

# DEVELOPMENT AND CHARACTERISATION OF IRAP MARKERS FROM EXPRESSED RETROTRANSPOSON-LIKE SEQUENCES IN *Pinus sylvestris* L.

Angelika Voronova and Dainis Runģis

Genetic Resource Centre, Latvian State Forest Research Institute „Silava”, Rīgas iela 111, LV-2169, Salaspils, LATVIA;  
angelika.voronova@silava.lv

Contributed by Dainis Runģis

*Conifer genomes are large and stably diploid, in contrast to angiosperms, which are more variable both in genome size and ploidy. Conifer genomes are characterised by multiple gene families and pseudogenes, contain large inter-gene regions and a considerable proportion of repetitive sequences. All members of plant retrotransposon orders have been identified in gymnosperm genomes, however active elements have not been described. Investigation of transposable elements in Scots pine (*Pinus sylvestris* L.) could offer insights into transposon-mediated reorganisation under stress conditions in complex and ancient plant genomes. Nine *Pinus sylvestris* specific markers were developed to hypothetical long terminal repeats (LTRs) from differentially expressed retrotransposon-like fragments after heat stress and insect damage. Genetic diversity of 150 trees from a naturally regenerated pine stand was investigated using the IRAP method. The developed markers revealed high levels of genetic diversity and were able to distinguish subpopulations growing in long-term differential environmental conditions. Somaclonal variation was also investigated using these markers and polymorphic fragments were identified between ramets of Scots pine clones growing in two different plantations, possibly indicating evidence of recent transposition events. Sequencing of the polymorphic fragments identified two groups of sequences containing LTR sequences of an unknown retrotransposon with homology to the LTRs of the Copia-17-PAb-I element.*

**Key words:** IRAP, transposable elements, retrotransposons, Scots pine.

## INTRODUCTION

Transposable elements (TEs) or mobile genetic elements are sequences that have the ability to change their location in a genome (transpose) within a single cell. These elements can be activated in stress conditions or in specific tissue types or developmental stages (McClintock, 1984; Wessler, 1996; Grandbastien, 1997; Kumar and Bennetzen, 1999; Capy *et al.*, 2000). TEs are classified into two classes according to the mode of transposition (RNA or DNA mediated): class I or retrotransposons (RE) and class II or DNA transposons (Finnegan, 1989; Capy, 2005; Wicker *et al.*, 2007). Both class I and II TEs are found in plant genomes, with the LTR REs are the most common (Kumar and Bennetzen, 1999). Large plant genomes contain many LTR RE families, which can contain over ten thousand copies in haploid genomes. Sequencing of the *Picea abies* genome revealed 1773 different repeat sequences, of which approximately half were similar to known TEs (Nystedt *et al.*, 2013). REs containing long terminal repeats are classified as LTR REs and according to the structure of the coding region are further classified into superfamilies (e.g. *Gypsy* and *Copia*). Long terminal repeats (LTRs) are non-coding direct repeats that

contain sequences regulating transcription. The length of LTRs can range from 85 bp in length (*FRetro129* from rice (Gao *et al.*, 2012)) up to 5kb (*Ogre* from pea (Neumann *et al.*, 2003)). The LTR sequence is specific to each RE family but contains some conserved sequences that regulate the transposition process. For example, reverse transcription is primed by various tRNA molecules that bind to the PBS region (Mak and Kleiman, 1997), which is highly conserved among RE families (Kalendar *et al.*, 2010). The U3 region of LTRs is variable between RE families and contains specific regulatory sequences that ensure expression in particular tissues, conditions or developmental stages. As in other plant species, LTR REs are the most common class of TEs in *P. abies*, and *Gypsy* REs are the most commonly represented REs (Nystedt *et al.*, 2013). Evidence of recent transposition bursts was not detected in the *P. abies* genome, and the ratio of single LTRs to full-length REs was found to be 1:9, indicating a low frequency of RE deletion in comparison to angiosperm genomes. Many diverged elements have been identified in the *Pinus taeda* genome that could be classified to novel families (Kovach *et al.*, 2010). Possible transpositional activity of conifer-specific elements has been previously analysed. The *Ty-Copia TPE1* RE has been

identified in *Pinus elliottii* (Kamm *et al.*, 1996), but it probably cannot transpose autonomously due to mutations in protein coding sequences (Brandes *et al.*, 1997). The *Gypsy IFG* RE was isolated from *Pinus radiata*, which contains a single stop codon between the gag and AP domains (Kossack and Kinlaw, 1999). *Spcl* from *Picea glauca* contains only one stop codon before the int domain, indicating the recent origin of this element (L'Homme *et al.*, 2000). A low-copy *Gypsy* RE *PpRT1* has been identified in *P. pinaster*, which is homologous to *IFG* (Rocheta *et al.*, 2007). *PpRT1* transcripts are also found in EST databases, indicating that it is expressed (Miguel *et al.*, 2008). Activation of REs in stress conditions may result in increased genotypic variation (Wessler, 1996; Murray, 1998; Murray, 2005). Experiments have shown that novel stress response networks can be activated due to RE transposition (Ito *et al.*, 2011). Transposition of REs can induce mutations, however most TEs have lost the ability to transpose, and only a few species-specific elements are active.

RE sequences are widely used for development of molecular markers in plants (Kumar *et al.*, 1997; Kumar and Hirochika, 2001; Schulman *et al.*, 2004; Schulman 2006; Schulman *et al.*, 2012). A number of features make REs advantageous as molecular markers: REs are well distributed and represented in all chromosomes, and are characterised by a high level of polymorphism. Polymorphism derived from RE markers reflects large structural rearrangements, from hundreds of bp to several kb in length. Additionally, RE induced mutations are irreversible in subsequent generations, overcoming the problem of homoplasia, which can affect other marker systems such as SSRs. RE-based markers can also be more sensitive than other marker systems. For example, they can differentiate breeding lines and clones, which are indistinguishable by SSR analysis (Tam *et al.*, 2005; D'Onofrio *et al.*, 2010; Baranek *et al.*, 2012; Carvalho *et al.*, 2012; Castro *et al.*, 2012; Subudhi *et al.*, 2013). Another characteristic feature of RE markers is that a relatively small amount of sequence information is required to obtain large amounts of genotypic data.

One RE based method is IRAP (Inter-Retrotransposon Amplified Polymorphism), where PCR fragments are amplified

from genomic regions between RE sequences in inverted orientation to each other (Kalendar *et al.*, 1999). RE tend to form clusters in plant genomes, therefore usually IRAP results in a high number of polymorphic bands. The disadvantages of IRAP are that it is a dominant and non-specific marker system. However, the benefit of IRAP is the high level of polymorphism, the possibility of automating genotyping and relatively low costs. To implement IRAP for analysis of a new species, information about species specific RE sequences should be available. Copy number of the particular RE family will affect the number of loci obtained by the markers (Kalendar and Schulman, 2007). In this study we used previously identified expressed RE-like sequences (Voronova *et al.*, 2011) to develop *P. sylvestris* specific RE markers and utilised these markers to assess polymorphism and differentiation in a natural pine population, as well as to investigate somaclonal variation in grafted pine clones.

MATERIALS AND METHODS

Differentially expressed RE-like fragments were isolated from stressed pine seedlings as described previously (Voronova *et al.*, 2011), using an iPBS-based method (Kalendar *et al.*, 2010). Nine Scots pine specific primers were developed (Table 1), based on hypothetical LTRs identified in differentially expressed RE-like fragments (Voronova *et al.*, 2011). IRAP analyses were conducted as described by Kalendar and Schulman (2007). Fragments were visualised on ethidium bromide stained gels with the UV transillumination system AlphaDigiDocRT (JH BIO Innovations) and Olympus sp-500uz camera. Images were processed with AlphaEaseFC v. 4.0.0. Binary genotypic data were analysed using GenAEx v. 6 (Peakall and Smouse, 2006) and FAMD v.1.30 (Schlüter and Harris, 2006).

Scots pine needle samples were harvested from 150 trees (approximately 36 years old) growing in a natural pine population (56° 46' 8'' N; 24° 33' 27'' E). The population could be divided into subpopulations according to growing conditions: 50 trees were growing on a ridge (30.4 m above sea level) and 50 trees in swampy conditions in the valley (Kausu swamp), and 50 trees were collected from the slope between the ridge and the swamp.

Table 1

PRIMER SEQUENCES AND GENETIC DIVERSITY PARAMETERS OF THE IRAP MARKERS

Primer	Sequence (5'-3')	Accession number of differentially expressed sequence	Average fragment no. per individual	No. of polymorphic fragments	Number of fixed fragments
IR_1	AACTTGGGGTTGACATACAACAGGCATCGA	JZ389987.1	11.60 ± 1.90	12	4
IR_2	GGGTTGTCAATTTGTATATTGCCGTGATA	JZ390004.1	13.72 ± 9.12	28	0
IR_3	GTTCAAAATCCCTTGTGGCT	JZ390006.1	8.40 ± 2.56	13	0
IR_4	TTCCAAAGTTTGGGGCCAAGTTCGGGTT	JZ390008.1	11.26 ± 3.68	19	1
IR_5	TTGAAGAAGGGCCAGAACCTAGTTCCTCCG	JZ390015.1	13.18 ± 4.18	19	0
IR_6	GACTGGTACATGGAATTGAAGAGAGATTCTA	JZ390027.1	22.20 ± 12.86	37	0
IR_7	CTGTGTGCGCCTATATTCCA	JZ390015.1	7.72 ± 1.99	14	0
IR_8	TCCATCTGTGGTGTTCGTCG	JZ389987.1	9.83 ± 1.13	13	3
IR_9	ATCTGGGGCTTCTTTGGACT	JZ390027.1	13.07 ± 2.65	19	2

Principal coordinate analysis (PcoA) was conducted using genetic distances calculated from binary data with GenAlEx v.6. In order to assess possible genetic differentiation detected with the IRAP markers, PcoA analyses were conducted on data from only the two subpopulations growing in the most differentiated conditions: on the ridge (50 samples) and in the swamp (50 samples). The 50 individuals collected from the slope between the ridge and the swamp were excluded from this analysis. For assessment of somaclonal variation, needles from 14 ramets of four 26 year old grafted pine clones were collected from two pine plantations ("Saviena" and "Dravas").

DNA was isolated from pine needles using a CTAB-based method (Porebski *et al.*, 1997). Genotyping with pine SSR markers PTTX3107, PTTX4001 and PTTX4011 (Soranzo *et al.*, 1998) was performed in order to confirm the clonal identity of the Scots pine ramets. IRAP products from the somaclonal variation study were analysed by electrophoresis in 1.7% TopVision Low Melting Point agarose (Thermo Scientific) for 8 h at 90 V. Polymorphic bands were isolated and purified using the Qiaquick Gel Extraction Kit (Qiagen). Fragments were re-amplified with a touch-down PCR programme from 55 to 47 °C with a final elongation step of 30 minutes. The PCR reaction mixture contained 30 ng purified PCR product, 1x Dream Taq buffer (Thermo Scientific), 2 mM MgCl<sub>2</sub>, 0.2 mM dNTP mix, 1 mM primer, 0.8 U DreamTaq polymerase (Thermo Scientific). Re-amplified fragment sizes were checked by electrophoresis, purified with Sephadex-G50 (Sigma) and ligated into the plasmid pTZ57R/T using the InsTAclone PCR Cloning Kit (Thermo Scientific). Competent *E. coli* cells were transformed with constructs using heat shock (Inoue *et al.*, 1990). Plasmids were isolated using an alkaline lysis method (Birnboim and Doly, 1979) with modifications (lysis buffer contained 100 mM NaOH, an additional chloroform:isoamylalcohol extraction was performed). Insert sizes were confirmed with double digestion with *Eco*RI and *Sa*II (Thermo Scientific). Sequencing was performed with M13

primers and BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). Sequencing reactions were analysed in POP7 polymer with ABI-Prism 3130x Avant Genetic Analyzer (Applied Biosystems).

Sequence analyses and mobile element identification was performed as proposed by Wicker *et al.* (2007). Searches were done in the NCBI data base using BLASTn against nucleotide sequences, Expressed Sequence Tags (ESTs), reference RNA and genomic sequences (<http://www.ncbi.nlm.nih.gov/BLAST/>). The mobile element database Repbase (<http://www.girinst.org/>) (Kohany *et al.*, 2006) was used to identify TEs. The identified *Copia-17-PTa-I* from *Pinus taeda* BAC sequence was submitted to Repbase. TE motifs were analysed using NSITE-PL (Solovyev, 2002), TSSP (Solovyev and Shahruradov, 2003), and RegRNA (Mignone *et al.*, 2005). Open reading frames were predicted using the online tool StarOrf (<http://star.mit.edu/>). Primers were designed to LTR RE-like structures using Primer-BLAST (Ye *et al.*, 2012).

## RESULTS

**Characterisation of IRAP markers in a natural pine population.** Markers for IRAP analysis (Kalendar *et al.*, 1999) were developed from LTR sequences of expressed RE-like sequences identified previously (Voronova *et al.*, 2011). Differentially expressed fragment sequences were analysed using homology searches to REs in available databases as well as based on presence of conservative motifs near the PBS site as described by Kalendar *et al.* (2010). Twenty markers were screened for amplification quality using four Scots pine individuals. Nine markers with the best amplification quality were used for further analysis (Table 1). In order to assess the informativeness and utility of these markers, 150 individuals were genotyped using the IRAP technique (Kalendar *et al.*, 1999) with the 9 IRAP markers, as well as with three SSR markers. Samples with inadequate amplification quality were excluded from analysis (Fig. 1).

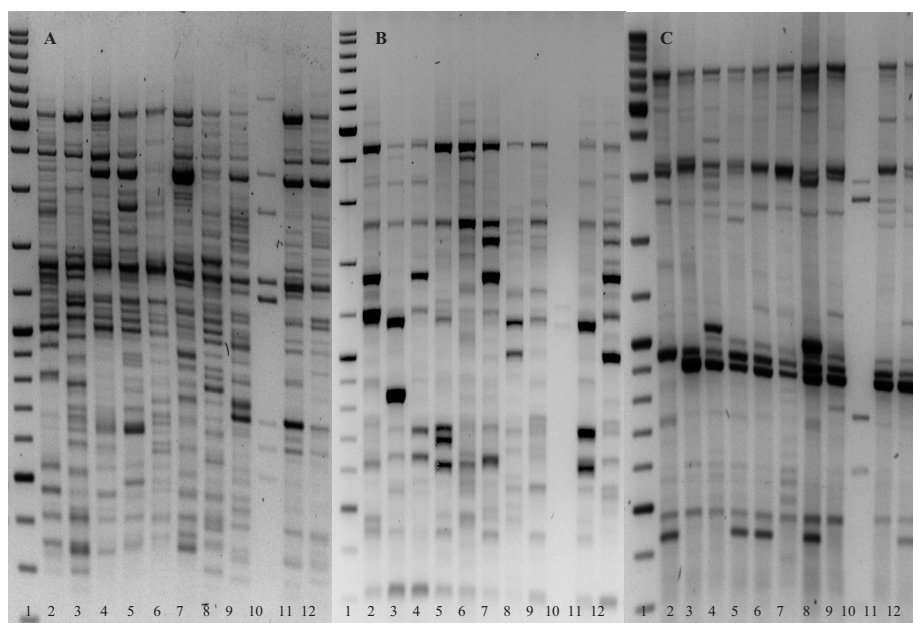


Fig. 1. Amplification with markers (A) – IR\_6; (B) – IR\_2; (C) – IR\_4. First lane – GeneRuler DNA Ladder Mix (Thermo Scientific). Lanes 2–12 – eleven *P. sylvestris* samples from a natural population. Low-quality samples such as in lane 10 were excluded from analysis.



Only clearly visible bands were analysed. 184 scorable bands were analysed in individuals from the natural pine population: 174 fragments were polymorphic and 10 were monomorphic. On average, 111 fragments were amplified from each individual. SSR analysis was conducted with the same samples and data were analysed in binary mode, resulting in 33 polymorphic loci. Fine-scale genetic differentiation between sub-populations derived from the total population was identified by the IRAP markers. Two sub-populations of 50 individuals each were identified, consisting of individuals growing in differing moisture conditions at the top of a ridge, and in a swamp. The 50 individuals growing in intermediate conditions on the slope connecting the ridge and the swamp were excluded from this analysis. Genetic differentiation (AMOVA) between these sub-populations obtained with both marker systems was comparable (Fig. 2). While analysis of molecular variance (AMOVA) did not identify a large differentiation between the subpopulations, Principal Coordinates Analysis (PCoA) of the RE marker data differentiated the ridge sub-population samples from the swamp sub-population samples (Fig. 3), while the SSR data did not indicate such a division. The first three coordinates of PCoA explain 56.46% of the IRAP data variation, and 61.15% of the SSR data variation.

**Assessment of somaclonal variation.** The developed markers were used to analyse fourteen 26 year old pine ramets from four different clones, growing in two different plantations, “Dravas” and “Saviena”. The ramets from one Scots pine clone were identical according to SSR marker genotyping (data not shown). Of the nine IRAP markers, there were one (IR\_2) amplified polymorphic fragments from one ramet of each of two pine clones (Fig. 4). Seven polymorphic fragments were excised, but only five were successfully cloned, and 17 plasmid inserts were sequenced. The IR\_2 primer sequence was identified at the ends of each sequenced fragment. Five unique sequences were identified that were derived from three groups of similar sequences and two individual sequences. One sequence group and the two individual sequences did not identify any homology in any of the databases queried. A second sequence group (869 bp in length) contained 98% identical sequences derived from two similarly sized excised polymorphic fragments from different clones (R-J17-II and Ug6-F). This sequence was similar to a small part of the LTR of the *Copia-17-PAb-I* element from *Picea abies* (GenBank accession no. MA\_172979). The identified sequence had a high simi-

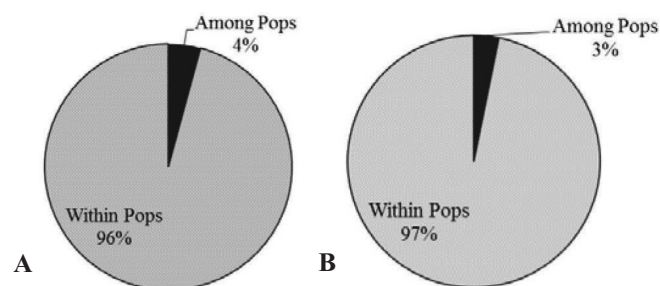


Fig. 2. Molecular variance (AMOVA) values utilising SSR (A) and IRAP marker data (B).

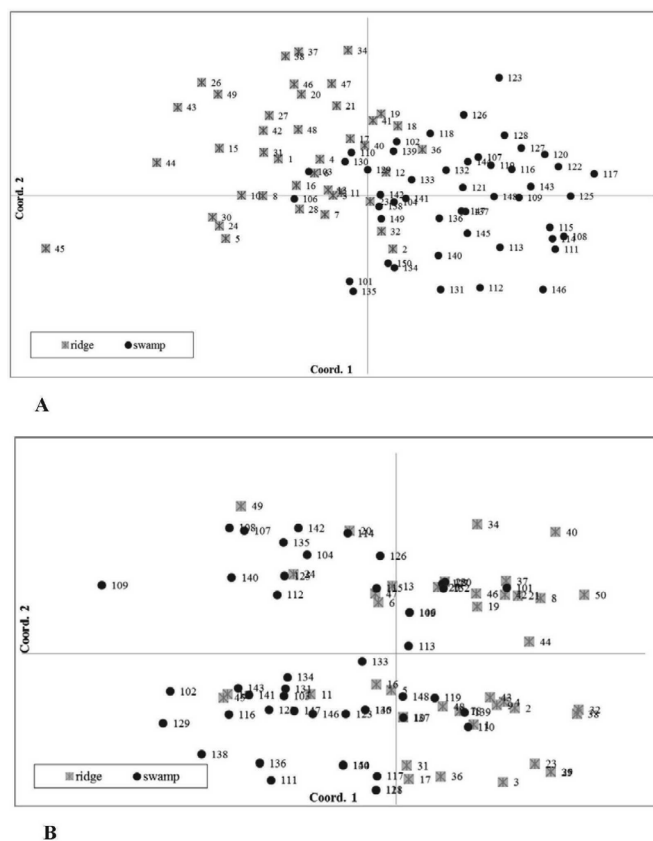


Fig. 3. (A) PCoA of IRAP data from the sub-populations (ridge, swamp) (B) PCoA of SSR data from the sub-populations (ridge, swamp).

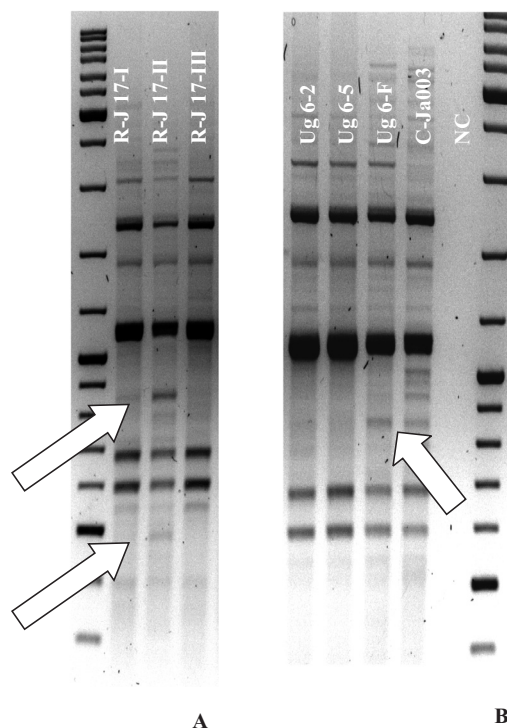


Fig. 4. IRAP amplification product polymorphism obtained with primer IR\_2 among pine ramets (A) – R-J 17, (B) – Ug 6; (C) – Ja03 – control of different genotype; NC-negative control.

larity level with a *Pinus taeda* BAC sequence (AC241282.1), from which a novel, previously unannotated 5986 bp putative full-length RE sequence homologous to *Copia-17-PAb-I* was identified and which was designated as *Copia-17-PTa-I* and submitted to the Repbase database. The internal sequences of both elements (*Copia-17-PAb-I* and *Copia-17-PTa-I*) are 72% identical (97% coverage), while the LTRs are 67-77% identical (25% coverage). 538 bp from the middle of the isolated polymorphic fragment from *P. sylvestris* was 80% similar to the 5' end of the *Copia-17-PTa-I* LTR. However, the flanking 235 bp and 97 bp at either end of the polymorphic fragment are not similar to the *Copia-17-PTa-I* element, suggesting that the 3' internal region of the *P. sylvestris* element differs from the *P. taeda* element. A 103 bp region at the 3' end of the *P. sylvestris* fragment is 99% similar to the 26S rRNA domain from various pine species. A third sequence group (507 bp in length), derived from one polymorphic fragment (from clone R-J17-II) was similar to *Copia-17-PAb-I*, spanning the 3' region of the internal domain and the 5' region of the LTR, and was also similar to *Copia-17-PAb-I*. Both fragment groups have a higher similarity level (77–80%) to different parts of the *Copia-17-PTa-I* element from *P. taeda*, but their sequences do not overlap each other.

It was not possible to use similar bioinformatic techniques to identify a putative full-length RE element from *P. sylvestris* due to the lack of BAC and other longer contiguous sequence data for this species. Therefore, the full-length sequence of the *Copia-17* RE in *Pinus sylvestris* has not been determined and in the absence of publicly available sequence resources, further sequencing is required to elucidate the complete structure of this element in *P. sylvestris* and to assess the possibility of transpositional activity in the *P. sylvestris* genome. However, the available partial *P. sylvestris* sequence was compared to the *Copia-17-PTa-I* and *Copia-17-PAb-I* sequences in order to identify functional motifs. The element *Copia-17-PAb-I* from *Picea abies* contains a 404 bp deletion in the reverse transcriptase domain. The polyadenylation motif (3'-AATAAA-5') was identified in the 5' LTR of the *Copia-17-PTa-I* element. The conserved plant RE promoter motif 3'-TATATA-5' (Vicient *et al.*, 2005; Benachenhou *et al.*, 2013) was only identified in the U3 region of the *P. sylvestris* LTR region of the sequenced polymorphic fragment. The particular *Copia-17-PTa-I* sequence found in the BAC clone sequence does not seem to be the result of a recent transposition event. The LTRs in the *Copia-17-PTa-I* element are 947 bp long and are 92% identical to each other, and the 5 bp Target site duplication (TSD) site at each end of the *Copia-17-PTa-I* element contain one SNP in comparison to each other. However, the *P. taeda* genome may contain other more functional copies of this element. The expression of sequences similar to this element seems to be widespread, with 293 EST sequences mostly from *Pinaceae* species (*P. taeda*, *P. banksiana*, *P. pinaster*, *Picea glauca*) found when the *Copia-17-PTa-I* sequence was queried against the green plants EST NCBI database with e-value

< 0.01. Both LTR and internal region sequences of the *Copia-17-PTa-I* element are present among the EST sequences.

## DISCUSSION

LTR REs are abundant and extremely diverse in eukaryotic genomes (Kumar and Bennetzen, 1999). Evolutionary studies of RE distribution and phylogeny among different plant species suggest that REs are evolutionally ancient, as homologous coding domains of RE pairs exist in distinct plant species (Xiong and Eickbush, 1990; Flavell *et al.*, 1994). In genomes of gymnosperms, REs of common origin have also been described (Berg and Howe, 1989; Stuart-Rogers and Flavell, 2001). The use of non-specific iPBS markers allows investigation of plant genomes with a low level of information about RE sequences (Kalendar *et al.*, 2010). In this study, previously identified stress-responsive transcripts of actively transcribed mobile elements or RE-like fragments (Voronova *et al.*, 2011) were utilised for IRAP marker development. Primer development relies on the assumption that the isolated differentially expressed fragments (Voronova *et al.*, 2011) contained a conservative PBS site in LTR regions (Kalendar *et al.*, 2010). Recent TE database updates indicated that some of the developed primers are located in the internal sequences of REs, rather than the LTRs, however, these primers still revealed a high level of polymorphism. Plant genomes often contain solo LTRs that remain in the genome after elimination of full-length REs via a non-homologous end joining process (Leitch and Bennett, 2004), and PCR amplification of regions between these solo LTRs may not reflect polymorphism of full-length REs. Some of the differentially expressed sequences from which the IRAP primers were developed were homologous to RE polyproteins or ORFs along their entire length, indicating that the PBS primers amplify these domains by annealing to sequences other than the primer binding sites (Voronova, unpublished data).

Polymorphic fragments derived from markers complementary to polyprotein domains may represent polymorphism derived from full-length RE structures of highly repeated RE families. Some polyprotein domains, e.g. reverse transcriptase, can be highly conservative among several RE families: therefore IRAP primers complementary to these domains could anneal to multiple element families. The IRAP method is non-specific and amplifies fragments between REs from multiple loci in a competitive manner (Kalendar *et al.*, 1999). The number of amplified fragments depends on the distribution and copy number of the specific RE family in a species' genome. Currently, IRAP markers have not been developed for any conifer species, therefore comparison of the amplification quality and informativeness of the developed IRAP markers was not possible. Analysis of natural pine populations with non-specific IRAP markers revealed the same high degree of diversity as SSR markers. Natural pine stands growing in long-term differential stress conditions are difficult to identify and to accurately characterise. The selected pine population was of natural origin

and sub-populations were identified which were growing in differing topographic conditions that indicate long-standing differences in moisture conditions. Principal coordinates analysis (PcoA) of the IRAP data differentiated the ridge and swamp populations. Alleles were equally represented in all populations, and the difference between the allele frequencies at each locus did not exceed 0.3. The genetic differentiation of the two subpopulations, which was detected with the IRAP markers, could be due to the transpositional activity of REs in the *P. sylvestris* genome in response to stress conditions. The IRAP markers reported in this study were developed from expressed RE-like sequences. Therefore, the polymorphism detected by at least some of the markers may reflect recent RE transposition events. It has been demonstrated that species with large genomes show a stronger association of genome size with environmental factors (Knight and Ackerly 2002). These results are consistent with the detailed wild barley *BARE-1* RE study, where the copy number of this widely distributed RE in the barley genome increased in populations growing in drought conditions, but PCoA of RE marker data differentiated populations growing in different microclimatic conditions (Kalendar *et al.*, 2000). Changes of the number of IRAP amplification products in relation to geographical origin is also described in other studies, such as the increased number of IRAP marker amplification products in north-western *Aegilops tauschii* populations (Saeid *et al.*, 2008). These results provide an insight into the relative degree of variation and diversity of natural pine population determined by RE distribution.

Previous reports on pine REs have not described transpositionally active elements. However, computational analyses of pine BACs have identified some elements with conserved LTRs that imply recent possible activity of these elements (Kovach *et al.*, 2010). There are many examples of transposition of non-autonomous TEs utilising proteins encoded by other REs (Flavell *et al.*, 1994; Feschotte *et al.*, 2002). Evidence of transcription of REs in *Pinaceae* is provided by the abundance of RE sequence fragments in EST databases, indicating that RE-like fragments are expressed in a wide range of conditions. Gymnosperm genomes are highly co-linear, and hybridisation experiments with TEs do not reveal large differences in genome organisation between distantly related gymnosperm species and strong reciprocal hybridisation signals are found even between distantly related species *Picea abies* and *Ginkgo biloba* (Friesen *et al.*, 2001). Therefore, it could be expected that a full-length homolog of the *Copia-17-PTa-I* RE is present also in the *P. sylvestris* genome.

The IRAP technique has been widely used to identify somaclonal variation (Asif and Othman, 2005; Lightbourn *et al.*, 2007; Bairu *et al.*, 2011; Campbell *et al.*, 2011; Bayram *et al.*, 2012). IRAP primers amplify the genomic region between two sufficiently close elements, and transposition of a TE can induce amplification of additional fragments (Kalendar *et al.*, 1999). IRAP analysis of somaclonal

variation between mature Scots pine ramets identified several polymorphic fragments. Sequencing of these identified the partial sequence of one RE that potentially could have induced the polymorphism by transposition. Two of the polymorphic fragment sequence groups were both homologous to the LTR *Copia-17-PTa-I* element from *Pinus taeda*, with partial sequence coverage, which may indicate that the same element may be inserted into different loci in different clones. Two elements could be assigned to one family if their internal sequence is more than 80% similar at the nucleotide sequence level, or if their LTRs are more than 60% similar (Wicker *et al.* 2007). We assigned the newly described element identified in a *P. taeda* BAC sequence to one family with *Copia-17-PAb-I* from *Picea abies*, as their LTRs share a high similarity level, although this coverage was only partial. In comparison to *Copia-17-PAb-I*, the newly reported *Copia-17-PTa-I* from *Pinus taeda* is longer and contains many polymorphisms as well as a large 404 bp insertion containing part of the reverse transcriptase domain. This indicates that, in *Pinus*, the *Copia-17* RE family may have a more complete structure than in *Picea abies*. Active REs are characterised by one continuous polyprotein ORF, and recently transposed TEs should have identical LTRs (Kumar and Bennetzen, 1999). The polyprotein domain of the *Copia-17-PTa-I* element contains stop codons, indicating that this element has lost autonomous transpositional activity, and that the LTRs are 92% homologous, and the TSD contains a SNP, which discounts the possibility of a recent transposition event for this copy of the element. However, taking into account the genome size and inter-species variation of REs, it is possible that the active element has not yet been identified in either *P. taeda* or *P. sylvestris*. Partial evidence for the expression of this element is provided by the presence of sequences homologous to parts of the *Copia-17-PTa-I* element among EST sequences for various pine species. Representation of sequences from the entire element (not only LTRs) in the ESTs indicates that these sequences could represent transcription of a full-length element.

It has been reported that the presence of only one active element family could produce proteins that can facilitate the transposition of non-autonomous elements (Berg and Howe, 1989; Flavell *et al.*, 1994; Feschotte *et al.*, 2002; Dewanieux *et al.*, 2003; Kazazian, 2004). While the full-length RE sequence has not been identified in the *P. sylvestris* genome, the U3 region of the LTR contains an intact TATATA motif that functions as a promoter and is conservative for plant REs (Vicent *et al.*, 2005; Benachenhou *et al.*, 2013). The IRAP marker results identifying sub-population differentiation and somaclonal variation provide an indication that this polymorphism may be due to RE transposition. Some of the IRAP fragments that were polymorphic between the pine ramets were homologous to a novel full-length RE identified in *P. taeda*. Further isolation and characterisation and full-length sequencing of the homologous RE in *P. sylvestris* is required to fully elucidate its structure and possible activity.



## ACKNOWLEDGMENTS

The research was supported by the European Social Fund Project “Importance of Genetic Factors in Formation of Forest Stands with High Adaptability and Qualitative Wood Properties” (NB.2009/0200/1DP/1.1.1.2.0/09/APIA/VIAA/146). The authors would like to thank Dr. silv. I. Baumanis for providing clonal material and Dr. chem. I. Veinberga for SSR genotyping of 14 pine clones.

## REFERENCES

- Asif, J. M., Othman, F. Y. (2005). Characterization of fusarium wilt-resistant and fusarium wilt-susceptible somaclones of banana cultivar Rastali (*Musa AAB*) by random amplified polymorphic DNA and retrotransposon markers. *Plant Mol. Biol. Rep.*, **23** (3), 241–249.
- Bairu, M. W., Aremu, A. O., Staden, J. V. (2011). Somaclonal variation in plants: Causes and detection methods. *Plant Growth Regul.*, **63**, 147–173.
- Baranek, M., Meszaros, M., Sochorova, J., Cechova, J., Raddova, J. (2012). Utility of retrotransposon-derived marker systems for differentiation of presumed clones of the apricot cultivar Velkopavlovická. *Sci. Hortic.*, **143**, 1–6.
- Bayram, E., Yilmaz, S., Hamat-Mecbur, H., Kartal-Alacam, G., Gozukirmizi, N. (2012). Nikita retrotransposon movements in callus cultures of barley (*Hordeum vulgare* L.). *Plant Omics Journal (POJ)*, **5** (3), 211–215.
- Benachenhou, F., Sperber, G. O., Bongcam-Rudloff, E., Andersson, G., Boeke, J. D., Blomberg, J. (2013). Conserved structure and inferred evolutionary history of long terminal repeats (LTRs). *Mobile DNA*, doi: 10.1186/1759-8753-4-5.
- Berg, D. E., Howe, M. H. (eds.) (1989). *Mobile DNA*. Washington, D.C.: American Society for Microbiology Press.
- Birnboim, H. C., Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucl. Acids Res.*, **7** (6), 1513–1523.
- Brandes, A., Heslop-Harrison, J. S., Kamm, A., Kubis, S., Doudrick, R. L., Schmidt, T. (1997). Comparative analysis of the chromosomal and genomic organization of *Ty1-copia*-like retrotransposons in pteridophytes, gymnosperms and angiosperms. *Plant Mol. Biol.*, **33** (1), 11–21.
- Campbell, B. C., LeMare, S., Piperidis, G., Godwin, I. D. (2011). IRAP, a retrotransposon-based marker system for the detection of somaclonal variation in barley. *Mol. Breed.*, **27**, 193–206.
- Capy, P. (2005). Classification and nomenclature of retrotransposable elements. *Cytogenet. Gen. Res.*, **110** (1–4), 457–461.
- Capy, P., Gasperi, G., Biemont, C., Bazin, C. (2000). Stress and transposable elements: Co-evolution or useful parasites? *Heredity*, **85**, 101–106.
- Carvalho, A., Guedes-Pinto, H., Lima-Brito, J. E. (2012). Genetic diversity in old Portuguese durum wheat cultivars assessed by retrotransposon-based markers. *Plant Mol. Biol. Rep.*, **30**, 578–589.
- Castro, I., D’Onofrio, C., Martın, J. P., Ortiz, J. M., De Lorenzis, G., Ferreira, V., Pinto-Carnide, O. (2012). Effectiveness of AFLPs and retrotransposon-based markers for the identification of Portuguese grapevine cultivars and clones. *Mol. Biotechnol.*, **52** (1), 26–39.
- D’Onofrio, C., De Lorenzis, G., Giordani, T., Natali, L., Cavallini, A., Scalabrelli, G. (2010). Retrotransposon-based molecular markers for grapevine species and cultivars identification. *Tree Genet. Gen.*, **6**, 451–466.
- Feschotte, C., Jiang, N., Wessler, S. R. (2002). Plant transposable elements: Where genetics meets genomics. *Nat. Rev. Genet.*, **3**, 329–341.
- Finnegan, D. J. (1989). Eukaryotic transposable elements and genome evolution. *Trends. Genet.*, **5**, 103–107.
- Flavell, A. J., Pearce, S. R., Kumar, A. (1994). Plant transposable elements and the genome. *Curr. Opin. Genet. Dev.*, **4**, 838–844.
- Friesen, N., Brandes, A., Heslop-Harrison, J. S. (2001). Diversity, origin and distribution of retrotransposons (*gypsy* and *copia*) in conifers. *Mol. Biol. Evol.*, **18** (7), 1176–1188.
- Gao, D., Chen, J., Chen, M., Meyers, B. C., Jackson, S. (2012). A highly conserved, small LTR retrotransposon that preferentially targets genes in grass genomes. *PLoS One*, doi:10.1371/journal.pone.0032010.
- Grandbastien, M. A., Lucas, H., Morel, J. B., Corinne, M. C., Vernhettes, S., Casacuberta, J. M. (1997). The expression of the tobacco *Tnt1* retrotransposon is linked to plant defense responses. *Genetica*, **100**, 241–252.
- Inoue, H., Nojima, H., Okayama, H. (1990). High efficiency transformation of *Escherichia coli* with plasmids. *Gene*, **96**, 23–28.
- Ito, H., Gaubert, H., Bucher, E., Mirouze, M., Vaillant, I., Paszkowski, J. (2011). An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature*, **472**, 115–118.
- Kalendar, R., Antonius, K., Smykal, P., Schulman, A.H. (2010). iPBS: A universal method for DNA fingerprinting and retrotransposon isolation. *Theor. Appl. Genet.*, doi:10.1007/s00122-010-1398-2.
- Kalendar, R., Grob, T., Regina, M., Suoniemi, A., Schulman, A. (1999). IRAP and REMAP: Two new retrotransposon-based DNA fingerprinting techniques. *Theor. Appl. Genet.*, **98**, 704–711.
- Kalendar, R., Schulman, A. H. (2007). IRAP and REMAP for retrotransposon-based genotyping and fingerprinting. *Nat. Protoc.*, **1** (5), 2478–2484.
- Kamm, A., Doudrick, R. L., Heslop-Harrison, J. S., Schmidt, T. (1996). The genomic and physical organization of *Ty1-copia*-like sequences as a component of large genomes in *Pinus elliottii* var. *elliottii* and other gymnosperms. *Proc. Natl. Acad. Sci. USA*, **93**, 2708–2713.
- Knight, C. A., Ackerly, D. D. (2002). Variation in nuclear DNA content across environmental gradients: A quantile regression analysis. *Ecol. Lett.*, **5**, 66–76.
- Kohany, O., Gentles, A. J., Hankus, L., Jurka, J. (2006). Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics*, doi:10.1186/1471-2105-7-474.
- Kossack, D. S., Kinlaw, C. S. (1999). IFG, a *gypsy*-like retrotransposon in *Pinus* (*Pinaceae*), has an extensive history in pines. *Plant Mol. Biol.*, **39**, 417–426.
- Kovach, A., Wegrzyn, J. L., Parra, G., Holt, C., Bruening, G. E., Loopstra, C. A., Hartigan, J., Yandell, M., Langley, C. H., Korf, I., Neale, D. B. (2010). The *Pinus taeda* genome is characterized by diverse and highly diverged repetitive sequences. *BMC Genomics*, doi: 10.1186/1471-2164-11-420.
- Kumar, A., Bennetzen, J. L. (1999). Plant Retrotransposons. *Annu. Rev. Genet.*, **33**, 479–532.
- Kumar, A., Hirochika, H. (2001). Applications of retrotransposons as genetic tools in plant biology. *Trends Plant Sci.*, **6** (3), 127–134.
- Kumar, A., Pearce, S. R., McLean, K., Harrison, G., Heslop-Harrison, J. S., Waugh, R., Flavell, A. J. (1997). The *Ty1-copia* group of retrotransposons in plants: Genomic organisation, evolution, and use as molecular markers. *Genetica*, **100** (1–3), 205–217.
- L’Homme, Y., Seguin, A., Tremblay, F. M. (2000). Different classes of retrotransposons in coniferous spruce species. *Genome*, **43**, 1084–1089.
- Leitch, I. J., Bennett, M. D. (2004). Genome downsizing in polyploid plants. *Biol. J. Linn. Soc.*, **82**, 651–663.
- Lightbourn, G. J., Jelesko, J. G., Veilleux, R. E. (2007). Retrotransposon-based markers from potato monoloids used in somatic hybridization. *Genome*, **50** (5), 492–501.
- Mak, J., Kleiman, L. (1997). Primer tRNAs for reverse transcription. *J. Virol.*, **71** (11), 8087–8095.
- McClintock, B. (1984). The significance of responses of the genome to challenge. *Science*, **226**, 792–801.
- Mignone, F., Grillo, G., Licciulli, F., Iacono, M., Liuni, S., Kersey, P. J., Duarte, J., Saccone, C. Pesole, G. (2005). UTRdb and UTRsite: A collec-

- tion of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. *Nucl. Acids Res.*, **33**, D141–146.
- Miguel, C., Simoes, M., Oliveira, M. M., Rocheta, M. (2008). Envelope-like retrotransposons in the plant kingdom: Evidence of their presence in Gymnosperms (*Pinus pinaster*). *J. Mol. Evol.*, **67**, 517–525.
- Murray, B.G. (1998). Nuclear DNA amounts in gymnosperms. *Ann. Bot.*, **82**, 3–15.
- Murray, B.G. (2005). When does intraspecific C-value variation become taxonomically significant? *Ann. Bot.*, **95**, 119–125.
- Neumann, P., Pozárková, D., Macas, J. (2003). Highly abundant pea LTR Retrotransposon *Ogre* is constitutively transcribed and partially spliced. *Plant. Mol. Biol.*, **53** (3), 399–410.
- Nystedt, B., Street, N. R., Wetterbom, A., Zuccolo, A., Lin, Y. C., Scofield, D. G., Vezzi, F., Delhomme, N., Giacomello, S., Alexeyenko, A., Vicedomini, R., Sahlin, K., Sherwood, E., Elfstrand, M., Gramzow, L., Holmberg, K., Hällman, J., Keech, O., Klasson, L., Koriabine, M., Kucukoglu, M., Käller, M., Luthman, J., Lysholm, F., Niittylä, T., Olson, A., Rilakovic, N., Ritland, C., Rosselló, J.A., Sena, J., Svensson, T., Talavera-López, C., Theißen, G., Tuominen, H., Vanneste, K., Wu, Z. Q., Zhang, B., Zerbe, P., Arvestad, L., Bhalerao, R., Bohlmann, J., Bousquet, J., Garcia, G. R., Hvidsten, T. R., de Jong, P., MacKay, J., Morgante, M., Ritland, K., Sundberg, B., Thompson, S. L., Van de Peer, Y., Andersson, B., Nilsson, O., Ingvarsson, P.K., Lundeberg, J., Jansson, S. (2013). The Norway spruce genome sequence and conifer genome evolution. *Nature*, doi:10.1038/nature12211.
- Peakall, R., Smouse, P. E. (2006). GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes*, **6**, 288–295.
- Porebski, S., Bailey, G. L., Baum, B. R. (1997). Modification of a CTAB DNA extraction protocol for plants containing high polysaharide and polyphenol componenets. *Plant Mol. Biol. Rep.*, **15** (1), 8–15.
- Rocheta, M., Cordeiro, J., Oliveira, M., Miguel, C. (2007). *PpRT1*: The first complete gypsy-like retrotransposon isolated in *Pinus pinaster*. *Planta*, **225**, 551–562.
- Saeidi, H., Rahiminejad, M. R., Heslop-Harrison, J. S. (2008). Retroelement insertional polymorphisms, diversity and phylogeography within diploid, D-genome *Aegilops tauschii* (*Triticeae*, *Poaceae*) sub-taxa in Iran. *Ann. Bot.*, **101** (6), 855–861.
- Schlüter, P. M., Harris, S. A. (2006). Analysis of multilocus fingerprinting data sets containing missing data. *Mol. Ecol. Notes*, **6**, 569–572.
- Schulman, A. H. (2007). Molecular markers to assess genetic diversity. *Euphytica*, **158**, 313–321.
- Schulman, A. H., Flavell, A. J., Ellis, T. H. (2004). The application of LTR retrotransposons as molecular markers in plants. *Meth. Mol. Biol.*, **260**, 145–173.
- Schulman, A. H., Flavell, A. J., Paux, E., Ellis, T. H. (2012). The application of LTR retrotransposons as molecular markers in plants. *Meth. Mol. Biol.*, **859**, 115–153.
- Solovyev, V. V. (2002). Structure, properties and computer identification of eukaryotic genes. In: *Bioinformatics Genomes to Drugs, Basic Technologies* (59–111 pp.). Lengauer, T. (ed.). Wiley.
- Solovyev, V. V., Shahmuradov, I. A. (2003). PromH: Promoters identification using orthologous genomic sequences. *Nucl. Acids Res.*, **31** (13), 3540–3545.
- Soranzo, N., Provan, J., Powell, W. (1998). Characterization of micro-satellite loci in *Pinus sylvestris* L. *Mol. Ecol.*, **7**, 1247–1263.
- Stuart-Rogers, C., Flavell, A. J. (2001). The evolution of *Ty1-copia* group retrotransposons in gymnosperms. *Mol. Biol. Evol.*, **18** (2), 155–163.
- Subudhi, P., Magpantay, G., Karan, R. (2013) A retrotransposon-based probe for fingerprinting and evolutionary studies in rice (*Oryza sativa*). *Genet. Res. Crop. Evol.*, **60** (4), 1263–1273.
- Tam, S. M., Mhiri, C., Vogelaar, A., Kerkveld, M., Pearce, S. R., Grandbastien, M. A. (2005). Comparative analyses of genetic diversities within tomato and pepper collections detected by retrotransposon-based SSAP, AFLP and SSR. *Theor. Appl. Genet.*, **110** (5), 819–831.
- Vicient, C. M., Kalendar, R., Schulman, A. H. (2005). Variability, recombination and mosaic evolution of the barley *BARE-1* retrotransposon. *J. Mol. Evol.*, **61**, 275–291.
- Voronova, A., Jansons, A., Ruņģis, D. (2011). Expression of retrotransposon-like sequences in Scots pine (*Pinus sylvestris* L.) in response to heat stress. *Environ. Exper. Biol.*, **9**, 121–127.
- Wessler, S. R. (1996). Plant retrotransposons: Turned on by stress. *Curr. Biol.*, **6** (8), 959–961.
- Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., Schulman, A. H. (2007) A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.*, **8** (12), 973–982.
- Xiong, Y., Eickbush, T. H. (1990). Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.*, **9**, 3353–3362.
- Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., Madden, T. (2012). Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, doi:10.1186/1471-2105-13-134.

Received 17 September 2013

## IRAP MARĶIERU IZVEIDOŠANA NO EKSPRESĒTĀM RETROTRANSPOZONIEM LĪDZĪGĀM SEKVENCĒM UN TO RAKSTUROŠANA *Pinus sylvestris* L.

Skujkoku genoms ir stabili diploīds un no vairākuma segsēkļu atšķiras ar genoma kompleksumu. Skujkoku genoms satur daudzkopiju gēnu ģimenes, pseidogēnus, lielus starpgēnu un gēnu intronu rajonus, kā arī ievērojamu transponējamo elementu frakciju. Kailsēkļu genoms ir atrasti visu galveno retrotranspozonu kārtu pārstāvji, taču transpozicionāli aktīvie elementi nav aprakstīti. Tāpēc pētījums par parastās priedes (*Pinus sylvestris* L.) genoma retrotranspozonu transkripciju varētu sniegt zināšanas par komplekso augu genomu pārkārtojumiem stresa apstākļos. Izmantojot hipotētiskus garus terminālus atkārtojumus (*Long Terminal Repeat*, LTR) no atšķirīgi ekspresētām retrotranspozoniem līdzīgām sekvencēm, izveidoti deviņi *P. sylvestris* specifiski marķieri. Izmantojot amplifikācijas starp retrotranspozonu sekvencēm (*Inter Retro Transposon Amplified Polymorphism*, IRAP) metodi, analizētas 150 priedes no dabiski atjaunojušās populācijas. Izveidotie marķieri atklāja augstu ģenētiskās daudzveidības līmeni, kā arī spēja nodalīt divas priedes subpopulācijas, augošas paugurā un pārpurvojumā. Ar izveidoto marķieru palīdzību analizēja parastās priedes klonu rametus, kas aug dažādās plantācijās. Rezultātā atklāja polimorfus fragmentus, kas, iespējams, norāda uz nesenu transpozīcijas gadījumu. Fragmentu sekvenēšanas rezultātā identificēja trīs sekvenču grupas, kas satur līdz šim neaprakstīta retrotranspozona *Copia-17-PAb-I* homologus rajonus.