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 being 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 (LTRs) are non-coding direct repeats that contain sequences regulating transcription. The length of LTRs can range from 85 bp in length (*F Retro129* 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* was isolated from *Pinus radiata*, which contains a single stop codon between the gag and AP domains (Kossack and Kinlaw, 1999). *Spel* 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 *Pinus* *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 TE families 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 homoplasy, 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 marker systems. For example, they can differentiate breeding lines and clones, which are indistinguishable by SSR marker systems. 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 TE families have lost the ability to transpose, and only a few species-specific elements are active.

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). REs 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 GenAlEx 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 (Kaus swamp), and 50 trees were collected from the slope between the ridge and the swamp.

### Table 1

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5’-3’)</th>
<th>Accession number of differentially expressed sequence</th>
<th>Average fragment no. per individual</th>
<th>No. of polymorphic fragments</th>
<th>Number of fixed fragments</th>
</tr>
</thead>
<tbody>
<tr>
<td>IR_1</td>
<td>AACTTGGGGTTGACATAACAACAGGCATCGA</td>
<td>JZ389907.1</td>
<td>11.60 ± 1.90</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>IR_2</td>
<td>GGTGTTGCAATTTGTATATGCGGTTGATA</td>
<td>JZ390004.1</td>
<td>13.72 ± 9.12</td>
<td>28</td>
<td>0</td>
</tr>
<tr>
<td>IR_3</td>
<td>GGTCGAAATCCCTGTGCTGTGCT</td>
<td>JZ390006.1</td>
<td>8.40 ± 2.56</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>IR_4</td>
<td>TTGCGAAAGGGCCATCGGTTGGT</td>
<td>JZ390008.1</td>
<td>11.26 ± 3.68</td>
<td>19</td>
<td>1</td>
</tr>
<tr>
<td>IR_5</td>
<td>TGAGGAAAGGGCCAGACCTAGTCCCG</td>
<td>JZ390015.1</td>
<td>13.18 ± 4.18</td>
<td>19</td>
<td>0</td>
</tr>
<tr>
<td>IR_6</td>
<td>GACGTGGTAATGGGATGGAAGAGGATTCTATA</td>
<td>JZ390027.1</td>
<td>22.20 ± 12.86</td>
<td>37</td>
<td>0</td>
</tr>
<tr>
<td>IR_7</td>
<td>CTGTTGGCCTATATATGCA</td>
<td>JZ390015.1</td>
<td>7.72 ± 1.99</td>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td>IR_8</td>
<td>TCCACGTGTTGCTGTTGTCG</td>
<td>JZ390987.1</td>
<td>9.83 ± 1.13</td>
<td>13</td>
<td>3</td>
</tr>
<tr>
<td>IR_9</td>
<td>ATCTGGGGCTCTTTGGACT</td>
<td>JZ390027.1</td>
<td>13.07 ± 2.65</td>
<td>19</td>
<td>2</td>
</tr>
</tbody>
</table>
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 (Qia-gen). 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 MgCl₂, 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 EcoRI and XhoI (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 Shahmuradov, 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 sub-populations, 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-
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* element from *P. taeda* and *Copia-17* sequences in order to identify functional motifs. The element *Copia-17* from *Picea abies* contains a 404 bp deletion in the reverse transcriptase domain. The polyadenylation motif (3'-AATAAA-5') was identified in the 5' end of the *Copia-17* 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* 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* 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* 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* 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* 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 evolutionarily 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 (Kalander 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 (Kalander 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 (Kalander 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-I* 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 transposition 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 *Ginko 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; Dewannieux 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 (Vicient 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/VI/AA/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


Received 17 September 2013

IRAP MARĶĶIERU IZVEIDOŠANA NO EKSPRĒSĒTĀM RETROTRANSPOZONIEM LĪDZĪGĀM SEKVENCEM UN TO RAKSTUROŠANA *Pinus sylvestris* L.