINEs (e.g. Tanaka et al. 2001), only the Tyl/copia-like elements of the LTR retrotransposons have been reported in sweetpotato 57 58 (Table 1).

59 The continuous use of clonal propagation has been implicated in the decline in storage 60 root yield of cultivars over time and this has been attributed to the accumulation of TEs (La 61 Bonte 2001). Active retrotransposons (Table 1) have been shown to cause morphological 62 aberrations such as skin colour mutations (Tanaka et al. 2001). The identification of TEs prevalent in sweetpotato is vital to study their nature and role in genome instability, and 63 64 ultimately clonal degeneration. In this paper, we report the detection of high copy numbers of sequence fragments of Gypsy-like, LINE and LTR retrotransposons in sweetpotato that may 65 66 have accumulated over generations of clonal propagation.

67

68 Materials and Methods

Total genomic DNA was extracted from two sweetpotato cultivars (Beauregard
and Northern Star) using a CTAB-based method (Connolly et al. 1994), and quantified
using a NanoDrop[®] (ND-1000) spectrophotometer (Thermo Fisher Scientific,
Wilmington, U.S.A.).

73 Database searches for nucleotide sequences of TEs reported in sweetpotato were 74 performed with BLASTn at the National Centre for Biotechnology Information (NCBI, 75 Bethesda, Md.) website: http://www.ncbi.nlm.nih.gov/BLAST/. Multiple sequence 76 alignments were performed using ClustalW (Thompson et al. 1994) at the European 77 Bioinformatics Institute website: http://www.ebi.ac.uk/Tools/clustalw2/. Primers (Table 2) 78 were then designed manually based on consensus sequences from alignments or by 79 submitting to the program Primer3 (Rozen and Skaletsky 2000), available at 80 http://fokker.wi.mit.edu/primer3/. Additionally, the degenerate primers designed by Vicient et al. (2001) to amplify the envelop (ENV) (KS1F, KS2R) and RT (KS3p1F, KS4p2R) 81 82 domains of Gypsy-like elements were also tested for sweetpotato. All the primers were 83 synthesized by Proligo Australia Pty Ltd (Lismore, Australia).

PCR amplifications were performed using a MJ-Research PTC-200[™] Thermal Cycler 84 85 (MJ Research Inc. Watertown, MA, USA). The 25 µl PCR reaction mixtures contained approximately 20 ng template DNA, 1X PCR Buffer, 2.5 mM MgCl₂, 0.25 mM dNTP, 0.2 86 87 µM of each primer and 1 U Taq DNA polymerase. The generic PCR program used comprised 88 a cycle of initial denaturation at 95 °C for 2 mins, 35 cycles of denaturation at 95 °C for 50 s, annealing at 50-60 °C for 50 s and extension at 72 °C for 1 min, and a final cycle of extension 89 90 at 72 °C for 5 mins. A 10 µl aliquot of the PCR product was electrophoresed through 1.5% 91 agarose, stained with ethidium bromide and photographed under UV light using a Molecular Imager[®] Gel Doc[™] XR System (BioRad Laboratories, Segrate (Milan), Italy). In cases where 92 93 a single band was amplified, the PCR products were purified directly using the QIAquick 94 PCR Purification Kit (Qiagen, Valencia, CA, USA. Cat. No. 28104). When multiple bands 95 were amplified, each band was purified from the gel using the QIAquick Gel Extraction Kit 96 (Qiagen, Valencia, CA, USA. Cat. No. 28704). The PCR products were ligated to the pCR[®]4-TOPO® (Invitrogen, Carlsbad, CA, USA. Cat. No. K4575-01) vector for cloning. The 97 98 transformed *Escherichia coli* colonies were each transferred onto a fresh plate, and in a 10 µl 99 of the PCR reaction mixture to verify their transformation. At least three confirmed colonies from each PCR product were selected (as replicates) and purified using PureLink[™] Quick 100 101 Plasmid Miniprep Kit (Invitrogen, Carlsbad, CA, USA. Cat. No. K2100-10). Three replicates 102 each of the purified PCR products and plasmids were sequenced at the Australian Genome Research Facility (AGRF), Brisbane Node, The University of Queensland, using BigDye[™] 103 104 Terminator Cycle Sequencing Ready Reaction Kit version 3.1 (Applied Biosystems, Foster 105 City, CA, USA). Sequence analyses (BLASTn and BLASTp) against the sequences in the 106 GenBank database, alignment with ClustalW and phylogenetic tree construction were 107 performed using the default settings of software Geneious version 4.7 (Drummond et al. 108 2009). The dendrogram was constructed using the neighbour-joining method (Saitou and Nei 109 1987) with 500 bootstrap replicates analysed; the default distance based on the proportion of
110 nucleotide and amino acid substitution was used, respectively. In cases where a significant
111 protein family was identified, further searches were conducted using Pfam version 24.0
112 (Wellcome Trust Sanger Institute at: <u>http://www.pfam.sanger.ac.uk/search/sequence.nces.</u>).

113 Dot-blot hybridization was used to determine the number of apparent full-length and 114 all detectable copies of the TEs in the sweetpotato genome. Serial dilutions of the mobile genetic element-containing plasmid and genomic DNA from the sweetpotato samples were 115 spotted onto Hybord N⁺ membranes as three replicates. The plasmid standards contained 116 117 0.001, 0.005, 0.01, 0.05, 0.1, 0.5 and 1 ng DNA, whereas genomic DNA dots contained 250, 500, 750 and 1000 ng DNA. The PCR products were purified using the Qiagen[®] PCR 118 119 Purification kit (Qiagen) and labelled with digoxigenin, DIG-dUTP using the PCR DIG 120 Probe Synthesis Kit (Roche Applied Science, Mannheim, Germany). Hybridization was 121 carried out according to the manual provided for the DIG Kit with a high-stringency wash: 122 twice in 2X SSC/0.1% SDS followed by a single wash in 0.5X SSC/0.1% SDS and 0.1X SSC/0.1% SDS at 65 °C for 10-15 mins. The hybridization signal intensities were quantified, 123 after subtracting the backgrounds, using a Molecular Imager[®] Gel Doc[™] XR System. The 124 volume (i.e. the sum of the intensities of the pixels within the volume boundary \times pixel area) 125 126 of each dot blot was determined, and this was used to calculate the concentration of the 127 hybridised probes using the equation derived from the standard curve. The relative 128 concentration of the nuclear genomic DNA attaching to the probe was determined by 129 dividing the concentration of the hybridised probes to the respective concentration of 130 genomic DNA. The copy numbers of the TE sequences present in the sweetpotato genome 131 (size = 1597 Mb/1C, where 1C is the haploid genome (Arumuganathan and Earle 1991)) 132 were then calculated using the formula: Copy number = (size of the haploid genome \times 133 average proportion of nuclear genomic DNA hybridizing to the probe)/ size of element probe.

135 **Results and Discussion**

136 Partial sequences of various retrotransposons including Gypsy- and copia-like LTR 137 retrotransposons and LINEs were found to be present in the sweetpotato genome (Table 3). 138 Two Gypsy-like elements were amplified: IbGy-1, a 128 bp partial ENV fragment and IbGy-139 2, a 364 bp partial Gypsy-like RT fragment (Fig. 1). A contiguous BLAST search found IbGy-1 to share similarities (Fig. 2a) with Bagy-2 cDNA sequences for the spliced ENV sub-140 141 genomic RNAs of barley (AJ298028-AJ298032, AJ298072), Gossypium barbadense 142 (accessions DQ109564, DQ10966, DQ109571, DQ109572) and G. hirsutum (Accession 143 AY257164). On the other hand, the partial RT sequence fragment, IbGy-2, was found to be 144 74.5% similar to Ty3/gypsy-like sequences from Zea mays L. (AJ295132), Brassica napus L. 145 (AJ421232) and accessions AF378016 and AF378017 of Oryza sativa L. (Fig. 2b). These 146 sequences (IbGy-1 and IbGy-2) were abundant in cv. Beauregard with estimated copy 147 numbers of 2100 and 500 in the sweetpotato haploid genome size of 1597 Mb (Table 3). 148 Although this is the first report of *Gypsy*-like retroelements for sweetpotato, they are widely 149 transcribed in flowering plants (Vicient et al. 2001b). A complete characterization of the 150 partial sequences IbGy-1 and IbGy-2, however, is required to verify their identity and 151 phylogeny. Nonetheless, their similarities to the active Bagy-2 Gypsy-like retroelement of 152 barley (Vicient et al. 2001a), plausibly indicates the occurrence of a Bagy-2-like family in 153 sweetpotato. Further, IbGy-1 was found to have over 2100 copies in the genome, and such a 154 multitude of copies may advocate its constitution of the ENV-like gene, albeit that only a 155 contiguous BLAST search was able to pick up related sequences. Conforming to earlier 156 observations (Abdel-Ghany and Zaki 2002; Vicient et al. 2001a; Wright 1998), the similarity 157 of these sweetpotato Gypsy-like elements with those of unrelated monocot species is 158 suggestive of their origins long before the divergence of dicots and monocots.

159 As anticipated, the PCR amplified fragments IbLi-1 (258 bp) and IbLi-2 (352 bp) 160 were found to be highly homologous to the Lib LINEs of sweetpotato (Yamashita and Tahara 161 2006, Fig. 1, 3a and 3b). A few base substitutions leading to amino acid substitutions in the 162 IbLi-1 and IbLi-2 sequences caused slight deviations from the Lib LINE sequences. They were also found to have relatively higher copy numbers in the genome of cv. Beauregard with 163 164 more than 4100 for IbLi-1 and ~600 for IbLi-2 (Table 3). Whereas numerous 165 retrotransposons have been shown to inhabit the sweetpotato genome, only a few have been 166 reported to actively mobilise. Tahara et al. (2004) demonstrated that the abundant Ty1/copia-167 like element Rtsp-1 can be actively mobilised in callous tissues. Yamashita and Tahara 168 (2006) reported on the active members of the LINE family Lib to cause spontaneous 169 mutations in callus and meristem stem cells in sweetpotato. We observed similar sequence 170 fragments (IbLi-1 and IbLi-2) having higher copy numbers of up to ~4100 in the sweetpotato 171 cv. Beauregard. Generally, the abundance of LINEs in plant genomes have been attributed to 172 the fact that they can proliferate by participating in double-strand DNA break repair besides 173 TPRT (as reviewed in Eickbush 2002). By their very nature, LINEs have transposition rates that exceed their rates of decay and loss or excision (Arkhipova and Meselson 2005). In 174 sweetpotato and other asexually propagated crop species (McKey et al. 2010), the 175 176 accumulation of actively mobilising retrotransposons dictated by such molecular mechanisms 177 can be extraneously enhanced by continuous clonal propagation over seasons.

The partial LTR sequence fragments, IbLtr-1 (205 bp) and IbLtr-2 (151 bp), showed high homology (97% and 96% identity, respectively) to similar sequences of sweetpotato from which their primers were designed (Fig. 1, 4*a* and 4*b*). The small variations between the sequences were mainly due to base substitutions, and also a single arginine duplication in IbLtr-2 resulting in a stop codon and frameshift. The dot blot hybridization revealed genome copy numbers of ~50 and 1600 for the partial LTR sequences IbLtr-1 and IbLtr-2,

184 respectively. Although LTR retrotransposons are abundant in plant genomes, only a few 185 remain active. Most are either transcriptionally silenced in highly methylated regions of the genome (Kumar and Bennetzen 1999), functionally inactivated by the presence of stop 186 187 codons and frameshifts in the coding regions or, particularly LTR retrotransposons, may be 188 depleted by excision leaving behind solo LTRs (Arkhipova and Meselson 2005; Soleimani et 189 al. 2006). Considering that full length elements are flanked by two LTRs, the estimated copy 190 numbers of IbLtr-2 (>1600) may in effect be ~800, or considerably lower when discounting 191 solo LTRs.

192 The 190 bp product of the partial RT sequence IbRt-1 matched a portion of the 193 conserved region of the RT domain ranging from residue 6 to residue 63 (Fig. 5). This RT 194 domain falls under the RVT_2 superfamily of the RNA-dependent DNA polymerase (RdDP, 195 CL0027) clan (Xiong and Eickbush 1990). Apart from a few amino acid substitutions, the 196 sequence was found to be homologous (73.8% identity, Fig. 5) to *copia*-like *pol* polyproteins 197 of angiosperms such as mungbean (Vigna radiata, accessions AAT90446, AAT90451, etc.) 198 and rice (Oryza sativa, accessions AAP46197, ABF96216, and T03664), and three 199 homologous sequences AAF37863, AAF37864 and AAF61082 from sweetpotato. Its 200 similarity to those from various dicot and monocot species supports the view of earlier 201 horizontal transfers as proposed by Xiong and Eickbush (1990). Further, the detection of 202 \sim 800 genomic copies (Table 3), and also other similar sequences from sweetpotato may 203 indicate the presence of a Ty1/copia-like element lineage apart from those that have been 204 reported (Tahara et al. 2004; Villordon et al. 2000) as RT sequences have been shown to be 205 strongly correlated with the terminal structure of the elements (Xiong and Eickbush 1990). 206 The genome copy number of IbRt-1 based on cv. Beauregard was estimated to be ~800 207 (Table 2).

The retrotransposons identified in this study were also PCR-amplified from several other sweetpotato cultivars from various countries in our working collection; an indication that they are widespread in the species. Although such elements have been demonstrated to be up-regulated by stress associated with tissue culture (Tahara et al. 2004; Yamashita and Tahara 2006) and virus infection (Villordon et al. 2000), the question remains, however, as to whether the isolated partial sequences reported here represent elements that are transcriptionally active under normal and/or stressful conditions.

215

216 Acknowledgement

For timely assistance in our lab work we thank Dr. Rosamond Godwin and Mrs Sharon Williams. TO was supported by an Australian Centre for International Agriculture Research scholarship under the John Allwright Fellowship scheme.

220

221 **References**

- Abdel-Ghany, A.A., and Zaki, E.A. 2002. Cloning and sequencing of an envelope-like gene
 in Gossypium. Planta 216: 351-353.
- Arkhipova, I., and Meselson, M. 2005. Deleterious transposable elements and the extinction
 of asexuals. BioEssays 27(1): 76-85. doi:10.1002/bies.20159.
- Arumuganathan, K., and Earle, E. 1991. Nuclear DNA content of some important plant
 species. Plant Molecular Biology Reporter 9(3): 208-218-218.
 doi:10.1007/bf02672069.
- Chen, Z.J. 2007. Genetic and epigenetic mechanisms for gene expression and phenotypic
 variation in plant polyploids. Annu. Rev. Plant Biol. 58: 377-406.
 doi:10.1146/annurev.arplant.58.032806.103835.

- Connolly, A.G., Godwin, I.D., Cooper, M., and DeLacy, I. 1994. Interpretation of randomly
 amplified polymorphic DNA marker data for fingerprinting sweet potato (Ipomoea
 batatas L.) genotypes. Theoretical and Applied Genetics 88: 332-336.
- Drummond, A.J., Ashton, B., Cheung, M., Heled, J., Kearse, M., Moir, R., Stones-Havas, S.,
 Thierer, T., and Wilson, A. 2009. Geneious v4.7.
- Eickbush, T.H. 2002. Repair by retrotransposition. Nat Genet **31**(2): 126-127.
- Eickbush, T.H., and Jamburuthugoda, V.K. 2008. The diversity of retrotransposons and the
 properties of their reverse transcriptases. Virus Research 134(1-2): 221-234. doi:DOI:
 10.1016/j.virusres.2007.12.010.
- Hawkins, J.S., Hu, G., Rapp, R.A., Grafenberg, J.L., and Wendel, J.F. 2008. Phylogenetic
 determination of the pace of transposable element proliferation in plants: copia and
 LINE-like elements in Gossypium. Genome 51: 11-18.
- Kejnovsky, E., Hobza, R., Cermak, T., Kubat, Z., and Vyskot, B. 2009. The role of repetitive
 DNA in structure and evolution of sex chromosomes in plants. Heredity **102**(6): 533541.
- Kumar, A., and Bennetzen, J.L. 1999. PLANT RETROTRANSPOSONS. Annual Review of
 Genetics 33(1): 479-532. doi:10.1146/annurev.genet.33.1.479.
- La Bonte, D.R., Villordon, A.Q., and Fajardo, D.S. 2001. Mutations in sweetpotato. *In* Proceedings of International Workshop on Sweetpotato Cultivar Decline. *Edited by* Y.
 Nakazawa and K. Ishiguro. Kyushu National Agriculture Experimental Station,
- 252 Miyakonojo, Miyazaki, Japan. pp. 1-9.
- McKey, D., Elias, M., Pujol, B., and Duputié, A. 2010. The evolutionary ecology of clonally
 propagated domesticated plants. New Phytologist 186(2): 318-332.
 doi:10.1111/j.1469-8137.2010.03210.x.

- Rozen, S., and Skaletsky, H.J. 2000. Primer3 on the WWW for general users and for biologist
 programmers. *In* Bioinformatics Methods and Protocols: Methods in Molecular
 Biology. *Edited by* S. Krawetz and S. Misener. Humana Press, Totowa, NJ. pp. 365386.
- Saitou, N., and Nei, M. 1987. The neighbor-joining method: A new method for
 reconstructing phylogenetics trees. Mol. Biol. Evol. 4: 406–425.
- Sakai, H., Tanaka, T., and Itoh, T. 2007. Birth and death of genes promoted by transposable
 elements in Oryza sativa. Gene **392**(1-2): 59-63. doi:DOI:
 10.1016/j.gene.2006.11.010.
- Soleimani, V.D., Baum, B.R., and Johnson, D.A. 2006. Quantification of the retrotransposon
 BARE-1 reveals the dynamic nature of the Barley genome. Genome 49: 389-396.
- Tahara, M., Aoki, T., Suzuka, S., Yamashita, H., Tanaka, M., Matsunaga, S., and Kokumai,
 S. 2004. Isolation of an active element from a high-copy-number family of
 retrotransposons in the sweetpotato genome. Molecular Genetics and Genomics 272:
 116-127
- Tanaka, M., Yoshinaga, M., Nakazawa, Y., and Nakatani, M. 2001. Identification of a novel
 SINE-like sequence in retrotransposon fragments of sweetpotato. Plant Biotechnology
 Reports 18: 157-161.
- Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: improving the
 sensitivity of progressive multiple sequence alignment through sequence weighting,
 position specific gap penalties and weight matrix choice. Nucleic Acids Research 22:
 4673-4680.
- Vicient, C., Kalendar, R., and Schulman, A. 2001a. Envelope-class retrovirus-like elements
 are widespread, transcribed and spliced, and insertionally polymorphic in plants.
 Genome Res 11: 2041 2049.

- Vicient, C.M., Jaaskelainen, M.J., Kalendar, R., and Schulman, A.H. 2001b. Active
 Retrotransposons Are a Common Feature of Grass Genomes. Plant Physiol. 125(3):
 1283-1292. doi:10.1104/pp.125.3.1283.
- Villordon, A.Q., Jarret, R.L., and LaBonte, D.R. 2000. Detection of Ty1-copia-like reverse
 transcriptase sequences in Ipomoea batatas (L.) Poir. Plant Cell Reports 19(12): 12191225. doi:10.1007/s002990000260.
- Wright, D.A., and Voytas, D.F. 1998. Potential retroviruses in plants: Tat1 is related to a
 group of Arabidopsis thaliana Ty3/gypsy retrotransposons that encode envelop-like
 proteinsGenetics, 149: 703-715. Genetics 149: 703-715.
- Xiong, Y., and Eickbush, T.H. 1990. Origin and evolution of retroelements based upon their
 reverse transcriptase sequences. EMBO Journal 9: 3353-3362.
- Yamashita, H., and Tahara, M. 2006. A LINE-type Retrotransposon Active in Meristem Stem
 Cells Causes Heritable Transpositions in the Sweet Potato Genome. Plant Molecular
 Biology 61(1): 79-84. doi:10.1007/s11103-005-6002-9.



Fig. 1. Detection of LTR and non-LTR retrotransposons studied in sweetpotato cultivars Beauregard and Northern star. L= DNA ladder; M5 = IbLtr-1; M27 = M28 = IbGy-2; M29 = IbLi-1; M30 = IbLi-2; M35 = IbLtr-2; and M36 = Ibrt-1.

•

a)

AJ279072 AJ298028 AJ298029 AJ298030 AJ298030 AJ298032 AJ298032 AY257164 DQ109566 DQ109566 DQ109571	1	10	20	30	40	50	οņ		
DQ109572	CCAACC	TOTATCCCA	CTTCCAACCC			CATCTTACACO	TTAACAC		
1509-1	CCAAGG	70	80	90	100	110	120 122		
AJ279072 AJ298028 AJ298030 AJ298030 AJ298031 AJ298032 AY257164 DQ109564 DQ109566 DQ109571 DQ109572		1	1		A G C A T T G G T A A G C A T T G G T A A G C A T T G G A G C A T T G G T A A G C A T T G G T A A G C A T T G G T A A G C A T T G G T A A G C A T T G G T A A G C A T T G G	TATGGGCAATC TATGGGCAATC TATGGGCAATC TATGGGCAATC TATGGGCAATC TATGGGCAATC TATGGGCAATC TATGGGCAATC TATGGGCAATC TATGGGCAATC TATGGGCAATC	CCCTTC CCCCTTC CCCCTTC CCCCTTC CCCCTTC CCCCTTC CCCCTTC CCCCTTC CCCCTTC		
IbGy-1	CTACAG	GAACTAGCI	AAGGGGACAA	GAGGAAAAA	TAAGCATTGG	TATGGGCAATC	CCCTTG		
b)	1	10 20	30	40	50	80 70	80	90 100	110 120
AF378016 AF378017 AJ295132	I		ľ	TTCRGTARC TTTRGTARC	CTEGATGGSTAT CTEGATGGATAT	ICAGGAMATCATCA ITCGGAMATCATCA	AATTCCCATCCAT AATTCCCATCCAT	ICCGG AGG ACCAGAG TAA ICCGG AGG ACTAGAG TAA	GA CHACG THCACANGH CCAN GA CHACG THCACANGH CCAN MA CCACG DHCMCANGH CCMM
lbGv-2	CAAAACTAA	CTTGCTTGAAG	ATGTCTTGCACAAA	TATTTTCCTTC	CTAGATGGITHT	CAGGGTATTTCCA	GATTICATCCA	CCAG ANG ANCAAG AAAA	GACCACG TTCACATGT CAAT
	130	140	150	160	170 180	190	200	210 220	230 240
AF378016 AF378017 AJ295132	atgcacca Atgcacca	ATGCGTATCGT ATGCGTATCAT	AGGA TGCCCT TTGC AGGA TGTCCT TTGC	ACTG TGCAACAC ACTG TGCAACGC	Test geate true Test geate true	CAAAGGTGTATGA CAA-GGTGTATGA	ngtigtäntinneti Ngtigtäntinneti	GGACANGANCGAGGATA GGACANGANCGAGGATA	TCATGGAAGTCTTCATGGAT TCATGGAAGTCTTCATGGAT
AJ421232 lbGy-2	ACGGAACAT	ACCCCTACAGG ATCCATTCAAA	a a a a tgCCTTTCGG Agga tgTCATTTGG	GOTG TGMAANGC CMTG TGMAACGC	TCCAGCGACCTTI GCCGACCACCTTI GCCGACCACCTTI	CAACCTGCATGA CAACGATGCATGA	TGTCAATTTTCAC TGGCAGTCTTTT	TG ACCTA ATT AAA GACA CG AG CTTGTGGAGGAAA	TAATGGAAGTMTTCATGGAC TAATAGAAGTGTTCATGGAT
	250	260	270	280 29	u 300	310	320	330 340	350 360 364
AF378016 AF378017 AJ295132 AJ421232 IbGv-2	GACTICICG GACTICICG GATICAGC	encharggaaa Gecearggaaa Gecearggaa Gecearggaa	gantingggncath Gaitincggncath Cticintagngtca Cticintagngtca	GTCNGCAGAATO GTCNGCAAAATO GTTNGTCAAACT	MAGACAAAGACTI RAGACAAAGACTI MGTGCAGGCATCI MMCACAGAGCATCI	VACAACGATGCCAA VACAACGATGCCAA VGAAGGATGCGAG	GMAAAGGMCCHAA GMAAAGGMCCHGA CHCA GMGAAGCATCHCA AAGCACANDCHCA	The CTITA A CTIGGA A A A A A TTC CTITA A CTIGGA A A A A TTC CTICA A CTIGGA GA A A A TTC CTICA A TTIGGGA A A A A	ng Channa Cangga Rgacanna Cangga Rgachar Canga Rgacana Cangga Rgacanna Cangga Canna Cangga Rgacana Cangga Rgacanana Cangga Rgacana Cangga Rgacana Cangga Rga

Fig. 2. Nucleotide sequence alignment showing the relationships of other reverse transcriptase sequences of retrovirus-like elements with *a*) IbGy-1 and *b*) IbGy-2 detected in sweetpotato. Gaps and stop codons are indicated as (-) and (*) respectively. The degree of shading indicate degree of nucleotide conservation ranging from black (100% conservation), dark grey (\geq 80%), light grey (\geq 60%), and no shading (<60%).

a)

	1 10	20	30	40	50	60	70	80 84
AB231837	AKMDRSHRMPVF	LKTASCFLITFK	RLKSPTFFAK.	ATNVRISWPT	SVNLHLGEF	RLFWNGRPMI [*]	GFFSRGTR*A	*PLADGDNSR
AB231838						PMI -	GFFSRGTR*A	*PLADGDNSR
AB231839	AKMDRSHRMPVF*	LKTASCFLITFK	RLKSPTFFAK.	ATNVRISWPT	SVNLHLGEF	RLFWNGRPMI	GFFSRGTR*A	*PLADGDNSR
AB231840	AKMDRSHRMPVF*	LKTASCFLITFK	RLKSPTFFAK.	ATNVRISWPT	SVNLHLGEF	RLFWNGRPMI [*]	GFFSRGTR*A	*PLADGDNSR
AB231841	AKMDRSHRMPVF*	LKTASCFLITFK	RLKSPTFFAK.	ATNVRISWPT	SVNLHLGEF	RLFWNGRPMI [*]	GFFSRGTR*A	*PLADGDNSR
AB231842	AKMDRSHRMPVF*	LKTASCFLITFK	RLKSPTFFAK.	ATNVRISWPT	SVNLHLGEF	RLFWNGRPMI"	GFFSRGTR*A	*PLADGDNSR
AB231843	AKMDRSHRMPVF*	LKTASCFLITFK	RLKSPTFFAK.	ATNVRISWPT	SVNLHLGEF	RLFWNGRPMI"	GFFSRGTR*A	*PLADGDNSR
AB231844	AKMDRSHRMPVF*	LKTASCFLITFK	RLKSPTFFAK.	ATNVRISWPT	SVNLHLGEF	RLFWNGRPMI*	GFFSRGTR*A	*PLADGDNSR
AB231845	AKMDRSHRMPVF*	LKTASCFLITFK	RLKSPTFFAK.	ATNVRISWPT	SVNLHLGEF	RLFWNGRPMI [*]	GFFSRGTR*A	*PLADGDNSR
AB231846	AKMDRSHRMPVF*	LKTASCFLITFK	RLKSPTFFAK.	ATNVRISWPT	SVNLHLGEF	RLFWNGRPMI*	GFFSRGTR*A	*PLADGDNSR
Ibli-1	AKMDRSHRM*AI*	L-DCSCFLITFK	RLKSPTFFAK	ATNVRISWPT	SVNLHLGEF	LFWNGRPMI	RFFSRGTR*A	*PLADGDNSR
<i>b</i>)								
<i>b</i>)							-	
<i>b</i>)	1 10	20	30	40	50	60	70	80
<i>b)</i> АВ231841	1 LWLWSGRASRTSI	20 LPPR* TNCTGNDGY	30 YW* QH*NCGGYG	40 GTTTC*RSGR	50 CRNATSRTP:	60 SIWTIDDCDQK	70 TTSHYKQBTSA	80 AGHEQITKSEQ*
b) AB231841 AB231839	1 <u>LWLWSGRASRTS</u> LWLWSGRASRTS	20 LPPR* TNCTGNDGY LPPR* TNCTGNDGY	30 YM* QH* NCGGYG YM* QH* NCGGYG	40 36 <u>TTTC</u> * <u>RSGR</u> 36TTTC*RSGR	50 CRNATSRTP: CRNATSRTP:	80 SLWTLDDCDQK SLWTLDDCDQK	70 TTSHYKOPTSA TTSHYKOPTSA	80 A <u>ĠHEQTKSEQ</u> * AGHEQTKSEQ*
b) AB231841 AB231839 AB231840	1 10 LWLWSGRASRTS LWLWSGRASRTS LWLWSGRASRTS	20 LPPR* Inctender LPPR* Inctender LPPR* Inctender	30 YW* QH* NCGGYG YW* QH* NCGGYG YW* OH* NCGGYG	40 36 TTTC * RSGR 36 TTTC * RSGR 36 TTTC * RSGR	50 CRNATSR TP CRNATSR TP CRNATSR TP	60 SLWTLDDCDQK SLWTLDDCDQK SLWTLDDCDQK	70 TTSHYKOPTSA TTSHYKOPTSA TTSHYKOPTSA	80 AGHEOTKSEQ* AGHEOTKSEQ* AGHEOTKSEO*
b) AB231841 AB231839 AB231840 Ibli-2	1 10 LWLWSGRASRTS LWLWSGRASRTS LWLWSGRASRTS LWLWSGRASRTS	20 LPPR* IN CIGNDGY LPPR* IN CIGNDGY LPPR* IN CIGNDGY LPPR* IN CIGNDGY	30 YM+QH+NCCCYC YM+QH+NCCCYC YM+QH+NCCCYC YM+QL+ICCCYI	40 GTTTC*RSGR GTTTC*RSGR GTTTC*RSGR DRTTTC*RSGR	50 CRNATSR TP: CRNATSR TP: CRNATSR TP: CCNATSR TP:	60 SLWTLDDCDQK SLWTLDDCDQK SLWTLDDCDQK SLWTLDDCDQK	70 TTSHYKOBTSA TTSHYKOBTSA TTSHYKOPTSA TTSHYKOTTSA	80 AGHEQTKSEQ* AGHEQTKSEQ* AGHEQTKSEQ* AGHE*TKSDO*
b) AB231841 AB231839 AB231840 Ibli-2	1 10 LWLWSGRASRTS LWLWSGRASRTS LWLWSGRASRTS 90 100	20 LPPR* TNCTGNDGY LPPR* TNCTGNDGY LPPR* TNCTGNDGY LPPR* TNCTGNDGY 110	30 XM*QH*NCGGYG XM*QH*NCGGYG XM*QH*NCGGYG XM*QL*ICGGYI 120	40 GTTTC*RSGR GTTTC*RSGR GTTTC*RSGR DRTTTC*RSGR 130	50 CRNATSRTP: CRNATSRTP: CRNATSRTP: CCNATSRTP: 140	60 SLWTLDDCDQK SLWTLDDCDQK SLWTLDDCDQK 150	70 TTSHYKOBTSA TTSHYKOBTSA TTSHYKOTTSA 180	80 AGHEOTKSEQ* AGHEOTKSEQ* AGHEOTKSEQ* AGHE*TKSDQ* 170 177
b) AB231841 AB231839 AB231840 Ibli-2	1 10 LWLWSGRASRTS LWLWSGRASRTS LWLWSGRASRTS 200 100	20 LPPR* TNCTGNDGY LPPR* TNCTGNDGY LPPR* TNCTGNDGY LPPR* TNCTGNDGY 110	30 XM+ OH+ NCGGYG XM+ QH+ NCGGYG XM+ QH+ NCGGYG XM+ QL+ ICGGYI 120	40 GGTTTC+RSGR GGTTTC+RSGR GGTTTC+RSGR DRTTTC+RSGR 130 DRUD+DTCC+0	50 CRNATSRTP: CRNATSRTP: CRNATSRTP: 140	60 SLWTLDDCDQK SLWTLDDCDQK SLWTLDDCDQK 150	70 TTSHYKOPTSA TTSHYKOPTSA TTSHYKOPTSA TTSHYKOTTSA 160	80 AGHEOTKSEO* AGHEOTKSEO* AGHEOTKSEO* AGHE*TKSDO* 170 177
<i>b)</i> AB231841 AB231839 AB231840 Ibli-2 AB231841	1 10 LWLWSGRASRTS LWLWSGRASRTS LWLWSGRASRTS 100 KHITATSTEHYNI	20 LPPR* INCIGNDGY LPPR* INCIGNDGY LPPR* INCIGNDGY 110 R*SHDEOWRGEG*	30 YW+ OH+ NCGGYG YW+ OH+ NCGGYG YW+ OH+ NCGGYG YW+ OL+ ICGGYI 120 TTDNRVAKHOAC	40 GGTTTC+RSGR GGTTTC+RSGR GGTTTC+RSGR DRTTTC+RSGR 130 RKR+DTGGAC	50 CRNATSRTP: CRNATSRTP: CRNATSRTP: CCNATSRTP: 140 W* LOR* VSY	60 SLWTLDDCDQK SLWTLDDCDQK SLWTLDDCDQK 150 SYLPICSCPTE	70 TTSHYKQETSA TTSHYKQETSA TTSHYKQETSA 160 FQEWQRELGEH	80 AGHEOTKSEO* AGHEOTKSEO* AGHEOTKSEO* AGHE*TKSDO* 170 177 VNSSLSLVFE
<i>b)</i> AB231841 AB231839 AB231840 Ibli-2 AB231841 AB231841 AB231839	1 10 LWLWSGRASRTS LWLWSGRASRTS LWLWSGRASRTS 20 100 KHITATSTEHYNI KHITATSTEHYNI	20 LPPR* INCIGNDG LPPR* INCIGNDG LPPR* INCIGNDG LPPR* INCIGNDG 110 R*SHDEQVRGEG* R*SHDEQVRGEG*	30 YW+QH+NCGGYG YW+QH+NCGGYG YW+QH+NCGGYG YW+QL+ICGGYI 120 TTDNRVAKHQAC	40 GTTTC * RSGR GTTTC * RSGR GTTTC * RSGR 130 RKR * DIGGAC RKR * DIGGAC	50 CRNATSRTP: CRNATSRTP: CRNATSRTP: CRNATSRTP: 140 M* LOR* VSY: W* LOR* VSY:	60 SLWTLDDCDQK SLWTLDDCDQK SLWTLDDCDQK 150 SYLPICSCPTE SYLPICSCPTE	70 TTSHYKQPTSA TTSHYKQPTSA TTSHYKQPTSA TTSHYKQTTSA 160 FQPWQRELGEH F*PWQRELGEH	80 AGHEQTKSEQ* AGHEQTKSEQ* AGHEQTKSEQ* AGHE*TKSDQ* 170 177 VNSSLSLVFF VNSSLSLVFF
<i>b)</i> AB231841 AB231839 AB231840 Ibli-2 AB231841 AB231839 AB231840	1 10 LWLWSGRASRTS LWLWSGRASRTS LWLWSGRASRTS 90 100 KHITATSTEHYNI KHITATSTEHYNI KHITATSTEHYNI	20 LPPR* INCIGNDGY LPPR* INCIGNDGY LPPR* INCIGNDGY 110 R*SHDEQVRGEG* R*SHDEQVRGEG* R*SHDEQVRGEG*	30 YM+QH+NCGGYG YM+QH+NCGGYG YM+QH+NCGGYG YM+QL+ICGGYI 120 PTDNRVAKHQAC TTDNRVAKHQAC	40 GTTTC+RSGR GTTTC+RSGR DRTTTC+RSGR 130 RKR+DIGGAC RKR+DIGGAC RKR+DIGGAC	$ \begin{array}{c} 50\\ \hline \mathbf{CRNATSRTP}\\ \hline \mathbf{CRNATSRTP}\\ \hline \mathbf{CRNATSRTP}\\ \hline \mathbf{CRNATSRTP}\\ \mathbf{CNATSRTP}\\ 140\\ \hline \mathbf{W}^{+} \mathbf{LQR}^{+} \mathbf{VSY}\\ \hline \mathbf{W}^{+} \mathbf{LQR}^{+} \mathbf{VSY}\\ \hline \mathbf{W}^{+} \mathbf{LQR}^{+} \mathbf{VSY} \end{array} $	60 SLWTLDDCDQK SLWTLDDCDQK SLWTLDDCDQK 150 SYLPICSCPTE SYLPICSCPTE SYLPICSCPTE	70 TTSHYKQETSA TTSHYKQETSA TTSHYKQETSA TTSHYKQTTSA 160 FQEWQRELGEH F* PWQRELGEH F* PWQRELGEH	80 AGHEQTKSEQ* AGHEQTKSEQ* AGHETKSEQ* 170 177 VNSSLSLVFF VNSSLSLVFF VNSSLSLVFF

Fig. 3. Sequence alignment of amino acids showing the relationship of the partial LINE sequences *a*) IbLi-1 and *b*) IbLi-2 to various *Lib* LINEs of sweetpotato. The degree of shading indicate degree of residue conservation ranging from black (100% conservation), dark grey (\geq 80%), light grey (\geq 60%), and no shading (<60%).

<i>a</i>)									
AB167515 AB167520	1 CLDMVRVRDGPMLLV	20 VRGVPGLSAEN	30 IVGSS H IR*H	40 TKFSKLQYLN	GVRNVRHLQ LGTPK	60 CLNKIIXCIV V*TKYLLI	70 VPTE*AN VGPYILG*		
IbLtr-1	CLDMVRVRDGPMLLV	VRGVLAVGRKC	WQQPH *VAH	-KV*QATIFE	WS*EC*APT	MFEQN-YLLI	GPNILG*		
<i>b</i>)									
	1 10	20	30	40	50	60	70	80	90 91
AB167518	LGKSKGV <mark>R</mark> HĠV-RHL	PKYLLFGP*TI	_D * L Q N	I	D R Y		SSTINŔ – – – – E	VICC	2SCHPKS
AB167519	LGKSKGV <mark>R</mark> HGV-RHL	PKYLLFGP*TI	LD*LQN	I	\mathbf{PL}		FLNYK*GGHLP	FLSSQI	
AB167513	LGKSKGV <mark>G</mark> HGV-RHL	PKYLLFGP*TI	_D*LQN	I	DPL		FLNYK*GGHLS	FLSSQI	
AB167517			IYCL	IPKLWTNYKM	F PI PQL*IGF	RSFAILVIPN			
lbLtr-2	LGKSKGVRHGVRRYL	PKYLLFGP*TI	D*LOND		-BY	S	-STINRE	VFCC	SCHPKS

Fig. 4. Sequence alignment of amino acids showing the level of similarity between *a*) IbLtr-1 and *b*) IbLtr-2 and other known LTR sequences of sweetpotato. Gaps are indicated as (-). The degree of shading indicate degree of nucleotide conservation ranging from black (100% conservation), dark grey (\geq 80%), light grey (\geq 60%), and no shading (<60%).



Fig. 5. A dendrogram showing the relationship between IbRt-1 and similar sequences found in other plant species based on the amplified Tycopia RT domain. Confidence of groupings was based estimated in percentage using 500 bootstrapping replication with a threshold value of 50 %. The bootstrap values are shown above the branches. The scale bar represents the proportion of amino acid substitution per site.

Name	GenBank	Retrotransposon	Transposition/ insertion	Estimated copy	Reference
	accession	group		number	
Tib1-Tib 31	AF223309-	LTR Tyl/copia	Near DFR-B gene; active in	NR	Tanaka et al (2001)
	AF223326,		callus and normal plant cells		
	AF223337-AF223345				
IBRT1-IBRT5	AF152900-AF152904	LTR, Ty1/copia	Activated by virus infection	NR	Villordon et al. (2000)
Rtsp-1	AB162659	LTR Ty1/copia	Active in callus cells	396	Tahara et al. (2004)
Lib	AB231839, B231840,	Non-LTR LINE	Active in callus and normal	108	Yamashita and Tahara (2006)
	AB231842-		plant cells		
	AB231846				
IPSE1	AF295596	Non-LTR SINE	Within retrotransposon	NR	Tanaka et al. (2001)
			fragment		

 Table 1 Various active retrotransposons identified in sweetpotato^{*}.

*LINE = Long interspersed element; SINE = Short interspersed element; LTR = Long terminal repeat; NR = not reported.

Table 2 The sequences of primers used in the study.

Name	Retrotransposon		Primer	Annealing	Author
	group/ domain*	Code	sequence (5'3')	temperature (°C)	
IbGy-1	Gypsy-like, ENV	KS1F	CCAAGGTCTATGGGACTTGGAACC	60.0	Vicient <i>et al.</i> (2001)
		KS2R	CAAGGGGATTGCCCATACCAATGC		
IbGy-2	Gypsy-like, RT	KS3p1F	AARGAYCAYTWYCCIYTICCITT	56.0	Vicient et al. (2001)
		KS4p2R	ACCATRAARTGRCAYTTYTCCCARTT		
IbLi-1	LINE	M29F	GCAAAGATGGACCGGTCACAC	60.0	This study
		M29R	CAAGTCCCGGGAGTTATCGCCG		
IbLi-2	LINE	M30F	CTTTGGTTGTGGAGTGGTAGG	60.0	This study
		M30R	CGAAGAATACGAGGGAGAGG		
IbLtr-1	LTR	M5F	TGCTTAGACATGGTTAGGGTC	57.7	This study
		M5R	GTTAGCCTAAAATGTAGGGACCA		
IbLtr-2	LTR	M35F	AGGCAAGTCAAAAGGAGTTAGG	60.0	This study
		M35R	GATTTGGGATGACAAGATTGG		
IbRt-1	RT	M36F	CCAACCTGAGGGTTTTTCAG	60.0	This study
		M36R	AACTTTCTCCCACTGACCTTC		

*ENV = envelop; RT = reverse transcriptase; LINE = Long interspersed element; LTR = Long terminal repeat.

Name	Sequence length (base pairs)	Proportion of genomic DNA hybridising to probe $(\times 10^{-4})^*$	Copy number estimate	Sequence domain [†]	Putative retroelement group
IbGy-1	128	1.681	2100	Partial ENV	Gypsy-like retroelement
IbGy-2	364	1.237	540	Partial RT	Gypsy-like retroelement
IbLi-1	258	6.585	4100	Partial LINE	Non-LTR LINE
IbLi-2	532	2.049	600	Partial LINE	Non-LTR LINE
IbLtr-1	205	0.058	50	Partial LTR	LTR retrotransposon
IbLtr-2	151	1.569	1600	Partial LTR	LTR retrotransposon
Ibrt-1	190	0.924	800	Partial RT	LTR retrotransposon

 Table 3 Retroelements detected in sweetpotato.

*Sweetpotato haploid genome size = 1597 Mb (Arumuganathan and Earle 1991).

 † ENV = envelope; RT = reverse transcriptase; LINE = Long interspersed element; LTR = Long terminal repeat.

References

- Arumuganathan, K., and Earle, E. 1991. Nuclear DNA content of some important plant species. Plant Molecular Biology Reporter **9**(3): 208-218-218. doi:10.1007/bf02672069.
- Tahara, M., Aoki, T., Suzuka, S., Yamashita, H., Tanaka, M., Matsunaga, S., and Kokumai, S. 2004. Isolation of an active element from a highcopy-number family of retrotransposons in the sweetpotato genome. Molecular Genetics and Genomics **272**: 116-127
- Tanaka, M., Yoshinaga, M., Nakazawa, Y., and Nakatani, M. 2001. Identification of a novel SINE-like sequence in retrotransposon fragments of sweetpotato. Plant Biotechnology Reports **18**: 157-161.
- Vicient, C., Kalendar, R., and Schulman, A. 2001. Envelope-class retrovirus-like elements are widespread, transcribed and spliced, and insertionally polymorphic in plants. Genome Res **11**: 2041 2049.
- Villordon, A.Q., Jarret, R.L., and LaBonte, D.R. 2000. Detection of Ty1-copia-like reverse transcriptase sequences in Ipomoea batatas (L.) Poir. Plant Cell Reports **19**(12): 1219-1225. doi:10.1007/s002990000260.
- Yamashita, H., and Tahara, M. 2006. A LINE-type Retrotransposon Active in Meristem Stem Cells Causes Heritable Transpositions in the Sweet Potato Genome. Plant Molecular Biology **61**(1): 79-84. doi:10.1007/s11103-005-6002-9.

	1 10	20	30	40	50	60 63
Consensus Identity	SNLXXESSKGK	EHLVCKLNKSIY(<u> SLKQASRQWY</u>		EENIMDQCL	QKVSGSKX
1. AAF37863	FSVEGKI	EHMVCKLNRSIY	LRQASRQWY	LKENDTVISFGI	KENTVDQCI	LKVSGSKF
2. AAF 37864 3. AAF 61082	ESVEGKI	EHMVCKLNKSIY	LKQASRQWY	IKFDGVISSFGI	FAENPLDHCI	ICKVSGRKE
4. AAG44348			LKQASRQWY	LKFDEVIRSEG	FEENKIDRCI	LKVSGSKF
5. AAP46197 6. AAT90446	EVMEGKI FTTGKI	JHLACKLKKSIY. Enmvckikksiy(LKQASRQWY I KOASROWY	LKEDELIKKEG	FVENEVDNCI	TKVSCSKE
7. AAT90451	FSSKDSI	EHLVCKLNKSIY	LKQASRQWY	LKFDDVITSFG	FEENIMDQCI	QKVSGSKI
8. AAT90455	F S S K D S I	EHLVCKLNKSIY(<u> </u>	LKFDDVITSFG	FEENVMDQCI	IQKVSGSKI
10. AAT90456	FTITGKI	ENLVCKLKKSIY(JIKOASROWI	LKFNDTITSYG	TVENTVDRCI	TKVSGSKF
11. AAT90459	ESSKDSI	EHLVCKLNKSIY(JLKQAFRQWY	LKFDDVITSFG	FEENIMDQCI	QKVSGSK
12. AAT90466 13. AAT90468	ESSKDSI	EHLVCKINKSIY(J L K Q A S R Q W Y	LKFDNVINSFG	FEENIMDQCL	IQKVSGSKI IOKVSG
14. AAT90470	ESSKDSI	EHLVCK L NKSIY(LKQASRQWY	LKFHDVITSFG	FEENVMDQCI	QKVSGSKI
15. AAT90471	Reeknei	UT VOUT NUCTY	LKQASRQWY	LKFHDVITSFG	EENVMDQCI	IQKVSGSKI Iokvscski
17. AAT90484	LOSVDSI	CUTACV T MV2TI	LKQASRQWY	LKFDDVITSFD	FEENIMDQCI	OKVSGSKI
18. AAT90485	ETISGKI	ENLVCKLKKSIY(LKQASRQWY	LKFKDTITSYG	VENTVDRCI	INVNGSKF
19. AA190486 20. AAT90489	ESSKDSI	CHLVCK NKSTY(LKQASRQWY TKOASHOWY	LKFHDVITSFG	FRENTMOOCT	IQKVSGSKI IOKVSGSKI
21. AAT90492	FSSKDS	GHLVCKLNKSIY(LKQASRQWY	LKFDDVITSFG	FQENIMDQCI	QKVSGSKI
22. ABF96216	EVMEGKI	DHLACKLKKSIY!	ELKQASRQWY I KOASOOWY	LKFDEIIKRFG	FKENEVDNCI	TKTKGGKF
24. ABW74556	LUVOVI	CNIVCK M KKSII(LKQASROWY	FKFNDIITSYG	IVEIIVDRCI	YKVSGR
25. BAB47197			LKQASRQWY	FKFHEVISSEG	VENPMDQCI	OKISGSK
26. BAB47198 27. BAB47203			LKQASRQWY LKOASROWY	EKFHEVISSEG IKFHDTISSEG	FVENPMDQCI FVENVMDOCI	OKVSGSK
28. BAB47204	LESESVKGKI	EHMVCKLKKLIY	LKQASRQWY	LKFNNTITSFG	FQENTIDRCI	MKVSGSK F
29. BAB83551 30. BAG72096	ESSKDGI	EHLGCKLNKSIY(J L K Q A S R Q W Y	KKEHKVISSEGI	FEENIMDQCI	ILKVSGSKI HKVSGSK
31. CAA04615	ESCKGKI	ЕНМУСК Ц ККЅLҮІ	RLKQASRQWY	LKFNETIVTFE	FKKNTVDRCI	LKVSG
32. CAN80961	R TMRCRI		LKQASRQWY	LKEDRIITQNG IKEDRITKREG	FKENTVDRCI KENENDNCI	ILRVSGSSY TEURCOER
34. XP 002263890	LUFCUL		LKQASROWY	LKFHNIISSFGI	FVENVMDQCI	LKVSGSK
35. XP_002273044	EKEKGKI	ENMVCRLKRSIY	LKQASRQWY	LKFDKIVTSFG	FIENKFDOCV	MKVNGSKY
36. XP_002276920 37. lbRt-1	SNLRVFOLKERI	UTWCVNLTDRYTI	DIDKASRQWY	LKFHNLLSSEG LKFNDTVMSEG	FKENTVDOCT	LKVSGSK
	RVT 2 superfa	milv				

Fig. S1. Alignment of amino acid residues of IbRt-1 and similar sequences from *Vigna radiata* (AAT90446, AAT90451, etc.), *Diospyros kaki* (BAB47197, BAB47198, BAB47203, BAB47204, BAB83551), *Glycine max* (BAG72096), *Vitis vinifera* (XP_002263, XP_002273, XP_002276, CAN80961), *Oryza sativa* (AAP46197, ABF96216; T03664), *Spiranthes hongkongensis* (AAG44348, ABU94831), *Solanum chilense* (CAA04615) and sweetpotato (AAF37863, AAF37864 and AAF61082) corresponding to the RT domain of *Copia*-like retrotransposons. The degree of shading indicate degree of residue conservation ranging from black (100% conservation), dark grey (\geq 80%), light grey (\geq 60%), and no shading (<60%).