Theor Appl Genet (2002) 105:862–869
DOI 10.1007/s00122-002-1015-0

M. Berenyi · S. T. Gichuki · J. Schmidt · K. Burg

Ty1-copia retrotransposon-based S-SAP (sequence-specific amplified polymorphism) for genetic analysis of sweetpotato

Received: 13 July 2001 / Accepted: 8 March 2002 / Published online: 30 July 2002

Abstract *RNaseH*-LTR regions of the *Ty1-copia* retrotransposon were amplified and cloned from the sweetpotato genome using *RNaseH* gene-specific degenerate primers and restriction site-specific adaptor primers. Ninety clones out of the 240 sequenced were identified with a variable degree of homology to the *Ty1-copia RNaseH* gene. Three (*Str6*, *Str85*, *Str187*) of the 90 had characteristic *RNaseH*-gene, stop codon, polypurine track and putative 3′ LTR sequence elements. Analysis of nine selected genotypes representing Africa, South and Central America, as well as Papua New Guinea, by the established S-SAP technique revealed that the majority of the *Ty1-copia* transposon insertions were unique (33 to 64%) and only few common bands were detected. Analysis of 177 East African varieties further supported this finding and showed that most of the *copia* retrotransposon locations were represented only by some genotypes. Considering that sweetpotato has been present in the East African region for only about 500 years, and the number of genotypes introduced was possibly limited, a surprisingly high level of genetic variability of the transposon insertion sites was detected. These findings may indicate the putative activity of the retrotransposon in sweetpotato in the recent past. Comparison of the *copia* retrotransposon insertion-based S-SAP method to AFLP and RAPD showed that the majority of the markers were more polymorphic (97–99%) in the case of S-SAP in comparison to AFLP (70–90%) and RAPD (88%). Thus demonstrating the transposon-based molecular marker system was very efficient for genotyping.

Communicated by C. Möllers

M. Berenyi (✉) · J. Schmidt · K. Burg
Austrian Research Centre Seibersdorf, A2444 Seibersdorf, Austria
e-mail: Maria.Berenyi@arcs.ac.at
Tel.: +43-50550-3524, Fax: +43-50550-3653

S.T. Gichuki
Kenya Agricultural Research Institute, Biotechnology Centre,
P.O Box 57811, Nairobi, Kenya

Keywords S-SAP · Sweetpotato [*Ipomoea batatas* (L.) Lam.] · Retrotransposon · Molecular marker · RAPD · AFLP

Introduction

After Columbus introduced the sweetpotato to Spain it spread to Africa, India, Asia and Oceania, and became an important crop in those parts of the world. Most probably, the spread of sweetpotato outside tropical America was restricted to a limited number of genotypes. In contrast, a wide variety of genotypes are found all over the word today. This could be the consequence of the high level of heterozygosity found in sweetpotato. The sweetpotato is an out-crossing hexaploid and the variation due to sexual reproduction and somatic mutation can be maintained through vegetative propagation.

Several marker systems were developed during the last decade for genotyping, such as RAPDs (Connolly et al. 1994; Zhang et al. 1998), SSRs (Tautz 1988) and AFLPs (Zabeau and Vos 1993; Vos et al. 1995). The amplified fragment length polymorphism (AFLP) and simple sequence repeats (SSR), or microsatellites, have recently become popular in fingerprinting and phylogenetic studies. Buteler et al. (1999) reported low amplifiable loci in sweetpotato using microsatellites and commented that their utility in sweetpotato genetic studies may be limited.

Recently, retrotransposon-based *marker systems* proved useful for developing multiplex DNA-based *marker systems* because of their ubiquitous distribution, high copy number and widespread chromosomal dispersion in plants (Lee 1990; Kumar 1996). A considerable advantage of the retrotransposon-based polymorphic marker system is dependant on the fact that these retrotransposons transpose via an RNA intermediate, which is converted to DNA by reverse transcription before reinsertion, whereas the parental retrotransposon remains fixed in the genome (Boeke and Corces 1989; Kumar and Bennetzen 1999). This means that the inserted retrotrans-

poson does not change its position during the evolution of the genome but every insertion elevates the polymorphism and the size of the genome. (Pearce et al. 1996; SanMiguel et al. 1996; SanMiguel and Bennetzen 1998)

Two types of retrotransposons are known, e.g. those that contain long terminal repeats (LTRs) and those that do not. Although the 5′ and 3′ LTR sequences are identical at the time of the insertion, they may differ in time through mutations. Solo LTR sequences found in different genomes indicate that unequal crossing over and/or intrachromosomal recombination events could also delete inserted retrotransposon sequences (Shirasu et al. 2000). The most-studied group of LTR retrotransposons is the *Ty1-copia* group, named after the best-studied elements in *Saccharomyces cerevisiae* and *Drosophila melanogaster* (Boeke and Corces 1989; Grandbastien et al. 1989; Schmidt et al. 1996). Plant *Ty1-copia* retrotransposons show a considerable degree of sequence heterogeneity and insertional polymorphism, both within and between species (Boeke and Corces 1989; Flavell et al. 1992a) compared to the Drosophila *copia*, the fungal *Ty1* or animal retrotransposons. The chromosomal distribution of the *Ty1-copia* group of retrotransposons in plants has been studied by in situ hybridisation on metaphase chromosomes. These studies revealed that these elements are dispersed preferentially in the heterochromatin region but are also found in the euchromatin region of plant chromosomes (Pearce et al. 1996; Schmidt et al. 1996; Heslop-Harrison et al. 1997; Bennetzen 2000).

Phylogenetic analyses of the retrotransposon sequences show that the degree of sequence divergence in *Ty1-copia* retrotransposon populations between any pair of species is generally proportional to the evolutionary distance between those species (Flavell et al. 1992b). Several authors have also hypothesised that transposition events could increase the genetic variability necessary for organisms to adapt to different environmental conditions and that they may be a major factor in the evolution of higher plants (McClintock 1984; Schwarz-Sommer and Saedler 1988; Wendel and Wessler 2000). Stress and environmental challenges are known to stimulate the expression or the transposition of mobile elements (Grandbastien et al. 1997; Mhiri et al. 1997).

Several retrotransposon-based marker systems have recently been reported. Using restriction-site polymorphism, Purugganan and Wessler (1995) analysed a limited region of the Magellan retrotransposon and was able to discriminate between even closely related *Zea mays* (L.) subspecies. Waugh et al. (1997) published the S-SAP method on barley and found that the level of polymorphism is about 25% higher than that revealed by AFLP. Ellis et al. (1998) amplified sequences between the polypurine track of the pea *PDR1* retrotransposon and a flanking *Taq* adaptor downstream to the 3′ LTR sequence, while Pearce et al. (2000) used the original S-SAP technique with two other pea retrotransposon sequences (*Tps12* and *Tps19*). Both experiments resulted in a detailed picture of the intra- and inter-species relationship within the *Pisum* genus. Ellis et al. (1998) compared the phylogenetic diversity trees generated by S-SAP and AFLP, and found that these trees possibly reveal the same underlying divergence. Yu and Wise (2000) combined the AFLP, RAPD and S-SAP markers to make a saturated map of diploid *Avena*. They found that the S-SAP-generated markers were more evenly distributed across the *Avena* genome, similar to the results of Waugh et al. (1997) on barley.

In the present study, we describe the partial cloning of the sweetpotato *Ty1-copia* LTR regions and the establishment of an S-SAP system based on this sequence information. The established S-SAP system is compared to other dominant marker systems like AFLP and RAPD, and the frequency of the insertions in East African accessions is also discussed.

Materials and methods

Sweetpotato materials and DNA purification

Lyophilised or silica gel dried leaf samples were used for all the analyses. Sixty seven landraces were obtained from the Kenya Agricultural Research Institute gene banks at the University of Nairobi, field station, Kabete; 59 landraces were obtained from the Ugandan National Agricultural Research Institute, Namulonge, and from Mrs. Erna Abidini of Arapai Agricultural College, Soroti. Forty four landraces were obtained from the Tanzanian Ministry of Agriculture, Agricultural Research Institute, Tengeru. Single accessions from Kenya (Mafuta, Simama), Uganda (Kyebandula, Wagabolige), Columbia (Camote Amarillo), Peru (Japonese), Mexico (No.221), Brazil (Santo Amaro) and Papua New Guinea (Naveto) were obtained from the International Potato Centre germplasm collection at Kabete, Kenya.

Total DNA was isolated and purified with the ‘DNeasy plant minikit’ (QIAGEN, Germany) following the original protocol.

PCR amplification of Ty1-copia retrotransposon LTRs

*Mse*I or *Eco*RI restriction enzyme-digested sweetpotato genomic DNA was amplified in two steps with 5′ biotin-labelled degenerate *RNaseH* gene-specific (5′MGNACNAARCAYATHGA) and adaptor-specific flanked PCR primers (*Mse*I: 5′GATGGATCCTGAGTAA; E01: 5′GACTGCGTACCAATTCA) as reported by Pearce et al. (1999). Streptavidin-coated magnetic Dynabeads particles were used to select the biotinilated PCR products, which were used as a template in the second PCR with the nested *RNaseH* primer (5′GCNGAYATNYTNACNAA).

The degenerate *RNaseH* primers (Donated by A.J. Flavell, Department of Biochemistry, University of Dundee) were designed by sequence homologies of the known retrotransposon original *RNaseH* genes. The amplified fragments were cloned into the *Topo 4 TA* cloning vector (Clontech) and sequenced.

S-SAP method

The procedure described by Waugh et al. (1997) was adapted with modifications. Genomic DNA was digested only with a single, rear-cutting enzyme (*Eco*RI) and ligated with the specific adaptor in one reaction. Two PCR reactions were performed. The first preselective PCR amplification was made with Dynazyme *Taq DNA polymerase* in 50-µl reactions (95 °C 50 s, 52 °C 1 min, 72 °C 1 min, 30 cycles). The 5′ LTR specific primers without extension (*Str187*/0: 5′AGACTAAGAGTCCTAACA; *Str6*/0: 5′AGAGTCCTAATACTCTAATA; Str85/0: 5′ATATGATATTACTGCTATTA)

and the E01-adaptor primer were used for this pre-selective PCR.

The second, selective PCR amplification was made with the Qiagen Hotstar *Taq DNA polymerase* in 25-µl reactions. Touch-down amplification conditions (95 °C 50 s, 70 °C 1 min – 0.7 °C/cycle, 72 °C 1 min, 20 times, followed by 95 °C 50 s, 55 °C 1 min, 72 °C 1 min in another 20 cycles) were used. FAM-labelled transposon primers with different selective nucleotides (*Str187*/G and GC) were combined with the E01 or E44 (5′GACTGCGTACCAATTCATC) adapter-specific primers.

Amplified samples were separated on an ABI Prism 373 automated sequencer (Applied Biosystems) using a 6% denaturing polyacrylamide gel.

AFLP analysis

The AFLP methodology was essentially as described by Vos et al. (1995) and adapted for sweetpotato with fluorescent labelling. Two restriction enzymes, *Mse*I and *Eco*RI, were used to fragment the genomic DNA with subsequent ligation to specific adaptors. Pre-amplification PCR was done for 45 cycles (92 °C 1 min, 60 °C 30 s, 72 °C 1 min). In the second selective PCR amplification the *Eco*RI selective primers were ABI-FAM fluorescent labelled (E33: 5′GACTGCGTACCAATTCAAG; E36: 5′GACTGCGTACCAATTCACC). The second PCR was made as: 94 °C 30 s, 65 °C – 1 °C/cycles 45 s, 72 °C 2 min, 10 cycles, then 94 °C 30 s, 56 °C 30 s, 72 °C 2 min, 30 cycles. The samples were loaded on a 6% denaturing polyacrylamide gel and separated on an ABI Prism 373 sequencer for 10 h.

RAPD analysis

RAPD amplification were carried out as described by Williams et al. (1990) with a few modifications. A 25-µl PCR reaction was set up as follows: 0.2 µM of RAPD primer; 0.2 mM for each dNTP; 0.5 U of Dynazyme Thermostable DNA Polymerase (Finzymes); 2.5 µl of 10 × PCR buffer [10 mM Tris–HCl, pH 8.8, 1.5 mM $MgCl_2$, 50 mM KCl, 0.1% Triton X-100); 5 µl genomic DNA template (5 ng/µl) and dH_2O]. The reaction was overlaid with one drop of mineral oil. Amplification was done in 96-well micro-test plates using a PTC-100 programmable thermal cycler (M.J. Research). The amplification conditions were 46 cycles of: denaturation at 94 °C for 1 min; annealing at 36 °C for 1 min and polymerisation at 72 °C for 2 min, and a final extension step of 72 °C for 8 min. Amplification products were analysed by electrophoresis of a 10-µl PCR product with 1 × loading buffer (0.25% Bromophenol blue, 0.25% Xylene Cyanol and 40% sucrose in water) on a 1.5% agarose gel in 1 × TBE and stained with ethidium bromide (0.5 µg/ml). The gel was run for $4\frac{1}{2}$ h at 100 V in 1 × TBE buffer. A 100-bp DNA ladder molecular-weight marker (Gibco) was used to determine the size of the amplified products.

Gel analysis for S-SAP and AFLP

Data were analysed with Genotyper 2.5 program (Applied Biosystems). Peaks, corresponding to an amplified retrotransposon insertion or an AFLP amplified fragment, were designated into categories. The tolerance of a category was chosen to be ±0.5 bp, meaning that two amplified fragments with a bigger difference than 1 bp were considered as non-identical ones. Banding patterns were subsequently converted to 0/1 strings for statistical analysis. Informative products typically fall within 50–450 bp, (Sharbel 1999). Only categories between 50–400 bp were utilised for AFLP data analysis while categories up to 500 bp were used for S-SAP analysis.

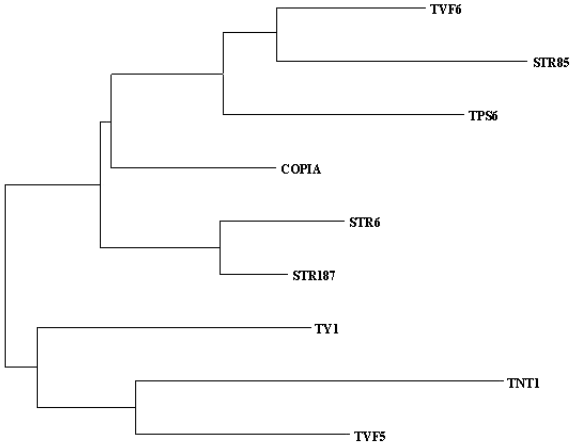
RAPD gel analysis and band scoring was done using the RFLPSCAN (SCANALYTICS) computer programme. All varieties were scored for presence (1) or absence (0) of a fragment for each selected band of the amplified products. Only distinct and polymorphic bands were scored. Faint bands were omitted from the score.

The dendograms were generated using the TREECON program [Yves Van de Peer, University of Antwerp (UIA), Antwerp, the Netherlands] with UPGMA cluster analysis based on the genetic distance by Nei and Li (1979) and bootstrapping 1,000 times.

The Mantel test was based on the same genetic distance data and executed with the program PopTools (Greg Hood, Pest Animal Control Co-operative Research Centre, Wildlife and Ecology, CSIRO, Canberra, Australia.)

Results

Putative *Ty1-copia RNaseH*-LTR regions have been amplified and cloned from the sweetpotato genome using the degenerate *RNaseH* primers (Pearce et al. 1999) and restriction-site specific adaptor primers. Two hundred and forty clones were isolated and sequenced. Altogether 90 clones out of the 240 were identified with a variable degree of homology to the *Ty1-copia RNaseH* gene, but only three [*Str6* (AF350442), *Str85* (AF350443), *Str187* (AF350441)] showed the characteristic *RNaseH* gene, stop codon, polypurine track and putative 3′ LTR sequence elements (Fig. 1).




```
                RnaseH gene                              PPT      IR  LTR
Str187   ADIFTKPLAA-RFAFLRDKLQVVPPCA*     (+1bp)  GAGGGGAAGTGTTAGGACTCTTAGTCTAG
Str6     ADMFTKALPTPRFTFLRDKLQVTALPCA*    (-1)    GAGGGGGAGTATTAGGACTCT-A-TGTATCT
Copia    ADIFTKPLPAARFVELRDKLGLLQDDQSNAE*(+199bp)GAGGGGGCGTGTT-GGAATAT-A--CTATTC

Str85    ADIFTKALGQRQLQYFIRKLGIRDLHAPT*   (-1bp)  GAGGGGGGGTAATAGCAGTAATATCATAT
Tvf6     ADIFTKALDANQLENLRGKLGICQHEKL*    (-9bp)  GAGAAA--TTG-TAGCAATTAATTTTGCTCA
```

Fig. 1 UPGMA cluster analysis of relatedness between the deduced amino-acid sequences of the *RNaseH* genes of *Ty1-copia* type retrotransposons. *TVF5, TVF6* (*V. faba*, CAB45147, CAB45148); *Str187, Str6, Str85* (*I. batatas*, AF350441, AF350442, AF350443); *TPS6*, (*P. sativum*, CAB44716); *Copia* (*D. melanogaster* P04146); *Ty1* (*S. cerevisiae* BAA09237, CAB45147); *TNT1* (*N. tabacum*, P10978). For comparison the sequence of the *D. melanogaster copia* (P04146) and *V. faba* (*Tvf6* CAB45148) sequences are also presented. The amino-acid sequence of the nested *RNaseH* primer is *underlined*. The identical amino acids and the IR sequence are in *bold*. *Asterisks* represent the identified the stop codon of the *RNaseH* gene. The number of nucleotides between the *RNaseH* stop codon and the polypurine track are shown in *brackets*

The putative inverted repeat (IR) of the LTR region is different in the three sweetpotato sequences and only the *Str187* clone contains the most-frequently found TGTT motif. In the putative LTR region of the *Str6* clone, a 34-bp direct repeat was recognised after the TATT inverted repeat sequence. We were unable to determine the starting point of the 3′ LTR sequences for the rest of the sequenced clones since they did not contain a recognisable polypurine-track after the *RNaseH*-gene stop codon. In many cases, the *Mse*I restriction site used to fragment the DNA for cloning terminated the sequence. We obtained longer clones using the six-cutter enzyme *Eco*RI for DNA fragmenting, but still the presence of the LTR sequence could not be verified.

The deduced amino-acid sequence of *Str85* is similar to the known *RNaseH TVF6* (*Vicia faba*) clone. The *Str187* and *Str6* clones are closely related and form a distinct group in the dendogram (Fig. 1). All three clones also cluster with the *Drosophila copia* (PO4146) sequence as well. The *Saccharomyces Ty1* and the *Nicotiana tabacum TNT1* retrotransposon sequences form another main cluster along with the *V. faba* (*TVF5*) retrotransposon clone.

Adaptation of the S-SAP method to sweetpotato

In preliminary experiments, primers based on the three sweetpotato LTR sequences were tested in S-SAP analysis. The *Str187*/0 and *Str6*/0 detected the highest number of transposon insertion sites (37 and 31, respectively), while only 2 to 7 amplification products were obtained with the *Str85*/0 primer (data not shown). In the case of the *Str6*/0 primer we observed that some DNA samples were not amplified; therefore in the further experiments the *Str187* LTR primer was extended either by one (G) or two (GC) selective nucleotides and the *Eco* adaptor primers E01 and E44 were used.

Nine sweetpotato varieties were selected and analysed from Africa, South and Central America, and Papua New Guinea, in order to evaluate the performance of these primer combinations in S-SAP reactions. The E44 adaptor primer, in combination with the *Str187*/GC or *Str187*/G primers, gave 39 and 173 polymorphic bands, respectively, representing individual retrotransposon insertions (Table 1). Reducing the number of the selective nucleotides on the LTR primer increased the number of bands by 4–5-fold as expected. The two-basepair extension of the *Eco*RI adaptor primer resulted only in a mild decrease of about 30% of the number of the amplification products (Table 1).

Analysing the frequency of individual insertions represented by a specific amplification product length, it was found that the majority of the insertion sites are unique (found only in a single genotype) in the nine selected genomes. In the case of the *187*/GC primer and the two different *Eco*RI adaptors, 25 and 32 bands out of 39 and 51 bands, respectively, were found in only one of the nine genomes tested, representing 64 and 63% of the total number of insertion sites (Table 1). The primer with a single base extension (*Str187*/G) yielded a somewhat lower portion of unique bands (46 and 33%) in the two adaptor primer combinations tested. In three primer combinations we found a single common band, while in the fourth (E01-*Str187*/G) we found six (Table 1).

Table 1 Comparison of the different primer combinations in S-SAP analysis of nine sweetpotato varieties

Frequency[a]	E44/187GC		E01/187GC		E44/187G		E01/187G	
	N[b]	%	*N*	%	*N*	%	*N*	%
1	25	64	32	63	79	46	86	33
2	3	8	4	8	40	23	51	20
3	3	8	4	8	24	14	41	16
4	4	10	3	6	11	6	28	11
5	3	8	1	2	10	6	19	7
6	0	0	2	4	4	2	7	3
7	0	0	4	8	2	1	10	4
8	0	0	0	0	2	1	12	5
9	1	2	1	2	1	1	6	2
Total	39		51		173		260	

[a] Frequency, means that a given insertion of the Str187 retrotransposon is detected in one, two or all of the nine genomes
[b] *N* represents the number of the insertions, which are present in one, two or all of the nine genomes

S-SAP, AFLP and RAPD analysis of sweetpotato accessions

The same sweetpotato genotypes, representing different geographic regions, were analysed by the AFLP and *Str187* sequence-based S-SAP techniques. Additionally, RAPD analysis of the same clones was also included in the UPGMA relatedness analysis, using the Nei and Li (1979) genetic distance (Fig. 2).

Two out of the three South American clones (Japonese, Camote Amarillo) form a separate cluster by all three marker systems, however with different bootstrap values. The highest value was obtained with RAPD (100%) and SSAP (89%) while AFLP gave only 36%. The Brazilian clone (Santo Amaro) was related to the two other South American ones only in the S-SAP analysis, while the AFLP and RAPD techniques allocated it to the group dominated by the African clones. This finding was also confirmed by both primer combinations used, E01-*Str187*/G and E44-*Str187*/G (data not shown). The S-SAP system also grouped the four African clones in the same sub-cluster (Fig. 2).

Estimating the association of the three genetic distance matrices obtained in S-SAP, RAPD and AFLP analysis by the Mantel test, we could show that the correlation of S-SSAP to RAPD was 0.46 (probability $P = 0.01$), while S-SAP to AFLP was 0.33 ($P = 0.07$). The correlation of AFLP to RAPD was however only 0.06 ($P = 0.37$).

The highest ratio of the polymorphic to monomorphic loci was in the S-SAP analysis (97.4–99.4%) where nearly all of the 260 amplified insertions obtained by a single

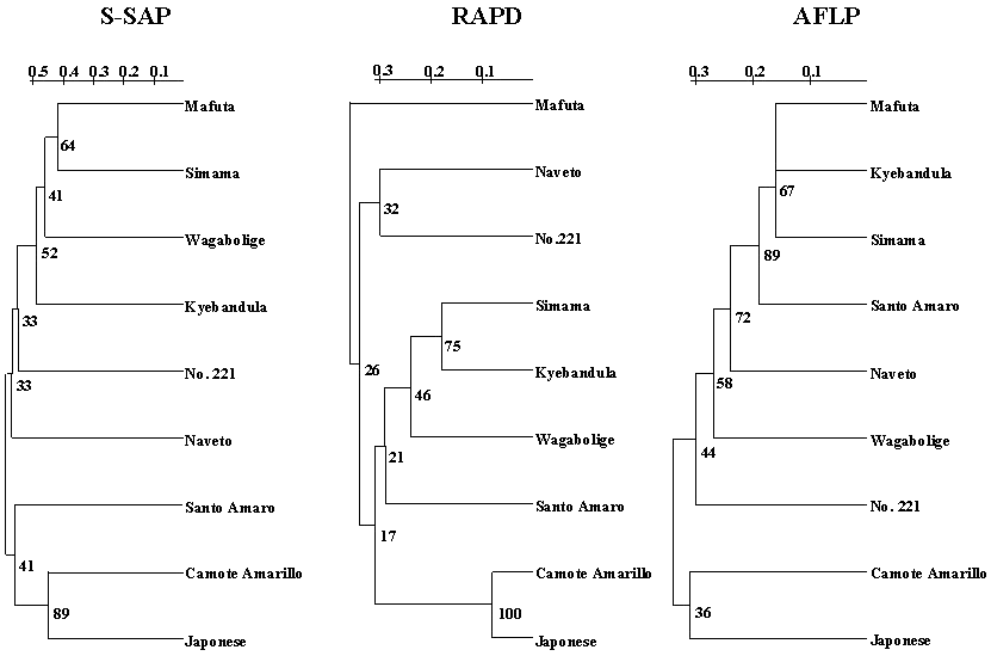


Fig. 2 UPGMA cluster analysis of the genetic relatedness of the three dominant marker systems, S-SAP, RAPD and AFLP. The bootstrap values are given in percent

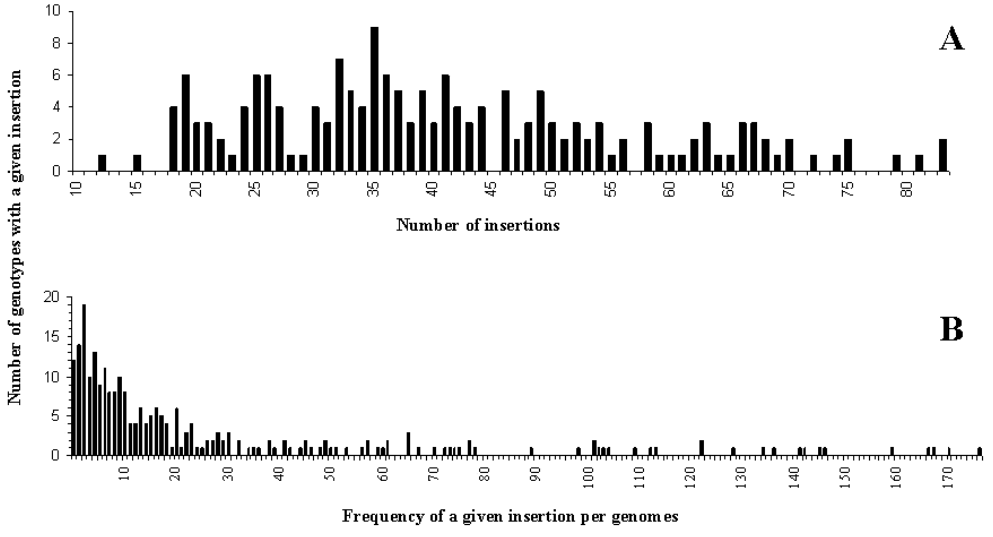


Fig. 3A, B Distribution of the *STR187* retrotransposon insertions in 177 African sweetpotato genotypes. **A** Distribution of the sweetpotato genotypes according to the number of *Str187* retrotransposon insertions. **B** Frequency distribution of individual insertions across the total number of genomes

Table 2 Comparison of RAPD, AFLP and S-SAP marker analysis for nine sweetpotato genotypes

Item	Primer combinations	Total number of assays	Total number of amplification products	Number of polymorphic products	% Polymorphic loci
RAPD		12	74[a]	65	87.8
AFLP	M40/E33	1	135	95	70.37
AFLP	M40/E36	1	93	84	90.32
S-SAP	E01/187G	1	260	254	97.7
S-SAP	E01/187GC	1	51	50	98.0
S-SAP	E44/187G	1	173	172	99.4
S-SAP	E44/187GC	1	39	38	97.4

[a] Only distinct bands that demonstrated polymorphism were scored for RAPD

primer pair proved polymorphic (Table 2). AFLP amplification of the samples, in both primer combinations, yielded a lower ratio of polymorphic to monomorphic bands (70.4 to 90.3%). RAPD fragments were intermediate (87.8%) in comparison with S-SAP and AFLP analysis.

Estimation of the frequency of the insertion sites

We analysed 177 East-African accessions to determine the number of insertion sites per genotype, detected by the E01-*Str187*/G primer combination. The analysis was

based on the assumption that identical fragment size in the S-SAP patterns represent identical, while distinct fragment size represent different, transposon insertions in the genomes. The study revealed a total number of 253 insertion sites of the *Str187* retrotransposon, averaging about 41 insertions per genome ranging between 12 and 83. The distribution of the number of insertions per genome in the analysed plants is shown in Fig. 3A.

Estimating the frequency of a particular transposon insertion in the sample set, we found that insertions present less than ten times were the most abundant. The maximum was found with insertions present three times in the 177 genotypes analysed. We found only a single transposon location, which was present in all the East African genomes analysed (Fig. 3B).

Discussion

Retrotransposon based sequence-specific amplified polymorphism of sweetpotato

Sweetpotato DNA sequences were amplified with degenerate oligonucleotide primers corresponding to conserved domains of the *Ty1-copia* retrotransposon *RNaseH* gene and flanking adaptor primers. The amplified fragments were cloned as described in Materials and methods. Two hundred and forty randomly selected clones were sequenced, yielding 90 clones with a recognisable *RNaseH* gene sequence. Only three (*Str6*, *Str85*, *Str187*) out of these 90 clones contained both the stop codon of the *RNaseH* gene and the characteristic polypurine track followed by the putative 3′ LTR region. The three isolates have three different IR and putative LTR regions, indicating the presence of several classes of *copia* elements in the sweetpotato genome (Fig. 1). Generally, the polypurine track is located at minus 1 to plus 40 bp downstream from the stop codon of the *RNaseH* gene (assumption based on sequence database data). The *D. copia* retrotransposon is an exception at 199 bp. In the three sweetpotato clones, the polypurine track was found in close vicinity (–1 + 1 bp) to the stop codon. The other 87 clones were sequenced after the stop codon for another 200–500 bp, but the putative PPT-IR-LTR region was not identifiable. The low incidence of the *RNaseH*-LTR complete sequence among the isolated clones may indicate that in sweetpotato the majority of the *Ty1* elements are already mutated. It is known that plant genomes have evolved mechanisms to repress uncontrolled retrotransposon expansion, such as DNA methylation (Liu and Wendel 2000), deleterious mutations (Heslop-Harrison et al. 1997; Nuzhdin 1999), unequal crossing over and/or intrachromosomal recombination between LTRs (Shirasu et al. 2000). These findings could explain why only a small portion of the clones has a complete sequence of the *RNaseH*-LTR region.

Since primers planned for the *Str85* region yielded only few amplificates and the *Str6* primers did not consistently amplify all DNA samples in SSAP analysis, only the primers designed for *Str187* were used for subsequent experiments. Increasing the number of selective nucleotides on the LTR primer, the amounts of the detected insertions were reduced 4–5 times per nucleotide as expected. However only a slight reduction on the number of the bands was observed when the adaptor primers were extended (Table 1). This latter observation could explain why Waugh et al. (1997) had to use 2–3 selective nucleotides on the adaptor primer to reduce the amount of the amplified polymorphic bands to a scorable number, concurrent with the non-selective LTR primers. In our case, the best results were obtained by extending the LTR primer by only one nucleotide.

In the original S-SAP protocol (Waugh et al. 1997) the genomic DNA was cut with two restriction enzymes, a six and a four cutter, as is usual in AFLP (Vos et al. 1995). Ellis et al. (1998) and Pearce et al. (2000) published a modification of the method using only a four-cutter enzyme to analyse diverse *Pisum* accessions but without pre-amplification of the LTR-adaptor fragments. However, adapting the S-SAP technique to sweetpotato, digesting the genomic DNA only with a six-cutter enzyme (*EcoR1*) instead of two enzymes, improved the number and length of the polymorphic bands (data not shown). Additionally, pre-amplifying the adapted DNA with the non-extended adaptor and LTR primers resulted in more robust bands on the gels after the second amplification with extended primers.

Comparison of the RAPD, AFLP and S-SAP marker systems

The *Ty1-copia* retrotransposon-based S-SAP analysis is a dominant marker system yielding a multiband pattern. Each individual band of this pattern represents a unique retrotransposon integration site.

Estimating the frequency of polymorphic bands/insertions of the three marker systems (Table 2) revealed that the majority of the markers were polymorphic in the case of S-SAP (97–99%) while this was lower in the case of AFLP (70 to 90%) and RAPD (88%). In the barley genome a 25–30% increase in the rate of polymorphism was observed with retrotransposon-based S-SAP, in comparison to AFLP analysis (Kumar 1996; Waugh et al. 1997; Yu and Wise 2000). In our case, this difference was less pronounced (19%).

The retrotransposon-based S-SAP system monitors the polymorphisms of the transposon insertion sites, while RAPD and AFLP are dependent on DNA sequence variability. Therefore, we were interested to determine whether the three dominant marker systems S-SAP, RAPD and AFLP yield the same genetic relatedness of the genomes. Comparing the genetic distance data by the Mantel test we found a significant correlation between S-SAP and RAPD marker systems (0.46, $P = 0.01$), a marginal correlation between S-SAP and AFLP (0.33, $P = 0.07$) and no correlation between RAPD and AFLP (0.06, $P = 0.37$). This latter result could indicate that

RAPD and AFLP represent different regions/sequence elements of the genome (Powell et al. 1997), while S-SAP overlaps with both. AFLP amplification sites were reported to form clusters on genetic linkage maps (Becker et al. 1995). In UPGMA cluster analysis, the two South American clones (Camote Amarillo and Japonese) always formed one group with all the three systems; S-SAP and RAPD yielding high (89 and 100%) bootstrap values, while in the case of AFLP it was only 36%. In the case of the S-SAP analysis, the third South American clone Santo Amaro was also confined to this group (Fig. 2). The other two systems grouped the Brazilian clone with the African ones. Even within the African group, only the S-SAP method could distinguish the two Ugandan clones (Wagabolige and Kyebandula) from the Kenyan ones (Mafuta and Simama). This indicates the potential of the S-SAP marker system to reflect the geographic origin of the clones. This result suggests area-specific transposon insertions. Ellis et al. (1998) published a similar finding for pea.

Frequency of the transposon insertions

In order to assess the quality of the established S-SAP technique we compared the insertion patterns of nine selected clones representing Africa, South and Central America, as well as Papua New Guinea (Table 1). These clones were analysed by the *Str187* primers and resulted in 33 to 64% of genome-specific, unique insertion sites dependent on the primer combination used (Table 1). In this sample set, there were only a few insertion sites, which were common in all the tested genomes.

Additional analysis of 177 East-African sweetpotato accessions revealed a high divergence (12–83) in the number of retrotransposon insertions in the individual genomes (Fig. 3A). Furthermore, we found that insertion sites present in a small number of the genomes are more frequent than those that are present in the majority of the genotypes. In fact, only a single insertion was present in all the genomes analysed (Fig. 3B). Considering that sweetpotato has been present in the East African region for only about 500 years, and the number of genotypes introduced was possibly limited, a surprisingly high level of genetic variability of the transposon insertion sites was detected. It is known that different biotic and abiotic stresses can induce the mobility of the retrotransposon (Grandbastien et al. 1997; Mhiri et al. 1997). It is plausible that retrotransposon elements of the sweetpotato genomes introduced into East Africa were mobilised during adaptation of the species to the new environment. Therefore, the presence of the high portion of unique and the low level of common insertions could indicate the high mobility, as well as a high mutation rate, of the *Ty1-copia* type retrotransposons in sweetpotato. Consequently, the Str187 sweetpotato retrotransposon sequence may represent a still-mobile transposon. In this case local geographic-region specific transposon insertions can accumulate as suggested by Ellis et al. (1998). Considering that the evolution of the transposon insertions could be faster than the evolution of the rest of the genome (e.g. household genes), the genetic-diversity measures based on retrotransposon marker systems may overestimate the existing genetic diversity. Our results also support this idea since the highest level of polymorphism was found with the retrotransposon-based S-SAP markers.

The important factors in the choice of a genetic marker includes development time and cost, capital outlay, the amount and quality of DNA required, prior knowledge of DNA sequence, required technical expertise, robustness, informativeness, genome coverage and reproducibility (Vos et al. 1995; Powell et al. 1996; Milbourne et al. 1997, 1998). The S-SAP markers require a higher initial cost of development compared to both RAPD and AFLP due to the need to isolate the LTR sequence of the retrotransposon. The S-SAP was demonstrated to be superior to both RAPD and AFLP in terms of the number of amplification products and the number of polymorphic loci (Table 2). The four S-SAP primer combinations tested revealed a mean of 130 insertions per assay (Table 2). Considering that up to 12 RAPD, or at least 2 AFLP, assays were required to achieve approximately the same level of polymorphism, it is evident that on a per-assay basis the S-SAP procedure may be the best of the three methods for genetic analysis and the characterisation of the sweetpotato at a comparable cost. Compared to fluorescent AFLP we found that the S-SAP peaks were more distinct. Though both the AFLP and S-SAP markers are dominant, the high multiplex ratio (describes the number of different genetic loci that can be simultaneously analysed per experiment) of the S-SAP indicates that they are more informative. An additional advantage of a system anchoring one primer in the nuclear genome is that accidental contamination of the DNA sample by infective agents should not influence the resulting banding pattern.

Acknowledgements This study was done in the framework of a special International Potato Center/Austrian Research Center, Seibersdorf, collaborative project on sweetpotato germplasm. We are grateful to the sweetpotato scientists from the Kenya Agricultural Research Institute, the National Agricultural Research Organisation (Uganda) and the Ministry of Agriculture, Tanzania, in addition to Mrs. Erna Abidini of Arapai Agricultural College, Uganda, for providing plant material of the East African varieties. We acknowledge Dr. A. Flavell (University of Dundee, Scotland) for technical assistance and providing the degenerated *RNaseH* primers.

References

Boeke JD, Corces VG (1989) Transcription and reverse transcription of retrotransposons. Annu Rev Microbiol 43:403–434

Bennetzen JL (2000) Transposable element contribution to plant gene and genome evolution. Plant Mol Biol 42:251–269

Becker J, Vos P, Kuiper M, Salamini F, Heun M (1995) Combined mapping of AFLP and RFLP markers in barley. Mol Gen Genet 249:65–73

Buteler MI, Jarret RL, LaBonte DR (1999) Sequence characterization of microsatellites in diploid and polyploid *Ipomoea*. Theor App Genet 99:123–132

Connolly AG, Godwin ID, Cooper M, DeLacy IH (1994) Interpretation of randomly amplified polymorphic DNA marker data for fingerprinting sweetpotato (*Ipomoea batatas* L.) genotypes. Theor App Genet 88:332–336

Ellis THN, Poyser SJ, Knox MR, Vershinin AV, Ambrose MJ (1998) Polymorphism of insertion sites of *Ty1-copia* class retrotransposons and its use for linkage and diversity analysis in pea. Mol Gen Genet 260:9–19

Flavell AJ, Smith DB, Kumar A (1992a) Extreme heterogeneity of *Ty1-copia* group retrotransposons in plants. Mol Gen Genet 231:233–242

Flavell AJ, Dunbar E, Anderson R, Pearce SR, Hartley R, Kumar A (1992b) *Ty1-copia* group retrotransposons are ubiquitous and heterogeneous in higher plants. Nucleic Acids Res 20:3639–3644

Grandbastien MA, Spielmann A, Caboche M (1989) *Tnt1* a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. Nature 337:376–380

Grandbastien MA, Lucas H, Morel JB, Mhiri C, Vernhettes S, Casacuberta JM (1997) The expression of the tobacco *Tnt1* retrotransposon is linked to plant defense responses. Genetica 100:241–252

Heslop-Harrison JS, Brandes A, Taketa S, Schmidt T, Versinin AV, Alkhimova EG, Kamm A, Doudrick RL, Schwarzacher T, Katsiotis A, Kubis S, Kumar A, Pearce SR, Flavell AJ, Harrison GE (1997) The chromosomal distributions of *Ty1-copia* group retrotransposble elements in higher plants and their implications for genome evolution. Genetica 100:197–204

Kumar A (1996) The adventures of the *Ty1-copia* group of retrotransposons in plants. Trends Genet 12:4143

Kumar A, Bennetzen J (1999) Plant retrotransposons. Annu Rev Genet 33:479–532

Lee D, Ellis THN, Turner L Hellens RP, Cleary WG (1990) A copia-like element in *Pisum* demonstrates the uses of dispersed repeated sequences in genetic analysis. Plant Mol Biol 15:707–722

Liu B, Wendel JF (2000) Retrotransposon activation followed by rapid repression in introgressed rice plants. Genome 43:874–880

McClintock B (1984) The significance of responses of the genome to challenge. Science 226:792–801

Milbourne D, Meyer R, Bradshaw JE, Baird E, Bonar N, Provan J, Powell W, Waugh R (1997) Comparison of PCR-based marker systems for analysis of genetic relationships in cultivated potato. Mol Breed 3: 127–136

Milbourne D, Meyer RC, Collins AJ, Ramsay LD, Gebhardt C, Waugh R (1998) Isolation and characterisation and mapping of simple sequence repeat loci in potato. In: Karp A, Isaac PG, Igram DS (eds) Molecular tools for screening biodiversity. Chapman and Hall, London, pp 371–381

Mhiri C, Morel J-N, Vernhettes S, Casacuberta JM, Lucas H, Grandbastien MA (1997) The promoter of the tobacco *Tn1* retrotransposon is induced by wounding and abiotic stress. Plant Mol Biol 33:257–266

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA 76:5269–5273

Nuzhdin SV (1999) Sure facts, speculations, and open questions about the evolution of transposable element copy number. Genetica 107:129–137

Pearce SR, Harrison G, Li D, Heslop-Harrison J.S, Kumar A, Flavell AJ (1996) The *Ty1-copia* group retrotransposons in *Vicia* species: copy number, sequence heterogeneity and chromosomal localisation. Mol Gen Genet 250:305–315

Pearce SR, Stuart-Rogers C, Knox MR, Kumar A, Ellis THN, Flavell AJ (1999) Rapid isolation of plant *Ty1-copia* group retrotransposon LTR sequences for molecular marker studies. Plant J 19:711–717

Pearce SR, Knox M, Ellis THN, Flavel AJ, Kumar A (2000) Pea *Ty1-copia* group retrotransposons: transpositional activity and use as markers to study genetic diversity in *Pisum*. Mol Gen Genet 263:898–907

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tinge S, Rafalski A (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol Breed 2:225–238

Powell W, Thomas WTB, Baird E, Lawrence P, Booth A, Harrower B, Mcnicol JW, Waugh R (1997) Analysis of quantitative traits in Barley by the use of amplified fragment length polymorphism. Heredity 79:48–59

Purugganan MD, Wessler SR (1995) Transposon signatures: species-specific molecular markers that utilize a class of multiple-copy nuclear DNA. Mol Ecol 4:265–269

SanMiguel P, Bennetzen Jl (1998) Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. Ann Bot 81:37–44

SanMiguel P, Tikhonov A, Jin Y-K, Melake-Berhan A, Springer PS, Edwards KJ, Avramova Z, Bennetzen JL (1996) Nested retrotransposons in the intergenic regions of the maize genome. Science 274:765–768

Schmidt T, Kubis S, Heslop-Harrison JS (1996) Analysis and chromosomal localisation of retrotransposons in sugarbeet (*Beta vulgaris*): LINEs and Ty1-copia-like elements as major components of the genome. Chrom Res 3:335–345

Sharbel TF (1999) Amplified fragment length polymorphism: a non-random PCR-based technique for multilocus sampling. In: Epplen JT, Lubjuhn T (eds) DNA profiling and DNA fingerprinting. Birkhäuser Verlag, Basel/Switzerland, pp 177–194

Shirasu K, Schulman AH, Lahaye T, Schulze-Lefert P (2000) A contiguous 66-kb barley DNA sequence provides evidence for reversible genome expansion. Genome Res 10:908–915

Swarz-Sommer Z, Saedler H (1988) Transposition and retrotransposition in plants. In: Nelson O (ed) Plant transposable elements. Plenum Press, New York, pp 175–187

Tautz D (1988) Hypervariability of simple sequence repeats as a general source for polymorphic DNA markers. Nucleic Acids Res 17:6463–6471

Vos P, Hogers R, Bleeker M, Reijans M, Van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414

Waugh R, Mclean K, Flavell AJ, Pearce SR, Kumar A, Thomas BBT (1997) Genetic distribution of *Bare-1*-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphism (S-SAP). Mol Gen Genet 253:687–694

Wendel JF, Wessler SR (2000) Retrotransposon-mediated genome evolution on a local ecological scale. Proc Natl Acad Sci USA 97:6250–6252

Williams JGK, Kubelik AR, Livak KJ, Rafaski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18:6531–6535

Zabeau M, Vos P (1993) Selective restriction fragment amplification: a general method for DNA fingerprinting. European Patent Application 924026297

Zhang D, Ghislain M, Huamán Z, Golmirzaie A, Hijmans R (1998) RAPD variation in Sweetpotato [*Ipomea batatas* (L.) Lam] cultivars from South America and Papua New Guinea. Genet Resour Crop Evol 45:271–277

Yu G-X, Wise RP (2000) An anchored AFLP- and retrotransposon-based map of diploid *Avena*. Genome 43:736–749