Occurrence of LINE, *gypsy*- and *copia*-like retrotransposons in the clonally-propagated sweetpotato (*Ipomoea batatas* L.)

Tom Okpul[1], Robert M. Harding[2], Mark J. Dieters[1] and Ian D. Godwin[1]

[1]The University of Queensland, School of Agriculture and Food Sciences, St. Lucia 4072, QLD, Australia.

[2]Centre for Tropical Crops and Biocommodities, Queensland University of Technology, Brisbane, QLD, Australia.

Correspondence

T. Okpul, The University of Queensland, School of Agriculture and Food Sciences, St. Lucia 4072, Brisbane, QLD, Australia.

Email: tokpul@uq.edu.au

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 expansive replication strategies. We screened and isolated sequence fragments of long terminal repeat (LTR), reverse transcriptase (RT) and *Gypsy*-like RT and envelop (ENV)-like domains of LTR retrotransposons, and two partial sequences of non-LTR long interspersed element (LINE) in the clonally-propagated allohexaploid sweetpotato (*Ipomoea batatas*) genome. Using dot blot hybridisation, these elements were found to be present in the ~1597 Mb haploid sweetpotato genome with copy numbers ranging from ~50 to ~4100 as observed in the partial LTR (IbLtr-1) and LINE (IbLi-1) sequences, respectively. The continuous clonal propagation of sweetpotato may have contributed to such a multitude of copies of some of these genomic elements. Interestingly, the isolated *Gypsy*-like ENV and RT sequence fragments, IbGy-1 (~2100 copies) and IbGy-2 (~540 copies) respectively, were found to be homologous to the *Bagy-2* cDNA sequences of barley (*Hordeum valgare*). Although the isolated partial sequences were found to be homologous to other transcriptionally active elements, future studies are required to determine whether they represent elements that are transcriptionally active under normal and/or stressful conditions.

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

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 protein-coding domains (Eickbush and Jamburuthugoda 2008). Non-LTR retrotransposons comprise both long and short interspersed elements, abbreviated as LINEs and SINEs respectively.

The proliferation of TEs can be detrimental to the survival of their host especially in asexual species where the mechanisms regulating TE activity (such as through mechanisms of DNA repair) are lax. Polyploids generally have been shown to undergo rapid and extensive genomic changes (Chen 2007). Therefore, it is not surprising that evidence of genomic instability in a clonally-propagated polyploid such as sweetpotato (2*n*=6x=90), *Ipomoea batatas* (L.) Lam. (Convolvulaceae) is accumulating. These genomic instabilities have been associated with TEs, particularly retrotransposons (Table 1; La Bonte 2001). Apart from LINEs (e.g. Yamashita and Tahara 2006) and SINEs (e.g. Tanaka et al. 2001), only the *Ty1*/*copia*-like elements of the LTR retrotransposons have been reported in sweetpotato (Table 1).

The continuous use of clonal propagation has been implicated in the decline in storage root yield of cultivars over time and this has been attributed to the accumulation of TEs (La Bonte 2001). Active retrotransposons (Table 1) have been shown to cause morphological 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 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 have accumulated over generations of clonal propagation.

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.).

Database searches for nucleotide sequences of TEs reported in sweetpotato were performed with BLASTn at the National Centre for Biotechnology Information (NCBI, Bethesda, Md.) website: http://www.ncbi.nlm.nih.gov/BLAST/. Multiple sequence alignments were performed using ClustalW (Thompson et al. 1994) at the European Bioinformatics Institute website: http://www.ebi.ac.uk/Tools/clustalw2/. Primers (Table 2) were then designed manually based on consensus sequences from alignments or by submitting to the program Primer3 (Rozen and Skaletsky 2000), available at 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) domains of *Gypsy*-like elements were also tested for sweetpotato. All the primers were synthesized by Proligo Australia Pty Ltd (Lismore, Australia).

PCR amplifications were performed using a MJ-Research PTC-200™ Thermal Cycler (MJ Research Inc. Watertown, MA, USA). The 25 µl PCR reaction mixtures contained approximately 20 ng template DNA, 1X PCR Buffer, 2.5 mM $MgCl_2$, 0.25 mM dNTP, 0.2 µM of each primer and 1 U *Taq* DNA polymerase. The generic PCR program used comprised 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 at 72 °C for 5 mins. A 10 µl aliquot of the PCR product was electrophoresed through 1.5% 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 a single band was amplified, the PCR products were purified directly using the QIAquick PCR Purification Kit (Qiagen, Valencia, CA, USA. Cat. No. 28104). When multiple bands were amplified, each band was purified from the gel using the QIAquick Gel Extraction Kit (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 transformed *Escherichia coli* colonies were each transferred onto a fresh plate, and in a 10 µl 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 Plasmid Miniprep Kit (Invitrogen, Carlsbad, CA, USA. Cat. No. K2100-10). Three replicates 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™ Terminator Cycle Sequencing Ready Reaction Kit version 3.1 (Applied Biosystems, Foster City, CA, USA). Sequence analyses (BLASTn and BLASTp) against the sequences in the GenBank database, alignment with ClustalW and phylogenetic tree construction were performed using the default settings of software Geneious version 4.7 (Drummond et al. 2009). The dendrogram was constructed using the neighbour-joining method (Saitou and Nei

1987) with 500 bootstrap replicates analysed; the default distance based on the proportion of nucleotide and amino acid substitution was used, respectively. In cases where a significant protein family was identified, further searches were conducted using Pfam version 24.0 (Wellcome Trust Sanger Institute at: http://www.pfam.sanger.ac.uk/search/sequence.nces.).

Dot-blot hybridization was used to determine the number of apparent full-length and 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 spotted onto Hybond N^+ membranes as three replicates. The plasmid standards contained 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 Purification kit (Qiagen) and labelled with digoxigenin, DIG-dUTP using the PCR DIG Probe Synthesis Kit (Roche Applied Science, Mannheim, Germany). Hybridization was carried out according to the manual provided for the DIG Kit with a high-stringency wash: 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, after subtracting the backgrounds, using a Molecular Imager® Gel Doc™ XR System. The volume (i.e. the sum of the intensities of the pixels within the volume boundary × pixel area) of each dot blot was determined, and this was used to calculate the concentration of the hybridised probes using the equation derived from the standard curve. The relative concentration of the nuclear genomic DNA attaching to the probe was determined by dividing the concentration of the hybridised probes to the respective concentration of genomic DNA. The copy numbers of the TE sequences present in the sweetpotato genome (size = 1597 Mb/1C, where 1C is the haploid genome (Arumuganathan and Earle 1991)) were then calculated using the formula: Copy number = (size of the haploid genome × average proportion of nuclear genomic DNA hybridizing to the probe)/ size of element probe.

Results and Discussion

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

As anticipated, the PCR amplified fragments IbLi-1 (258 bp) and IbLi-2 (352 bp) were found to be highly homologous to the *Lib* LINEs of sweetpotato (Yamashita and Tahara 2006, Fig. 1, 3a and 3b). A few base substitutions leading to amino acid substitutions in the 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 more than 4100 for IbLi-1 and ~600 for IbLi-2 (Table 3). Whereas numerous retrotransposons have been shown to inhabit the sweetpotato genome, only a few have been reported to actively mobilise. Tahara et al. (2004) demonstrated that the abundant *Ty1/copia*-like element *Rtsp*-1 can be actively mobilised in callous tissues. Yamashita and Tahara (2006) reported on the active members of the LINE family *Lib* to cause spontaneous mutations in callus and meristem stem cells in sweetpotato. We observed similar sequence fragments (IbLi-1 and IbLi-2) having higher copy numbers of up to ~4100 in the sweetpotato cv. Beauregard. Generally, the abundance of LINEs in plant genomes have been attributed to the fact that they can proliferate by participating in double-strand DNA break repair besides 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 sweetpotato and other asexually propagated crop species (McKey et al. 2010), the accumulation of actively mobilising retrotransposons dictated by such molecular mechanisms 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
186 genome (Kumar and Bennetzen 1999), functionally inactivated by the presence of stop
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 ~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.

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.

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.

232 Connolly, A.G., Godwin, I.D., Cooper, M., and DeLacy, I. 1994. Interpretation of randomly
233 amplified polymorphic DNA marker data for fingerprinting sweet potato (Ipomoea
234 batatas L.) genotypes. Theoretical and Applied Genetics **88**: 332-336.
235 Drummond, A.J., Ashton, B., Cheung, M., Heled, J., Kearse, M., Moir, R., Stones-Havas, S.,
236 Thierer, T., and Wilson, A. 2009. Geneious v4.7.
237 Eickbush, T.H. 2002. Repair by retrotransposition. Nat Genet **31**(2): 126-127.
238 Eickbush, T.H., and Jamburuthugoda, V.K. 2008. The diversity of retrotransposons and the
239 properties of their reverse transcriptases. Virus Research **134**(1-2): 221-234. doi:DOI:
240 10.1016/j.virusres.2007.12.010.
241 Hawkins, J.S., Hu, G., Rapp, R.A., Grafenberg, J.L., and Wendel, J.F. 2008. Phylogenetic
242 determination of the pace of transposable element proliferation in plants: copia and
243 LINE-like elements in Gossypium. Genome **51**: 11-18.
244 Kejnovsky, E., Hobza, R., Cermak, T., Kubat, Z., and Vyskot, B. 2009. The role of repetitive
245 DNA in structure and evolution of sex chromosomes in plants. Heredity **102**(6): 533-
246 541.
247 Kumar, A., and Bennetzen, J.L. 1999. PLANT RETROTRANSPOSONS. Annual Review of
248 Genetics **33**(1): 479-532. doi:10.1146/annurev.genet.33.1.479.
249 La Bonte, D.R., Villordon, A.Q., and Fajardo, D.S. . 2001. Mutations in sweetpotato. *In*
250 Proceedings of International Workshop on Sweetpotato Cultivar Decline. *Edited by* Y.
251 Nakazawa and K. Ishiguro. Kyushu National Agriculture Experimental Station,
252 Miyakonojo, Miyazaki, Japan. pp. 1-9.
253 McKey, D., Elias, M., Pujol, B., and Duputié, A. 2010. The evolutionary ecology of clonally
254 propagated domesticated plants. New Phytologist **186**(2): 318-332.
255 doi:10.1111/j.1469-8137.2010.03210.x.

256 Rozen, S., and Skaletsky, H.J. 2000. Primer3 on the WWW for general users and for biologist
257 programmers. *In* Bioinformatics Methods and Protocols: Methods in Molecular
258 Biology. *Edited by* S. Krawetz and S. Misener. Humana Press, Totowa, NJ. pp. 365-
259 386.
260 Saitou, N., and Nei, M. 1987. The neighbor-joining method: A new method for
261 reconstructing phylogenetics trees. Mol. Biol. Evol. **4**: 406–425.
262 Sakai, H., Tanaka, T., and Itoh, T. 2007. Birth and death of genes promoted by transposable
263 elements in Oryza sativa. Gene **392**(1-2): 59-63. doi:DOI:
264 10.1016/j.gene.2006.11.010.
265 Soleimani, V.D., Baum, B.R., and Johnson, D.A. 2006. Quantification of the retrotransposon
266 BARE-1 reveals the dynamic nature of the Barley genome. Genome **49**: 389-396.
267 Tahara, M., Aoki, T., Suzuka, S., Yamashita, H., Tanaka, M., Matsunaga, S., and Kokumai,
268 S. 2004. Isolation of an active element from a high-copy-number family of
269 retrotransposons in the sweetpotato genome. Molecular Genetics and Genomics **272**:
270 116-127
271 Tanaka, M., Yoshinaga, M., Nakazawa, Y., and Nakatani, M. 2001. Identification of a novel
272 SINE-like sequence in retrotransposon fragments of sweetpotato. Plant Biotechnology
273 Reports **18**: 157-161.
274 Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: improving the
275 sensitivity of progressive multiple sequence alignment through sequence weighting,
276 position specific gap penalties and weight matrix choice. Nucleic Acids Research **22**:
277 4673-4680.
278 Vicient, C., Kalendar, R., and Schulman, A. 2001a. Envelope-class retrovirus-like elements
279 are widespread, transcribed and spliced, and insertionally polymorphic in plants.
280 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): 1219-1225. 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.

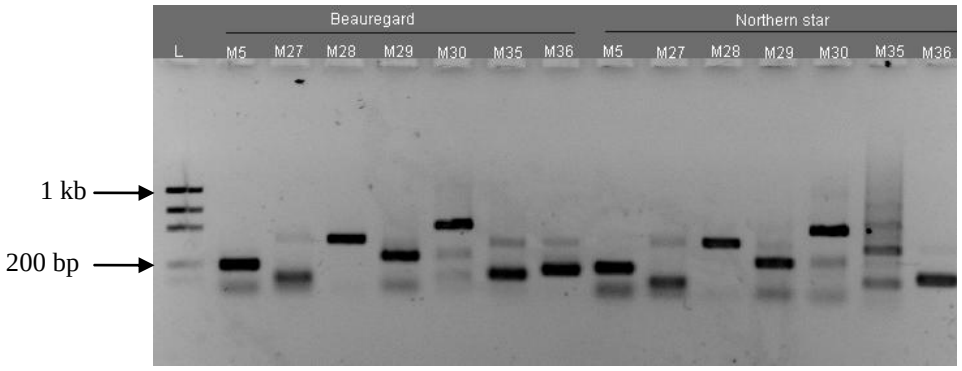
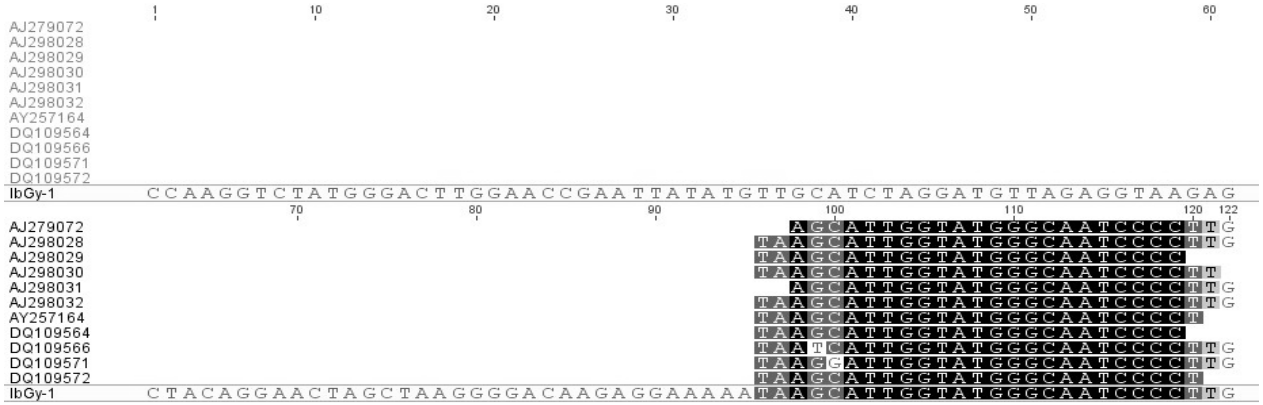

Beauregard
Northern star
L
M5
M27
M28
M29
M30
M35
M36
M5
M27
M28
M29
M30
M35
M36
1 kb
200 bp

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)

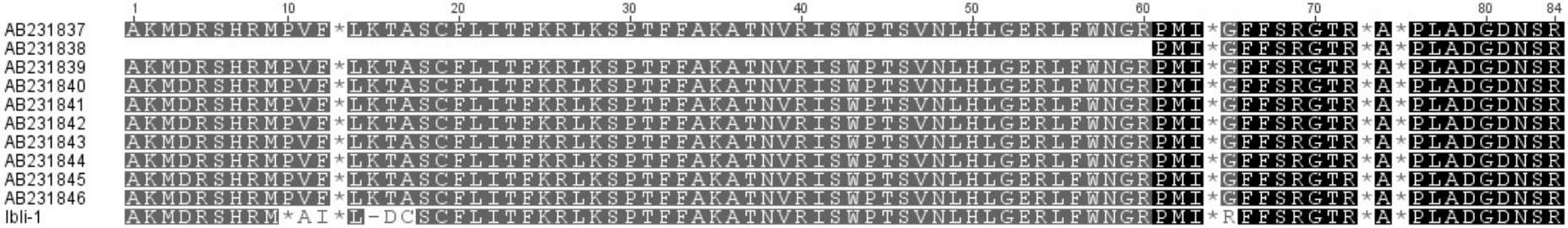

AJ279072
AJ298028
AJ298029
AJ298030
AJ298031
AJ298032
AY257164
DQ109564
DQ109566
DQ109571
DQ109572
IbGy-1

b)


AF378016
AF378017
AJ295132
AJ421232
IbGy-2

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 (≥80%), light grey (≥60%), and no shading (<60%).

a)

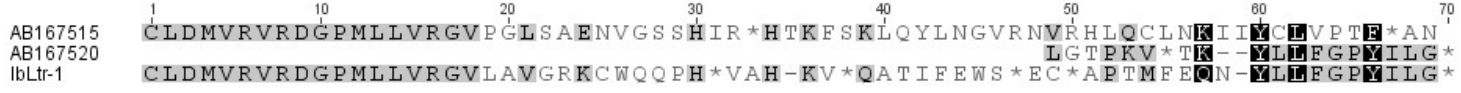

AB231837
AB231838
AB231839
AB231840
AB231841
AB231842
AB231843
AB231844
AB231845
AB231846
Ibli-1

b)


AB231841
AB231839
AB231840
Ibli-2

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 (≥80%), light grey (≥60%), and no shading (<60%).

a)


AB167515 CLDMVRVRDGPMLLVRGVPGLSAENVGSSHIR*HTKFSKLQYLNGVRNVRHLQCLNKIIYCLVPTF*AN
AB167520 LGTPKV*TK--YLLFGPYILG*
IbLtr-1 CLDMVRVRDGPMLLVRGVLAVGRKCWQQPH*VAH-KV*QATIFEWS*EC*APTMFEQN-YLLFGPYILG*

b)

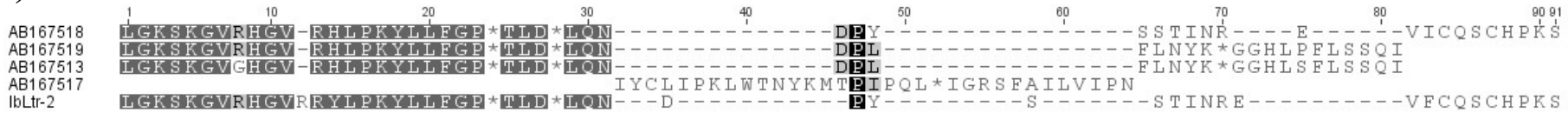

AB167518 LGKSKGVRHGV-RHLPKYLLFGP*TLD*LQN----------DPY----------SSTINR----E------VICQSCHPKS
AB167519 LGKSKGVRHGV-RHLPKYLLFGP*TLD*LQN----------DPL----------FLNYK*GGHLPFLSSQI
AB167513 LGKSKGVGHGV-RHLPKYLLFGP*TLD*LQN----------DPL----------FLNYK*GGHLSFLSSQI
AB167517 IYCLIPKLWTNYKMTPIPQL*IGRSFAILVIPN
IbLtr-2 LGKSKGVRHGVRRYLPKYLLFGP*TLD*LQN---D----------PY---------S-------STINRE----------VFCQSCHPKS

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 (≥80%), light grey (≥60%), and no shading (<60%).

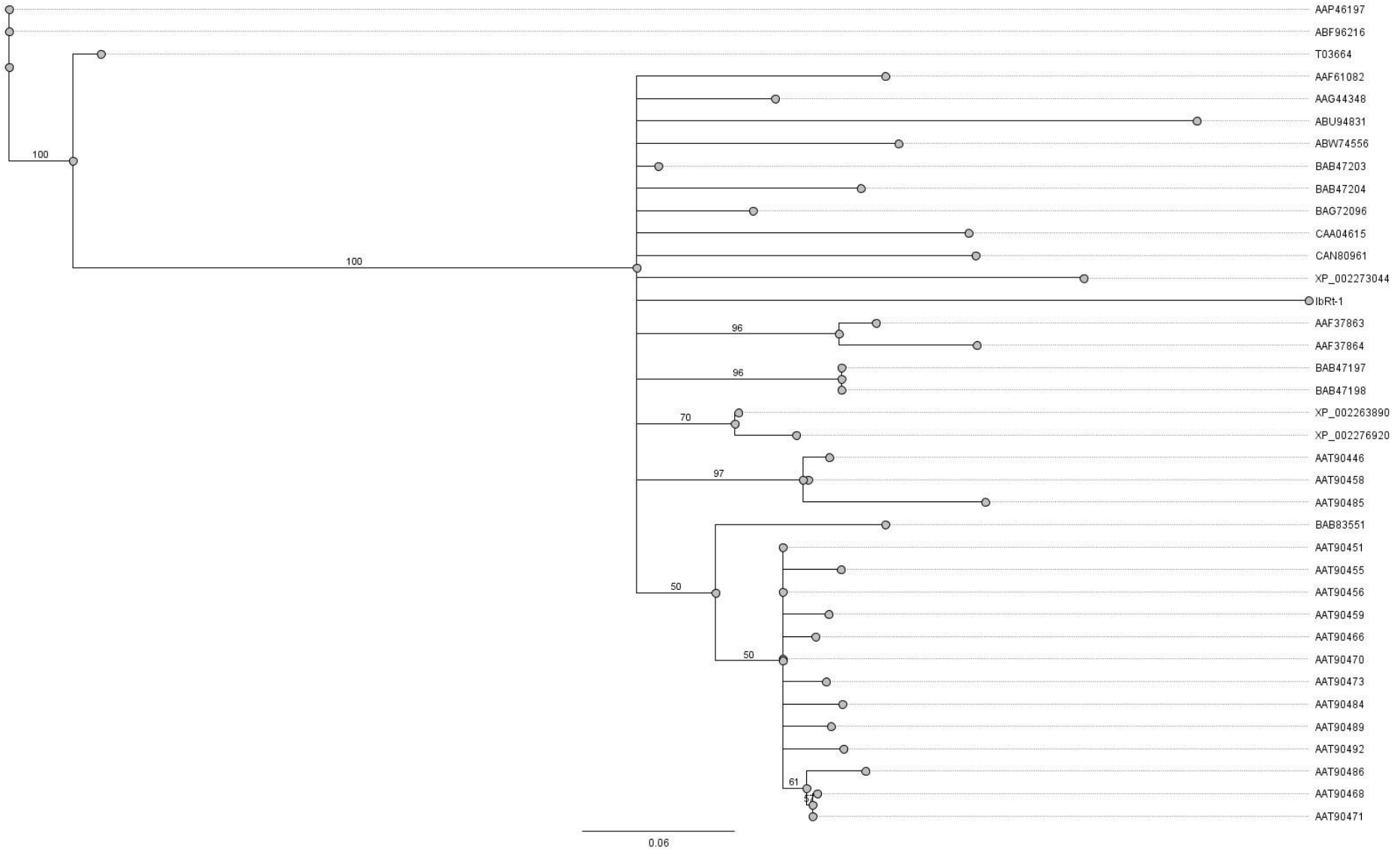

AAP46197
ABF96216
T03664
AAF61082
AAG44348
ABU94831
ABW74556
BAB47203
BAB47204
BAG72096
CAA04615
CAN80961
XP_002273044
IbRt-1
AAF37863
AAF37864
BAB47197
BAB47198
XP_002263890
XP_002276920
AAT90446
AAT90458
AAT90485
BAB83551
AAT90451
AAT90455
AAT90456
AAT90459
AAT90466
AAT90470
AAT90473
AAT90484
AAT90489
AAT90492
AAT90486
AAT90468
AAT90471
100
100
96
96
70
97
50
50
61
0.06

Fig. 5. A dendrogram showing the relationship between IbRt-1 and similar sequences found in other plant species based on the amplified Ty-copia 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.

Table 1 Various active retrotransposons identified in sweetpotato[*].

Name	GenBank accession	Retrotransposon group	Transposition/ insertion	Estimated copy number	Reference
Tib1-Tib 31	AF223309-AF223326, AF223337-AF223345	LTR *Ty1/copia*	Near *DFR-B* gene; active in callus and normal plant cells	NR	Tanaka et al (2001)
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, AB231842-AB231846	Non-LTR LINE	Active in callus and normal plant cells	108	Yamashita and Tahara (2006)
IPSE1	AF295596	Non-LTR SINE	Within retrotransposon fragment	NR	Tanaka et al. (2001)

[*]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 group/ domain*	Primer		Annealing temperature (°C)	Author
		Code	sequence (5'…3')		
IbGy-1	*Gypsy*-like, ENV	KS1F	CCAAGGTCTATGGGACTTGGAACC	60.0	Vicient *et al.* (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.

Table 3 Retroelements detected in sweetpotato.

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

*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 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.

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.

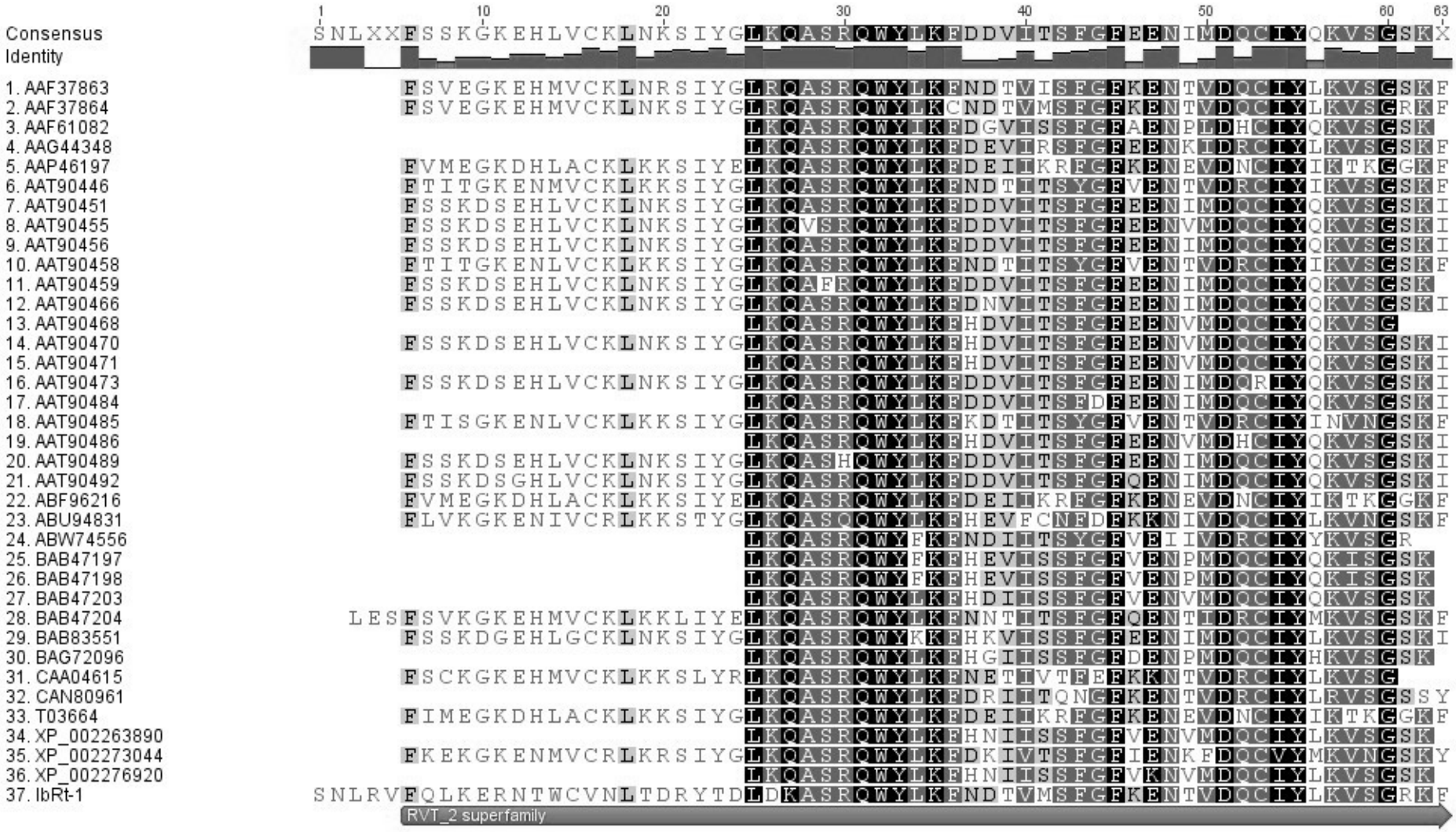

Consensus
Identity
1. AAF37863
2. AAF37864
3. AAF61082
4. AAG44348
5. AAP46197
6. AAT90446
7. AAT90451
8. AAT90455
9. AAT90456
10. AAT90458
11. AAT90459
12. AAT90466
13. AAT90468
14. AAT90470
15. AAT90471
16. AAT90473
17. AAT90484
18. AAT90485
19. AAT90486
20. AAT90489
21. AAT90492
22. ABF96216
23. ABU94831
24. ABW74556
25. BAB47197
26. BAB47198
27. BAB47203
28. BAB47204
29. BAB83551
30. BAG72096
31. CAA04615
32. CAN80961
33. T03664
34. XP_002263890
35. XP_002273044
36. XP_002276920
37. IbRt-1
RVT_2 superfamily

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 (≥80%), light grey (≥60%), and no shading (<60%).