

Latvijas Universitāte  
Bioloģijas fakultāte

Angelika Voronova

RETROTRANSPOZONU STRUKTŪRA PARASTĀS PRIEDES  
(*PINUS SYLVESTRIS* L.) GENOMĀ UN TO EKSPRESIJA

Promocijas darba kopsavilkums  
Doktora grāda iegūšanai bioloģijas nozarē  
Apakšnozare: molekulārā bioloģija

Rīga, 2014



Promocijas darbs izstrādāts :  
Latvijas Valsts mežzinātnes institūtā (LVMI) „Silava”  
laika posmā no 2009. līdz 2014. gadam.



2009.-2012. gados darbs realizēts ar Eiropas Sociālā  
fonda projekta „Ģenētisko faktoru nozīme adaptēties spējīgu  
un pēc koksnes īpašībām kvalitatīvu mežaudžu izveidē”  
atbalstu (Nr. 2009/0200/1DP/1.1.1.2.0/09/APIA/VIAA/146).

Darbs sastāv no ievada, 7 nodaļām, literatūras saraksta un 2 pielikumiem.

Darba forma: disertācija bioloģijas nozarē, molekulārās bioloģijas apakšnozarē

Darba zinātniskie vadītāji: 1) Dainis Ruņģis, Dr.biol., vad. pētn. LVMI Silava;  
2) Nils Rostoks, Dr.biol., vad.pētn., Latvijas Universitāte.

Darba recenzenti:

- 1) Īzaks Rašals, Dr.hab.biol., prof., Latvijas Universitāte.
- 2) Gunārs Lācis, PhD, vad.pētn., Latvijas Valsts Augļkopības institūts.
- 3) Agnese Kolodinska Brantestam, PhD, vad.pētn., Ziemeļvalstu Ģenētisko  
resursu centrs

Promocijas darba aizstāvēšana notiks Latvijas Universitātes bioloģijas nozares  
promocijas padomes atklātā sēdē 2014. gada 4. februārī, plkst. 10:00 LU Bioloģijas  
fakultātē (Kronvalda bulv. 4., Rīga) 2. klausītavā.

Ar promocijas darbu var iepazīties Latvijas Universitātes Akadēmiskās bibliotēkas  
filiālē, Rīgā, Raiņa bulv. 19-203.

LU bioloģijas zinātņu nozares promocijas padomes

priekšsēdētājs \_\_\_\_\_ /Prof., Dr.hab biol. Ģederts Ieviņš/

Promocijas padomes sekretāre \_\_\_\_\_ /Daina Eze/

## ANOTĀCIJA

Promocijas darbs izstrādāts Latvijas Valsts mežzinātnes institūtā „Silava”. Pētījuma mērķis bija identificēt aktīvos retrotranspozonus (RE) parastās priedes (*Pinus sylvestris* L.) genomā un raksturot RE-līdzīgo sekvenču transkripcijas līmeni dažādu stresa veidu ietekmē kontrolētos apstākļos. Pētījumā izmantoti nespecifiskie RE marķieri balstīti uz augsti konservatīvām tRNS piesaistes vietām, kas ļauj pētīt RE sekvenču bez iepriekšējām zināšanām par organisma genoma sekvenci. Inovatīvi šos marķierus izmantoja pētot transkriptomu starp viena klona priedes rametiem, kas praktiski izslēdza genotipu variāciju starp dažādiem indivīdiem. Rametus pakļāva dabā izplatītam karstuma stresam, kukaiņu invāzijai (priežu hermess, (*Pineus pini* L.)), kā arī mākslīgi izraisītai augu hormonu (abscīzskābes un salicilskābes) ietekmei. Atšķirīgi ekspresētie fragmenti sekvenēti ar Sangerā sekvences noteikšanas metodi, kas šobrīd ir visprecīzākā zināmā nukleotīdu sekvences noteikšanas metode. Darba rezultāti sakrīt ar citu pētījumu publicētajām jaunākajām atziņām par RE un ar tiem saistīto sekvenču transkripciju, kā arī ar kailsēkļu genomu strukturālajiem pētījumiem. Darba rezultātā identificēti pārsvarā I klases RE pārstāvji (LTR *Gypsy* un *Copia*), kā arī tos saturošie himēriskie transkripti, kas norāda uz to, ka RE sekvenču ir izplatītas priedes genomā un RE-līdzīgie transkripti tiek ekspresēti atbildot uz stresu. Noskaidrota arī atrasto sekvenču homoloģija ar iepriekš aprakstīto citu augu sugu RE sekvencēm. Atšķirīgo elementu ģimeņu pārstāvji tika identificēti dažādu ietekmju rezultātā, starp eksperimentiem pārklājās tikai daži fragmenti. Fragmenti atšķīrās savā starpā arī pēc konstrukcijas (RE *pol*, LTR, LTR-LTR, RE *pol*-hloroplasts), kā arī RE domēnu reprezentācijas (AP, RT, gag). Identificēto RE-līdzīgo fragmentu ekspresijas līmeni noteica ar reālā laika PĶR metodi. Ievērojamu ekspresijas pieaugumu novēroja hloroplastu sekvencēm līdzīgajiem himēriskajiem fragmentiem, kas skaidrojams ar šīs sekvences daļas reprezentāciju arī hloroplastu genomā un/vai šo fragmentu homoloģiju ar citām sekvencēm kompleksajā *P.sylvestris* genomā. Identificēja četrus diferenciāli ekspresētus fragmentus, kas satur gan RE homologas sekvenču, gan hloroplastu sekvencēm līdzīgas daļas. B153TE LTR *Gypsy Cereba*-līdzīga fragmenta ekspresija paaugstinās 45-73 reizes kukaiņu invāzijas rezultātā un 5,62 reizes pēc karstuma stresa, salīdzinot ar normāliem apstākļiem. Savukārt hloroplastam līdzīgā B153C fragmenta ekspresija paaugstinās 6888-9373 reizes pēc kukaiņu uzbrukuma un 255 reizes pēc karstuma stresa. Vairākām noteiktām šī fragmenta sekvencēm, un īpaši RE daļai, ir novēroti polimorfismi, kas liecina par hloroplastam līdzīgā lokusa vairākām kopijām nukleārajā priedes genomā. Šī veida transkripti varētu piedalīties regulatīvajos procesos ar RNS interferences mehānisma starpniecību. No cDNS ieguva pilna izmēra 5796 bp elementu ar *Gypsy* LTR RE struktūrelementiem, to sekvenēja un raksturoja (psTE283). Tāpat, dažiem no priedes transkriptoma identificētiem RE homoloģiem izdevās iegūt RE iekšējo domēnu sekvenču. Tas liecina par to, ka diferenciāli ekspresēto fragmentu vidū atrodamas ne tikai himēriskas sekvenču un izjaukto RE sekvenču transkripti. Turpmāk ir jānoskaidro, vai mūsu pētījumā identificētās sekvenču ir lokalizētas ģēnu rajonos, vai arī tie ir autonomi RE. Šī darba gaitā ir izveidoti priedes RE specifiskie DNS marķieri, kuri ir izmantoti vairāku eksperimentu veikšanai. Izmantojot IRAP metodi, ir pētīti dažādos apstākļos augoši dabīgās *P.sylvestris* audzes indivīdi. Novērota augsta ģenētiskā daudzveidība, kas ir salīdzināma ar neitrālo SSR marķieru datiem.

Galveno koordināšu analīze ar IRAP datiem nodalīja krasi atšķirīgos reljefa apstākļos augošus indivīdus. Noskaidrots, ka paugurā augošām priedēm ir statistiski būtisks amplificēto fragmentu skaita palielinājums, salīdzinot ar mitros apstākļos augošajiem kokiem. Izmantojot IRAP metodi analizēja Latvijas austrumu un rietumu pusēs atšķirīgos klimatiskajos apstākļos esošās plantācijās augošus 26 gadus vecus priedes rāmetus. Tikai viens marķieris no desmit uzrādīja polimorfismu rāmetu vidū, kas norāda uz iespējamu nesen notikušu transpozīciju šiem indivīdiem. Polimorfu fragmentu sekvenēšana atrada homologiju ar līdz šim priedē neaprašstītu *Copia* LTR sekvenci diviem atšķirīgiem fragmentiem. Vēl divi fragmenti saturēja priedē neaprašstīta DNS transpozona homologus rajonus. Trešais eksperiments veltīts identificēto RE-līdzīgo fragmentu izplatības pētījumiem, kas ir veikts gan ar specifiskiem fragmentu praimeriem, gan ar IRAP praimeriem. Pētījumā iekļāva dažādas skujkoku (*Pinales*) sugas: 16 priedes, 14 egles, 7 baltegles, kā arī citus attālāku kailsēkļu taksonu pārstāvjus. Izmantojot RE marķierus, veikta kailsēkļu filoģenētiskā analīze. Ar specifisko RE amplifikāciju, kā arī ar nespecifisku RE amplifikāciju pēc IRAP metodes iegūtie dati varēja nodalīt priedes apakšģintis *Pinus* un *Strobus*. IRAP dati identificēja priežu, egļu un baltegļu sugu klāsterus, lai gan zarojumu varbūtības parametri nebija augsti.

## SATURS

ANOTĀCIJA .....	3
SATURS.....	5
VISPĀRĪGAIS DARBA RAKSTUROJUMS.....	6
1. PĒTĪJUMA TEORĒTISKAIS PAMATOJUMS .....	9
2. MATERIĀLS UN METODES .....	14
3. GALVENIE REZULTĀTI.....	17
4. DISKUSIJA.....	30
SECINĀJUMI .....	36
PATEICĪBAS .....	36
IZMANTOTĀS LITERATŪRAS SARAKSTS.....	39

## VISPĀRĪGAIS DARBA RAKSTUROJUMS

### ***Tēmas aktualitāte un pētījuma novitāte***

Kopš TE atklāšanas tie ilgu gadus tika uzskatīti par funkcionālo slodzi nenesošām sekvencēm, par parazītiem, kas savtīgos nolūkos pavairojas genomā, izjaucot kodējošas gēnu sekvenču (Orgel, Crick 1980; Doolittle, Sapienza 1980; Kidwell, Lisch 2001). Taču šis uzskats mainījās līdz ar vairāku genomu sekvenēšanu un to detalizētiem pētījumiem. Tagad ir zināms, ka TE sekvenču ir plaši izplatītas visu eikariotu, baktēriju un arheju genomos un, atšķirībā no konservatīviem gēniem, ievērojami variē gan pēc skaita, gan sastāva (Kazazian 2004, 2011). TE dalība gēnu regulācijā, hromosomu struktūru uzturēšanā un organismu evolūcijā pašlaik ir nepārprotama (Wessler 1996; Kumar, Bennetzen 1999; Kazazian 2004, Madlung, Comai 2004; Kazazian 2011; Rebollo *et al.* 2012). Augu genomi satur ievērojami vairāk TE, to sekvenču var sastādīt līdz pat 90% no visa genoma daļiem liliņu dzimtas pārstāvjiem (Feschotte *et al.* 2002). RE ir visplašāk pārstāvētā TE grupa augu genomos. Balstoties uz RE sekvencēm izveidotas vairākas molekulāro marķieru sistēmas. RE augu genomos parasti ir izplatītas, plaši sastopamas, reprezentētas visās hromosomās, to insercijas ir neatgriezeniskas un raksturojas ar augstu polimorfisma līmeni. Šīs īpašības padara RE par vērtīgu izpētes objektu ar plašu praktiskās izmantošanas potenciālu.

Transpozīcijas aktivitāte parasti genomā tiek nomākta izmantojot hromatīna modifikācijas, un abu mobilu elementu klašu aktīvie pārstāvji transponējas tikai stresu ietekmē vai noteiktos attīstības posmos un specifiskos organisma audos. Transpozīcija ir viens no ievērojamākajiem genoma nestabilitātes mehānismiem, kas iniciē pārkārtojumus un līdz ar to arī jaunu genotipu veidošanos, ļaujot pielāgoties mainīgajiem vides apstākļiem. TE var ietekmēt gēnu aktivitāti tiešā veidā transponējoties netālu no kodējošiem rajoniem un izjaucot tos, vai arī ietekmēt gēnu darbību ar savām regulatīvajām sekvencēm. Turklāt, RE var piedalīties regulatīvajos procesos ar RNS interferences mehānisma starpniecību, producējot dažāda tipa regulatīvās RNS molekulas. Pašlaik nav skaidrs, cik lielā mērā transpozīcijas izsuktā mutaģenēze ietekmē augu stresa apstākļos un atjaunošanās stadijā. Priedes genomā atrasti visu galveno TE ģimeņu pārstāvji, taču transkripcionāli vai transpozicionāli aktīvie TE nav pētīti. Kailsēkļu grupas pārstāvju genomi atšķiras no vairākuma eikariotu ar genoma lielumu un augstu gēnu un nekodējošo sekvencu atkārtojumu saturu. Tāpēc pirmā kailsēkļu - parastās egles (*Picea abies*) - sekvence tika publicēta tikai 2013. gada pavasarī (Nystedt *et al.* 2013). Skujkoku genomi satur daudzkopiju gēnu ģimenes, pseidogēnus, tiem ir lieli starpgēnu un intronu rajoni. Tāpēc pētījums par *P.sylvestris* genoma RE transkripciju varētu sniegt zināšanas par komplekso augu genomu pārkārtojumiem stresa apstākļos, kā arī par ar RE saistītām regulatīvām sekvencēm.

### ***Promocijas darba mērķi un uzdevumi***

Promocijas darba mērķis bija identificēt aktīvos retrotranspozonus parastās priedes (*Pinus sylvestris* L.) genomā un raksturot retrotranspozoniem-līdzīgo sekvencu transkripcijas līmeni dažādu stresa veidu ietekmē kontrolētos apstākļos.

Izvirzītie darba uzdevumi :

1. Identificēt un raksturot ar retrotranspozoniem saistītus transkriptus ekspresētus karstuma ietekmē, priežu hermesa (*Pineus pini* L.) invāzijas ietekmē un pēc apstrādes ar abscīzskābi un salicilskābi.

2. Salīdzināt identificēto fragmentu ekspresiju pētamo stresoru ietekmē, kā arī noteikt to ekspresijas līmeni.
3. Noskaidrot identificēto fragmentu piederību himēriskiem transkriptiem vai pilna izmēra retrotranspozoniem. Izolēt pilna izmēra retrotranspozonu sekvenci un raksturot tā transkripcijas un transpozīcijas iespējas.
4. Izveidot *P.sylvestris* retrotranspozonu molekulāros marķierus un ar tiem raksturot dabīgas *P.sylvestris* audzes daudzveidību, kā arī analizēt priedes rametu somaklonālās variācijas iespējas.
5. Raksturot identificēto fragmentu un retrotranspozonu izplatību kailsēkļu genomos, kā arī analizēt retrotranspozonu izsaukto daudzveidību citu kailsēkļu genomos.

#### ***Aizstāvēšanai izvirzītās tēzes***

1. *P.sylvestris* genoms satur transkripcionāli un transpozicionāli aktīvus retrotranspozonus.
2. Transkripcionāli aktīvie retrotranspozoni un ar tiem saistītās sekvenču diferenciāli ekspresējas biotiskā un abiotiskā stresa ietekmē.
3. Nelabvēlīgo vides apstākļu ietekmē *P.sylvestris* genomā notiek jaunas retrotranspozīcijas un, iespējams, arī elementu ekspansija.
4. *P.sylvestris* retrotranspozoni ir nevienmērīgi izplatīti dažādu skujkoku sugu genomos.

#### ***Īss metožu raksturojums***

Priedes rametu vienveidības pārbaudei visos eksperimentos izmantoja SSR (*Simple Sequence Repeat*) marķierus. RE-līdzīgo sekvenču identificēšanai izmantoja iPBS metodi (Kalendar *et al.* 2010), kas izmanto tRNS piesaistes vietas sekvencēm komplementārus praimerus (*Primer Binding Site, PBS*), kas ir konservatīvi dažādām RE ģimenēm. Fragmentu sekvenēšanai tos liģēja ar plazmīdu, kam sekoja transformēšana *E.coli* šūnās (Inoue *et al.* 1990). Insertus sekvenēja ar Sangera metodi. Sekvenču analizēja izmantojot nukleotīdu sekvenču datubāzes, kā arī specifiskas TE sekvenču datubāzes pēc Wicker *et al.* 2007 piedāvātās shēmas. Praimeru konstruēšanai izmantoja primer-BLAST (Ye *et al.* 2012). Priedes klonālās un naturālās daudzveidības pētījumos izmantoja pētījuma gaitā izveidotus marķierus pēc IRAP metodes (Kalendar *et al.* 1999). Molekulāro marķieru datus analizēja izmantojot programmas *GenalEx* v. 6 (Peakall, Smouse 2006), *FAMD* v.1.30 (Schlüter, Harris 2006). Kailsēkļu sugu paraugus analizēja ar pētījuma gaitā izveidotiem fragmentu, RE un IRAP praimeriem.

#### ***Darba rezultāti tika prezentēti starptautiskajās zinātniskās konferencēs:***

1. 19.-22.10.2012. Voronova A, Jansons Ā, Ruņģis D. Identification of retrotransposon-like sequences in *Pinus sylvestris* L. activated in response to stress conditions. 5th Baltic Congress of Genetics, Kaunas, Lithuania.
2. 16.-18.10.2012. Voronova A, Jansons Ā, Ruņģis D. Activation of retrotransposon-like sequences in *Pinus sylvestris* in response to stress conditions. NovelTree Conference, Helsinki, Vantaa, Finland.
3. 3.-5.10.2012. Voronova A, Jansons Ā, Ruņģis D. Retrotransposon-like sequences expressed under various stressors in pine genome shows different similarity level to known mobile genetic elements. AdaptCar

- conference: Genetic aspects of adaptation and Mitigation: forest health, wood quality and biomass production, p.24., Riga, Latvia.
4. 08.02.2012-11.02.2012. Voronova A, Jansons Ā, Ruņģis D. Expression of retrotransposon-like sequences in Scots pine (*Pinus sylvestris* L) clones under stress conditions. Molecular Mapping & Marker Assisted Selection, Vīne, Austrija. VIPCA Conference Abstract Book app.1
  5. 04.-07.02.2012. Voronova A, Jansons Ā, Ruņģis D. Expression of retrotransposon-like sequences in Scots pine (*Pinus sylvestris*. L) clones under stress conditions. Molecular Ecology, Vīne, Austrija. VIPCA Conference Abstract Book p.41.
  6. 22.-24.04.2009. Žuka A, Ruņģis D, Schulman AH, Kalendar R. Retrotransposon variability in Scots Pine (*Pinus sylvestris* L.) genome. 5th International Conference Research and Conservation of Biological Diversity in Baltic Region, Daugavpils, Latvia. Book of Abstracts, Daugavpils University Academic Press Saule, p.156.
  7. 19.-21.03.2009. Zhuk A, Rungis D, Kalendar R, Schulman AH. Retrotransposon variability in Scots pine (*Pinus sylvestris* L.) genome. FEBS Workshop, Adaption Potential in Plants, Vīne, Austrija.

#### ***Vietējas konferences un kongresi:***

1. 13.02.2013. Voronova A, Beļēviča V, Jansons Ā, Ruņģis D. Stresa apstākļu inducējamu retrotranspozoniem līdzīgo sekvenču transkripcionālā aktivēšanās priedes genomā. Latvijas Universitātes 71. konference. Molekulārās bioloģijas sekcija, Rīga, Latvija.
2. 01.02.2012. Voronova A, Jansons Ā, Ruņģis D. Mobilēm ģenētiskiem elementiem līdzīgo sekvenču izplatība parastās priedes (*Pinus sylvestris* L.) genomā. Latvijas Universitātes 70. konference. Augu bioloģijas sekcija, Rīga, Latvija.
3. 24.-27.10.2011. Ruņģis D, Šķipars V, Voronova A, Krivmane B, Gaile I, Beļēviča V, Korica A, Kānberga-Siliņa K, Veinberga I. Apvienotais Pasaules latviešu zinātnieku 3. kongress. Molekulārās ģenētikas metodes Latvijas mežzinātnē. Jelgava, Latvija.
4. 04.02.2010. Voronova A, Ruņģis D. Mobilēm ģenētiskiem elementiem līdzīgo sekvenču ekspresija stresoru ietekmē parastās priedes (*Pinus sylvestris* L.) genomā. Latvijas Universitātes 68. konference. Augu bioloģijas sekcija, Rīga, Latvija.
5. 30.11.2011.-1.11.2011. Voronova A, Jansons Ā, Ruņģis D. Genome plasticity of Scots pine (*Pinus sylvestris* L.) under different stress conditions. Genetic variation of adaptive traits, Salaspils, Latvija.
6. 4.02.2009. Žuka A. Retrotranspozonu variācija parastās priedes (*Pinus sylvestris* L.) genomā. Latvijas Universitātes 67. konference. Augu bioloģijas sekcija, Rīga, Latvija.

#### ***Publikācijas:***

1. Voronova A, Belevich V, Jansons A, Rungis D (2013) Stress induced transcriptional activation of retrotransposon-like sequences in the Scots pine (*Pinus sylvestris* L.) genome. (29.05.2013 iesniegts publicēšanai žurnālā *Tree Genetics&Genomes*; apstiprināts ar labojumiem 12.11.2013).



2. Voronova A, Rungis D. (2013) Development and characterisation of IRAP markers from expressed retrotransposon-like sequences in *Pinus sylvestris* L. (Pieņemts publicēšanai).
  3. Voronova A, Jansons Ā, Ruņģis D. (2011) Expression of retrotransposon-like sequences in Scots pine (*Pinus sylvestris*) in response to heat stress. *Environmental and Experimental Biology* 9: 121-127 (Publicēts).
  4. Lielākā atkārtojumu sekvenču datubāzē „*Repbase*” publicētas triju retrotranspozonu sekvenču: *PsAppalachian-I*, *Silava-Pta-I*, *Copia-17-Pta-I* (Publicēts).
  5. Darba rezultātā 126 sekvenču tika publicētas NCBI EST datu bāzē ar ģēnu bankas numuriem JZ389987-JZ390112 (Publicēts).
- 2009.-2012. pētījums ir saistīts ar ESF projektu: „Ģenētisko faktoru nozīme adaptēties spējīgu un pēc koksnes īpašībām kvalitatīvu mežaudžu izveidē” (Nr. 2009/0200/1DP/1.1.1.2.0/09/APIA/VIAA/146).

### ***Darba apjoms un struktūra***

Promocijas darba apjoms ir 202 lpp. Darbā ietvertas 20 tabulas, 56 attēli, 2. pielikumi. Darbs sastāv no ievada, 7 nodaļām (Literatūras apskats, Materiāls un metodes, Rezultāti, Diskusija, Secinājumi, Aizstāvēšanai izvirzītās tēzes, Pateicības). Darba nobeigumā formulēti 9 secinājumi. Darbā izmantoti 446 literatūras avoti.

## **1. PĒTĪJUMA TEORĒTISKAIS PAMATOJUMS**

### ***1.1. Mobilo ģenētisko elementu definīcija un klasifikācija***

Transponējamie elementi (TE) vai arī mobilie ģenētiskie elementi ir sekvenču, kas spēj mainīt savu lokalizāciju genomā (transponēties) vienas šūnas robežās. Pārsvarā mobilie ģenētiskie elementi ir zaudējuši savu spēju transponēties un tikai daži sugas specifiskie ģimeņu pārstāvji ir aktīvi, taču šī aktivitāte izpaužas organismam nelabvēlīgos apstākļos vai arī specifiskos audos un attīstības stadijās (McClintock 1984; Wessler 1996; Grandbastien *et al.* 1997; Kumar, Bennetzen, 1999).

TE klasificē atkarībā no RNS vai DNS molekulu iesaistīšanas transpozīcijā, izšķirot divas mobilo ģenētisko elementu klases: I klase, jeb retrotranspozoni un II klase, jeb DNS transpozoni (Finnegan 1989; Curcio, Derbyshire 2003; Capy 2005; Wicker *et al.* 2007). Retrotranspozoni (RE) pēc savas struktūras un integrēšanās mehānisma ir līdzīgi retrovīrusiem, taču retrotranspozoni nespēj inficēt citas šūnas. Transpozīcijai RE izmanto RNS molekulas starpniecību ar sekojošu reverso transkripciju un komplementārās DNS integrāciju jaunajā genoma vietā. Tādēļ līdz ar RE pārvietošanos citā hromosomas pozīcijā genomā notiek arī to kopiju skaita palielināšanās. RE iedala kārtās, virsgimenēs un ģimenēs (Wicker *et al.* 2007). RE, kuri satur tiešos garos terminālos atkārtojumus, ietilpst LTR RE kārtā, elementi bez šī tipa atkārtojumiem tiek iedalīti autonomās LINE vai neautonomās SINE kārtās. Pašlaik RE tiek klasificēti arī *Pseudoviridae* (*Ty1*, *Copia* tipa) un *Metaviridae* (*Ty3*, *Gypsy* tipa) vīrusu dzimtās, kas nav pieskaitāmi nevienai no vīrusu kārtām (Boeke *et al.* 2006).

### ***1.2. LTR RE struktūra***

LTR ir proteīnus nekodējoša sekvenču, kas satur transkripciju regulējošas sekvenču. LTR garums var sastādīt no 85 bp *F<sub>R</sub>Retro129* rīsiem (Gao *et al.* 2012), līdz pat 5 kb *Ogre* zirņu genomā (Neumann *et al.* 2003). LTR ir specifisks katrai RE ģimenei. Izšķiramas šādas konservatīvas LTR sekvenču: aptuveni 7 bp gari invertētie

atkārtojumi (IR) ar īpaši konservatīviem 5' TG un 3' CA nukleotīdiem (int piesaistes vieta), TATATA motīvs, kurš pilda polimerāzes II promotera funkciju augiem (Vicent *et al.* 2005; Benachenhou *et al.* 2013). Bieži LTR satur arī poliadenilēšanas signālu. U3 rajons vairāk variē dažādās RE grupās, kas ir skaidrojams ar to, ka šajā rajonā atrodas specifiskas transkripciju regulējošas sekvenču, kas nodrošina ekspresiju noteiktos audos / apstākļos / attīstības stadijā. 5'LTR uzsākas RE transkripcija, bet beidzas 3'LTR atkārtotumā (Kumar, Bennetzen 1999). Tā kā 5' un 3'LTR ir vienādi, tad 3'LTR var iniciēt blakus novietoto sekvenču transkripciju. Vīrusiem 3' LTR promotera aktivitāte parasti ir nomākta, bet var tikt aktivēta 5'LTR regulatīvo sekvenču izjaukšanas gadījumā (Boerkoel, Kung 1992; Gama Sosa *et al.* 1994; Klaver, Berkhout 1994). Autonomie elementi parasti satur divus gēnus - *gag* un *pol*; *gag* gēns kodē kapsīdamlīdzīgus proteīnus, bet *pol* gēns kodē poliproteīnu, kurš turpmāk tiek sašķelts ar paša kodētu proteāzi (pr), veidojot reverso transkriptāzi (RT) ar RNāzes H aktivitāti un integrāzi (int). Env-līdzīgs domēns retāk ir atrodams RE (Eickbush, Malik 2002). RE var saturēt arī papildus ORF ar nezināmu funkciju, piemēram, rīsu *RIRE2* (Kumekawa *et al.* 1999). RE var iegūt gēna daļas, piemēram, kukurūzas *Bs1* satur ATPāzes gēna daļu (Jin, Bennetzen 1994). Citoplazmā mRNS transkripti tiek izmantoti gan par matricu proteīnu veidošanai, gan reversajai transkripcijai dubultspirāles cDNS sintēzei. Reversās transkripcijas iniciācijai tiek izmantotas dažāda tipa tRNS, kura piesaistās PBS rajonam (Mak, Kleiman 1997). PBS vairāku RE ģimeņu vidū ir ļoti konservatīvs (Kalendar *et al.* 2010).

### **1.3. TE izplatība augu genomos**

Augu genomi satur TE elementus no abām klasēm ar LTR RE virsgimenes pārstāvju ievērojamu pārsvaru (Kumar, Bennetzen 1999). Augu genomiem RE sekvenču un to izplatība variē un var sastādīt no 15% (*Arabidopsis thaliana*) līdz pat 90% (*Liliaceae*) no visa genoma lieluma (Sabot, Schulman 2006; Morgante *et al.* 2007). Kompleksie augu genomi raksturojas ar t.s. kodējošo gēnu salām atkārtotumu jūrā (SanMiguel *et al.* 1996; Panstruga *et al.* 1998; Shirasu *et al.* 2000; Rostoks *et al.* 2002; Kejnovsky *et al.* 2012). Salīdzinoša vairāku graudaugu genomu sekvenēšana atklāja augstu variācijas pakāpi starpgēnu un gēnus kodējošajos rajonos. Tāpēc viena individuālā genoma pabeigta sekvenču nevar atspoguļot visas pastāvošās variācijas populācijā. Variējošajā daļā ietilpst gan TE, gan pseidogēni, gan arī individuālie gēni. Pēdējo gadu pētījumi rāda, ka līdzīga genoma organizācija ir raksturīga visiem augstākajiem augiem (Brunner *et al.* 2005; Piegu *et al.* 2006; Hawkins *et al.* 2006; Neumann *et al.* 2006). Turklāt augiem bieži RE rada atšķirīgu vidi pirms konservatīvajiem gēnu blokiem (Brunner *et al.* 2005; Morgante *et al.* 2007).

### **1.4. RE klāsterizācija un daudzveidība**

LTR TE ir bieži sastopami neparastas konstrukcijas kompleksajos augu genomos. Labi izpētītā miežu RE *BARE-1* gadījumā viens LTR miežu genomā ir sastopams 7 līdz 42 reizes biežāk salīdzinot ar pilna garuma elementiem un sastāda  $13.7 \times 10^3$  kopijas haploīdā genomā (Vicent *et al.* 1999). Struktūra ar diviem iekšējiem domēniem, flankētu ar vienu rekombinantu LTR, tika atrasta miežu haploīdā genomā  $4.6 \times 10^3$  kopijās (Vicent *et al.* 2005). Matricas nomaiņa ar RT paver iespējas jauno rekombinanto elementu veidošanai (Sabot, Schulman 2007). RE transpozīcijas mehānismā ir iesaistīta RT, kuras precizitāte, salīdzinot ar citām šūnas polimerāzēm, ir zemāka (Preston 1996; Gabriel, Mules 1999). Lielākas mutācijas un pārkārtojumi LTR

RE rodas nevienādās homologās rekombinācijas procesā, pateicoties to kopiju daudzumam un šo sekvenču homologijai (Vicent *et al.* 2005; Hawkins *et al.* 2009). LTR-LTR nevienādā homologā rekombinācija var veidot genomā vienu LTR bez elementa, vai arī vairākus LTR, vai arī vairākas iekšējo domēnu sekvenču pēc kārtas (Mieczkowski *et al.* 2006). Turklāt, bieži vien RE preferenciāli ievietojas atkārtotumu rajonos izjaucot iepriekš ievietotos RE. Rezultātā veidojas genoma rajons, kas sastāv no sadrumstaloto RE sekvenču daļām un neaktīvo elementu klāsteriem (Kumar, Bennetzen 1999).

### **1.5. Genoma izmēra saistība ar TE proliferāciju**

TE uzkrāšanās dažādu augu genomos varētu atspoguļot preferenciālu atlasītiem lielākiem genomiem (Murray 1998, 2005). Lielu augu genomi satur daudz LTR RE ģimenes, kas var sasniegt vairāk kā 10 000 kopiju haploīdā genomā. Lai gan galvenās TE grupas ir sastopamas visu augu genomos, relatīvais TE sastāvs un TE klašu un kārtu proporcijas ievērojami variē (Kejnovsky *et al.* 2012). Izplatītākās RE ģimenes *BARE-1* kopiju daudzums korelē ar genoma lielumu miežu genomā (Vicent *et al.* 1999). Ir parādīta arī *BARE-1* ekspansija savvaļas miežiem atkarībā no to augšanas vietas mitruma pakāpes (Kalendar *et al.* 2000). Datu analīze parādīja arī genoma lieluma saistību ar ģeogrāfisko stāvokli (Kalendar *et al.* 2000). Iespējams, pastāvīgā stresa rezultātā TE delēciju skaits ir zemāks, kas noved pie pilna garuma elementu uzturēšanas un genoma palielinājuma (Nevo 2001). *BARE-1* kopiju skaits, iespējams, atspoguļo adaptīvu atlasītiem lielākiem genomiem ar augstāku RE aktivitāti (Kalendar *et al.* 2000; Nevo 2001; Wessler 2001). Turklāt, RE transpozīcija var radīt jaunas funkcionālas mutācijas un līdz ar to arī jaunus pielāgojumus stresa apstākļu pārvarēšanai (Madlung, Comai 2004).

### **1.6. RE aktivēšanās stresa apstākļos**

RE aktivēšanās rezultātā izsauktais mutāģenēzes process varētu dot priekšroku augu genomiem, sniedzot iespēju nelabvēlīgos vides apstākļos ātri palielināt genotipu variāciju (Wessler 1996; Murray 1998; Murray 2005). Turklāt, eksperimentāli pierādīts, ka RE augu genomā var radīt jaunus stresa atbildē iesaistītus gēnu tīklus (Ito *et al.* 2011). Arī rauga genomā *Tfl* elements, pēc karstuma šoka tiek aktivēts un ar savu integrāciju ietekmē stresa atbildē jau iesaistīto gēnu, padarot tā transkripciju vēl efektīvāku (Feng *et al.* 2012). Tabakas *Tnt1* RE ģimenes pētījums apliecina, ka atšķirības TE 3'UTR sekvenču nosaka elementu transkripcionālu aktivēšanos noteikto stresa asociēto signālmolekulu ietekmē (Beguiristain *et al.* 2001). RE *Osen* tiek ekspresēts karstuma stresam pakļauto *Arabidopsis* augu pēcnācējiem (Ito *et al.* 2011). Savukārt, RE *Tam3* no lielās lauvmutītes (*Antirrhinum majus*) genoma (Hashida *et al.* 2003) un *Tcs1* no apelsīnu genoma (Butelli *et al.* 2012) ekspresija pieaug pēc zemo temperatūru iedarbības. *Ty1 Copia TLC1* ģimene no tomātu genoma tiek ekspresēta pēc ievainojuma, protoplastu izveides, augstas sāļu koncentrācijas; bet motīvi U3 rajonā, padara *TLC1.1* elementus inducējamus ar etilēnu (Tapia *et al.* 2005). Melones *Copia RE Reme1* tika atrasts UV starojumam pakļautajās lapās (Ramallo *et al.* 2008). LINE bez-LTR TE *Karma* ekspresējas reģenerēto augu audos un kultivētajās šūnās (Komatsu *et al.* 2003). Arī biotiskie stresori iniciē virkni RE ekspresiju: *Oare-1*, vīrusu infekcijas kartupeļos; *Tto1*, tabakas mozaīkas vīruss (Hirochika 1993); *Tnt1*, tabakas sēnīšu infekcijas (Melayah *et al.* 2001). Korķa ozola (*Quercus suber*) *Corky* ir ekspresēts dažādās attīstības stadijās, kā arī pēc ievainošanas (Rocheta *et al.* 2012). Zirņu *Psr*

ģimenes RE aktivējas atbildot uz protoplastu izveidi un uz apstrādi ar sēnīšu elicitoriem (Kato *et al.* 1999). Šūnapvalka hidrolāžu iespaidā aktivējas tabakas *Tnt1* ekspresija (Pouteau *et al.* 1991). RE ekspresija bieži vien ir arī audu specifiska, piemēram, sakņu audiem specifiska ekspresija ir *Tnt1* (Grandbastien *et al.* 1994) un *TLC1* (Tapia *et al.* 2005); *BARE-1* ekspresējas lapās (Suoniemi *et al.* 1996), bet kviešu *PREM-2* ekspresējas mikrosporās (Turcich *et al.* 1996).

### 1.7. RE regulatīvā nozīme

RE insercija gēnu rajonā var ietekmēt gēnu transkripciju vairākos veidos (Rebollo *et al.* 2012). RE LTR var saturēt promoterus, kas darbojas abos virzienos, transkripcijas faktorus, piesaistes saitus, terminācijas saitus, insulatorus, splaisingu regulējošas sekvenču. RE transpozīcijas augu genomos veido daudz mutāciju. RE *Bs1* kukurūzas genomā pēc vīrusa infekcijas sākotnēji tika atrasts insertēts *Adh1* alkohola dehidrogenāzes gēnā (Johns *et al.* 1985). Līdzīgi, *Tnt1* tika identificēts tabakas nitrāta reduktāzes gēnā (Grandbastien *et al.* 1989), *Tos17* insertēts rīsu *PHYA* (fitohroma A) gēnā (Takano *et al.* 2001). Kukurūzas *COMT* gēnā (O-metiltransferāze), kurš ir svarīgs lignīna biosintēzes dalībnieks, *B5*-līdzīga RE insercija producē īsāku himērisku cDNS (Vignols *et al.* 1995). Kukurūzas *waxy* gēna alternatīvās alēles dažādos intronos satur RE insercijas, kur ar alternatīvā splaisinga mehānismu tiek veidoti dažādi mRNS transkripti (Varagona *et al.* 1992). Vairāku gēnu promoteri un alternatīvie promoteri ir cēlušies no RE (Rebollo *et al.* 2012). Pat viens LTR var regulēt blakusesošo gēnu transkripciju, kā bija parādīts *Myb* transkripcijas faktora gēna gadījumā, pirms kura notika *Copia* LTR RE *Tsc1* un *Tsc2* insercija apelsīnu genomā (Butelli *et al.* 2012). Tāpat, RE insercija var inhibēt gēna transkripciju, piemēram, gadījumā ar *Gypsy* elementu *Gret1* vīnogu (*Vitis vinifera*) genomā (Kobayashi *et al.* 2004). Zirņu *PSPAL2* (fenilalanīna amonija liāzes gēns 2) un *PSCHS1* (1,3-difenil-2-propēn-1-ona sintēzes gēns) ekspresējas atbildot uz patogēnu invāziju, un aktivē auga aizsardzības mehānismus. Abi gēni savos 5' reģionos satur vienas apakšģimenes *Copia* RE *PsrC* insercijas (Kato *et al.* 1999). Amfiploīdos kviešos atrastie atšķirīgie himēriskie transkripti satur *Wis 2-1A* LTR RE, visbiežāk lokalizētu gēnu 3'UTR sekvencē. Transkriptiem ar palielinātu ekspresiju amfiploīdos, *Wis 2-1A* LTR bija novietots vienādā orientācijā ar blakusesošo gēnu. Autori norāda, ka no *Wis* elementa var būt atkarīgi aptuveni 208 gēni kviešu genomā (Kashkush *et al.* 2003).

RE var piedalīties regulatīvajos procesos netieši producējot garās nekodējošās RNS molekulas ar pagaidām nezināmām funkcijām, vai arī tās var tikt šķeltas ar RNS interferences mehānisma starpniecību, veidojot mikroRNS molekulas, kas piedalās RNS interferencē. Piemēram, cilvēka genomā atrasti 55 mikroRNS gēni, kas ir cēlušies no RE (Piriyaopongsa *et al.* 2007). Pretēji agrākiem uzskatiem, ka transkribēti tiek tikai gēni, līdz ar sekvenēšanas tehnoloģiju attīstību atklāts, ka lielākā daļa no genoma, kā, piemēram, zīdītāju genoma, var tikt transkribēta, veidojot nekodējošās RNS molekulas ar nezināmu funkciju (Mattick, Makunin 2006; Zhang *et al.* 2012; Ponting *et al.* 2009). Šī transkripcija nav tik intensīva, kā gēniem, tomēr uzskata, ka šīm molekulām varētu būt kompleksa regulatīva ietekme uz kvantitatīvo pazīmju variāciju, attīstības atšķirībām un kompleksām slimībām. Transkriptomu analīžu veikšanai RE transkriptus parasti neiekļauj, jo tie satur atkārtojumus un ir pārstāvēti daudzās kopijās, līdz ar to, tie ir grūti analizējami (Mourier, Willerslev 2009).

LTR atkārtojumu klāsteri piedalās heterohromatīna veidošanā, kas ir ļoti dinamisks un svarīgs process dažādās auga attīstības stadijās (Finnegan *et al.* 1998). Vairāku augu hromosomu centromēras satur centromēriskus satelītu atkārtojumus pamīšus ar specifiskiem *Ty3 Gypsy* RE, kurus sauc par centromēru RE (CR). Zināms, ka augos CR un centromēru satelīti tiek transkribēti zemā līmenī (Topp *et al.* 2004; Neumann *et al.* 2007). Iespējams, ka CR kalpo par promoteru avotu satelītu transkripcijai, kas, savukārt, ir nepieciešami centromēras atpazīšanai un/vai hromosomu segregācijai (May *et al.* 2005).

RT, kuras vienīgais avots genomā ir RE un telomerāzes, aktivitātes rezultātā var veidoties pseidogēni vai transkribēti retrogēni, kuri ir plaši sastopami kompleksajos augu genomos (Zhang *et al.* 2005, Wang *et al.* 2006). Šādu retrogēnu struktūra parasti ir himēriska, tā satur ar citu regulatoro sekvenci kombinētu gēna sekvenci (Betran *et al.* 2002; Wang *et al.* 2002), vai var būt kombinēta struktūra ar citiem eksoniem, ja insercija ir notikusi cita gēna sekvencē (Sayah *et al.* 2004; Zhang *et al.* 2004). Šis mehānisms ir nozīmīgs organismu adaptācijā un evolūcijā, jo ir atrasti funkcionāli gēni, kuri ir radušies pateicoties pseidogēnu diferenciacijai (Betran *et al.* 2002; Emerson *et al.* 2004).

### **1.8. TE raksturojums skujkoku genomos**

RE sekvenču pētījumi kailsēkļiem aizsākās pirms to genomu sekvenēšanas. *Ty-Copia TPE1* RE tika izolēts no *P.elliottii* (Kamm *et al.* 1996). *TPE1* transpozīcijas iespējas tika novērtētas kā ierobežotas, jo sekvenēšana atklāja mutācijas proteīnu kodējošajās sekvencēs (Brandes *et al.* 1997). No *P.radiata* izolēja un raksturoja skujkokos izplatīto *Gypsy IFG* RE, kuram starp gag un AP domēniem tika atrasts viens stop-kodons (Kossack, Kinlaw 1999). *Spcl* no *P.glauca* genoma satur tikai vienu stop-kodonu pirms int domēna, kas liecina par šī elementa neseno izcelsmi (L'Homme *et al.* 2000). Mazkopiju *Gypsy PpRT1* elements tika izolēts no *P.pinaster* genoma, tā sekvenca bija homologa iepriekš aprakstītajam *IFG* (Rocheta *et al.* 2007). *PpRT1* transkripts tika atrasts EST datubazē, kas norāda uz tā ekspresiju (Miguel *et al.* 2008).

*P.taeda* BAC sekvenēšana atklāja augstu TE sekvenču saturu un priedei specifiskas TE ģimenes (Morse *et al.* 2009). *Gymny* RE ir izolēts no *P.taeda* un atrasts arī citās *Pinus* sugās, bet netika atrasts *Picea*, *Abies*, *Tsuga* un *Ginkgo*, kas norāda uz *Gymny* ekspansiju pēc šo taksonu nodalīšanās (Morse *et al.* 2009). Līdzīgs pētījums *P.taeda* genomā atklāja gan I, gan II klases TE visu kārtu pārstāvjus, bet zinātnieki secināja, ka vairums elementu varētu tikt klasificēti jaunās ģimenēs (Kovach *et al.* 2010). TE līdzība nukleotīdu līmenī starp segsēkļu un kailsēkļu sugām atrasta no 37,7% līdz 99,4%. Hibridizācija ar dažādiem elementiem neuzrādīja krāsas atšķirības genoma organizācijā starp attāliem taksoniem, piemēram *P.abies* un *P.omorika*, stipra hibridizācija ar TE kloniem no *P.abies* tika novērota *Ginkgo biloba* DNS un otrādi (Friesen *et al.* 2001). *P.abies* genomā tika atrastas 1773 atkārtojumu sekvences, puse no tām līdzinās zināmām TE ģimenēm. LTR RE ir visizplatītākā TE kārtā, bet *Gypsy* virsģimene ir prevalējoša (Nystedt *et al.* 2013). *P.abies* genoma analīze neatrada nesen notikuša transpozīciju sprādziena pēdas. Viena LTR attiecība pret pilna izmēra RE ir 1:9, kas liecina par zemu TE izgriešanas intensitāti salīdzinot ar segsēkļu genomiem.

## 2. MATERIĀLS UN METODES

### 2.1. *Vispārējā eksperimentu struktūra*

Darbā aprakstītais pētījums sastāvēja no vairākiem posmiem. Lai noskaidrotu aktīvo elementu transkriptu esamību stresam pakļautajos priežu stādos, izmantoja nespecifiskus iPBS praimerus, jo kailsēkļu retrotranspozonu sekvences var atšķirties no publiski pieejamajām segsēkļu sekvencēm un neaktīvajiem retrotranspozonu homoloģiem, kas līdz šim ir aprakstīti kailsēkļiem. Lai izvairītos no genotipa izraisīto atšķirību identificēšanas, katrā eksperimentā izmantoja viena priežu klona rametus. TE reprezentāciju un ekspresiju pārbaudīja ar izveidotiem specifiskiem praimeriem. Pārbaudīja karstuma, priežu hermesa invāzijas, ABA (abscīzskābes) un SA (salicilskābes) apstrādes ietekmi, pēc sekojošas vienotas shēmas.

- ▶ Priedes rametus pakļauj pētāmā stresora ietekmei.
- ▶ Ievāc pa 4 skujām no katra stāda, ieliek iezīmētajos 2 ml stobriņos un nekavējoties sasaldē šķidrajā slāpekļī. Šos paraugus izmanto RNS izdalīšanai vēlāk.
- ▶ Pēc eksperimenta beigām, ievāc vēl pa 2 skujām. Šos paraugus izmanto DNS izolēšanai, ģenētiskās vienveidības pārbaudei ar SSR marķieriem u.c.
- ▶ Veic DNS izdalīšanu, ģenētiskās vienveidības pārbaudi ar SSR marķieriem.
- ▶ Veic RNS izdalīšanu, RNS attīrīšanu no DNS piemaisījumiem un kontaminācijas pārbaudi ar RNS PĶR.
- ▶ cDNS sintēze un analīze ar kontroles marķieriem.
- ▶ iPBS reakcija, fragmentu analīze un atšķirīgo fragmentu izolēšana un attīrīšana.
- ▶ Fragmentu klonēšana vektorā, šūnu transformācija.
- ▶ Transformēto šūnu pavairošana, atlase, vektoru izdalīšana, ieklonēto fragmentu garuma pārbaude, plazmīdu atlase sekvenēšanai.
- ▶ Sekvenēšana, hromatogrammu analīze un sekvenču salikšana.
- ▶ Sekvenču līdzības meklējumi datubāzēs, retrotranspozonu un to strukturālo elementu identificēšana, datu apkopošana.

Pēc identificēto TE-līdzīgo sekvenču analīzēm radās jautājums, vai kāds no identificētajiem TE-līdzīgiem fragmentiem atspoguļo pilna izmēra retrotranspozonu transkripciju stresoru ietekmē, vai arī visi šie fragmenti ir himēru transkriptu sastāvdaļas. Tādēļ nākamais pētījuma posms veltīts pilna izmēra retrotranspozonu izolēšanai, raksturošanai un lielāku sekvenču daļu iegūšanai no eksperimentālā cDNS.

Trešais pētījuma posms veltīts praktiskai fragmentu polimorfisma un izplatības noteikšanai. Lai saprastu, cik lielā mērā TE-līdzīgās sekvences ir sastopamas citu genotipu parastās priedes genomos, izveidoja specifiskus praimerus, kurus pārbaudīja dabiskās priežu audzēs ar IRAP metodi (Kalendar *et al.* 1999). Šos marķierus izmēģināja arī klonālās daudzveidības noteikšanai, par materiālu izvēloties atšķirīgās plantācijās augošus viena klona parastās priedes rametus. Lai saprastu, cik lielā mērā atrastie fragmenti ir konservatīvi un izplatīti citu kailsēkļu sugās, veica specifisku amplifikāciju ar dažādu kailsēkļu dzimtu sugu pārstāvjiem.

## 2.2. Stresa apstākļu inducēšana

32 parastās priedes divgadīgos stādus, potētus viena kлона rametus, nodrošināja AS „Latvijas valsts meži” Kalsnavas kokaudzētava (Jaunkalsnava, Latvija).

**Karstuma stresa iniciācija.** Karstuma ietekmes eksperimentu veica izmantojot klimata kameras ar gaisa, temperatūras un mitruma kontroli HOTCOLD-GL (*SELECTA*). Divus JA003 rametus pakļāva nepārtrauktam karstuma stresam (40 °C 24 stundas), ar 10 stundu fotoperiodu. Divus JA003 rametus ievietoja otrā klimata kamerā ar identisku fotoperiodu ar temperatūru 22 °C dienā un 14 °C naktī.

**Priežu hermesa invāzija.** 16 TU-28 rametus izmantoja priežu hermesa (*Pineus pini*) invāzijas pētījumā. Invāziju konstatēja pēc vairākām pazīmēm (sveķošana, balti vaska plankumi, dzeltējošas skujas un zaru kalšana). Tāpat ārējo barošanas apstiprināja pēc baltā tinuma mikroskopēšanas.

**Augu hormonu ietekme.** Izmantoja 1mM ABA (*Sigma*) šķīdumu dejonizētā destilētā ūdenī ar 1,92% etilspirta un eksperimenta laikā to uzglabāja 4 °C. No 1mM ABA katru dienu pirms apsmidzināšanas pagatavoja 10 μM ABA darba šķīdumu, ar kuru apstrādāja POP12 rametu stādus. Ūdens šķīdumu ar ekvivalentu etanola daudzumu pagatavoja kontroles rameta apstrādei. POP21 rametus izmantoja SA ietekmes pētījumā, kur divus rametus vienu nedēļu apsmidzināja un aplaistīja ar 5 mM SA (*Sigma*) ūdens šķīdumu. Kontroles rametu apsmidzināja un aplaistīja ar ūdeni.

## 2.3. Diferenciāli ekspresēto fragmentu identificēšana

RNS izdalīja ar CTAB-metodi (Chang *et al.* 1993) ar modifikācijām. DNS kontamināciju novērsa, vairākas reizes attīrot ar *Turbo DNA-free* kitu (*Ambion*) ar sekojošu fenola:hloroforma:izoamilsipirta (24:24:1) ekstrakciju, līdz standarta PĶR amplifikācija neuzrādīja produktu esamību. RNS koncentrāciju mērīja ar *Quibit* fluorimetru (*Invitrogen*). Stresa iniciāciju pārbaudīja ar kontroles marķieru amplifikāciju: *dhn3* (dehidrīna III), *abaH* (abscīzskābes inducējamais), *lea* (vēlā embriogēnēzē dominējošs), *P5CS2* (delta 1-pirolīna-5-karboksilāta sintetāze 2). RNS paraugu kvalitāti novērtēja pēc endogēnās kontroles gēna GAPDH (glicerāldehīda-3-fosfāta dehidrogenāze) amplifikācijas. Reverso transkripciju veica ar *TaqMan* kitu (*Applied Biosystems*) ar 1 μg RNS un heksamēru praimeriem. Tipiska PĶR reakcija ar cDNS veikta 20 μl kopējā reakcijas tilpumā ar sekojošu sastāvu: 1x Taq buferis (*Fermentas*), 1,5-2 mM MgCl<sub>2</sub>, 0,2 mM dNTP mix, 0,1 μM no katra praimesa, 0,7 U Taq polimerāze (*Fermentas*). Programma sastāvēja no denaturēšanas posma 94 °C 3 min, tai sekoja 38 cikli ar denaturāciju 94 °C 30 sek, praimesu pielipšanu 50-62 °C 30 sek, sintēzi 72 °C 40 sek, un beigu elongācijas posmu 72 °C 10 min. Produktus vizualizēja ar elektroforēzes metodi 1xTAE buferī, ar etīdija bromīda šķīdumu (*Merck*) iekrāsotā 1-1,5% agarozes gēlā, 1,5-2 stundas 70-94 V. Visus elektroforēzes attēlus vizualizēja izmantojot UV-staru transiluminatoru ar fotosistēmu *AlphaDigiDocRT* (*JH BIO Innovations*) un fotoaparātu *Olympus sp-500uz*. Attēlus apstrādāja ar *AlphaEaseFC v. 4.0.0* programmu. RE-līdzīgo sekvenču identificēšanai izmantoja iPBS metodi (Kalendar *et al.* 2010). Diferenciāli ekspresētus fragmentus noteica, salīdzinot amplifikāciju starp stresam pakļauto klonu paraugiem un kontroles klonu paraugiem. Izolētus fragmentus attīrīja no agarozes gēla ar *Qiaquick Gel Extraction Kit* (*Qiagene*). Fragmentus pāramplificēja ar gradienta PĶR (60 °C uz 47 °C). PĶR produktus attīrīja ar *Sefadex-G50* (*Sigma*) kolonām izmantojot *MultiScreen* filtru plates (*Millipore*) un *Multiscreen Assay system* (*Millipore*). Produktus liģēja ar plazmīdu pTZ57R/T

izmantojot *InsTAclone PCR Cloning Kit (Fermentas)*. Kompetentās *E.coli* šūnas sagatavoja pēc Inoue *et al.* (1990). Plazmīdas izdalīja ar alkalīna lizēšanas metodi (Birnboim, Doly 1979) ar modifikācijām. Insertu garuma atbilstību pārbaudīja pēc šķelšanas ar *EcoRI* un *SaII* endonukleāzām (*Thermo Scientific*). Sekvenēšanai izmantoja M13 praimerus un *BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems)*.

Sekvences analīzei izmantoja *megablast*, *blastn* un *blastx* rīkus (Altschul *et al.* 1990) meklēšanai NCBI sekvenču datubāzēs (<http://www.ncbi.nlm.nih.gov>); kā arī TE datubāzes: *GrainGenes TREP* (<http://wheat.pw.usda.gov/>), *GypsyDatabase 2.0* (<http://gydb.org/>, Llorens *et al.* 2011), *Repbase* (<http://www.girinst.org/>, Kohany *et al.* 2006). Sekvenču salīdzināšanu veica ar CLUSTALW v.1.83 rīku (Thompson *et al.* 1994). Dendrogrammas vizualizēja ar programmu MEGA v.5 (Tamura *et al.* 2011). EST sakrītību skaitu aprēķināja izmantojot NCBI *blastn* zaļo augu (*Viridiplantae*) EST meklēšanu, ņēma vērā ierakstus ar e-vērtību  $\leq 0,01$ . Sekvences publicētas NCBI EST datu bāzē ar gēnu bankas numuriem JZ389987-JZ390112.

#### **2.4. TE-līdzīgo fragmentu ekspresijas kvantitēšana un salīdzināšana**

Izmantojot Primer-BLAST programmu (Ye *et al.* 2012) izveidoja 27 specifiskus praimeru pārus. Veica salīdzinošo PĶR ar cDNS, kā arī ekspresijas līmeņa noteikšanu ar reālā laika PĶR. Reālā laika PĶR reakcijas veica ar *Maxima SYBR Green/ROX qPCR Master Mix (Thermo Scientific)* un *StepOnePlus™ Real-Time PCR System (Applied Biosystems)* un amplikonu uzkrāšanas dinamiku analizēja ar *StepOne v.2.2.2 (Applied Biosystems)*. Analīzei izmantoja relatīvās kvantitēšanas  $2^{-\Delta\Delta Ct}$  metodi, normalizējot paraugus pret GAPDH amplifikācijas rādītājiem (Schmittgen, Livak 2008). PĶR efektivitāti katram praimerim un arī endogēnajai kontrolei salīdzināja izmantojot standartlīkņu analīzi.

#### **2.5. Pilna izmēra retrotranspozonu izolēšana, marķieru konstruēšana**

Pilna izmēra retrotranspozonu izolēšanai izmantoja identificēto homologo retrotranspozonu daļu izolēšanu, garās distances PĶR izmantojot identificētās sekvences, citu priēžu sugu BAC sekvenču analīzi ar sekojošu retrotranspozonu daļu izolēšanu no *P.sylvestris*. Uzdevuma īstenošanai ar Primer-BLAST programmu (Ye *et al.* 2012) kopumā izveidoja ap 60 dažādām retrotranspozonu daļām komplementāru praimeru. Izveidoti priedes RE specifiskie praimeru variējošo sekvenču amplifikācijai ar IRAP (pēc Kalendar *et al.* 1999). IRAP reakcijai ar priedes gDNS izmantoja *Dream Taq* reaģentus (*Fermentas*). Reakcijas maisījums saturēja gDNS līdz 100 ng uz reakciju; 1x buferis; 0,5 mM MgCl<sub>2</sub>; 1 μM praimeris; 0,2 mM dNTP mix; 0,7 U *dreamTaq* polimerāze (*Fermentas*); 0,2 U *Pfu* polimerāze (*Fermentas*). PĶR programma: denaturācija 94 °C 3 min; 38 × (94 °C - 40 sek; 50 °C - 40 sek; 68 °C - 1 min); 72 °C - 10 min.

#### **2.6. Priedes klonālās un dabīgās daudzveidības analīzes**

Atšķirīgos apstākļos augošas priēžu populācijas analīzei genotipēja 150 priēžu indivīdus. SSR un IRAP datus analizēja izmantojot programmas *GenalEx v. 6* (Peakall, Smouse 2006), *FAMD v.1.30* (Schlüter, Harris 2006). Biometrisko mērījumu un ģenētisko analīžu datu kopu atšķirību novērtēšanai izmantoja vidējo vērtību salīdzināšanu, to standartnovirzes noteikšanu, rezultātu ticamības koeficientu un F-kritērija salīdzināšanu.



Klonālās variācijas identificēšanā izmantoja dažādās klimatiskajās zonās ierīkotajās plantācijās „Sāviens” un „Dravas” augošus 14 parastās priedes individuus, kas reprezentē četru klonu rametus. Paraugus analizēja ar 9 praimeriem pēc IRAP metodes principa.

60 Latvijas florai eksotisko un reti sastopamo kailsēkļu sugu paraugi iegūti no Latvijas Universitātes Botāniskā dārza un Nacionālā botāniskā dārza. Analizēja diferenciāli ekspresēto RE-līdzīgo fragmentu izplatību, kā arī dažu priedei specifisko RE izplatību. Veiktas arī IRAP reakcijas ar 6 marķieriem un datu analīzes, kā aprakstīts iepriekš.

### 3. GALVENIE REZULTĀTI

#### 3.1. Diferenciāli ekspresēto fragmentu analīze

Visu eksperimentu rezultātā identificēja 126 fragmentus, kas reprezentē katra stresora izsaukto retrotranspozoniem-līdzīgo sekvenču ekspresiju. Kopumā izolēja 112 diferenciāli ekspresētus fragmentus, sekvenēja 335 plazmīdu insertus. 40 sekvenču neuzrādīja līdzību ne ar vienu ierakstu izmantotajās datubāzēs. 28 identificētie fragmenti saturēja augsti homologus TE rajonus, vai līdzinājās TE visā fragmenta garumā. Pārējie 98 fragmenti neuzrādīja līdzību TE sekvencēm, vai arī līdzība bija zem pieļaujamā sliekšņa. ABA un SA pētījumos identificētais TE homologu skaits ir ievērojami zemāks. Karstuma un hormonu apstrādes rezultātā nav identificēta neviena līdzīga sekvenču. Priežu hermesa invāzijas rezultātā un karstuma stresa rezultātā identificēja *Copia-2 TA-I* līdzīgās retrotranspozona sekvenču (B104 un T30). Endogēnam retrovirusam (ERV) līdzīga sekvenču atrasta gan kukaiņu invāzijas rezultātā, gan pēc apstrādes ar SA, tāpat šajās grupās bija līdzīgi divi neanotēti fragmenti (B103 un SA24). Divus līdzīgus anonīmus transkriptus, kā arī vienu *Ogre*-līdzīgu transkriptu (ABA18 un B125) atrada gan pēc priežu hermesa invāzijas, gan pēc apstrādes ar ABA. Dažādos eksperimentos identificētie līdzīgie fragmenti bija līdzīgi arī pēc garuma, bet nukleotīdu sekvenču līdzība bija lielāka par 90%.

#### 3.2. TE-līdzīgo transkriptu analīze

No cDNS izolētās 28 TE-līdzīgās sekvenču pēc savas struktūras bija daudzveidīgas. Identificētie fragmenti saturēja: sekvenču, pilnīgi vai daļēji homologas TE ORF, poliproteīna rajonam vai UTR; viena elementa LTR ar fragmentu pārklājumu tuvu 100%; sekvenču ar augsti homologu LTR rajonu kombinācijā ar nezināmu sekvenču; divus LTR, kas abos galos orientēti pretējos virzienos. Lielākā TE-līdzīgo sekvenču grupa (17) saturēja tieši kodējošo domēnu rajonus. Astoņas sekvenču saturēja LTR un trīs TE nekodējošās daļas. Identificēja divas himēriskās sekvenču, kur viena daļa fragmenta līdzinājās TE, bet otra daļa līdzinājās hloroplasta sekvenču.

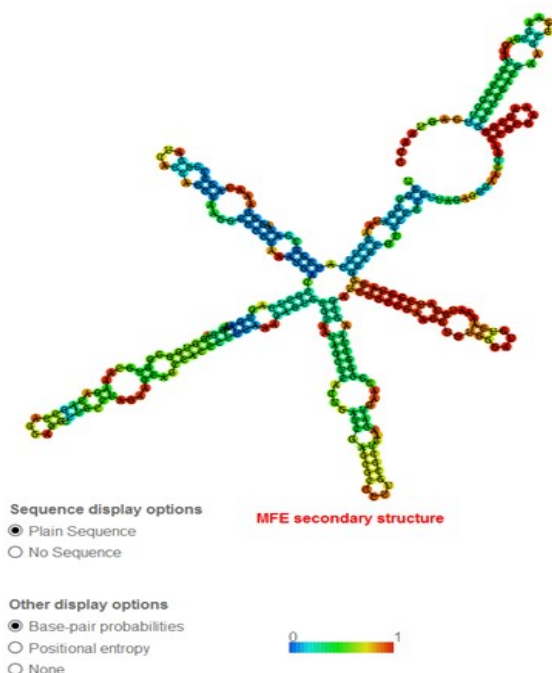
Dažādu stresoru ietekmē atrasti četri *Gypsy* virsgimenes LTR (*Gypsy-6*, *Ogre*, *Appalachian*, *PIGY-1*) un divi *Copia* virsgimenes LTR (*Copia-2-TA\_I*, *HORPIA*) transkripti, kas ir pilnā garumā nukleotīdu līmenī homologu RE sekvenču. Karstuma ietekmes rezultātā identificēja homologās retrotranspozonu pārstāvju sekvenču no dažādām kārtām (bez LTR LINE un LTR) un no deviņām LTR RE ģimenēm. Papildus identificēja dažus DNS transpozoniem *EnSPM* (T49) homologus transkriptus ar augstu līdzību TE proteāzes domēnam nukleotīdu sekvenču līmenī.

Pēc priežu hermesa invāzijas eksperimenta izolēja salīdzināmu RE-līdzīgu transkriptu skaitu: tie ir trīs LTR RE virsgimeņu pārstāvji (*Gypsy*, *Copia*, *TRIM*), kuri

pieder astoņām ģimenēm, kā arī atrada attālu provīrusam līdzīgu elementu (ERV). Liela fragmentu grupa bija līdzīga *Ogre* RE dažādiem domēniem. Arī pēc hormonu ietekmes identificēja transkriptus ar dažādu līdzību *Ogre* elementam (ABA18, SA12). Transkripts B166 satur augsti līdzīgu sekvenci TRIM *Cassandra* LTR RE. TRIM ir īsi transkripcionāli aktīvi neautonomie RE, izplatīti augu vidū. Šo elementu insercijas bieži ir atrodamas gēnu intronos un promoteru rajonos (Witte *et al.* 2001). Tomēr, līdz šim šie elementi skujkoku genomam nav aprakstīti. Augu hormonu ietekmē diferenciāli ekspresēto fragmentu skaits bija ievērojami zemāks. Tomēr kopumā identificēja piecām RE ģimenēm (*Gypsy* un *Copia*) līdzīgus transkriptus, kā arī vienu provīrusam līdzīgu fragmentu. Identificēti vairāki transkripti ar *Copia-18-BD* RE līdzīgu sekvenci.

### 3.3. Himērisko transkriptu analīze

Detektēja divus transkriptus, kas satur gan TE daļu, gan hloroplasta genoma sekvences daļu (T51 un B153). 49% no T51 fragmenta līdzinājās bez LTR *LINE Karin\_TA* poliproteīnam, bet otra fragmenta daļa bija augsti homologa priedes hloroplasta genoma divām dažādām daļām (ar 16% un 24% pārklājumu). Otrs himēriskais transkripts B153 bija līdzīgs *Gypsy Cereba* poliproteīnam, kā arī nepārtrauktai priedes hloroplasta sekvencei. B153C hloroplastam homologā daļa līdzinās lielskaitam ierakstu no augu EST datubāzes (2465 sakrītību), bet RE līdzīgā daļa B153TE sakrita tikai ar vienu ierakstu no EST datubāzes. Šo hloroplastam-līdzīgo sekvenci atrada starp citām diferenciāli ekspresētām sekvencēm, tikai bez *Cereba*-līdzīgas daļas (SA01). Pie kam, šīs sekvences (B153C un SA01) hloroplasta genomā ir novietotas blakus ar citu transkriptu grupu (SA07, ABA03, ABA02). 33% no katras no šo transkriptu sekvencēm satur augsti homologu rajonu no RE *Copia-18\_BD* 3' rajona ar 98% līdzību, tāpēc šīs sekvences iekļāva TE-līdzīgo sekvencu analīzē. Četri līdzīgie fragmenti ar līdzību 83-94% identificēti pēc apstrādes ar ABA un viens pēc SA. BLAST meklēšana ar *Copia 18\_BD\_I* elementu atrada, ka 313-315 bp garš fragments 3' nekodējošā retrotranspozona daļā pirms LTR ir ļoti līdzīgs vairāku augu sugu hloroplastu genomiskajai sekvencei, ieskaitot *A.lyrata* (ORF64c). Pie kam, hloroplastam homologā *Copia 18-BD* sekvence atrodas apgrieztā orientācijā. Identificētais fragments sakrīt ar 1930 sekvencēm EST, bet *Copia 18\_BD-I* TE sekvence sakrīt ar 2500 EST sekvencēm.



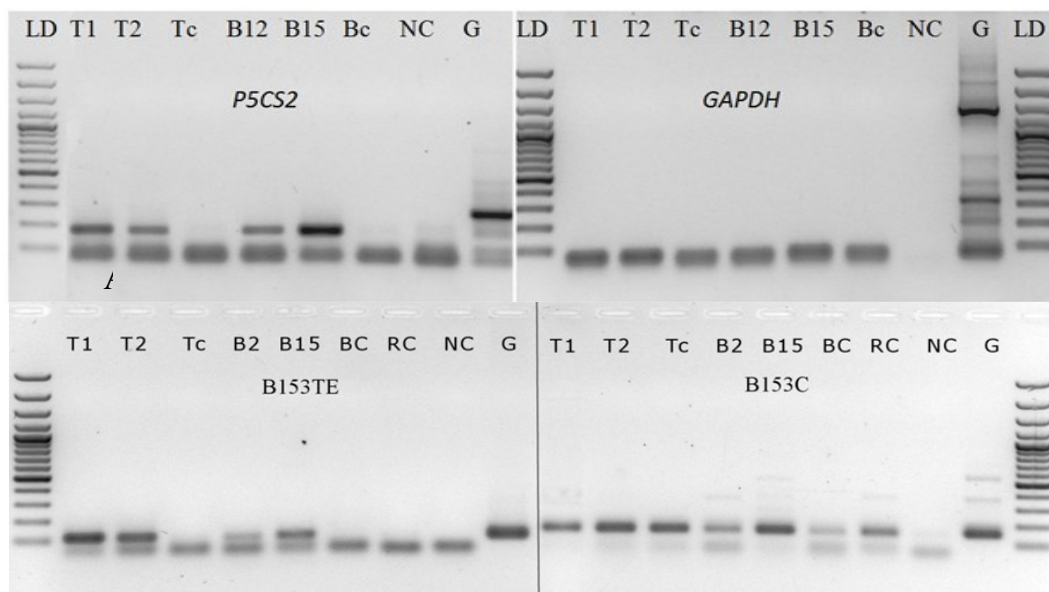
1. attēls. *Copia-18-BD* mRNS sekundārā struktūra un rajoni, kuri varētu veidot cilpveida struktūras ar iespējamību tuvu 1 (sarkans). Struktūras minimālā brīvā enerģija – 99,20 kcal/mol.

Interesanti, ka ir atrasti dažāda garuma un struktūras transkripti, kas varētu lokalizēties dažādos genoma lokusus, bet identificētie fragmenti var būt arī viena TE atkarīgās struktūras ekspresijas rezultāts. Abas identificētās transkriptu grupas atrodas 45 bp attālumā viena no otras, ko var izskaidrot ar PBS praimeru komplementaritāti šim hloroplasta sekvences rajonam.

Hloroplasta sekvencei homologo *Copia-18BD* 315 bp rajonu analizēja sīkāk, jo šis konservatīvais rajons varētu saturēt nozīmīgus regulatīvus motīvus. Meklējot augu regulatīvo motīvu datubāzē (REGSITE DB) ar programmu NSITE (Softberry Inc.) v.2.2004, atrada septiņu regulatīvo elementu motīvus ar homoloģiju  $\geq 80\%$ . Viens no tiem (TTTAAAG, BPBF no *H.vulgare*) bija palindroms (atrodas abās DNS ķēdēs), abās orientācijās nesatur nevienu SNP (e-vērtība=0,0094 un 0,0061). Tāpat, bez SNP atrada divus nezināmus kodola faktoros 5'-CCCGGATC-3' no kviešu (e-vērtība=0,00395) un 5'-AAACCTCC-3' no spinātu genoma. Sekvences analīze ar mikro RNS meklēšanas programmu *FindMiRNA* (Softberry Inc.) prognozēja šim rajonam sešu pre-miRNS gēnus tiešā orientācijā un piecus komplementārajai ķēdei, kas veido cilpveida struktūru ar brīvo enerģiju zemāk par -17 kcal/mol. 1. attēlā parādīta viena no prognozētām sekundārām RNS struktūrām ar RNAfold (Lorenz *et al.* 2011), kur vismaz viena cilpveida struktūra veidojās ar augstu iespējamību.

### 3.4. TE specifisko fragmentu amplifikācija

Lai salīdzinātu identificēto TE-līdzīgo fragmentu amplifikāciju dažādu stresoru ietekmē, konstruēja 27 specifiskus PĶR praimerus. Eksperimenta gaitā atkārtoti veikta amplifikācija ar kontroles gēniem (2. attēls A). Vienam no himēriskajiem fragmentiem (pB153) piemeklēja divus praimeru pārus katrai no sastāvdaļām: *Cereba*-līdzīgai daļai (pB153TE) un hloroplasta sekvencei līdzīgai daļai (pB153C). B153TE amplifikācijas produkti novēroti stresam pakļauto rametu cDNS paraugos, bet nebija kontroles paraugiem (2. attēls B). Visiem paraugiem novēroja B153C amplifikācijas produktus.

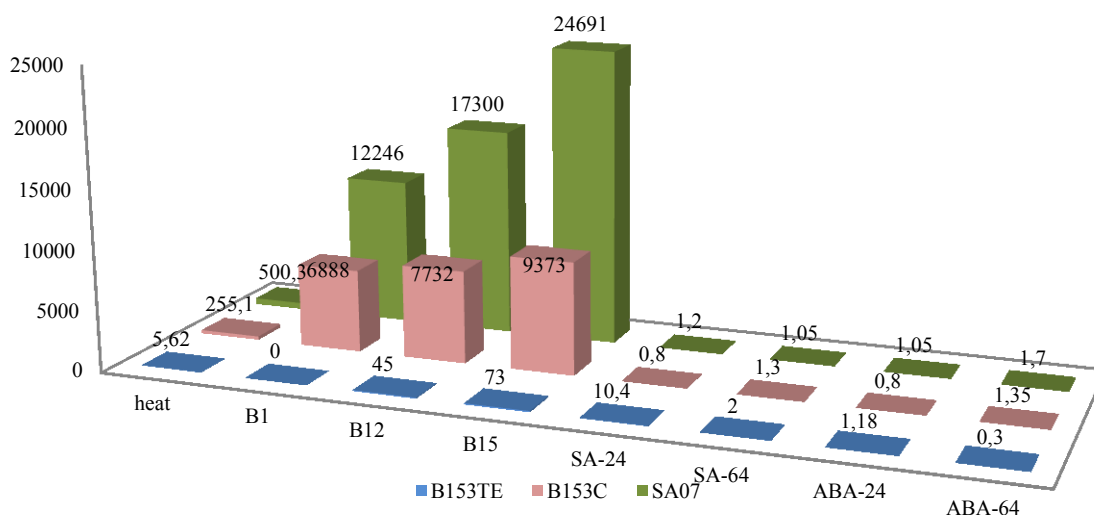


2. attēls. Salīdzinoša amplifikācija ar cDNS karstuma stresam (T1, T2), priežu hermesa invāzijai (B12, B15) pakļautiem un kontroles rametu paraugiem (Tc, Bc), negatīvā kontrole (NC) un gDNA (G). A - P5CS2-līdzīgā gēna produkti (190 bp un 250bp) un endogēnā kontrole GAPDH (70 bp). B - B153TE (110 bp) un B153C (120 bp).

### 3.5. Reālā laika PQR

Papildus standarta PQR reakcijām veica relatīvā ekspresijas līmeņa noteikšanu ar reālā laika PQR metodi. Standartizācijai visos gadījumos izmantoja endogēno kontroli GAPDH. Ievērojams transkriptu pieaugums stresam pakļautajos paraugos novērots SA07 fragmentam, kas bija līdzīgs *Copia-18-BD* un hloroplasta sekvencēm (3. attēls). SA07 amplifikācijas produktu kušanas līknes bija ar vienu izteiktu pīķi (aptuvenis  $T_m=84$  °C), standartlīknes slīpums bija no -2,9 līdz -3,68,  $R^2=0,65-0,90$ , efektivitāte=86,8-117%.

Ekspresijas līmeņa izmaiņas noteica himēriskā transkripta divām daļām (B153TE un B153C). B153C amplifikācijas produktu kušanas līknes bija aptuveni vienādas ( $T_m=82$  °C) un standartlīkne raksturojās ar slīpumu no -2,2 līdz -3,7;  $R^2=0,94-1$  un efektivitāti 85-121%. B153TE amplifikācijas produktu kušanas līknes nebija tik precīzas (aptuvenais  $T_m=78$  °C), standartlīkne raksturojās ar slīpumu -1,8,  $R^2=0,92$ , efektivitāti 241%. Vidēji, SA07 relatīvais ekspresijas līmenis bija divas reizes augstāks, salīdzinot ar B153C ekspresijas līmeni.



3. attēls. Relatīvā ekspresija *Copia-18-BD*-līdzīgam transkriptam SA07, hloroplastam-līdzīgam transkriptam B153C un *Cereba*-līdzīgam transkriptam B153TE.

### 3.6. Pilnizmēra RE izolēšana

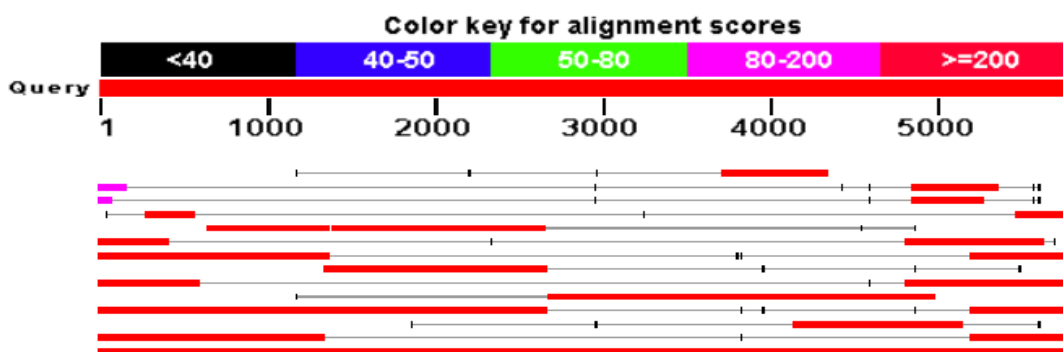
Analizējot diferenciāli ekspresētus fragmentus pamanīja, ka daži no tiem uzrāda augstu līdzības pakāpi skujkoku BAC sekvencēm, arī ja netika atrasta līdzība TE sekvencu datubāzēs. Piemēram, fragments T50 līdzinājās *P.taeda* BAC sekvencei PT\_7Ba4363K13 (AC241358.1). Analizējot sīkāk šo sekvenci, tai varēja identificēt LTR un līdzību TE sekvencu datubāzēs ar dažādu elementu domēniem. No *P.taeda* genoma BAC sekvences tika identificēts hipotētisks pilna izmēra LTR TE. Izveidotie praimeru rezultējās ar sagaidāmā garuma produktu izolēšanu un homologās sekvences iegūšanu no *P.sylvestris* gDNS. Izmantojot tos pašus praimerus, izdevās amplificēt un nosekvenēt sagaidāmā garuma fragmentus arī no cDNS

LTR TE-līdzīgu struktūru ptTE101 (ar līdzību T50) ir identificēts *P.taeda* BAC sekvencē PT\_7Ba4363K13 (gb|AC241358.1). ptTE101 *P.taeda* genomā ir 7697 bp garš, ar 949 bp gariem LTR ar 87% savstarpēji identiskiem nukleotīdiem un 5'-TG/

3'CA LTR galos. Identificēts TSD, kas ir 7 bp 5'-CATCATTT-3'. Šīs struktūras pilna izmēra anotēti homologi nav starp TE datubāzes ierakstiem, bet EST datubāžu analīze parāda, ka *P.taeda*, *P.contorta* un *P.glauca* ir atrodami TE iekšējai sekvencei līdzīgi transkripti. No gDNS un cDNS izolētas desmit LTR sekvences, kas bija homologas ptTE101 par 74-81%. Analīze *Repbases* atrada līdzību ar dažādu TE daļām, kā arī vāju RT/RH domēna līdzību nukleotīdu līmenī ar *Gypsy-18\_SMo-I* no *Selaginella moellendorffii*, kas ir senākais vaskulāro augu pārstāvis. BLASTX TREP datubāzē atrasta līdzība ar *BAGY-2* poliproteīnu (PTREP241), 36% identisku aminoskābju un 58% vienlīdzīgu aminoskābju, 183 nukleotīdu sakritību vērtība, e-vērtība= 2e-47. 4699-5481 pozīcijā pirms RT domēna atrodas int homologs rajons, kurš līdzinās *BAGY-2* un *Calypso-5* int domēniem proteīnu līmenī, kas norāda, ka šis elements pieder *Copia* virsgimenei.

ptTE121 (T73) retrotranspozonom-līdzīgs konstrukts ir identificēts *P.taeda* PT\_7Ba3500G20 (AC241315) sekvencē. ptTE121 ir 5631 bp garš ar 1536 bp LTR, un 2558 bp iekšējo sekvenci. LTR ir homologi viens otram par 89%, tie ir norobežoti ar 5'TG/3'CA sekvencēm, TE ir 5 bp TSD 5'- AACCC-3'. LTR satur interesantu poli-G mikrosatelītu (14bp) 5' rajonā. Identificēšanas laikā tam netika atrastas homologas sekvences TE sekvenču datubāzēs, taču 2013. gada 22. aprīlī publicētā sekvence *PtCumberland\_I* (Wegrzyn, *Repbases*) ar 72% pārklājuma rajonos līdzinās ptTE121, kā arī *PtCumberland\_I* satur vairākas lielas delēcijas. ptTE121 satur visas transkripcijai un translācijai nepieciešamās regulatīvās sekvences (uORF (*upstream ORF*), IRES (*Internal Ribosome Entry Site*), TATA, poliadenilēšanas motīvu). ptTE121 komplementārajā ķēdē atrodas četri ORF, kuri līdzinās dažādiem TE poliproteīniem. Izdevās iegūt no cDNS psTE121 LTR un 3' iekšējās daļas sekvenci.

Tāpat RE-līdzīga struktūra atrasta analizējot citos pētījuma posmos iegūtos fragmentus. Piemēram, psTE283 fragments darba izstrādāšanas gaitā neuzrādīja līdzību ne ar vienu RE datubāzēs, bet līdzinājās ar 77% nukleotīdu identitāti *P.glauca* BAC sekvencei, kuru analizējot atrada LTR RE-līdzīgu struktūru.



#### 4. attēls. *P.sylvestris* sekvenēto daļu pārklājums ar hipotētisku TE283 struktūru.

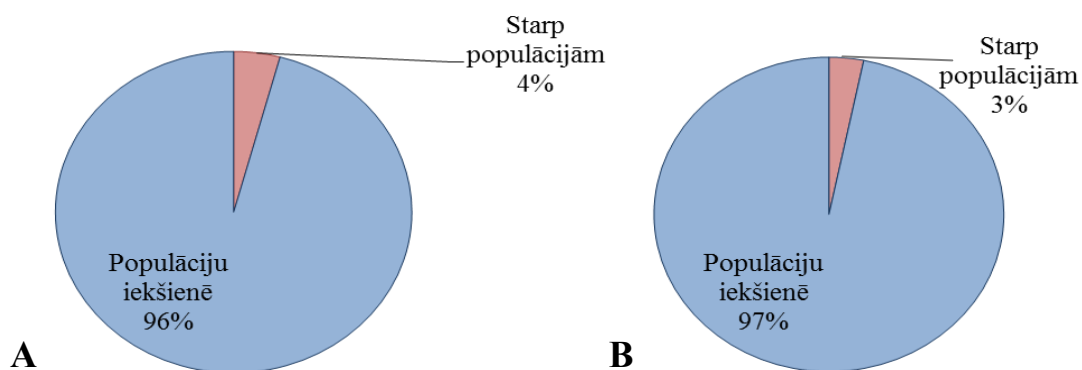
Izdevies identificēt kodējošos domēnus tikai ar EST datubāzes palīdzību kombinācijā ar ORF meklēšanas rīku, kas norāda, ka šis elements ir specifisks skujkokiem. Piemeklēja praimerus kodējošiem domēniem un no *P.sylvestris* cDNS ieguva šī RE daļas, kas sakrita ar homologu *P.glauca* struktūru. Šī RE sekvence līdzinās gan karstuma stresa rezultātā izolētajam T28 fragmentam, gan ar garās distances PĶR izolētajam fragmentam. Tomēr, 2013. gada 18. martā šī TE homologa sekvence tika publicēta *P.taeda* genomam ar nosaukumu *PtAppalachian\_I* (Wegrzyn, *Repbases*).



No *P.sylvestris* karstumam pakļautā eksperimentālā cDNS pa daļām izolēja pilna izmēra psTE283 retrotranspozonu (4. attēls) un izveidoja tā konsensus sekvenci. psTE283 ir 5796 bp garš ar 615 bp gariem LTR, kuri ir 80% līdzīgi viens otram. Izolētam psTE283 LTR tiek norobežoti ar 5'-TG/CA-3'. 5'LTR ir uORF, IRES, TATA-motīvs (5'-AATAATAAATA-3'), kā arī ir atrasta *P.sylvestris* regulatīvā sekvenca PtMYB4 5'-ACCACCC-3' (e-vērtība=0,007) un četri nezināmi kodola faktori. 3'LTR 391. pozīcijā ir poliadenilēšanas motīvs. psTE283 iekšējās sekvences 3'daļā atrodas polipurīnu motīvs (5'-AAGGGGGGAGGAA-3'). TSSP programma augu polimerāzes II promoteru noteikšanai (*Softberry*) psTE283 prognozēja promoteru tikai uz komplementārās ķēdes 5574. pozīcijā (3'LTR). Tiešās ķēdes psTE283 sekvencei atrasti divi ORF, kuru translētās aminoskābju secības uzrāda līdzību TE proteīniem. 354 aa garš ORF1 (2644-3709) ir līdzīgs RT/RH domēnam, bet 302 aa ORF2 (3669-4578) līdzinās int domēnam (pfam00665). Int domēns atrodas psTE283 3'daļā, kas norāda, ka psTE283 elements pieder *Gypsy* virsgimenei. Interesanti, ka psTE283 RT/RH domēns līdzinās *Cereba* poliproteīnam (TREP99) ar 43% identiskām aminoskābēm un 58% vienādām aminoskābēm. Līdzīga sakritība ar *Cereba* poliproteīnu bija B153 himēriskajam fragmentam. Tas varētu norādīt, ka šie transkripti reprezentē vienas ģimenes RE. Tāpat, TE283 ir arī par 77% līdzīgs B193 transkriptam, kuru izolēja priežu hermesa bojātam rametam. psTE283 sekvenca sakrīt ar 85 skujkoku EST sekvencēm NCBI datubāzē. *P.taeda PtAppalachian\_I* un *P.sylvestris* psTE283 izolētā homologa iekšējie domēni līdzinās par 91%, bet LTR par 81-88%. Analizējot pieejamu *PtAppalachian\_I* sekvenci no NCBI BAC PT\_7Ba2797E03 (AC241290.1), to LTR sakrīt par 90%, kas ir vairāk nekā izolētam *P.sylvestris* psTE283. Komplementārajā ķēdē *PtAppalachian\_I* pilna izmēra homologu sekvenci no BAC atrasts nepārtraukts 584 aa ORF, kas līdzinās RE poliproteīniem, bet otrs ORF ar 375 aa līdzinās gag domēniem. Visticamāk *P.sylvestris* psTE283 ir zaudējis transpozīcijas iespējas, jo tam nav pilna poliproteīna ORF un atrastais RT ORF ir īsāks par *PtAppalachian*. Taču transkripcijas funkcijas, iespējams, ir palikušas vairākām kopijām.

### 3.7. Atšķirīgos apstākļos augošas priežu populācijas genotipēšana

Izvēlējās dabiski atjaunojušos priežu audzi, kur vienas populācijas indivīdi aug paugurā (30,4 m), nogāzē un pārpurvojumā. Deviņu RE marķieru (1. tabula) fragmentu amplifikācija tika salīdzināta starp 150 indivīdiem pēc IRAP metodes principa (Kalendar *et al.* 1999). DNS ģenētisko analīžu rezultātā ieguva 184 lokusus ar IRAP metodi un 33 lokusus ar trim SSR marķieriem.



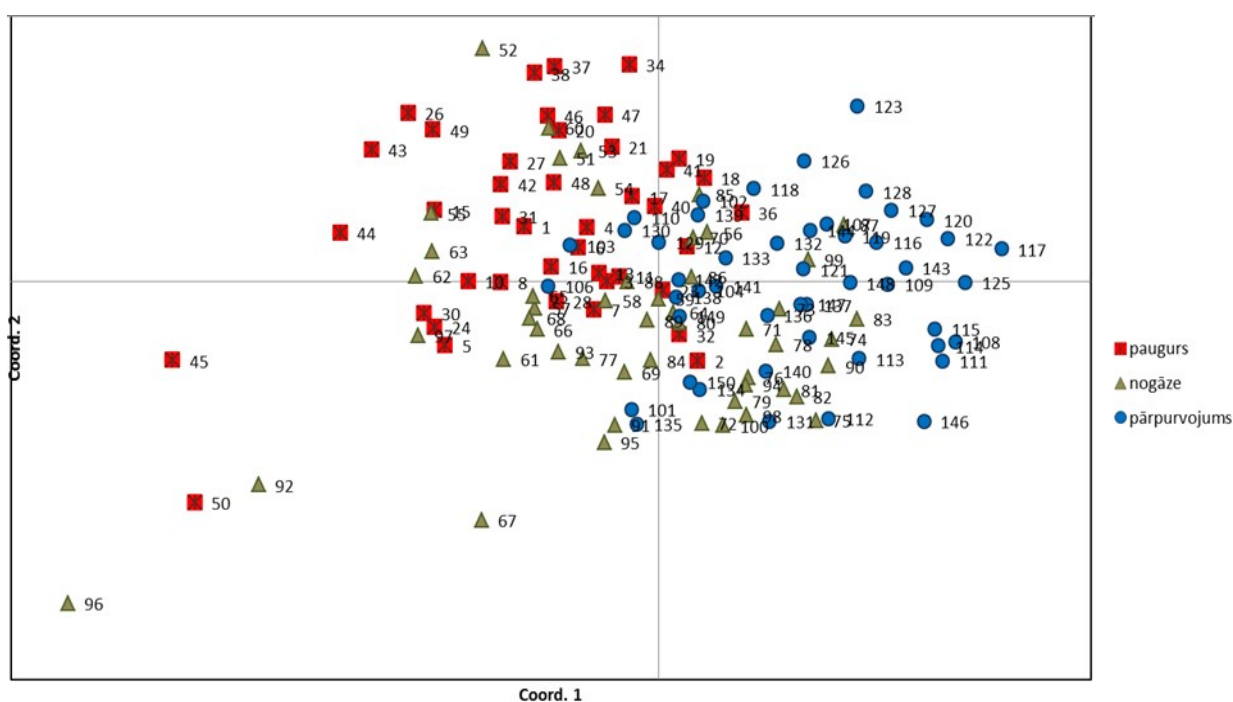
5. attēls. Molekulārās variācijas procentuālā daļa (AMOVA, GenAlEx) izmantojot SSR marķieru datus (A) un RE marķieru datus (B).

Molekulārās daudzveidības analīze (AMOVA) neatrada būtisku ģenētiskās daudzveidības atšķirību starp populācijām (5. attēls). Izmantojot indivīdu ģenētiskās distances matricu, veica galveno koordināšu analīzi jeb PCoA testu (*Principal Coordinates Analysis*), kas RE marķieru gadījumā nodalīja paugurā augošu populāciju no nogāzes un pārpurvojuma populācijām (6., 7. attēli). Pirmās trīs koordinātes izskaidroja 56,46% no variācijas starp IRAP datiem, un 61,15% SSR datiem.

Novērtēja katras populācijas indivīdiem amplificēto fragmentu skaita vidējās aritmētiskās vērtības. RE marķieru amplifikācijas fragmentu analīzei izmantoja tikai paraugus ar augstas kvalitātes amplifikāciju, tāpēc indivīdu skaits (n) bija atšķirīgs katrai populācijai.

1. tabula. Konstruēto IRAP marķieru amplifikācijas raksturojums.

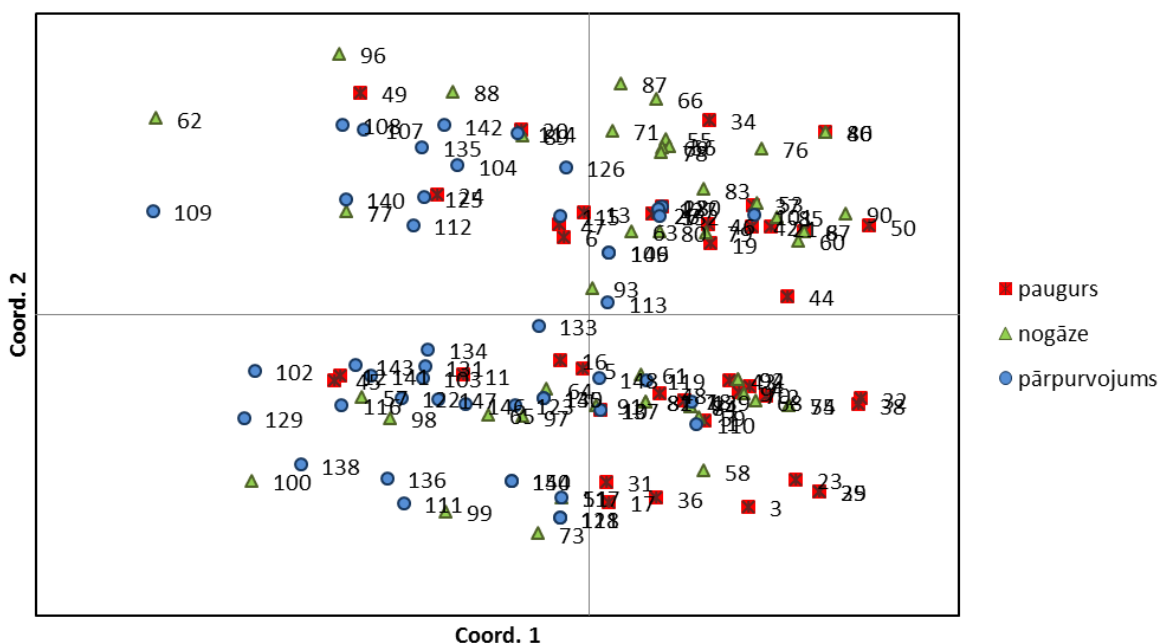
Praimeris	Izmantotās sekvenču numurs	Vidējais joslu skaits indivīdam	Polimorfo joslu skaits	Nepolimorfo joslu skaits
IR_1	JZ389987.1	11,60±1,90	12	4
IR_2	JZ390004.1	13,72±9,12	28	0
IR_3	JZ390006.1	8,4±2,56	13	0
IR_4	JZ390009.1	11,26± 3,68	19	1
IR_5	JZ390015.1	13,18±4,18	19	0
IR_6	JZ390027.1	22,20±12,86	37	0
IR_7	JZ390015.1	7,72±1,99	14	0
IR_8	JZ389987.1	9,83±1,13	13	3
IR_9	JZ390027.1	13,07±2,65	19	2



6. attēls. Dabīgās priedes audzes IRAP nespecifisko RE amplifikācijas datu galveno koordināšu analīze (PCoA tests).

Paugurā augošajai populācijai vērtēja 38 paraugus, nogāzē – 44 un pārpurvojumā - 45. Lai noskaidrotu, vai atšķirība starp paugura un nogāzes/pārpurvojuma populācijām ir būtiska, noteica vidējo vērtību starpību ( $d$ ), šīs starpības vidējo kļūdu ( $Sd$ ) un ticamības koeficientu ( $t$ ) (1. tabula). Starpība paugura un nogāzes populācijām ir vienāda ar  $6,78 \pm 1,42$ , bet paugura un pārpurvojuma starpība ir  $8,48 \pm 1,34$ . Tā kā ticamības koeficientam vērtības ir lielākas par vērtību trīs (1. tabula), starpību var vērtēt kā būtisku ar varbūtību 99% ( $P=0,01$ ).

Papildus ir noteikts F-kritērijs un novērtēta tā vērtība pēc tabulu vērtībām ticamības līmenim  $P=0,05$  un brīvības pakāpei  $df$  (40). Pēc F-kritērija novērtēšanas, paugura un pārpurvojuma lokusu skaita vidējo vērtību starpību var uzskatīt par būtisku, bet atšķirības starp paugura un nogāzes populācijām – netišas, dotajai ticamības pakāpei (2. tabula). Paugura populācijai ievērojami atšķirās gan amplificēto lokusu skaita vidējā aritmētiskā vērtība, gan standartnovirze. Lai gan variācija populāciju iekšienē ir augsta, var redzēt, ka amplificēto lokusu vidējais aritmētiskais skaits konstanti pieaug paugurā augošai populācijai veicot līdzīgu analīzi ar katra marķiera datiem.

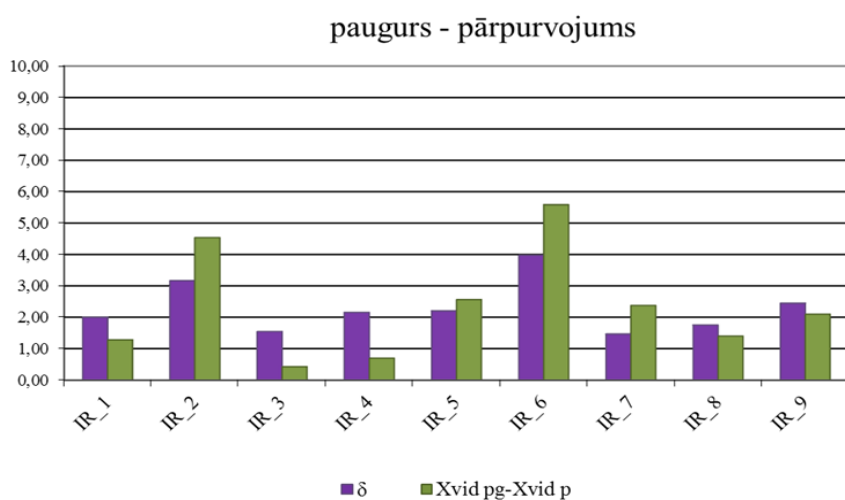
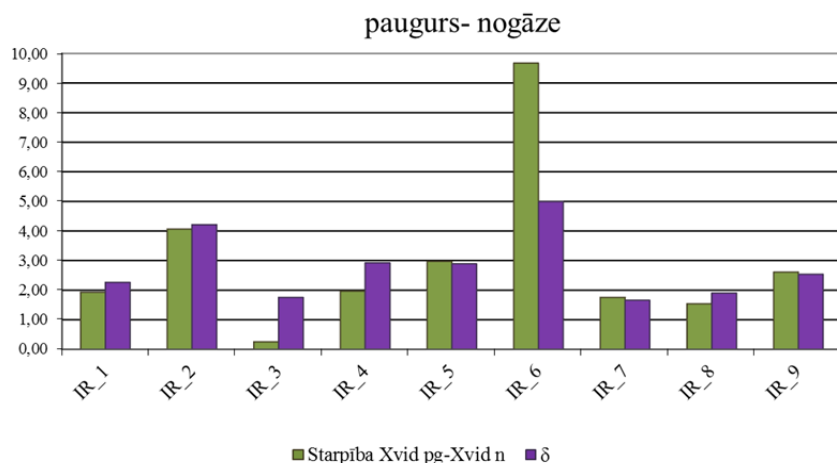


7. attēls. Dabīgās priedes audzes analīze ar PCoA testu izmantojot SSR lokusu datus.

2. tabula. Retrotranspozonu fragmentu amplifikācijas datu kopas analīzes statistiskie rādītāji. Populācija augoša paugurā (pg), nogāzē (n) un pārpurvojumā (p).

Pop	Vidējais aritmētiskais	Vidējā kvadrātiskā novirze	Vidējā kvadrātiskā kļūda	Starpība	Vidējā starpības kļūda	Ticamības koeficients	F-kritērijs
	Xvid	$\delta$	Sx	$d(X_{vpg} - X_{vn/p})$	Sd (pg-n/p)	$t > 3$	F, $P = 0,05$ , $df(40) \geq 1,51$
Pg	116,37	6,80	1,10	-	-	-	-
N	109,59	6,08	0,92	6,78	1,42	4,77	1,25
P	107,89	5,40	0,80	8,48	1,34	6,33	1,58



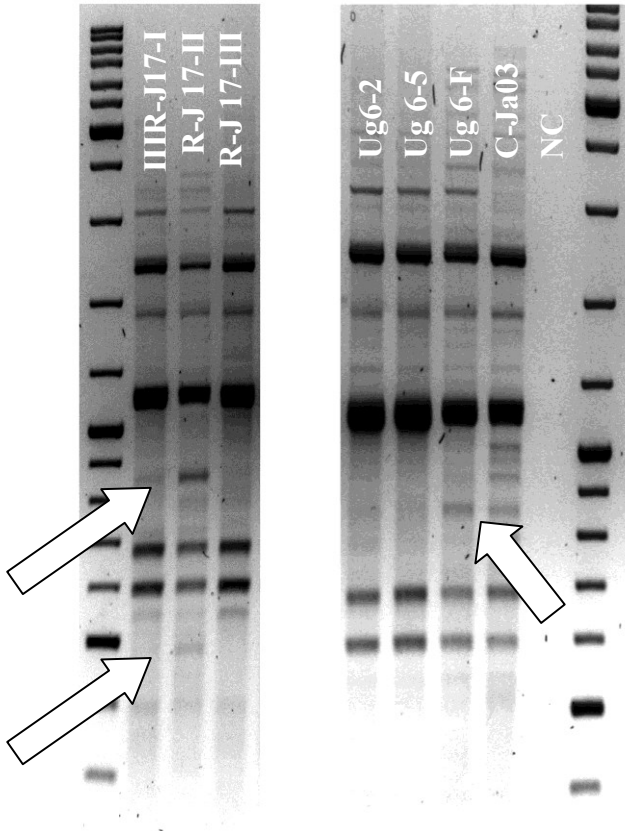


8. attēls. Marķieru amplifikācijas joslu skaita vidējo aritmētisko starpības salīdzinājums ar vidējo kvadrātisko novirzi ( $\sigma$ ).

Dažiem marķieriem (IR\_1, IR\_4, IR\_6) amplifikācija pieaug arī pārpurvojuma populācijā. Lai novērtētu starpības būtiskumu, tā tika salīdzināta ar vidējo kvadrātisko novirzi (8. attēls). Atšķirības amplifikācijā starp pauguru un pārpurvojumu ir būtiskas marķieriem IR\_2, IR\_5, IR\_6, IR\_7, bet starp pauguru un nogāzi tikai IR\_6.

### 3.8. Klonālās variācijas novērtēšana

Deviņus izveidotus IRAP praimerus pārbaudīja uz 26 gadus veciem parastās priedes kokiem, kas reprezentē dažādās klimatiskajās zonās ierīkotajās plantācijās „Drava” un „Sāviena” augošus četru klonu rametus. Pēc SSR datiem rameti neatšķirās. Astoņiem IRAP marķieriem amplifikācijas produkti bija vienveidīgi, neidentificēja skaidri atšķiramus polimorfus fragmentus. Tikai ar praimeru IR\_7h ieguva polimorfus produktus, kas bija skaidri atšķirami katrai rametu grupai (9. attēls). Polimorfie produkti novēroti gan „Sāvienas”, gan „Dravas” plantāciju paraugiem no dažādiem rametiem. Izolēja septiņus polimorfus fragmentus, sekvenēja 17 insertu. Visām sekvencēm to galos atrasts praimeris IR\_2. Ieguva trīs līdzīgas sekvenču grupas un divas individuālās sekvences. No 5 sekvencēm (869 bp) sastāvoša sekvenču grupa daļēji līdzinājās *P.taeda* BAC sekvencēm. Tās analizējot izdevies identificēt divus tiešos atkārtojumus iespējamus LTR, kas ar 555 bp līdzinās polimorfām sekvencēm. Šī struktūra nelīdzinās



9. attēls. IR\_7h (T73, Gypsy-I) IRAP amplifikācijas produktu polimorfisms priedes ramiem.

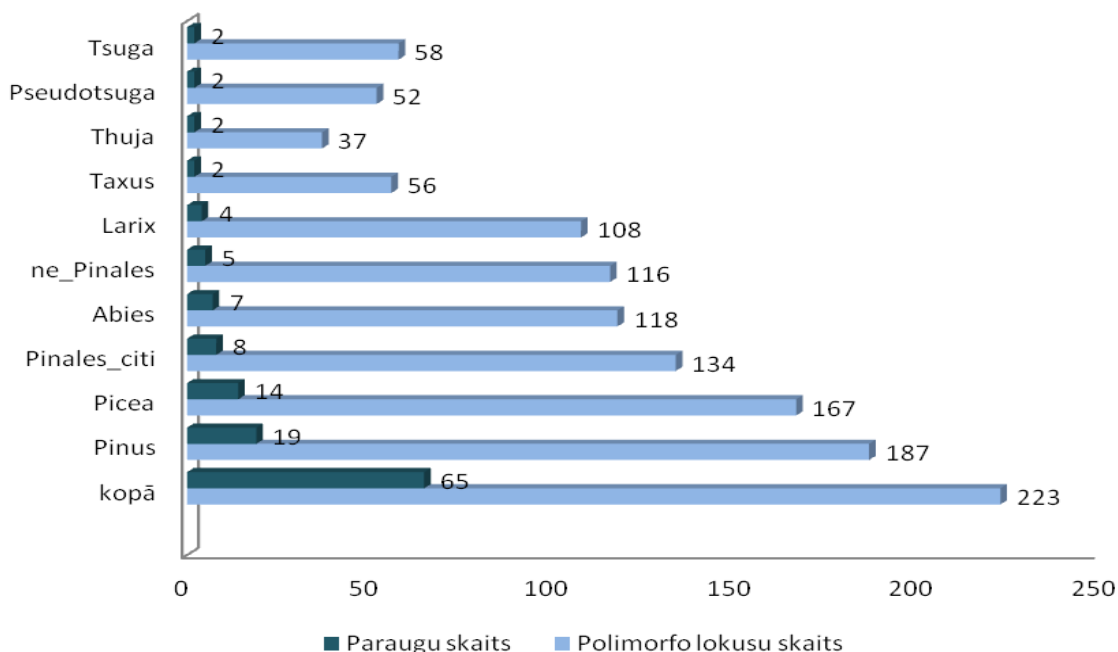
60 dažādu sugu kailsēkļu DNS paraugu analīzei izmantoja sešus IRAP marķierus, analīzē iekļāva arī četrus parastās priedes paraugus. Ieguva 241 fragmentu. Vidēji katram indivīdam bija 38,8 fragmenti, ar variāciju 134,59 un standartnovirzi 11,6. Datus sagrupēja lielākās grupās pēc piederības ģintij: baltegles (7), egles (14), priedes (19), citi *Pinales* (20) un citi kailsēkļi (5). Kailsēkļu ģintis nebija pārstāvētas ar vienādu sugu skaitu (10. attēls), tomēr *Thuja* raksturojas ar vismazāko amplificēto polimorfo lokusu skaitu (37), salīdzinot ar tikpat lielā mērā pārstāvētiem *Tsuga* (58), *Taxus* (56) un *Pseudotsuga* (52).

Pārējām skujkoku ģintīm polimorfo lokusu skaits bija lielāks par 100 un daļēji atkarīgs arī no analizēto paraugu skaita. Fragmentu reprezentācija indivīdā (%) jeb frekvence (11. attēls) ir aprēķināta kā fragmentu skaits indivīdam attiecībā pret kopējo lokusu skaitu reizināts ar 100. Vairākus paraugus saturošām ģintīm aprēķināja šo frekvenču vidējo aritmētisko, atdalot *P.sylvestris* paraugus atsevišķi no pārējām *Pinus* sugām. Augstākā fragmentu amplifikācija ir parastajai priedei, kuras genomam veidoja priēžu sugu zemāka fragmentu reprezentācija skaidrojama ar vairāku attālu priēžu sugu iekļaušanu analīzē (*P.banksiana*, *P.contorta*, *P.pumila*, *P.canariensis*). *Thuja* ir samazināts ne tikai polimorfo lokusu skaits, bet arī visu fragmentu frekvence (11. attēls). Īpaši zema amplifikācija novērota eksotiskajiem kailsēkļiem, kas neaug ziemeļu puslodē (*Wollemia nobilis*, *Zamia integrifolia*, *Araucaria*).

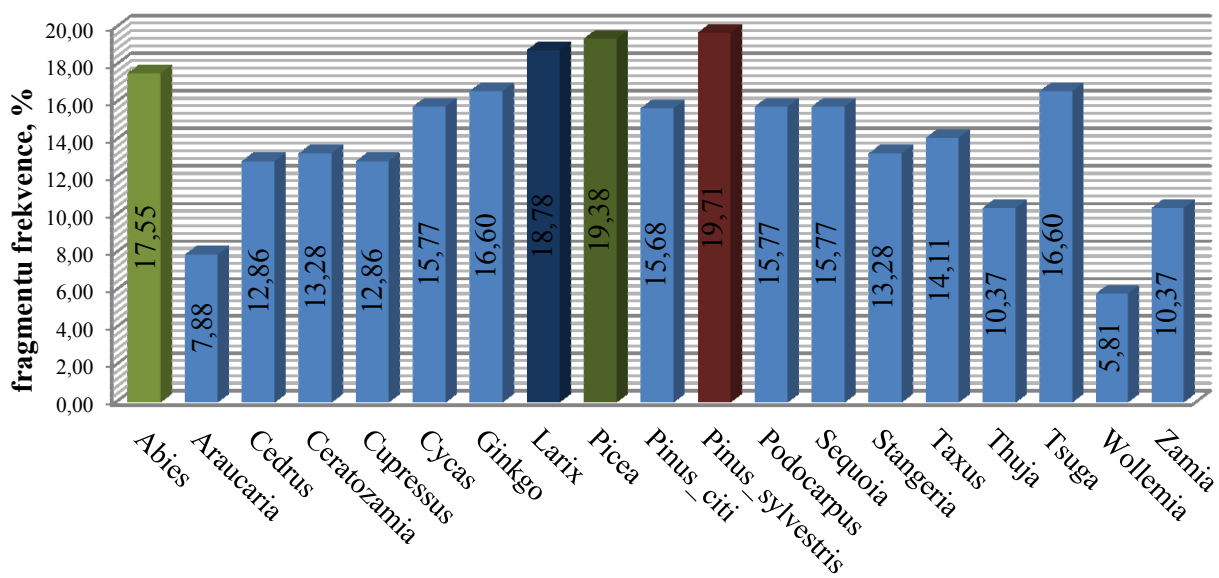
nevienam RE datubāzēs, bet daļēji līdzinās (par 70%) *Copia-17\_PAb-I* (MA\_172979) no *P.abies* iekšējam domēnam un nelielam LTR rajonam (63 bp). Identificētais *Copia-17* līdzīgais elements ir tikai 4052 bp garš ar 858 bp LTR, pa vidu ir ORF ar vāju līdzību *Copia* poliproteīnam.

NCBI datubāzē ir pilnas struktūras līdzība tikai ar sešām *P.taeda* BAC sekvencēm. 17 EST līdzība ir ar struktūras LTR rajoniem, bet 4 EST līdzinās arī iekšējai sekvencei ar zemāku līdzības pakāpi. Otras grupas sekvenses (507 bp) arī līdzinājās *Copia-17 Pab-I* un arī identificētās *P.taeda* RE LTR, kur bija līdzīgi 32 bp fragmenta beigās. Individuālas sekvenses nelīdzinājās datu bāžu ierakstiem. Kopumā trīs no sekvenētajiem polimorfajiem priedes ramiem fragmentiem līdzinājās jaunatklātajai struktūrai, līdzīgai *Copia-17* LTR RE, kura sekvence arī identificēta *P.taeda* BAC.

### 3.9. Kailsēkļu filoģenētiskā analīze ar IRAP



10. attēls. Polimorfo lokusu skaits dažādās kailsēkļu grupās un šo grupu lielums.

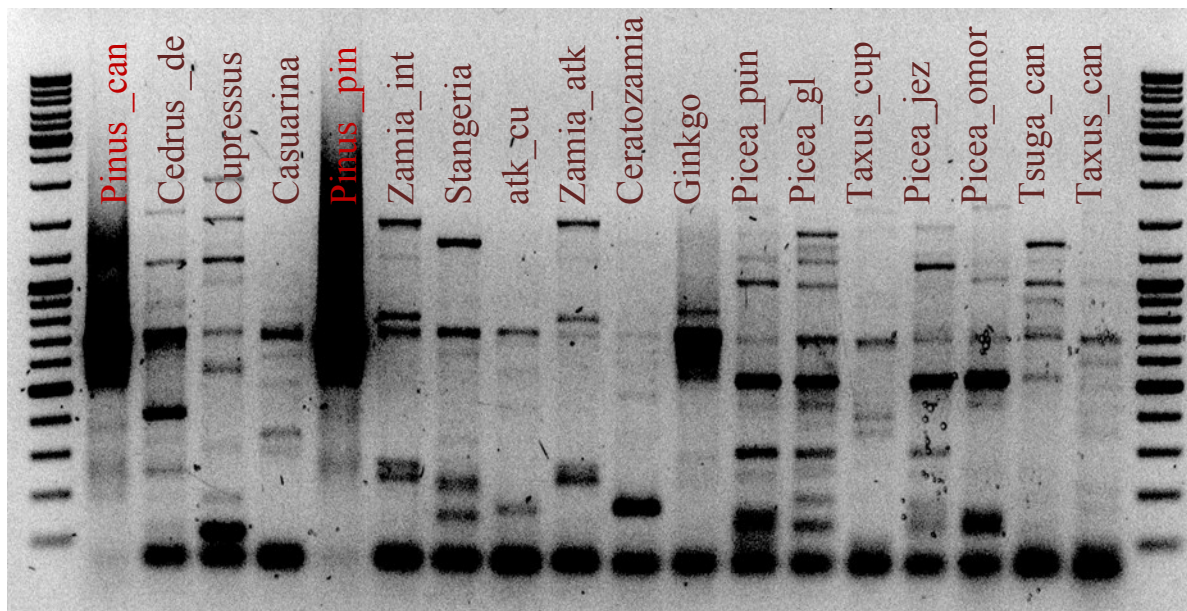


11. attēls. Vidējā (kur ir iespējams) un individuālā fragmentu reprezentācijas frekvence, %.

NJ-filoģenētiskā analīze sagrupēja parastās priedes paraugus atsevišķā zarā ar augstām 'bootstrap' vērtībām (12. attēls), kas ir sagaidāms, jo pārējās sugas ir pārstāvētas tikai ar vienu paraugu. Kā bija sagaidāms, dažādas kailsēkļu grupas savstarpēji ir diezgan līdzīgas, uz ko norāda katra zarojuma statistiskās ticamības līmenis. Balteglu ģints pārstāvji pārliecinoši nodalījās no pārējiem paraugiem, kas liecina par pētāmo TE atšķirībām šajā grupā.



diapazonā no 500 bp līdz 1500 bp (13. attēls). TE121 retrotranspozonomam amplificēja iekšējās sekvenču-LTR fragmentu (1608 bp). Fragmenti variēja no 1,2 kb līdz apmēram 2,5 kb *P.nigra*, *P.latifolia*, *P.nigra* var. *banatica*, *P.mugo*, *P.ponderosa*, *P.sylvestris*. Salīdzinoši specifiska un spilgta josla sagaidāmajā garumā novērota eglei *P.rubens*, vairākām priežu sugām *P.peuce*, *P.sibirica*, *P.koraiensis*, *P.canariensis*, *P.pinea*, kā arī attālākajiem *Ginkgo biloba* un *Araucaria heterophylla*.



13. attēls. *psTE101* retrotranspozona LTR-gag domēnu amplifikācija 32 paraugiem, sagaidāmais produkta garums 700 bp.

Amplifikācijas ar SA07 (ar *Copia-18BD*-līdzīgu rajonu) rezultējās vienā specifiskā produktā visiem kailsēkļu paraugiem, kas droši vien ir hloroplasta augsti konservatīvais rajons. Monomorfs produkts, kas arī līdzinās hloroplasta sekvenču, ieguva ar B153C praimeriem. Analīze parādīja, ka šie fragmenti ir sastopami visu pētījumā iekļauto paraugu vidū, kas apstiprina šī rajona izcelsmi no hloroplasta genoma. Savukārt *Cereba*-līdzīgā fragmenta B153TE amplifikācijas produkti variēja. Sagaidāmā garuma (120 bp) fragmenti novēroti 15 priedes, 10 egles, divām balteglu un vienai lapegļu sugai. Tāpat amplifikācijas produkts novērots ginkam, bet vairāki polimorfi fragmenti novēroti tikai *Stangeria eriopus*.

*Pinus* sugām biežāk bija sastopama izolēto TE polimorfa amplifikācija, attālākās kailsēkļu rindās tā samazinās. Šajā pētījumā no *P.sylvestris* izolētie ptTE101, ptTE121 un psTE283 ir arī *P.nigra*, *P.latifolia*, *P.mugo* un *P.ponderosa* genomos. Karstuma stresā izolētais psTE283, atšķirībā no pārējiem aprakstītajiem elementiem, ir atrodams arī attālajos *Cedrus deodara*, *Ginkgo biloba*, *Ceratozamia mexicana* genomos, kas liecina par šī elementa senu izcelsmi un saglabāšanos priedes sugām apejot egles ģinti. Izteikta psTE283 domēna amplifikācija nav novērota nevienai no pētījumā iekļautajām egļu, balteglu un lapegļu sugām. Visatšķirīgākie no *P.sylvestris* RE bija *Pseudotsuga menziesii* var. *glauca* un *Wollemia nobilis*. Interesanti, ka priežu ģintī pēc izmantotiem marķieriem no *P.sylvestris* ir atšķirīga *P. contorta*.



## 4. DISKUSIJA

### 4.1. TE-saistīto transkriptu identificēšana

Stresa apstākļos nespecifiski PBS praimeru (Kalendar *et al.* 2010) ļāva konstatēt diferenciālu TE-līdzīgo sekvenču ekspresiju, identificējot virkni zināmiem RE līdzīgas poliproteīnu sekvenču, kā arī citas, pagaidām neidentificētas sekvenču. Pretēji sagaidāmajai dominējošajai fragmentu struktūrai, kas ir aprakstīta genomiskās DNS amplifikācijas gadījumā, ir detektēti tikai daži diferenciāli ekspresēti transkripti ar pretēji orientētām LTR sekvenču to galos. Genomiskās DNS amplifikācijas gadījumā ar iPBS metodi ir iespējams iegūt 50-70% ar TE saistītu sekvenču (Kalendar *et al.* 2010). Zemāks identificēto TE-līdzīgo sekvenču skaits ir skaidrojams gan ar šo transkriptu daudz mazāku reprezentāciju transkribējamo sekvenču vidū, gan arī ar mazāku iespēju identificēt priedes TE-līdzīgos transkriptus. Priedes genoma RE sekvenču ir atšķirīgas no segsēkļu RE (Kovach *et al.* 2010; Nystedt *et al.* 2013), kā arī izmantotajās datubāzēs publicēto skujuķoku TE sekvenču skaits ir nesalīdzināmi mazāks nekā segsēkļiem. Vairāki diferenciāli ekspresēti fragmenti saturēja TE homologos rajonus visā tā garumā. Arī identificētajam DNS TE *EnSpm* izmantotā PBS praimera sekvenču bija komplementāra TE ORF izņemot pirmos trīs nukleotīdus no 5' praimera gala. Īpaši daudz sekvenču identificētas pateicoties TE kodējošo domēnu līdzībai, jo šie rajoni dažādām TE ģimenēm ir viskonservatīvākie (Wicker *et al.* 2007). Zems identificēto LTR homologu skaits ir skaidrojams ar zināmu LTR sekvenču variabilitāti starp TE ģimenēm, līdz ar to vairums no iespējamajām LTR sekvenču palika neidentificētas meklējot homologiju starp publicēto segsēkļu genomu TE sekvenču. Nesena papeles genoma TE analīze parādīja, ka tikai 42% no papeles TE ir homologu kādai publicētai TE sekvenču (Cossu *et al.* 2012). *P.abies* genomā tikai aptuveni puse atkārtoto sekvenču uzrādīja līdzību ar jau aprakstītiem TE. Senākas izcelsmes *Gypsy* virsģimenei vairāk identificēto transkriptu bija homologu, salīdzinot ar *Copia RE*. Vienā no nedaudziem bioinformātikajiem pētījumiem par augu RE transkriptiem, kur analizēja kukurūzas EST sekvenču sastāvu, novēroja skaitlisko *Gypsy* transkriptu pārsvaru pār LTR *Copia* un CACTA DNS TE transkriptu skaitu (Vicent 2010). Līdzīgi rezultāti ir novēroti kartupeļiem (Yadav, Singh 2013) un graudaugiem (Echenique *et al.* 2002). Tāpat, *P.abies* genomā identificēja *Gypsy* TE ir ar lielāku īpatsvaru salīdzinot ar *Copia* TE (Nystedt *et al.* 2013).

### 4.2. Stresa izsaukta TE-saistīto sekvenču ekspresija

Parastās priedes genomā tādi dabā sastopami stresa apstākļi, kā karstums un kukaiņu invāzija, veicina dažādu ar TE-saistīto sekvenču ekspresiju. Salīdzinājumā, apstrāde ar ABA un SA jau pirmajā pētījuma posmā izraisīja zemāku nespecifisku iPBS polimorfo lokusu amplifikāciju, un arī izolēto fragmentu vidū TE-saistīto sekvenču transkriptu sastopamība bija viszemākā. Tas var būt skaidrojams arī ar karstuma stresa un priežu hermesa invāzijas nespecifisku darbību uz vairākiem augu aizsardzību aktivējošiem signālceļiem, savukārt SA un ABA ietekmē tiek aktivēti specifiskie signālceļi, apejot primārā stresa iniciācijas ceļus. ABA ir augu hormons, kurš mitruma nepietiekamības gadījumā tiek iesaistīts atbildes reakcijā, un tas tiek sintezēts gan saknēs, gan lapās, bet var tikt uzņemts arī no apkārtējās vides. ABA receptori lokalizējas atvārsnīšu šūnu membrānā (Wilkinson, Davies 2002). ABA spēj uzkrāties simplasta rezervuārā un neizraisīt atbildes reakcijas augiem, kamēr lapās un saknēs nav panākts tipisks ūdens deficītam raksturīgs pH (Wilkinson, Davies 2002). Iespējams, tas

izskaidro tik nelielu ABA efektu uz priedes rametu TE-saistīto sekvenču transkripciju šajā eksperimentā, jo stādiem neradīja ūdens deficītu. Literatūrā ir aprakstīti gadījumi, kad RE transkribējas atbildot uz ABA apstrādi, piemēram, zemenēm *FaRel Copia* LTR TE (He *et al.* 2010). Miežu *Copia* LTR *BARE-1* sekvenca satur ABA inducējamu elementu, kas var izskaidrot šī elementa kopiju skaita palielināšanos sausuma stresa ietekmē (Kalendar *et al.* 2000).

Lielākā TE-līdzīgo sekvenču daudzveidība ir atrasta pēc karstuma stresa iniciācijas, kur ir atrasts pat II klases DNS TE transkripts, kā arī bez-LTR LINE RE sekvenca un *Gypsy* un *Copia* virsgimeņu pārstāvji. DNS TE CACTA ģimenes *EnSpm*-līdzīgu transpozonu aktivitāte stresa apstākļos ir pierādīta vairākām augu sugām: kukurūzai (Peterson 1953; McClintock 1954), kviešiem (Wicker *et al.* 2003), sorgo (Chopra *et al.* 1999), rīsiem (Jurka 2005), pākšaugiem (Altinkut *et al.* 2006). Nejauša *EnSpm*-līdzīgā RNS transkripta identificēšana norāda uz aktivēšanas iespēju stresa apstākļos šīs ģimenes pārstāvim arī *P.sylvestris* genomā. Tomēr bez pilnā garuma sekvences noteikšanas, nav iespējams precizēt, vai tiek aktivizēts DNS TE, vai tikai tā nefunkcionālā daļa. Nevienu līdzīgu fragmentu neatrada starp karstuma stresa eksperimentā izolētiem un augu hormonu apstrādes rezultātā izolētiem fragmentiem. Zināms, ka RE transkripcija ir atkarīga no *in cis* darbojošos regulatīvo sekvenču reprezentācijas LTR sekvencē (Vernhettes *et al.* 1998; Takeda *et al.* 1999; Butelli *et al.* 2012). Šīs regulatīvās sekvences ir atkarīgas no dažādām signālmolekulām, un tād arī dažādiem vides vai stresa faktoriem (Hirochika 1993; Beguiristain *et al.* 2001; Tapia *et al.* 2005; Ito *et al.* 2011).

Kukaiņu invāzijas eksperimentā liela fragmentu grupa līdzinājās dažādām *Ogre* domēnu sekvencēm. *Ogre* ir milzīgs transkripcionāli aktīvs *Gypsy* LTR RE, kurš sākotnēji ir aprakstīts zirņu genomā, bet atrasts arī citu divdīgļlapju sugām, piemēram, *Medicago truncatula* (Neumann *et al.* 2003). Dažus līdzīgus transkriptus atrada pēc kukaiņu invāzijas un pēc SA apstrādes, kas ir saskaņā ar zināmu SA funkciju. SA darbojas kā signālmolekula, kura inducē auga pretestības reakciju patogēnu uzbrukuma laikā (Malamy *et al.* 1990; Malamy, Klessig 1992; Klessig *et al.* 2000; Durrant, Dong 2004). Zināms, ka vairāku augu RE transkripcija ir inducējama ar SA apstrādi, piemēram, tomātu *TLC1.1* (Tapia *et al.* 2005), tabakas *Tnt1* (Grandbastien *et al.* 1997), auzu *OARE-1* (Kimura *et al.* 2001).

Analizējot TE-līdzīgus un arī anonīmus fragmentus, katram pētāmajam stresa tipam pārsvarā identificēja atšķirīgus transkriptus, tikai daži fragmenti bija kopīgi. Savstarpējā fragmentu salīdzināšana pārsvarā sakrita ar identifikācijas rezultātiem, tomēr īsāks amplifikācijas rajons var novest pie maldinošiem rezultātiem. Pirmkārt, kompleksajā priedes genomā šie lokusi ir pārstāvēti vairākās kopijās un dažādas transkripta daļas praktiski var būt līdzīgas augsti konservatīviem rajoniem, piemēram, hloroplasta sekvencai un tās insercijām kodola genomā. Otrkārt, zināms, ka vienas ģimenes dažādu tipu TE pārstāvji var transkribēties atšķirīgos apstākļos pateicoties stresam sensitīviem elementiem lokalizētiem TE LTR (Kumar, Bennetzen 1999; Takeda *et al.* 1999; Havecker *et al.* 2004; Beguiristain *et al.* 2001). Treškārt, homoloģija ar datubāzē esošajiem ierakstiem var būt maldinoša, jo vienai ģimenei kodējošiem domēniem homologie transkripti, priedes genomā var izrādīties dažādu TE ģimeņu pārstāvji. Piemēram, vairākkārt atrastie *Gypsy Ogre* ģimenei homologie transkripti bija ar augstu līdzību dažādiem domēniem pat nukleotīdu sekvenču līmenī, bet amplifikācija starp kodējošiem domēniem ar genomisko DNS nedeva rezultātus. Iespējams, šī līdzība

attiecas uz dažādu ģimeņu elementiem un ar šiem praimeriem tiek amplificēts anonīms genomiskais rajons starp elementiem. Iespējams arī, ka šis elements ir sens un priedes genomā tas vairs netiek transkribēts pilnā garumā. Tādā gadījumā, izkaisītās konservatīvās TE daļas ģenerē himērus transkriptus, bet pētījumā identificēja tikai to daļas. Tomēr, amplifikācijas produkti pat ar vienu no *Ogre-līdzīgo* transkriptu praimeriem ir ļoti polimorfī, droši vien, katra no šīm ģimenēm priedes genomā ir pārstāvēta vairākās kopijās.

#### **4.3. Himērisko transkriptu analīze**

Lielākā transkriptu grupa, kuru identificēja pēc apstrādes ar ABA, bija homologa *Copia-18\_BD* elementam no *Brachypodium*, bet šie transkripti ir arī homologu hloroplasta rajonam. Iespējams, hloroplasta rajons ir iekļauts *Copia-18\_BD* RE *Brachypodium* genomā. Nav zināms, vai šī *Copia-18\_BD* homologs ir atrodams priedes genomā, šajā pētījumā homologu sekvenci izolēt neizdevās. Iespējams, šī homologija ar *Copia-18-BD* ir nejauša, jo priedei un visiem augiem šī secība ir atrodama hloroplastu genomā. Izplatīta parādība augu genomos ir hloroplastu un mitohondriju DNS insercijas kodola genomā (Lin *et al.* 1999; Stupar *et al.* 2001; Guo *et al.* 2008; Ueda *et al.* 2005). Interesanti, ka šajos kodola genoma rajonos bieži ir sastopami TE ar hloroplasta vai mitohondrija DNS inserciju, un uzskata, ka TE piedalās šo rajonu pārstrukturēšanā (Ueda *et al.* 2005, Guo *et al.* 2008). Analizētā hloroplastu rajona insercija augiem ar sekvenētu genomu ir atrodama kodola genomā. Piemēram, atšķirīga skaita insercijas ir atrodamas aļģēm, *Arabidopsis lyrata*, pākšaugiem (*Glycine max*, *Medicago truncatula*), *Populus trichocarpa*, nakteņu dzimtas augiem (kartupeļi, tomāti). Viendīgļlapjiem, piemēram, rīsiem genomā ir vismaz astoņas šī reģiona insercijas, piecas insercijas atrodamas *Brachypodium distachyon*, bet divas *Sorghum bicolor*. Hloroplasta genomam homologu transkriptu identificēšana ar variējošām flankējošām sekvencēm, norāda uz vairākām šī rajona kodola insercijām arī *P.sylvestris* genomā. Vēl jo vairāk, citas transkriptu grupas identificēšana (B153) ar hloroplastam homologu daļu un nehomologu TE daļu, apstiprina šo iespēju. Visi šie transkripti (B153, SA01, SA07, ABA03, ABA02) varētu būt dažādu inserciju rezultāts, interesanti, ka šie transkripti ir atrodami pirms stresa iniciācijas un to ekspresijas līmenis strauji pieaug stresa apstākļu ietekmē, bet RE-līdzīgie amplifikācijas produkti ir atrodami tikai stresam pakļautajos augos. Tāpat, šo transkriptu homologu ir sastopami citu augu EST sekvencēs, norādot, ka šīs sekvences ir transkripcionāli aktīvas arī citās sugās, iespējams TE inserciju dēļ. Turpretī, līdzības meklēšana ar B153TE *Cereba*-līdzīgu rajonu rezultējās vienīgajā EST ierakstā Benksa priedei (*P.banksiana*), kas var norādīt uz šīs sekvences specifisku transkripciju dažām priedes sugām. Šo transkriptu motīvu analīze parādīja, ka šī sekvenca var saturēt mikroRNS gēnu, jo vismaz vienai sekvences daļai ar lielu iespējamību veidojas cilpveida struktūra. Rajona sekvences konservatīvāte, sastopamība, ekspresijas līmeņa izmaiņas dažādu stresa apstākļu ietekmē, norāda uz šī rajona iespējamu funkcionalitāti, kas ir jānoskaidro turpmākos pētījumos.

#### **4.4. Identificēto priežu TE izplatība kailsēkļiem**

Vairāku augu genomu RE izplatīšanās un struktūras salīdzinošie pētījumi atklāj, ka RE ir evolucionāri seni un homologas RT ģimenes var būt sastopamas attāliem augu taksoniem (Flavell *et al.* 1992; Voytas *et al.* 1992; Xiong and Eickbush 1990). Iepriekšējie *P.taeda* genoma pētījumi aprakstīja visu klašu un kārtu TE pārstāvju identificēšanu: DNS TE, endogēnos retrovīrusus, LTR un bez LTR RE (Friesen *et al.*



2001; Kovach *et al.* 2010). *Triticeae*, *Arabidopsis* un *Oryza* genomu *Copia* virsgimenes RE pētījumos atklāti seši evolucionāri atzari, kas izplatījās vēl pirms viendīgļlapju un divdīgļlapju nodalīšanas (Wicker, Keller 2007). Līdzīgs sojas, rīsu un kukurūzas genomu pētījums atrada septiņus *Copia* RE un piecus *Gypsy* evolucionārus atzarus (Du *et al.* 2010). RE, kas pieder noteiktajam priekšteču atzaram, raksturojas ar kopīgām iezīmēm, kā arī var saturēt negaidīti augstu homologiju DNS sekvences līmenī starp attāliem augu taksoniem, un var liecināt par horizontālu gēnu pārnesi šiem elementiem (Wicker, Keller 2007). Šajā pētījumā identificēja RE transkriptus ar homologiju poliproteīnu kodējošām sekvencēm *Copia* virsgimenes *Maximus* evolucionārajam atzaram (*Copia-2-TE* līdzīgie fragmenti B104 un T30), kā arī *Angela* un *Ivana* grupām līdzīgs transkripts (T12). *Gypsy* virsgimenes gadījumā identificēja *Tat* atzara pārstāvjus (*Ogre*-līdzīgie transkripti), *CRM* atzara pārstāvjus (*Appalachian*-līdzīgie fragmenti B193, T28) un *Athila* atzara pārstāvji (*PIGY-1*, B301). *Maximus*, *Athila* un *Tat* evolucionāro grupu pārstāvji satur endogēniem retrovīrusiem līdzīgus struktūrelementus, kas norāda uz šo elementu senu izcelsmi.

Trīs pilna izmēra LTR TE-līdzīgus konstruktus identificēja BAC sekvencēs no *P.taeda* (ptTE101, ptTE121) un *P.glauca* (psTE283), izmantojot līdzību ar stresa pētījumos identificētiem diferenciāli ekspresētiem fragmentiem. TE datubāžu papildinājumi ļāva pārlicināties, ka divi no identificētajiem konstruktiem tiešām pieder skujkokiem specifiskajiem RE. psTE121 ir homologs *P.taeda PtCumberland\_I* (Wegrzyn, *Repbase*), bet psTE283 ir homologs *PtAppalachian\_I* (Wegrzyn, *Repbase*). Pēc struktūras, ptTE101 un ptTE121 pieder *Copia* LTR RE, bet psTE283 pieder *Gypsy* virsgimenei. Visu identificēto elementu domēni raksturojas ar izteiktu dažāda garuma fragmentu ‘mākoņu’ amplifikāciju visiem *P.sylvestris* DNS paraugiem, kas norāda uz lielu daudzumu līdzīgu genomisko kopiju esamību. Līdzīga amplifikācija ir novērota vairākām priežu sugām, kurām, iespējams, šo elementu homologu ir tikpat izplatīti. Domēnu amplifikācijas analīzes norāda, ka psTE283 *PtAppalachian\_I*-līdzīgais retrotranspozons ir atrodams pat tik attālām sugām, kā *Ceratozamia mexicana* un *Ginkgo biloba*, un skujkoku klases Himalaju ciedram un Kanādas hemlokeglei, bet nav tik izplatīts radniecīgajām egļu, balteglu, lapegļu ģintīm. Savukārt, ptTE101 ir augsti izplatīts 10 priežu sugām, pie kam divām priežu sugām (*P.contorta* un *P.siberica*) ar šo praimeru nebija amplifikācijas. Ārpus priežu ģints šis elements ir sastopams vienai eglei (*P.rubens*) un vienai lapeglei (*L.kaemferi*). ptTE121 ir sastopams tikai 7 priežu ģints pārstāvjiem, pārējiem izteikta amplifikācija nav novērota. Pētot šo elementu izplatību kailsēkļos secināts, ka ptTE101 un ptTE121 ir evolucionāri jaunāki par psTE283, un ir salīdzinoši nesen izplatījušies priežu ģints sugās. Rezultāti sakrīt ar līdzšinējiem TE izplatības pētījumiem, par tās nevienmērīgumu dažāda līmeņa taksoniem (Capy *et al.* 1994; Langdon *et al.* 2000; Friesen *et al.* 2001; Du *et al.* 2010; Kejnovsky *et al.* 2012).

#### **4.5. Retrotranspozonu transkripcionālā un transpozicionālā aktivitāte**

Šajā pētījumā ir analizētas salīdzinoši īsas (zem 10 kb) diferenciāli ekspresētas sekvences, tādēļ nevar secināt par visu identificēto homologu patstāvīgu pilna izmēra transkripciju. Potenciāli, var notikt fosilo vai nefunkcionālo elementu un to daļu vai arī TE-gēnu himēru ekspresija (Vignols *et al.* 1995; Betran *et al.* 2002; Wang *et al.* 2002; Rebollo *et al.* 2012). Tik īsu fragmentu piederība RE var tikt pierādīta tikai analizējot līdzību ar publicētām TE sekvencēm, bet arī līdzības detektēšana nevar sniegt atbildi par fragmenta izcelsmi. Skujkoku genomu raksturojas ar augstu kolinearitāti. Zinot, ka

vairāki parastās egles gēnu rajoni satur lielus intronus (Nystedt *et al.* 2013), bet viena autonoma elementa sekvenca var būt tikai dažus simtus bp gara (Kumar, Bennetzen 1999), nav iespējams pierādīt RE aktivēšanos pat pēc pilna izmēra sekvences iegūšanas no cDNS. Tomēr, šajā pētījumā no stresam pakļautas priedes cDNS paraugiem ieguva divu elementu (ptTE101 un ptTE121) domēnu sekvences, kā arī vienu pilna izmēra LTR elementu psTE283. Tas liecina, ka šo elementu transkripti ir atrodami stresam pakļautajai priedei un tie nav tikai fosilo elementu palieku transkripcijas rezultāts. Pat ja šie elementi atrodas intronos, tie var ievērojami ietekmēt gēnu transkripciju un splaisingu (Vignols *et al.* 1995; Marillonnet, Wessler 1997). Tāpat, izolētam elementam (psTE283) ir atrasti divi ORF, viens no kuriem kodē reversās transkriptāzes domēnu, bet otrs integrāzes domēnu. Aktīvi elementi raksturojas ar vienu nepārtrauktu poliproteīna transkriptu, bet nesen transponētie elementi - ar vienādiem LTR atkārtojumiem (Kumar, Bennetzen 1999). Daudzkopiju TE raksturs traucē konkrētā RE kopijas izolēšanu, tādēļ sekvenētam elementam neizdevās salīdzināt tā LTR, jo tiek transkribētas vairākas kopijas un sekvences savā starpā satur polimorfismus. Taču novērtējot ORF un transkripciju regulējošus motīvus, psTE283 visdrīzāk ir palikusi tikai transkripcionāla aktivitāte. Šajā pētījumā aprakstīto elementu transkripcija ir identificēta karstuma stresa apstākļos. psTE283 transkripti ir identificēti karstuma stresa ietekmē, bet tie tiek ekspresēti arī kukaiņu invāzijas rezultātā un 64 stundas pēc apstrādes ar ABA un SA hormoniem. ptTE101 homologi rajoni identificēti karstuma stresa ietekmē, arī tā ekspresija pierādīta tikai karstumam pakļautajiem paraugiem. PtTE121 transkripti ir atrodami arī pēc kukaiņu invāzijas un 24, 64 stundas pēc apstrādes ar SA. Transkripcija ir tikai pirmais posms RE transpozīcijā, aktīvajam elementam ne vienmēr jākodē funkcionālie proteīni. Pastāv virkne piemēru, kad neautonomais RE aktīvi pārvietojas ar citu elementu kodētiem proteīniem (Berg, Howe 1989; Flavell *et al.* 1994; Feschotte *et al.* 2002; Dewannieux *et al.* 2003; Kazazian 2004).

Iepriekšējos kailsēkļu RE pētījumos nav aprakstīti aktīvie TE. Tomēr, bioinformātiskas *P.taeda* BAC sekvenču analīzes identificēja dažus elementus ar konservatīviem LTR, kas var norādīt uz šo elementu nesenu transpozīciju (Kovach *et al.* 2010). Priedes klonālās variācijas analīžu rezultātā izdevies identificēt divus elementus, kas potenciāli ir nesen transponējušies, izraisot polimorfismu dažādās plantācijās augošu rametu starpā. IRAP metode ir lietota somaklonālās variācijas pētījumos (Asif, Othman 2005; Lightbourn *et al.* 2007; Bairu *et al.* 2011; Campbell *et al.* 2011; Bayram *et al.* 2012). IRAP amplificē genomisko rajonu starp diviem pietiekami tuvu esošiem elementiem, bet amplifikācija nenotiek, ja genomiskā rajona izmērs ir pārāk liels (Kalendar *et al.* 1999). Līdz ar jaunas TE kopijas integrēšanos, var parādīties jauns amplifikācijas fragments. Sekvenēšanas rezultātā ieguva divas atšķirīgas sekvenču grupas ar homologiju vienam hipotētiskam LTR *Copia* elementam no vienas puses, kas tehniski bija sagaidāms, jo dažādiem rametiem polimorfismu varēja izraisīt aktīvā elementa integrācija dažādās genoma vietās. Tāpat, divām individuālām sekvencēm identificēja homologiju ar DNS TE, kas arī varētu piedalīties somaklonālā mutāģenēzē priedes genomā. Nepieciešami turpmāki pētījumi šo elementu izolēšanai un raksturošanai.

Priedes dabīgās populācijas analīze ar nespecifiskiem RE marķieriem pēc IRAP metodes principa uzrādīja tikpat augstu daudzveidības pakāpi, kā SSR marķieru analīze tiem pašiem indivīdiem. Izvēlēta priedes populācija bija dabīgas izcelsmes un auga dažādos reljefa apstākļos, kas norāda uz regulārām mitruma režīma atšķirībām. IRAP

amplificēto fragmentu skaita analīze parādīja statistiski būtisku vidējā fragmentu skaita pieaugumu subpopulācijai, kas auga paugurā, salīdzinot ar pārpuļļojumā augošajiem indivīdiem. Galveno koordināšu analīze (PcoA) nodalīja paugura un pārpuļļojuma subpopulācijas, kamēr nogāzes subpopulācijas indivīdi grupējās vidū. Alēles bija vienmērīgi reprezentētas visās subpopulācijās un atšķirība frekvencē katrā no lokusiem nepārsniedza 0,3. Šīs analīzes rezultāti saskan ar detalizētu savvaļas miežu *BARE-1* RE izplatības pētījumu, kur sausuma apstākļos dienvidu kanjonā augošai populācijai pieauga pilna izmēra elementu skaits, bet retrotranspozonu marķieru datu analīze ar PCA testu arī nodalīja miežu populācijas (Kalendar *et al.* 2000). IRAP amplifikācijas joslu skaita izmaiņas līdz ar ģeogrāfisko izcelsmi ir aprakstītas arī citos pētījumos, piemēram, *Aegilops tauschii* pētījumā noteica augstu daudzveidības pakāpi visām pētītajām sugām, bet amplifikācijas fragmentu skaits bija palielināts ziemeļrietumu izcelsmes sugām (Saeidi *et al.* 2008). Precīzai RE kopiju skaita noteikšanai izmanto citas metodes, tomēr, šie rezultāti sniedz ieskatu RE nosacītā variācijas pakāpē un daudzveidībā dabīgā priedes populācijā.

#### 4.6. Kailsēkļu filoģenētiskā analīze

Kailsēkļu dzimtu filoģenētiskās saistības molekulārajā līmenī ir labi izpētītas izmantojot rbcL (Chase *et al.* 1993; Hasebe *et al.* 1992), 18S rRNS (Chaw *et al.* 1997), 28S rRNS sekvenču (Stefanovic *et al.* 1998), kodola un plastīdu mazo subvienību rRNS (Troitsky *et al.* 1991), mitohondriju gēnu sekvenču salīdzināšanu (Bowe *et al.* 2000; Ran *et al.* 2010). Skujkoki ir arī augsti monofilētiski (Raubeson, Jansen 1992; Bowe *et al.* 2000; Kan *et al.* 2007). Tomēr, dzimtas iekšienē filoģenētiskās attiecības līdz galam nav noskaidrotas, piemēram, nav zināms kura ģints ir senākā (Raubeson, Jansen 1992; Bowe *et al.* 2000; Donoghue, Doyle 2000). Zināms, ka priedes, egles, *Cathaya*, duglāzijas un lapegles ģintis nodalās no balteglu, hemlokeglu, *Keteleeria*, *Nothotsuga*, *Pseudolarix* ģintīm, bet ciedru ģints veido atsevišķu atzarojumu (Wang *et al.* 2000). ITS struktūras, SSR pētījums nodalīja lapegļu un duglāziju klāsteri no priežu, egļu un *Cathaya* (Kan *et al.* 2007). Šajā pētījumā dažas no lapeglēm klāsterizējās savā starpā un ar balteglēm, taču neveidojās saistības ar duglāzijām, kuras bija pārstāvētas tikai ar vienu sugu.

Uz 18 kodola gēnu sekvencēm balstīts pētījums 10 priežu sugām (Palmé *et al.* 2009), sīkāk nodalīja *Pinus* un *Strobus* priežu apakšģintis. Saskaņā ar Palmé *et al.* (2009), filoģenētisko analīžu dendrogramās priežu *Pinus* apakšģintī, *Conkorta* nodalījuma *P.contorta* un *P.banksiana* klāsterizējas kopā, tāpat *P.sylvestris* un *P.nigra* veido klāsteri. Mūsu pētījumā *Strobus* apakšģints *Cembrae* nodalījums ir pārstāvēts ar *P.cembra*, *P.pumila*, *P.koraiensis*, *P.sibirica*, kas klāsterizējas kopā gan specifisko RE-līdzīgo marķieru datu dendrogramā, gan nespecifisko IRAP praimeru dendrogramā, kur klāsterim pievienojas arī *Strobus* nodalījuma *P.strobus* un *P.peuce*. Specifiskie TE-līdzīgo marķieru dati izšķīra tikai priedes ģinti, kas skaidrojams ar priedes marķieru specifiskumu un datu nelielo apjomu. IRAP nespecifiskie marķieri atsevišķos klāsteros nodalīja arī egles un baltegles, lai gan klāsteru varbūtības vērtības nebija augstas. Tas lielā mērā ir noteikts ar marķieru nespecifiskumu un nelielu daudzumu. Tāpat, IRAP marķieru konstruēšanā izmantoti hipotētiskiem LTR komplementāri praimeru, kas var neatbilst reālo un izplatītāko priežu RE LTR sekvencēm. Taču rezultāti rāda, ka uz RE sekvenču pamata veidotiem marķieriem kailsēkļu filoģenētiskajos pētījumos ir

potenciāls, jo metode atspoguļo nekodējošas genoma daļas polimorfismu un var sniegt papildus informāciju par taksonu filoģenētiskajām saistībām.

## SECINĀJUMI

1. Ar nespecifisko iPBS metodi ir iespējams identificēt retrotranspozonu transkriptus, kā arī tos saturošus himēriskus transkriptus, bez iepriekšējās informācijas par sugai specifiskām retrotranspozonu sekvencēm.
2. Retrotranspozoniem-līdzīgas sekvenču tiek diferenciāli ekspresētas *Pinus sylvestris* genomā karstuma, priežu hermesa invāzijas un hormonu apstrādes ietekmē.
3. Starp segsēkļiem un kailsēkļiem augsti konservatīvi retrotranspozoniem-līdzīgie transkripti nenorāda uz homologa retrotranspozona transkripciju *P.sylvestris* genomā. Uz skujkoku transkripcionāli aktīviem retrotranspozoniem norāda transkripti, kas ir vāji līdzīgi segsēkļu retrotranspozoniem.
4. Ir identificēti hloroplastam homologu transkripti (SA07, ABA03, ABA02, B153, SA01) ar izcelsmi no kodola genoma hloroplasta sekvenču insercijām. Hloroplastam homologām sekvencēm ir noteikts augsts ekspresijas līmeņa pieaugums priežu hermesa invāzijas rezultātā, kā arī karstuma stresa ietekmē. Ir identificēts himērisks transkripts B153 saturošs gan hloroplasta, gan retrotranspozona sekvenču.
5. Ir identificēts pilna izmēra elements psTE283, kurš tiek transkribēts karstuma stresa ietekmē, pēc kukaiņu invāzijas un pēc apstrādes ar ABA un SA. psTE283 pieder LTR *Gypsy* retrotranspozoniem un ir homologs *P.taeda PtAppalachian\_I* elementam. psTE283 pilna izmēra elementa transkripti nāk no dažādām vietām *P.sylvestris* genomā, bet tā hipotētiskais ORF satur mutācijas, kas norāda uz ierobežotām transpozīcijas iespējām šim elementam. Uz psTE283 elementa senu evolucionāru izcelsmi norāda tā izplatība attāliem kailsēkļu taksoniem, tādiem kā *Cycadales* un *Gingkoales* klašu pārstāvjiem.
6. Identificēti ptTE101 un ptTE121 homologu pilna izmēra transkripti arī tiek ekspresēti *P.sylvestris* karstumam pakļautiem stādiem un pēc struktūras pieder *Copia* LTR retrotranspozoniem. PtTE121 tiek ekspresēts arī priežu hermesa invāzijas un SA apstrādes ietekmē. Šie elementi ir īpaši izplatīti priežu ģints pārstāvjiem, kas norāda uz to neseno izcelsmi.
7. Tika konstruēti deviņi IRAP marķieri, kas *P.sylvestris* genomā atklāj augstu ģenētisko daudzveidību vienas populācijas iekšienē un spēj nodalīt indivīdus augošus atšķirīgos mitruma režīma apstākļos. Atšķirības amplifikācijā starp dažādām grupām ir būtiskas šajā pētījumā izstrādātajiem marķieriem IR\_2, IR\_5, IR\_6, IR\_7.
8. Viena klona dažādu indivīdu IRAP analīze norāda uz nesenu transpozīcijas gadījumu un aktīvo elementu esamību *P.sylvestris* genomā. Polimorfo

fragmentu sekvenēšana identificēja *Copia-17-Pab-I* LTR retrotranspozona daļas.

9. Izmantojot *P.sylvestris* specifiskus retrotranspozonu marķierus un IRAP marķierus ir iespējams atklāt jaunas filoģenētiskās saistības starp kailsēkļu taksoniem.

## PATEICĪBAS

Esmu pateicīga darba vadītājam Dr. biol. Dainim Ruņģim par līdzdalību darbā gūto rezultātu izprašanā un vērtīgiem padomiem darba izstrādāšanas gaitā.

Izsaku pateicību LVMI „Silava” Ģenētisko resursu centra darbiniekiem un īpaši Dr.chem. Ilzei Veinbergai par sniegto atbalstu un iedrošināšanu izstrādāt šo pētījumu, Dr. biol. Agnesei Gailītei, Mg.math. Anitai Gailei, Dr. silv. Vilnim Šķiparam, Mg. biol. Viktorijai Beļēvičai par sniegtajiem padomiem un praktisku palīdzību eksperimentu izstrādāšanas un darba uzrakstīšanas gaitā.

Pateicos LVMI „Silava” Meža selekcijas grupas darbiniekiem un īpaši Dr. silv. Imantam Baumanim par priežu rametu materiāla nodrošinājumu. Esmu pateicīga Dr. silv. Ārim Jansonam par iedziļināšanos pētījuma problemātikā un līdzdalību nepieciešamo priežu audžu identificēšanā un eksperimentālo paraugu iegūšanā. Pateicos Mg. biol. Mārtiņam Zepam par profesionālo palīdzību priežu uzmērīšanā.

Pateicos Dr. biol. Nilam Rostokam un viņa darba grupai par ātru un efektīvu palīdzību klonēšanas un transformēšanas metodikas apgūšanā, kā arī par sniegtajiem padomiem darba izstrādāšanas gaitā.

Pateicos Helsinku Universitātes profesoram un Augu genomikas laboratorijas vadītājam Dr. Alanam Šulmanim un viņa darba grupai par iedrošināšanu izstrādāt šo darbu un sniegtajām zināšanām par augu retrotranspozoniem un to analizēšanas iespējām. Pateicos docentam Ruslanam Kalendaram par atļauju izmantot viņa konstruētos nespecifiskos PBS praimerus pirms to publicēšanas.

Izsaku pateicību Nacionālā Botāniskā dārza Dendrofloras nodaļas vadītājai Lindai Strodei un darbiniecei Inesei Graudiņai par palīdzību skujkoku sugu noteikšanā. Pateicos Latvijas Universitātes Botāniska dārza personālam par palīdzību eksotisko kailsēkļu sugu noteikšanā.

Pētījums no 2009-2012. gadam ir finansēts ESF projekta „Ģenētisko faktoru nozīme adaptētās spējīgu un pēc koksnes īpašībām kvalitatīvu mežaudžu izveidē” (Nr. 2009/0200/1DP/1.1.1.2.0/09/APIA/VIAA/146) ietvaros.





University of Latvia  
Faculty of Biology

Angelika Voronova

RETROTRANSPOSON STRUCTURE AND EXPRESSION IN THE  
SCOTS PINE (*PINUS SYLVESTRIS* L.) GENOME

Summary of Doctoral Thesis  
Submitted for the degree of Doctor of Biology  
Subfield of Molecular Biology

Riga, 2014



The doctoral thesis was carried out in Latvian State Forest Research Institute „Silava” from 2009 to 2014.



Years 2009-2012 the research was supported by funds from the European Social Fund's Project “Importance of Genetic Factors in Formation of Forest Stands with High Adaptability and Qualitative Wood Properties” (Nr. 2009/0200/1DP/1.1.1.2.0/09/APIA/VIAA/146).

The thesis contains the introduction, 7 chapters, reference list and 2 appendices.  
Form of the thesis: dissertation in biology, molecular biology.

Supervisors:

- 1) Dainis Ruņģis, Dr.biol., leading researcher, LSFRI „Silava”
- 2) Nils Rostoks, Dr.biol., leading researcher, University of Latvia

Reviewers:

- 1) Īzaks Rašals, Dr.hab.biol., professor, University of Latvia.
- 2) Gunārs Lācis, PhD, leading researcher, Latvia State Institute of Fruit-Growing.
- 3) Agnese Kolodinska Brantestam, PhD, Senior Scientist, Nordic Genetic Resource Center, Sweden

The thesis will be defended at the public session of the Doctoral Committee of Biology, University of Latvia, at 10:00 on February 4, 2014, at the Faculty of Biology (Kronvalda bulv. 4, Riga), auditorium No. 2.

The thesis is available at the Library of the University of Latvia, Riga, Raiņa bulv. 19-203.

This thesis is accepted for the commencement of the degree of Doctor of Biology on December 19, 2013 by the Doctoral Committee of Biology, University of Latvia.

Chairman of the Doctoral Committee \_\_\_\_\_ / Prof, Dr.hab biol. Ģederts Ieviņš /

Secretary of the Doctoral Committee \_\_\_\_\_ /Daina Eze/

© University of Latvia, 2014

© *Angelika Voronova*, 2014



## SUMMARY

The research contained in this doctoral thesis was carried out in the Latvian State Forest Research Institute "Silava". The study aimed to identify active retrotransposons (RE) in the Scots pine (*Pinus sylvestris* L.) genome and to characterize transcription of RE like sequences in response to various controlled stress conditions. In the study non-specific primers complementary to conservative RE tRNA binding sites were used, that enable investigation of REs without previous sequence knowledge. In a novel utilisation, these markers were applied to the analysis of expressed sequences in ramets of one Scots pine clone, which minimised genotypic variation between different individuals. Pine ramets were subjected to naturally occurring heat stress, insect infestation (pine woolly aphid (*Pineus pini* L.)) as well as to plant hormone treatment (abscisic acid and salicylic acid). Differentially expressed fragments were sequenced, analysed and compared to various sequence databases. The results are consistent with other published studies on mobile genetic elements and related transcribed sequences as well as with gymnosperm structural genomic studies.

The majority of the identified differentially expressed fragments were Class I transposable elements, or LTR RE sequences (LTR *Gypsy* and *Copia*), as well as chimeric transcripts, which indicates that RE sequences are common in the Scots pine genome and RE-like transcripts are expressed in response to stress. Homology of detected sequences with previously described transposable elements from other plant species was established. Representatives of various transposable element families were identified after differing stress conditions, but only a few common fragments were identified between the experiments. Fragments differed from each other in structure (TE pol, LTR, LTR-LTR, TE pol-chloroplast) and in domain representation (AP, RT, GAG). Identified RE-like fragment expression levels were determined by real-time PCR. A significant increase in expression was observed for chimeric fragments containing chloroplast-like sequences, which can be explained by the homology of parts of the sequence to the chloroplast genome or/and representation of highly similar regions of homology in multiple copies in the *P.sylvestris* genome. Four differentially expressed transcript groups containing both TE homologous sequences and chloroplast-like sequences were identified. For one of these transcripts (B153TE), expression of the *Cereba* RE-like part increased 45-73 fold after insect infestation and 5.62 fold after heat stress when compared to normal conditions. Expression of the chloroplast-like part of the B153 fragment increased by 6888-9373 fold after insect attack and 255 fold after heat stress. Many sequence polymorphisms were observed in several identified transcripts of this chimeric fragment, and in particular in the RE-like sequences, suggesting that besides the chloroplast genome, the chloroplast-like locus is present in multiple copies in the nuclear genome of Scots pine. These types of transcripts could take part in regulatory processes involving RNA interference mechanisms.

A full-length 5796 bp element with structural elements characteristic of *Gypsy* LTR REs was sequenced and characterized from cDNA (psTE283). Utilising hypothetical RE sequences identified after similarity searches between *P.taeda*

BAC sequences and sequenced RE-like sequences, internal domains of these hypothetical RE were sequenced from cDNA isolated from stressed *P.sylvestris* clones, indicating that putatively functional RE were expressed in addition to chimeric or disrupted RE-like sequences. Subsequent research is necessary to establish whether sequences identified in this study are localized in gene regions (introns or UTR), or if they are autonomous transcribed elements.

Using the identified RE sequences, specific *P. sylvestris* RE markers were developed and utilised in several differing analyses. Using the IRAP technique individuals from a naturally regenerated population were genotyped. This population could be differentiated into sub-populations based on long-term differential growth conditions. High genetic diversity was identified among individuals, which is comparable to data obtained from neutral SSR markers. Principal coordinate analysis of the IRAP data differentially clustered individuals from subpopulations growing in different topographic conditions. Pine trees growing on the hill have a statistically significant increase in mean amplified fragment number when compared to the trees growing in the swamp. In a second experiment utilising the IRAP technique, 26-year-old ramets of four pine clones, growing in two separate plantations in the east and west of Latvia, were analysed. One of the nine analysed IRAP markers identified somaclonal polymorphism among ramets of one clone, which indicates that the *P. sylvestris* genome could contain transpositionally active elements. Polymorphic PCR fragments were sequenced and two separate sequence groups were homologous to a *Copia-17-PAb-I* LTR previously not described in *P. sylvestris*. This suggests evidence of a recent transposition event in these ramets, and further studies are necessary for isolation and description of the full-length element. The distribution of identified RE-like fragments in various conifer species (*Pinales*) was analysed, including 16 pine species, 14 spruce, 7 firs as well as representatives of other more distant gymnosperm taxa. RE markers were used for phylogenetic analysis of these gymnosperm species. Both specific RE amplification as well as non-specific RE primer amplification using the IRAP technique were able to distinguish the pine subgenus *Pinus* from *Strobus*. IRAP data was also able to distinguish clusters of spruces, pines and firs, but bootstrap values were low.

## CONTENTS

SUMMARY .....	41
CONTENTS .....	43
GENERAL DESCRIPTION OF THE WORK.....	44
1. THEORETICAL BASIS .....	47
2. MATERIALS AND METHODS .....	51
3. MAIN RESULTS .....	55
4. DISCUSSION .....	69
CONCLUSIONS .....	76
ACKNOWLEDGMENTS.....	77
IZMANTOTĀS LITERATŪRAS SARAKSTS/ REFERENCES .....	78

## GENERAL DESCRIPTION OF THE WORK

### *Actuality of the themes and research*

For many years after the discovery of TEs, they were regarded as either as non-functional or parasitic sequences (Orgel, Crick 1980; Doolittle, Sapienza 1980; Kidwell, Lisch 2001). However, with the sequencing of genomes, and more detailed research, this view has changed. Now it is known that TE sequences are widely distributed in all eukaryotic, bacterial and archeal genomes and in contrast to more highly conserved genes, vary both in number and composition (Kazazian 2004, 2011). TEs are involved in gene regulation, chromosome structure and genome evolution (Wessler 1996; Kumar, Bennetzen 1999; Kazazian 2004, Madlung, Comai 2004; Kazazian 2011; Rebollo 2012). Plant genomes contain a significantly higher proportion of TEs, and in the case of some lily species, can comprise of up to 90% of the genome (Feschotte *et al.* 2002). REs are the most widespread TE group in plant genomes, and a number of marker techniques have been developed based on RE sequences. In plant genomes, REs are common, widely distributed over all chromosomes, insertions are non-revertant, and they are highly polymorphic. The study of REs is both of theoretical and practical significance.

In general, the transpositional activity of TEs is suppressed via chromatic modification, and functional members of both classes of mobile elements are activated under stress conditions or at specific developmental stages or in specific tissue types. Transposition of mobile elements is one of the most significant factors in genome instability, initiating genomic restructuring and the formation of novel genotypes, enabling the adaptation to fluctuating environmental conditions. TEs can influence the function of genes either directly by insertional mutagenesis, or by changing their regulation. In addition, TEs can influence regulatory processes via RNA interference mechanisms by producing various RNA molecules. Currently it is not fully understood to what extent TE-induced mutagenesis influences plants during stress conditions and developmental stages. All members of TE families have been identified in the pine genome, however active TEs have not been investigated. In comparison to the majority of eukaryotes, conifer genomes are large and contain a large proportion of repetitive sequences. Therefore the sequence of the first conifer genome was only published in 2013 (Nystedt *et al.* 2013). Conifer genomes are characterised by large gene families and pseudogenes, and contain large inter-gene regions and introns. Therefore, investigation of transcription of REs in *P. sylvestris* could offer insights into reorganisation of complex genomes under stress conditions as well as RE-associated regulation mechanisms.

### *Aims and objectives*

The aim of this doctoral work was to characterise active retrotransposons in the Scots pine (*Pinus sylvestris* L.) genome and to quantify the expression of RE-like sequences under various stress conditions. The objectives were:

1. Identification and characterisation of RE-like transcripts induced by heat stress, pine woolly aphid (*Pineus pini* L.) infestation and after application of ABA and SA.
2. Comparison of induced fragments after various stresses, and determination of their expression levels

3. Identification of expressed fragments as chimeric fragments or full-length REs, and sequencing of full-length REs to determine potential transcription and transposition events.
4. Development of RE-based markers and utilisation of these to analyse natural pine populations as well as to investigate possible somaclonal variation between ramets of pine clones.
5. Characterisation of RE distribution and RE –derived diversity in gymnosperm genomes.

### ***Hypotheses***

1. The *P. sylvestris* genome contains transcriptionally and transpositionally active retrotransposons
2. Transcriptionally active retrotransposons and related sequences are differentially expressed after biotic and abiotic stresses
3. Transposition events occur as a result of environmental stress conditions as well as possible expansion of elements
4. Retrotransposons identified in the *P.sylvestris* genome are unevenly distributed in gymnosperm genomes

### ***Overview of methods***

SSR markers were utilised to confirm clonal identity of pine ramets. The iPBS technique, utilising primers complementary to conservative RE tRNA binding sites (Kalendar *et al.* 2010), was used to identify differentially expressed RE-like fragments. Differentially expressed fragments were ligated into plasmids, transformed into *E.coli* cells (Inoue *et al.* 1990), and Sanger sequenced. The obtained sequences were compared to nucleotide databases as well as specific TE databases following the scheme described in Wicker *et al.* (2007). Primer-BLAST was used for primer construction (Ye *et al.* 2012). IRAP markers developed as described in Kalendar *et al.* (1999) were used for analysis of natural pine populations and pine clone ramets. Genotypic data were analysed using *GenAlEx* v. 6 (Peakall, Smouse 2012) and FAMD v.1.30 (Schlüter, Harris 2006). RE-derived and IRAP markers, developed during this study, were utilised for phylogenetic analysis of gymnosperm species.

### ***Results presented at international scientific conferences:***

1. 19.-22.10.2012. Voronova A, Jansons Ā, Ruņģis D. Identification of retrotransposon-like sequences in *Pinus sylvestris* L. activated in response to stress conditions. 5th Baltic Congress of Genetics, Kaunas, Lithuania.
2. 16.-18.10.2012. Voronova A, Jansons Ā, Ruņģis D. Activation of retrotransposon-like sequences in *Pinus sylvestris* in response to stress conditions. NovelTree Conference, Helsinki, Vantaa, Finland.
3. 3.-5.10.2012. Voronova A, Jansons Ā, Ruņģis D. Retrotransposon-like sequences expressed under various stressors in pine genome shows different similarity level to known mobile genetic elements. AdaptCar conference: Genetic aspects of adaptation and Mitigation: forest health, wood quality and biomass production, p.24, Riga, Latvia.
4. 08.02.2012-11.02.2012. Voronova A, Jansons Ā, Ruņģis D. Expression of retrotransposon-like sequences in Scots pine (*Pinus sylvestris* L) clones

- under stress conditions. Molecular Mapping & Marker Assisted Selection, Vīne, Austrija. VIPCA Conference Abstract Book app.1
5. 04.-07.02.2012. Voronova A, Jansons Ā, Ruņģis D. Expression of retrotransposon-like sequences in Scots pine (*Pinus sylvestris* L.) clones under stress conditions. Molecular Ecology, Vīne, Austrija. VIPCA Conference Abstract Book p.41.
  6. 22.-24.04.2009. Žuka A, Ruņģis D, Schulman AH, Kalendar R. Retrotransposon variability in Scots Pine (*Pinus sylvestris* L.) genome. 5th International Conference Research and Conservation of Biological Diversity in Baltic Region, Daugavpils, Latvia. Book of Abstracts, Daugavpils University Academic Press Saule, p.156.
  7. 19.-21.03.2009. Zhuk A, Rungis D, Kalendar R, Schulman AH. Retrotransposon variability in Scots pine (*Pinus sylvestris* L.) genome. FEBS Workshop, Adaption Potential in Plants, Vīne, Austrija.

**Local conferences:**

1. 13.02.2013. Voronova A, Beļēviča V, Jansons Ā, Ruņģis D. Stresa apstākļu inducējamu retrotranspozoniem līdzīgo sekvenču transkripcionālā aktivēšanās priedes genomā. Latvijas Universitātes 71. Konference. Molekulārās bioloģijas sekcija, Rīga, Latvija.
2. 01.02.2012. Voronova A, Jansons Ā, Ruņģis D. Mobilēm ģenētiskiem elementiem līdzīgo sekvenču izplatība parastās priedes (*Pinus sylvestris* L.) genomā. Latvijas Universitātes 70. Konference. Augu bioloģijas sekcija, Rīga, Latvija.
3. 24.-27.10.2011. Ruņģis D, Šķipars V, Voronova A, Krivmane B, Gaile I, Beļēviča V, Korica A, Kanberga-Siliņa K, Veinberga I. Apvienotajais Pasaules Latviešu Zinātnieku 3. Kongress. Molekulārās ģenētikas metodes Latvijas mežzinātnē. Jelgava, Latvija.
4. 04.02.2010. Voronova A, Ruņģis D. Mobilēm ģenētiskiem elementiem līdzīgo sekvenču ekspresija stresoru ietekmē parastās priedes (*Pinus sylvestris* L.) genomā. Latvijas Universitātes 68. Konference. Augu bioloģijas sekcija, Rīga, Latvija.
5. 30.11.2011.-1.11.2011. Voronova A, Jansons Ā, Ruņģis D. Genome plasticity of Scots pine (*Pinus sylvestris* L.) under different stress conditions. Genetic variation of adaptive traits, Salaspils, Latvija.
6. 4.02.2009. Žuka A. Retrotranspozonu variācija parastās priedes (*Pinus sylvestris* L.) genomā. Latvijas Universitātes 67. Konference. Augu bioloģijas sekcija, Rīga, Latvija.

**Publications:**

1. Voronova A, Belevich V, Jansons A, Rungis D (2013) Stress induced transcriptional activation of retrotransposon-like sequences in the Scots pine (*Pinus sylvestris* L.) genome. (29.05.2013- submitted to journal "Tree Genetics&Genomes"; 12.11.2013- accepted with revisions).
2. Voronova A. Rungis D. (2013) Development and characterisation of IRAP markers from expressed retrotransposon-like sequences in *Pinus sylvestris* L. (Accepted for publishing in Proceedings of the Latvian Academy of Sciences. Section B.).

3. A.Voronova, Ā.Jansons, D.Ruņģis (2011) Expression of retrotransposon-like sequences in Scots pine (*Pinus sylvestris*) in response to heat stress. *Environmental and Experimental Biology* 9: 121-127. (Published).
4. Sequences of three retrotransposons were submitted to “*Rebase*” largest repeat database: *PsAppalachian-I*, *Silava-Pta-I*, *Copia-17-Pta-I*. (Published).
5. 126 sequences were submitted to NCBI EST database with gene bank numbers: JZ389987-JZ390112. (Published).

Years 2009-2012 this study was supported by the ESF project “Importance of Genetic Factors in Formation of Forest Stands with High Adaptability and Qualitative Wood Properties” (Nr. 2009/0200/1DP/1.1.1.2.0/09/APIA/VIAA/146).

### ***Length and structure of thesis:***

This thesis consists of 202 pages, containing 20 tables, 56 figures, 2 appendices. The thesis consists of an introduction and 7 sections (literature review, materials and methods, results, discussion, conclusions, hypotheses and acknowledgements). The thesis has 9 conclusions, and 446 references.

## **1. THEORETICAL BASIS**

### ***1.1. Definition and classification of mobile genetic elements***

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. Most TEs have lost the ability to transpose, and only a few species-specific elements are active. These elements can be activated in stress conditions or in specific tissue types or developmental stages (McClintock 1984; Wessler 1996; Grandbastien 1997; Kumar, 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 and class II or DNA transposons (Finnegan 1989; Curcio, Derbyshire 2003; Capy 2005; Wicker *et al.* 2007). Retrotransposons (REs) are similar to retroviruses in structure and function; however RE do not have the ability to infect other cells. REs transpose via an RNA intermediate, which is then reverse transcribed to DNA prior to integration at a new genomic location. Therefore, RE transposition entails an increase in copy number of the RE. REs are divided into orders, super-families and families (Wicker *et al.* 2007). REs containing long terminal repeats are classified as LTR REs, elements without these repeats are classified as autonomous LINEs or non-autonomous SINEs. REs are also classified as *Pseudoviridae* (*Ty1*, *Copia* type) and *Metaviridae* (*Ty3*, *Gypsy* type) virus families, separately from other viruses (Boeke *et al.* 2006).

### ***1.2 Structure of LTR REs***

Long Terminal Repeats (LTRs) are non-coding direct repeats that contain sequences regulating transcription. The length of LTRs can range from 85bp in length (*FRetro129* from rice (Gao *et al.* 2012)) up to 5kb (*Ogre* from pea (Neumann *et al.* 2003)). The LTR sequences are specific to each RE family but contain the following conserved sequences: approximately 7bp long inverted repeats (IRs) with highly conserved 5' TG and 3' CA nucleotides (int binding sites) and a TATATA motif, which functions as a polymerase II promoter in plants (Vicent *et al.* 2005; Benachenhou *et al.* 2013). Often, LTRs also contain

polyadenylation signals. The U3 region is variable between RE families and contains specific regulatory sequences that ensure expression in particular tissues, conditions or developmental stages. Transcription of RE is initiated in the 5' LTR, and concludes in the 3' LTR (Kumar, Bennetzen, 1999). As the 5' and 3' LTRs are identical in sequences and orientation, a 3' LTR can initiate transcription of adjacent sequences. This was shown for proviruses, where the promoter activity of 3' LTRs is usually suppressed, but can be activated if the 5' LTR regulatory functions are disrupted (Boerkoel, Kung 1992; Gama Sosa *et al.* 1994; Klaver, Berkhout 1994). Autonomous REs usually contain two genes: the *gag* gene encodes a capsid-like protein, and the *pol* gene encodes a polyprotein, which is subsequently cleaved by a self-encoded protease (*pr*) to form a reverse transcriptase (RT) with RNase H activity, and an integrase (*int*). Env-like domains are less-commonly found in REs (Eickbush, Malik 2002). REs can also contain additional ORFs with unknown function, e.g. the rice RE *RIRE2* (Kumekawa *et al.* 1999), or they can contain gene fragments, e.g. the maize RE *Bs1* contains a fragment of an ATPase gene (Jin, Bennetzen 1994). In the cytoplasm, mRNA transcripts are both translated to proteins and reverse transcribed to form double stranded cDNA. Reverse transcription is primed by various tRNA molecules, which bind to the PBS region (Mak, Kleiman 1997), which is highly conserved among RE families (Kalendar *et al.* 2010).

### **1.3 Distribution of TEs in plant genomes**

Both classes of TEs are found in plant genomes, with the LTR REs the most common (Kumar, Bennetzen 1999). RE sequences and distribution vary between species, and they can comprise from 15% (*Arabidopsis thaliana*) up to 90% (*Liliacea*) of the entire genome size (Sabot, Schulman 2006; Morgante *et al.* 2007). The genomes of higher plants are characterised by gene islands in a sea of repetitive sequences (SanMiguel *et al.* 1996; Panstruga *et al.* 1998; Shirasu *et al.* 2000; Rostoks *et al.* 2002; Kejnovsky *et al.* 2012). Comparative genome sequencing of cereals reveals high levels of polymorphism in intergenic and coding regions, and the genomic sequence of one individual does not reflect the range of polymorphism found in populations. The regions of polymorphism include TEs, pseudo-genes and coding genes. Recent studies have shown that all higher plants have a similar genome organisation (Brunner *et al.* 2005; Piegu *et al.* 2006; Hawkins *et al.* 2006; Neumann *et al.* 2006), and that REs can form polymorphic regions between blocks of more highly conserved genes (Brunner *et al.* 2005; Morgante *et al.* 2007).

### **1.4 Clustering and polymorphism of REs**

Often, LTR REs are found in alternate conformations in plant genomes. In the well-studied barley RE *BARE-1*, the frequency of single LTRs is 7-42 times higher than full-length elements, and the haploid genome can contain  $13.7 \times 10^3$  LTR copies (Vicent *et al.* 1999). Two internal domains flanked by a single recombinant LTR was found in  $4.6 \times 10^3$  copies in the haploid genome (Vicent *et al.* 2005). Template switching by reverse transcriptases (RTs) can form new recombinant elements (Sabot, Schulman 2007), and the fidelity of RTs is lower than of other polymerases (Preston 1996; Gabriel, Mules 1999). Larger mutations and rearrangements can be caused by unequal recombination, which is facilitated by



the high copy number of REs (Vicient *et al.* 2005; Hawkins *et al.* 2009). LTR-LTR non-homologous recombination can result in single LTRs, multiple sequential LTRs as well as repeats of internal domains (Mieczkowski *et al.* 2006). In addition, often REs preferentially insert into regions already containing REs, forming clusters of disrupted and inactive REs (Kumar, Bennetzen 1999).

### **1.5 Genome size and TE expansion**

The accumulation of TEs in plant genomes may reflect selection for larger genome sizes (Murray 1998, 2005). Large plant genomes contain many LTR RE families, which can contain over 10000 copies in haploid genomes. While all main TE groups are found in plant genomes, the proportion and representation of TEs can vary (Kejnovsky *et al.* 2012). Copy number of the *BARE-1* element is correlated with genome size in barley (Vicient *et al.* 1999), and proliferation of *BARE-1* can be initiated by water deficit (Kalendar *et al.* 2000). Genome size variations within species are also correlated with geographic distribution (Kalendar *et al.* 2000). Possibly, in prolonged stress conditions, the deletion of TEs is reduced leading to increase of genome size (Nevo 2001). The high copy number of the *BARE-1* element possibly reflects adaptive selection for larger genomes with increased RE activity (Kalendar *et al.* 2000; Nevo 2001; Wessler 2001). RE transposition can cause new functional variations, and thus increase fitness in stress conditions (Madlung, Comai 2004).

### **1.6 Activation of REs in stress conditions**

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). After heat stress in yeast, the *TF1* element is activated, which enhances the expression of stress response genes (Feng *et al.* 2013). Variation of the 3' UTR sequence of the tobacco *Tnt1* RE determines differential expression in response to stress-associated signal molecules (Beguiristain *et al.* 2001). The *Arabidopsis* RE *Olsen* is expressed in the offspring of heat stressed plants (Ito *et al.* 2011). In contrast, the expression of REs *Tam3* in the *Antirrhinum majus* genome (Hashida *et al.* 2003) and *Tcs1* in the orange genome (Butelli *et al.* 2012) increase at lower temperatures. Expression of the *Ty1 Copia TLC1* family in tomato increases in response to injury, protoplast formation and high salt concentration, and regulatory motifs in the U3 region of the *TLC1.1* RE are ethylene sensitive (Tapia *et al.* 2005). The *Copia* RE *Remel* in melon is activated in UV irradiated leaves (Ramallo *et al.* 2008). The non-LTR TE *Karma* is expressed in regenerated plant tissues and cultivated cells (Komatsu *et al.* 2003). Biotic stress also activates a range of REs: *Oare-1* – virus infection in potatoes; *Tto1* – tobacco mosaic virus (Hirochika 1993); *Tnt1* – tobacco fungal infection (Melayah *et al.* 2001). In oak (*Quercus suber*) *Corky* is expressed in various developmental stages, as well as after injury (Rocheta *et al.* 2012). The *Psr* family in pea is activated in response to protoplast formation and fungal elicitors (Kato *et al.* 1999). Cell wall hydrolase activates tobacco *Tnt1* (Pouteau *et al.* 1991). Expression of RE often is tissues specific, for example, *Tnt1* (Grandbastien *et al.* 1994) and *TLC1* (Tapia *et al.* 2005) expression is observed in roots; *BARE-1* is expressed in leaves (Suoniemi *et al.*

1996) and the wheat RE *PREM-2* is expressed in microspores (Turcich *et al.* 1996).

### **1.7 Regulatory significance of REs**

Insertion of REs can influence gene expression in various ways (Rebollo *et al.* 2012). RE LTRs can contain promoters, transcription initiation sites, termination sites, insulators and splicing regulators. Transposition of REs can induce mutations. After virus infection in maize, a *Bs1* RE insertion was found in the *Adh1* gene (Johns *et al.* 1985). Similarly, *Tnt1* was found in the tobacco nitrate reductase gene (Grandbastien *et al.* 1989) and *Tos17* inserts in the rice *PHYA* gene (Takano *et al.* 2001). A *B5*-like RE insertion in the lignin biosynthesis *COMT* gene in maize results in a shorter, chimeric mRNA (Vignols *et al.* 1995). In the maize *waxy* gene, RE insertions in introns lead to formation of alternative splicing variants (Varagona *et al.* 1992). Many gene promoters and alternative promoters are derived from REs (Rebollo *et al.* 2012). Even single LTRs can influence adjacent gene expression, as in the case of *Copia* LTR RE *Tsc1* and *Tsc2* insertions in before the *Myb* transcription factor in the orange genome (Butteli *et al.* 2012). RE insertion can also down-regulate gene expression e.g. the *Gypsy* RE *Gret1* in the grape (*Vitis vinifera*) genome (Kobayashi *et al.* 2004). The pea genes *PSPAL2* and *PCHS1* are activated in response to pathogens, and both these genes contain 5' *Copia* RE *PsrC* insertions (Kato *et al.* 1999). Differentially expressed chimeric transcripts in amphiploid wheat contain the *Wis 2-1A* LTR RE, usually located in the 3' UTR. In genes with increased expression, the *Wis 2-1A* LTR was in the same orientation as the adjacent gene. The expression of approximately 208 genes in the wheat genome may be influenced by the *Wis* RE (Kashkush *et al.* 2002).

REs may be indirectly involved in regulatory processes via long non-coding RNA molecules of unknown function, or microRNAs which are involved in RNA interference. The human genome contains 55 miRNA genes derived from REs (Piriyapongsa *et al.* 2007). Large proportions of genomes have been shown to be transcribed, forming non-coding RNA molecules with unknown function (Mattick, Makunin 2006; Zhang *et al.* 2012; Ponting *et al.* 2009). Expression levels are not as high as for protein-coding genes; however these transcripts may influence quantitative trait variation, developmental processes and complex diseases. Due to their high copy number and repetitive nature, usually RE sequences are excluded from transcriptome analysis (Mourier, Willerslev 2009).

LTR clusters are involved in heterochromatin formation, which is a dynamic and important process during various stages of plant development (Finnegan *et al.* 1998). Many plant centromeres consist of satellite repeats and specific *Ty3 Gypsy* RE, termed centromeric RE (*CR*). These satellite and *CR* sequences are transcribed at a low level (Topp *et al.* 2004; Neumann *et al.* 2007). Possibly, these *CR* sequences act as promoters for transcription of satellite sequences, which are required for centromeric recognition and chromosome segregation (May *et al.* 2005).

The only source of reverse transcriptase in genomes are REs and telomeres, and this enzyme can generate pseudogenes or transcribed retrogenes, which are widespread in plant genomes (Zhang *et al.* 2005, Wang *et al.* 2006). These retrogenes are usually chimeric, containing coding sequence in combination with

different regulatory elements (Betran *et al.* 2002; Wang *et al.* 2002), or they can consist of shuffled exons (Sayah *et al.* 2004; Zhang *et al.* 2004). These mechanisms are significant in evolution and adaptation, as functional genes have been identified that are a result of pseudogene differentiation (Betran *et al.* 2002; Emerson *et al.* 2004).

### **1.8 TEs in conifer genomes**

The *Ty-Copia TPE1* RE has been identified in *P.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 *P. radiata*, which contains a stop codon between the gag and AP domains (Kossack, Kinlaw 1999). *Spcl* from *P.glauca* contains only one stop codon before the int domain, indicating a recent origin (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).

Sequencing of *P. taeda* BACs has revealed a high proportion of TE sequences and pine-specific TE families (Morse *et al.* 2009). The *Gymny* RE from *P. taeda* is also found in other pine species, but not in *Picea*, *Abies*, *Tsuga* and *Ginkgo*, indicating the expansion of this RE after differentiation of these taxa (Morse *et al.* 2009). Both Class I and II TEs have been identified in the *P. taeda* genome, but many can be classified to novel families (Kovach *et al.* 2010). Comparisons between angiosperm and gymnosperm TEs has revealed between 37.7% - 99.4% nucleotide similarity. Hybridisation experiments with TEs do not reveal large differences in genome organisation between distantly related gymnosperm species (e.g. *P. abies* and *P. omorika*) and strong reciprocal hybridisation signals are found even between *P. abies* and *Ginkgo biloba* (Friesen *et al.* 2001). Sequencing of the *P. abies* genome revealed 1773 different repeat sequences, of which approximately half were similar to known TEs. LTR REs are the most common class, and *Gypsy* REs are the most commonly represented REs (Nystedt *et al.* 2013). Evidence of recent transposition burst was not detected in the *P. abies* genome, and the ratio of single LTRs to full-length REs is 1:9, indicating a low frequency of RE deletion in comparison to angiosperm genomes.

## **2. MATERIALS AND METHODS**

### **2.1 General experimental structure**

The experimental work was divided into several stages. To identify stress-induced expressed RE sequences, non-specific iPBS markers were utilised, as the transcriptionally active REs could be distinct from previously described RE in gymnosperms and REs from angiosperms. To minimise the effect of genetic background, ramets of one clone were used in each experiment. Distribution and expression of TEs was investigated using specific PCR primers, designed from the identified sequences. The influence of heat stress, pine woolly aphid infestation, ABA and SA treatments were investigated using the following scheme:

- ▶ Pine ramets were subjected to stress conditions.
- ▶ Four needles were collected from each individual, placed into 2mL tubes, and frozen in liquid nitrogen for later RNA extraction.

- ▶ After the end of each experiment, a further 2 needles were collected for DNA extraction and confirmation of clonal identity using SSR markers.
- ▶ RNA was extracted and possible DNA contamination assessed by PCR amplification. RNA was purified several times.
- ▶ cDNA was synthesised and analysed with stress-responsive control PCR primers.
- ▶ Genotyping of cDNA with iPBS markers, identification and excision of differentially expressed fragments.
- ▶ Cloning and transformation of competent *E.coli* cells, isolation of plasmids and sequencing of inserts.
- ▶ Sequence analysis and database comparison

Analysis of the differentially expressed TE-like sequences did not reveal the expression of full-length TEs, therefore in the second experimental stage full-length REs were isolated, analysed and at least partially sequenced from cDNA.

The third experimental stage utilised specific markers and primers developed from the identified sequences to analyse the distribution and polymorphism of TEs in a natural Scots pine population. Using the IRAP marker technique (Kalendar *et al.* 1999), natural pine stands were compared, and polymorphism was assessed between ramets of pine clones. Phylogenetic analysis and distribution of TEs was undertaken using a range of gymnosperm species.

## **2.2 Induction of stress conditions**

Two year old Scots pine ramets were provided by “Latvia’s State Forests” “Seeds and Plants” department in Jaunkalsnava, Latvia. Prior to the experiments, ramets were stored in the greenhouse. For each experiment, ramets of a different clone were used, within experiments all ramets were of one clone. DNA was isolated from pine needles using a CTAB-based method (Porebski *et al.* 1997). Genotyping with pine SSR (Simple Sequence Repeat) markers PTTX3107, PTTX4001 and PTTX4011 (Soranzo *et al.* 1998) was performed in order to confirm the clonal identity of the ramets.

Heat stress was induced by continuous +40°C treatment for 24 h, and was performed as described previously (Voronova *et al.* 2011) using the clone JA03. Samples from two stressed ramets (T1, T2) and three controls (Tc3-5) were stored in liquid nitrogen until extraction. For the aphid infestation experiment, it was noted that some ramets of the clone FU28 showed symptoms of infestation upon delivery from the greenhouse, while most ramets seemed visually healthy. After two weeks Pine Woolly Aphid (*Pineus pini*) infestation was identified by several properties (gummosis, white waxy flux, yellowed needles and abnormal needle cast), external feeding was confirmed microscopically. RNA from 4 pine ramets (B1-B4) was collected immediately and from 14 ramets after two weeks of injury (B1-B16 (excluding B5 and B7, which died after two weeks)). One sample showing no amplification with any of the stress inducible marker genes was used as a control (B4).

For treatment with abscisic acid (ABA), a 10 µM solution was prepared each day from 10 mM stock. Water solution with an equal quantity of ethanol was used to spray control pine ramets. The clone POP12 was used in the ABA experiment. Two pine ramets were foliar-sprayed and soil drenched each day with ABA and

one ramet was treated similarly with the control solution. The clone POP21 was used in the salicylic acid (SA) experiment. For spraying and watering with SA, 5 mM water solution was used; control ramets were treated similarly with water. Samples were collected before treatment (control-SA, control-ABA, water-SA, water-ABA), 24 h after treatment (SA-24, ABA-24, water-24-SA, water-24-ABA), 64 h (SA-64, ABA-64, water-SA-64, water-ABA-64) and one week after treatment (SA-W-1-2, ABA-W-1-2, water-W-ABA, water-W-SA). Two needles from treated and control ramets were collected at each time point and stored in liquid nitrogen until RNA extraction.

### **2.3. Identification of differentially expressed fragments**

RNA from pine needles was isolated using a CTAB-based method (Chang *et al.* 1993) followed by several purifications with the Turbo DNA-free kit (*Ambion*) and followed extraction with phenol:chloroform:isoalcohol (24:24:1) to ensure complete removal of DNA. Purification was continued until amplification of the extracted RNA using a standard PCR protocol with genomic primers showed no PCR products on ethidium bromide stained agarose gels. Stress initiation was confirmed by detection of amplification of genes whose expression has been shown to respond to a wide range of stress conditions: *P5CS2* (delta 1-pyrroline-5-carboxylate synthase-2), *dhn3* (dehydrin III), *AbaH* (abscisic acid and water-stress inducible protein) *LEA* (late embryogenesis abundant protein), and one housekeeping gene *GAPDH* (glyceraldehyde-3-phosphate dehydrogenase). Reverse transcription of the extracted RNA was performed with random hexamer primers using 1 µg of RNA with the TaqMan reverse transcription Kit (*Applied Biosystems*). A typical reaction was performed in a 20 µL reaction mixture containing cDNA, 1x Taq buffer (*Fermentas*), 1.5-2 mM MgCl<sub>2</sub>, 0.2 mM dNTP mix, 0.1 µM of each primer, 0.7 U Taq polymerase (*Fermentas*). Cycling conditions comprised of a denaturation step at 94°C for 3 min, followed by 38 cycles at 94°C for 30 s, 55-62°C for 30 s, 72°C for 40 s; and final elongation step at 72°C for 10 min. Amplification products were visualized after electrophoresis at 94 V for 2 hours in 1xTAE buffer in a 1.5% agarose gel stained with ethidium bromide. Inter PBS (iPBS) PCR amplification and separation was performed as described previously (Kalendar *et al.* 2010), excluding the separation step performed in low melting point agarose gels (*Fermentas*) for subsequent fragment purification. Differentially expressed fragments amplified from stressed tree samples, but not from control tree samples were excised and purified using the Gel Extraction Kit (*Qiagen*) following the manufacturer's protocol. Fragments were re-amplified with a touch-down PCR program 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 (*Fermentas*), 2 mM MgCl<sub>2</sub>, 0.2 mM dNTP mix, 1 µM PBS primer, 0.8 U DreamTaq polymerase (*Fermentas*). Re-amplified fragments were examined by electrophoresis, purified with Sephadex-G50 (*Sigma*) and ligated into the plasmid pTZ57R/T using the InsTAclone PCR Cloning Kit (*Fermentas*), following the manufacturers protocol. Competent *E. coli* cells were prepared using the method of Inoue *et al.* (1990) and transformed with the plasmids using the heat shock procedure described by the same authors. Plasmids were isolated using the alkaline lysis method. Insert sizes were confirmed with

double digestion with *EcoRI* and *SalI* (*Fermentas*). Sanger sequencing of an average of five plasmids from each isolated fragment with the expected insert size was performed with M13 primers and BigDye® Terminator v3.1 Cycle sequencing Kit (*Applied Biosystems*) and analysed using an ABI Prism 3130x-Avant Genetic Analyzer (*Applied Biosystems*).

Sequences were processed with the SequencingAnalysis v.5 software (*Applied Biosystems*). Multiple Sequence Alignment and similarity dendrograms were generated using CLUSTALW v.1.83 (<http://www.genome.jp/tools/clustalw/>). 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 (Refseq, Ref\_genomic) (<http://www.ncbi.nlm.nih.gov/BLAST/>). Several mobile element databases were used: GrainGenes Triticeae Repeat Sequence Database (<http://wheat.pw.usda.gov/>), Gypsy Database 2.0 (<http://gydb.org/>) (Llorens *et al.* 2011) and Repbase (<http://www.girinst.org/>) (Kohany *et al.* 2006). Count of matches among expressed sequences was performed using search in NCBI nucleotide EST database against *Viridiplantae* (green plants) database with specified algorithm parameters: 20 000 maximum number of aligned sequences and expect threshold equal or less than 0.01. Dendrograms were visualised using MEGA v.5 (Tamura *et al.* 2011). The obtained sequences were deposited in the NCBI EST database (accession numbers JZ389987-JZ390112).

#### **2.4 Expression analysis of TE-like fragments**

For evaluation and confirmation of shared transcribed RE-like sequences between studies, specific primer sets of 27 primer pairs for identified sequences were developed using Primer-BLAST (Ye *et al.* 2012). For quantitative PCR, the Maxima SYBR Green/ROX qPCR Master Mix (*Thermo Scientific*) was used. Reaction was performed on StepOnePlus™ Real-Time PCR System (*Applied Biosystems*) and analyzed using StepOne Software v.2.2.2 (*Applied Biosystems*). Relative quantification and comparative  $2^{-\Delta\Delta C_t}$  method was used using normalization against the endogenous control GAPDH (Schmittgen and Livak, 2008). PCR efficiencies for the target assay and endogenous control GAPDH assay were compared using standard curve analysis.

#### **2.5 Isolation of full-length REs and marker development**

To identify full-length REs several methods were applied: isolation of homologous RE; long-distance PCR using primers complementary to the identified sequences; and analysis of identified fragments similarity to related pine species BAC sequences, for identification of *P. sylvestris* homologs. Approximately 60 primer pairs were designed using Primer-BLAST (Ye *et al.* 2012).

Scots pine specific IRAP primers (Kalendar *et al.* 1999) were developed, based on the identified RE-like sequences. *DreamTaq* reagents (*Fermentas*) were used for IRAP genotyping of genomic DNA. PCR reactions contained up to 100ng DNA; 1x PCR buffer; 0.5 mM MgCl<sub>2</sub>; 1 μM primer; 0.2 mM dNTP mix; 0.7 U *DreamTaq* polymerase (*Fermentas*); 0.2 U *Pfu* polymerase (*Fermentas*); 3% DMSO. PCR cycling conditions: denaturation 94 °C 3 min; 38 x (94 °C - 40 sec; 50 °C - 40 sec; 68 °C - 1 min); 72 °C - 10 min.

## **2.6. Pine clonal and natural stand polymorphism analysis**

150 pine individuals from a natural Scots pine stand were genotyped with 9 IRAP markers. The population could be divided into sub-populations according to long-term differential growth conditions. SSR and IRAP data were analysed with *GenalEx* v. 6 (Peakall, Smouse 2006), *FAMD* v.1.30 (Schlüter, Harris 2006). Genotypic data were compared using AMOVA, PCoA tests; differences in height and diameter measurements were analysed using mean values, standard deviations were determined, significance of determined differences were tested using probability coefficient and F-criteria.

For assessment of somaclonal polymorphism, 14 individuals representing 4 pine clones were genotyped with 9 IRAP markers. The pine ramets were collected from two pine plantations – “Sāviena” and “Dravas”.

For phylogenetic and RE distribution analysis, individuals from 60 gymnosperm species were collected from the Botanical Garden of the University of Latvia, and the National Botanic Garden. The distribution of differentially expressed RE-like fragments and pine-specific REs was analysed. The gymnosperm species were also genotyped as described previously with 6 IRAP markers.

## **3. MAIN RESULTS**

### **3.1 Analysis of differentially expressed PBS fragments**

From all stress treatment experiments a total of 112 differentially expressed iPBS fragments were excised and cloned, and 335 plasmids were sequenced, yielding 126 unique fragments. BLAST analysis of the 126 unique sequences revealed that 28 of these sequences were highly similar to known REs either at the nucleotide or the protein level. The remaining 98 sequences were not found in the RE databases or the similarity level was low. Some of the unidentified sequences were similar to un-annotated plant EST or BAC sequences from various pine species; some were similar to hypothetical proteins, or partially similar to unknown proteins. 40 sequences showed no similarity to any sequences in any of the databases queried. 46 of the 98 un-annotated transcripts were moderately represented in the green plants EST database (1-31 hits), but 12 transcripts were similar to a large number of ESTs (108-8734 hits). Of the sequences highly represented in EST databases, four were similar to chloroplast and mitochondrial sequences, but eight transcripts were similar to un-annotated nuclear DNA sequences. The number of unique fragments identified in heat stress and ABA and SA treatment were comparable, but the number of TE-like sequences identified differs. A larger number of differentially expressed fragments were identified in the aphid infestation experiment. After treatment with SA, only five TE-like fragments were identified. After ABA treatment, only two TE-like fragments were found, both of which were also highly similar to fragments found in other stress conditions. Multiple sequence alignments of all unique fragments revealed that there was minimal overlap of identified sequences between experiments. No highly similar fragments were identified between the heat stress and plant hormone treatment experiments. Pine woolly aphid damage induced transcription of *Copia-2 TA-I* like sequences, which were also identified in the heat stress experiment (B104



and T30). One TE-like sequence (similar to a mammalian Endogenous Retrovirus-like, MSTA family LTR) and two unknown sequences were highly similar in both the aphid infestation experiment and treatment with SA (B103 and SA24). The ERV transcript identified after aphid infestation was shorter, but both transcripts were 97% identical. However, although the transcript is homologous to a mammalian element, one similar sequence was identified in *Hordeum vulgare* ESTs after infection with fungal pathogens. Two similar unidentified transcripts were found in the aphid infestation experiment and ABA treatment. One identified *Ogre*-like transcript was also highly similar in both these experiments (ABA18 and B125). One *TE Copia 18 BD*-like transcript group and three unidentified transcripts were expressed after both SA and ABA treatment. In total, four *Gypsy* superfamily homologs were found (*Gypsy-6*, *Ogre*, *Appalachian*, *PIGY-1*) and two *Copia* superfamily homologs (*Copia-2-TA\_1*, *HORPIA*). In addition, several transcripts of protease domain of DNA transposon belonging to CACTA family *EnSPM* was identified after heat shock.

### **3.2 Analysis of TE-like fragments**

In general, both ends of transcripts were completely complementary to the PBS primer sequences. Occasionally, the fragments contained one or more SNPs when compared to the PBS primer sequences. The iPBS fragments amplified from genomic DNA should contain LTR repeats from both sides in opposite directions and with genomic sequence between elements (Kalendar *et al.* 2010). The 28 differentially expressed fragments obtained in this study which were similar to previously reported REs could be divided into several classes: sequences having similarity to one LTR sequence (close to 100% fragment coverage); sequences having similarity to one LTR sequence plus a non-TE or unknown sequence at one end; composite sequences with similarity to two different LTRs in opposite directions; sequences with homology to RE ORFs, polyproteins or UTRs either along their entire length, or partially with a non-TE or unknown sequence at one end. Of the 28 identified RE fragments, 17 contained RE protein domains, eight contained LTRs, and three contained RE non-coding sequences. The majority of identified sequences were highly similar to various polyprotein domains of REs. This is probably due to the low levels of LTR-LTR transcripts, higher levels of sequence conservation within protein coding sequences, and the fact that there are relatively few pine-specific LTR sequences in the queried databases.

### **3.3 Analysis of chimeric transcripts**

Two transcripts were identified that combined TE-like and chloroplast-like sequences. One transcript (T51) has similarity to a non-LTR *LINE Karin\_TA* polyprotein at one end (49% of fragment coverage) and the other part of the sequence is similar to two separate regions of various *Pinus* species chloroplast genomes (16% and 24% coverage). A second chimeric transcript (B153) was similar to the *Gypsy Cereba* RE polyprotein coding sequence at one end, and the other half of the fragment was also similar to chloroplast genome sequences from various *Pinus* species. This part of the fragment with homology to chloroplast sequences was also similar to a large number of plant ESTs (2465 hits), while the RE-like part of the transcript was poorly represented in EST databases. Another interesting fragment (SA07) is similar to an internal 3' non-coding sequence of a

*Copia-18\_BD\_I* RE from *Brachypodium*, with query coverage of 33%. Four highly similar transcripts were identified after treatment with ABA and one was identified after SA treatment. Within the sequenced fragments there was observed 83-94% nucleotide identity and a size variation of 788 - 836 bp. Identified transcripts were also highly similar to *Pinus sp.* chloroplast sequences, with a maximum query coverage of 98%. This suggests that there is some homology between the *Copia-18\_BD\_I* TE from *Brachypodium* and chloroplast sequences. BLAST searches of genome sequences with the *Copia-18\_BD\_I* TE sequence indicated that there is an approximately 250 bp region in the non-coding 3' region of the TE which has homology with a chloroplast genome sequence from various species, including *Arabidopsis lyrata* (ORF64c). The highest nucleotide similarity was observed in the sequences homologous to the *Copia-18\_BD\_I* element and in the middle of the non-RE sequences, but flanking sequences were more dissimilar, indicating expression of this element from various loci in the pine genome. Entire fragments were found to match 1930 sequences in EST databases, while the *Copia-18\_BD\_I*-like sequence alone matches 2500 ESTs.

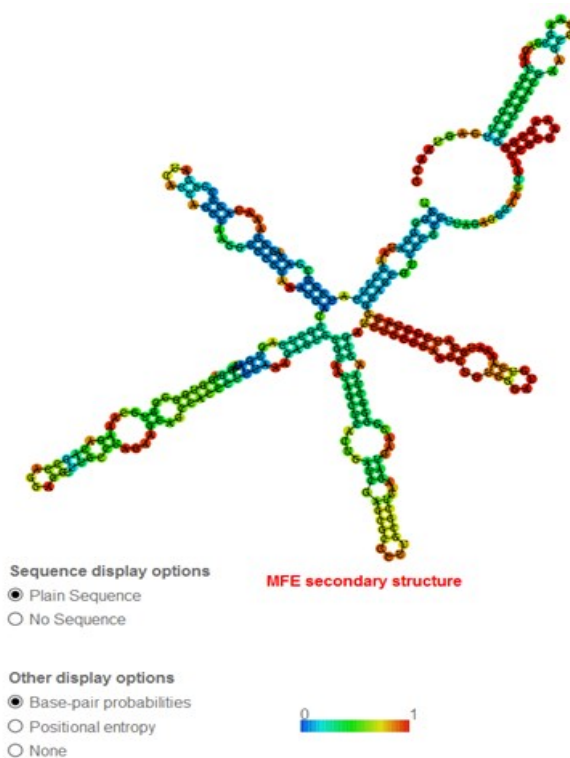


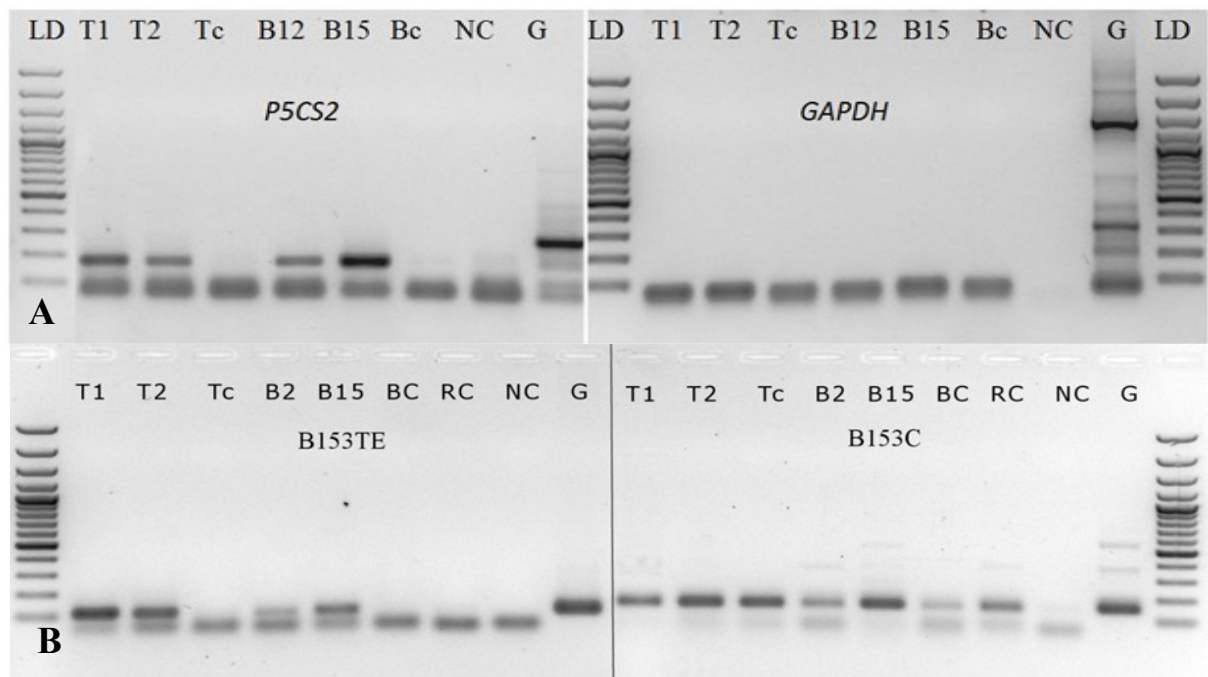
Figure 1. *Copia-18-BD* mRNA secondary structure. Regions with a probability of forming stem-loop structures approaching 1 are depicted in red. Minimal free energy -99.20 kcal/mol.

The 315 bp conserved chloroplast-like sequence found in *Copia-18BD* was analysed for potential regulatory motifs by searching the REGSITE database of regulatory motifs using NSITE NSITE-PL (Solovyev 2002), TSSP (Solovyev, Shahmuradov 2003), RegRNA (Mignone *et al.* 2005). Seven potential regulatory motifs were identified with homology  $\geq 80\%$ . One of these (TTTAAAG, BPBF from *H. vulgare*) was palindromic, and did not contain SNPs in either orientations (e-values=0.0094 and 0.0061). Two similarly conserved unknown nuclear factors were identified: 5'-CCCGGATC-3' from wheat (e-value=0.00395) and 5'-AAACCTCC-3' from spinach. Analysis of this sequence with the microRNA

detection software *FindMiRNA* (Softberry Inc.) identified six potential pre-miRNAs in direct orientation and 5 pre-miRNAs in reverse orientation, with a free energy value of less than -17 kcal/mol. *Figure 1* depicts one of the proposed secondary RNS structures predicted by *RNAfold* (Lorenz *et al.* 2011), with at least one high probability stem-loop structure.

### 3.4. Amplification of specific TE fragments

In order to compare and verify expression of the identified RE-like fragments between various stress treatments, 27 fragment-specific PCR primers were designed. Amplification of the control markers was also repeated (*Figure 2A*). For one of the chimeric fragments (B153), two primer pairs were developed, one pair amplifying the TE-like region (B153TE), and the other pair amplifying the chloroplast-like region (B153C). Differing patterns of amplification were observed for each primer pair: the B153TE primer pairs amplified PCR products only from cDNA of stressed trees, while the B153C primers amplified PCR products from all samples (*Figure 2B*).



*Figure 2. Comparative amplification of cDNA from heat stressed samples (T1, T2), pine woolly aphid affected samples (B12, B15), control ramets (Tc, Bc), negative control (NC) and genomic DNA (G). A- P5CS2-like gene fragments (190 bp and 250bp) and endogenous control GAPDH (70 bp). B- B153TE (110 bp) and B153C (120 bp).*

### 3.5. Real-time PCR

In addition to the standard PCR, the fragment-specific primers were also utilised for quantitative real-time PCR in order to obtain a more precise indication of differential expression of these fragments. A significant increase in transcription was detected for the SA07 fragment (similar to *Copia 18-BD*) after heat stress and also aphid damage (*Figure 3*). No significant increase of this transcript was observed after SA and ABA treatment. Expression levels were quantified with the

two separate primer pairs designed for the B153 fragment (B153TE and B153C). Expression of the B153TE (*Cereba*-like) fragments was elevated to a maximum 73-fold increase in insect damaged trees (Fig.5) and a 10-fold increase was observed after 24 hours of SA treatment, but after 64 hours expression drops to a 2-fold increase. After heat stress, a 5.62-fold increase was observed. Expression levels of the chloroplast-like sequence (B153C) were increased 9373-fold in aphid infested samples and 255-fold after heat stress. Transcription patterns of the B153C fragment and the SA07 fragment were similar. Both these sequences showed high similarity to chloroplast sequences. However, expression of the SA07 fragment was on average two-fold higher than the B153C fragment (bio15 = 2.6, bio-12 = 2.2, bio-1 = 1.7, heat stress = 1.96).

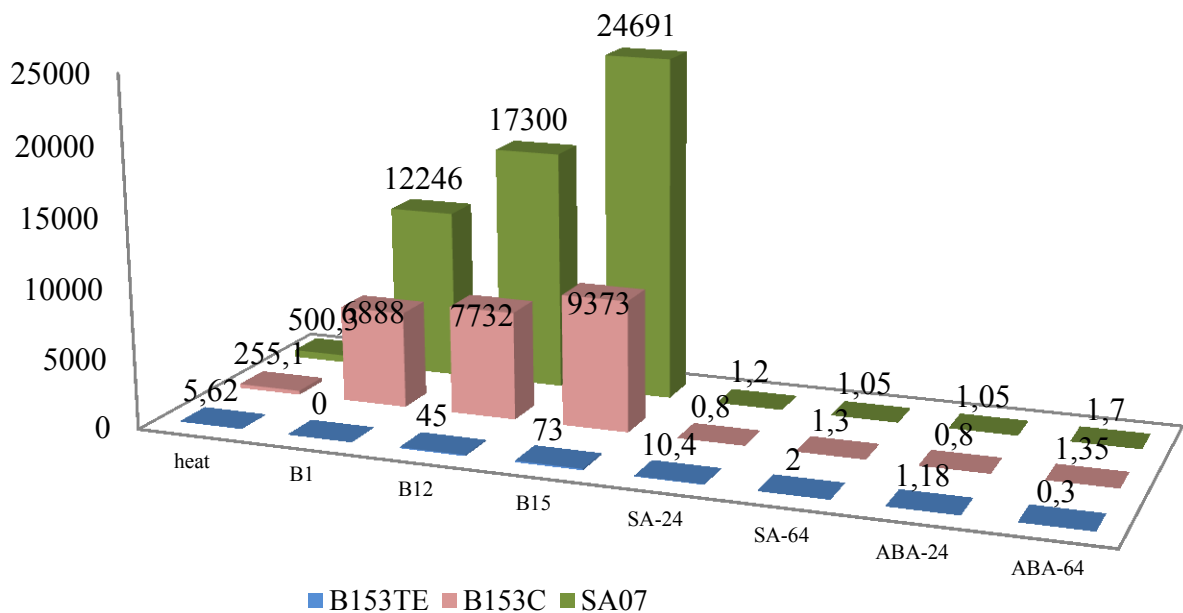


Figure 3. Relative expression levels of the *Copia*-18-*BD*-like transcript SA07, chloroplast-like transcript B153C and *Cereba*-like transcript B153TE.

### 3.6 Isolation of full-length REs

Analysis of some of the differentially expressed fragments revealed that some of them were highly similar to conifer BAC sequences; even if no homology was found with TE databases (e.g. fragment T50 is similar to the *P.taeda* BAC sequence PT\_7Ba4363K13 (AC241358.1). Further analysis of this BAC sequence identified LTRs and other internal TE domains, and a putative full-length TE was identified. PCR primers designed from the *P. taeda* BAC sequence amplified a similarly sized fragment from *P. sylvestris* genomic DNA. Using the same primers, it was also possible to amplify and sequence a fragment of the expected length from cDNA.

An LTR TE-like construct (ptTE101) with similarity to differentially expressed fragment T50 was identified from the *P.taeda* BAC sequence PT\_7Ba4363K13 (gb|AC241358.1). In the *P. taeda* genome, ptTE101 is 7697 bp long, containing 949 bp LTRs with 87% nucleotide identity and 5'-TG/ 3'CA nucleotides at the end of the LTRs. A 7bp TSD was identified (5'-CATCATTT-3'). A full-length homolog was not found in TE databases; however EST

transcripts with homology to the internal domains were identified from *P.taeda*, *P.contorta* and *Picea glauca*. Ten LTR sequences were isolated from *P.sylvestris* genomic DNA and cDNA with 74-81% similarity to ptTE101. Analysis of this sequence in *Repbase* identified a weak similarity with *Gypsy-18\_SMo-I* from *Selaginella moellendorffii*, a basal vascular plant species. BLASTX search of the TREP database revealed similarity to the *BAGY-2* polyprotein (PTREP241) - 36% identical amino acids, 58% synonymous amino acids (score 183, e-value= 2e-47). Nucleotides 4699-5481 are homologous to the *BAGY-2* and *Calypso-5* int domains at the protein level, indicating that this TE belongs to the *Copia* superfamily.

An LTR TE-like construct (ptTE121) with similarity to differentially expressed fragment T73 was identified from the *P.taeda* BAC sequence PT\_7Ba3500G20 (AC241315). In the *P.taeda* genome, ptTE121 is 5631 bp long, containing 1536 bp LTRs with 89% nucleotide identity and 5'-TG/ 3'CA nucleotides at the end of the LTRs. A 5bp TSD was identified (5'- AACCC -3'). The LTRs contain an interesting 14 bp poly-G SSR motif in the 5' region. At the time it was identified, no homology was found in TE databases, however subsequently homology was found with *PtCumberland\_I* (Wegrzyn, *Repbase*) (published 22/04/2013). The putative TE ptTE121 contains all motifs required for activity, including uORF (*upstream ORF*), IRES (*Internal Ribosome Entry Site*), TATA, polyadenylation motif. The complementary strand of ptTE121 contains four ORFs, which are similar to various TE polyproteins. The LTR and partial 3' internal sequence was obtained from *P.sylvestris* cDNA.

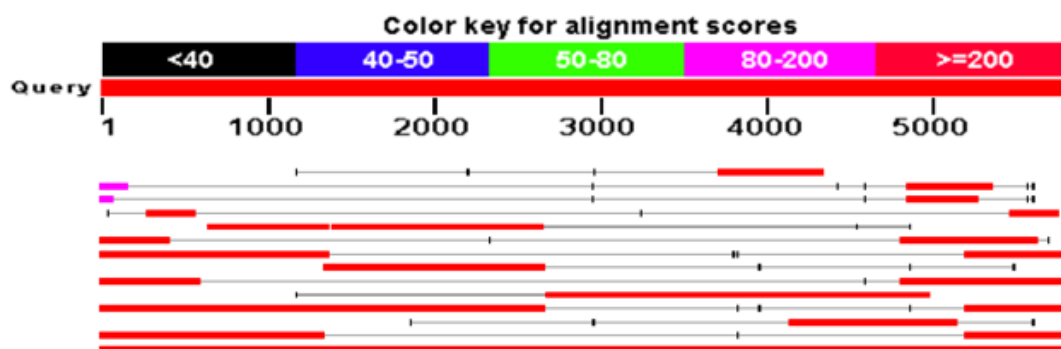


Figure 4. *P. sylvestris* sequence coverage of the putative RE TE283.

RE-like domains were identified in other differentially expressed fragments – e.g. initially no homology with fragment psTE283 was found in TE databases, but it had 77% nucleotide similarity to a *P. glauca* BAC sequence, which also contained a putative LTR RE. Identification of coding domains was only possible using similarity to EST databases and ORF detection tools, indicating that this RE is conifer-specific. The coding domains were able to be PCR amplified from *P. sylvestris* cDNA, which were homologous to the *P. glauca* structure. On the 18/03/2013 a homologous sequence from *P. taeda* was published - *PtAppalachian\_I* (Wegrzyn, *Repbase*). A full-length homolog (psTE283) was sequenced from *P. sylvestris* cDNA isolated from heat-stressed individuals (Figure 4). The putative RE psTE283 is 5796 bp long, containing 615 bp LTRs with 80% nucleotide identity and 5'-TG/ 3'CA nucleotides at the end of the LTRs. The 5' LTR contains a uORF, IRES, TATA-motif (5'-AATAATAAATA-3'), as well as a



*P.sylvestris* regulatory sequence PtMYB4 5'-ACCACCC-3' (e-value=0.007) and four unknown nuclear factors. The 3' LTR contains a polyadenylation motif at position 391.

The 3' end of the internal sequence of psTE283 contains a polypurine motif (5'-AAGGGGGGAGGAA-3'). The *TSSP* software for predicting plant polymerase II promoter sites only identified a potential promoter on the complementary sequence at position 5574 in the 3' LTR. Two ORFs were found in the forward sequence, whose translated amino acid sequences were similar to TE proteins. The 354 aa ORF1 (2644-3709) is similar to a RT/RH domain, and the 302 aa ORF2 (3669-4578) is similar to the int domain (pfam00665). The int domain is located at the 3' end, indicating that the psTE283 RE belongs to the *Gypsy* superfamily. Interestingly, the psTE283 RT/RH domain resembles the *Cerebra* polyprotein (TREP99) with 43% identical amino acids and 58% synonymous amino acids. The B153 chimeric fragment was also with the *Cerebra* polyprotein. This could indicate that these transcripts represent one RE family. Also, TE283 is about 77% similar to the B193 transcript isolated from pine woolly aphid damaged ramets. The psTE283 sequence is similar to 85 conifer EST sequences in the NCBI database. The *P.taeda PtAppalachian\_I* and *P.sylvestris* psTE283 internal domains are 91% similar, and the LTRs are 81-88% similar. Analysis of the *PtAppalachian\_I* sequence from NCBI BAC PT\_7Ba2797E03 (AC241290.1), indicates that the LTRs are 90% similar to each other, which is higher than the similarity of the *P.sylvestris* psTE283 LTRs. A continuous 584 aa ORF that resembles RE polyproteins was identified in the BAC sequence complementary to *PtAppalachian\_I*, and another 375 aa ORF is similar to gag domains. Probably the *P.sylvestris* RE psTE283 has lost the ability to autonomously transpose, as it does not contain a full polyprotein ORF and the identified RE ORF is shorter than the *PtAppalachian*. However, a number of copies may retain transcriptional function.

### 3.7 Genotyping of pine populations growing in differing conditions

A natural pine stand was identified, which could be divided into sub-populations growing on a hill, on a slope, and in a swamp. Due to its topographic location, the hill population had been subjected to long-term water deficit in comparison to the other populations.

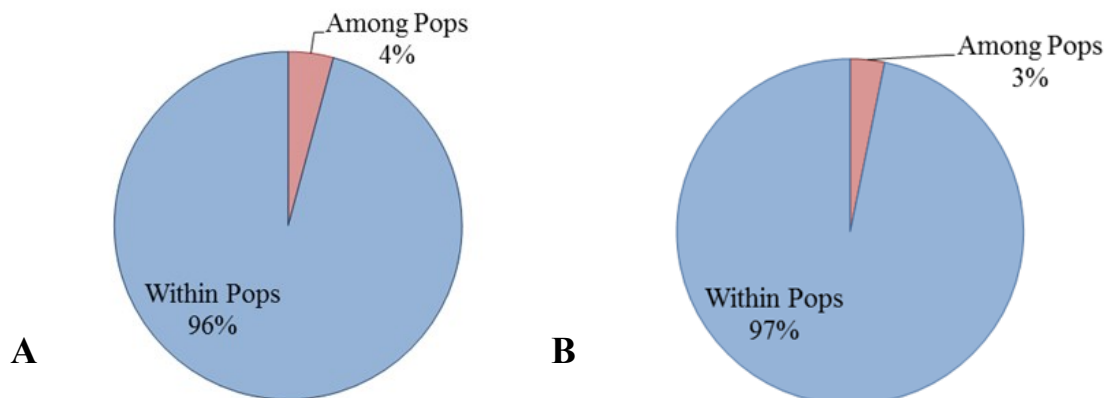


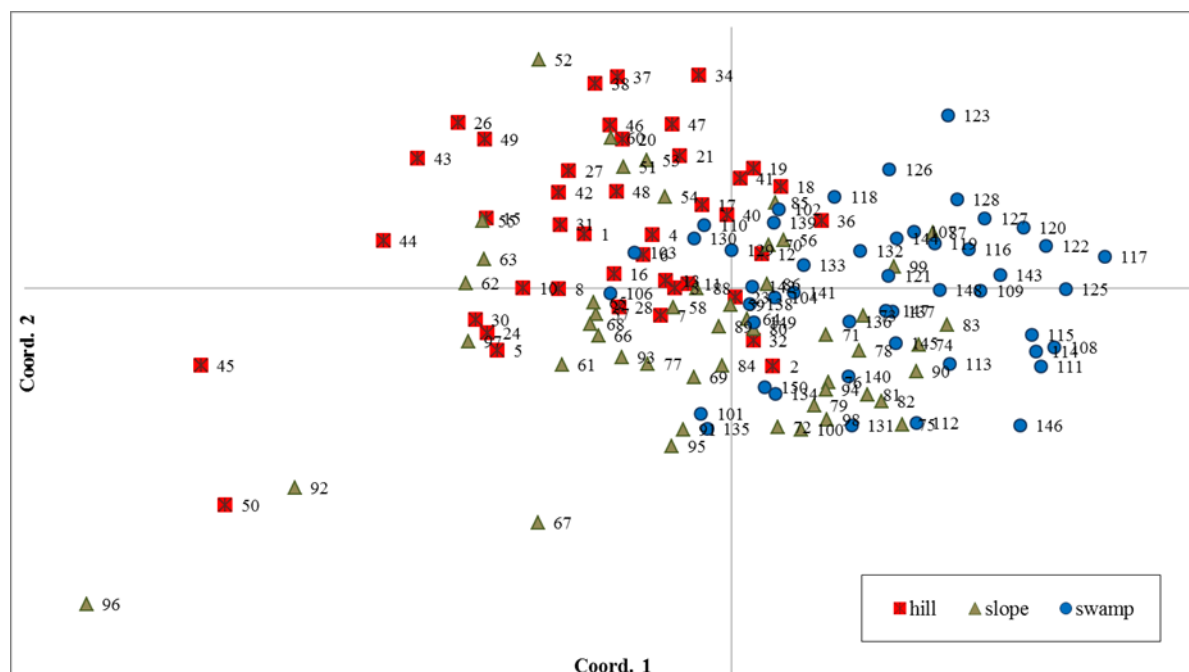
Figure 5. AMOVA utilising SSR marker data (A) and RE marker data (B)

150 individuals were genotyped using the IRAP technique (Kalendar *et al.* 1999), using 9 markers developed from the previously described experiments (*table 1*), as well as with three SSR markers.

*Table 1. Primer sequences and genetic diversity parameters of the RE markers.*

Primer	Accession number of expressed sequence	Average fragment no. per individual	No. of polymorphic fragments	Number of fixed fragments
IR_1	JZ389987.1	11.60 ± 1.90	12	4
IR_2	JZ390004.1	13.72 ± 9.12	28	0
IR_3	JZ390006.1	8.4 ± 2.56	13	0
IR_4	JZ390008.1	11.26 ± 3.68	19	1
IR_5	JZ390015.1	13.18 ± 4.18	19	0
IR_6	JZ390027.1	22.20 ± 12.86	37	0
IR_7	JZ390015.1	7.72 ± 1.99	14	0
IR_8	JZ389987.1	9.83 ± 1.13	13	3
IR_9	JZ390027.1	13.07 ± 2.65	19	2

As a result, 184 loci were identified with the IRAP primers and 33 with the SSR markers. Analysis of molecular variance (AMOVA) did not identify a large differentiation between the populations (*Figure 5*), however Principal Coordinates Analysis (PCoA) of the RE marker data differentiated the hill population from the other populations (*Figure 6*), while the SSR data did not indicate such a division (*Figure 7*). The first three coordinates explain 56.46% of the IRAP data variation, and 61.15% of the SSR data variation.



*Figure 6. PCoA of IRAP data from the three sub-populations*

The average number of fragments amplified from individuals from each subpopulation was calculated. Only individuals with high quality genotypes were utilised, therefore the number of individuals (n) analysed in each population varied: hill population n=38, slope population n=44 and swamp population n=45.

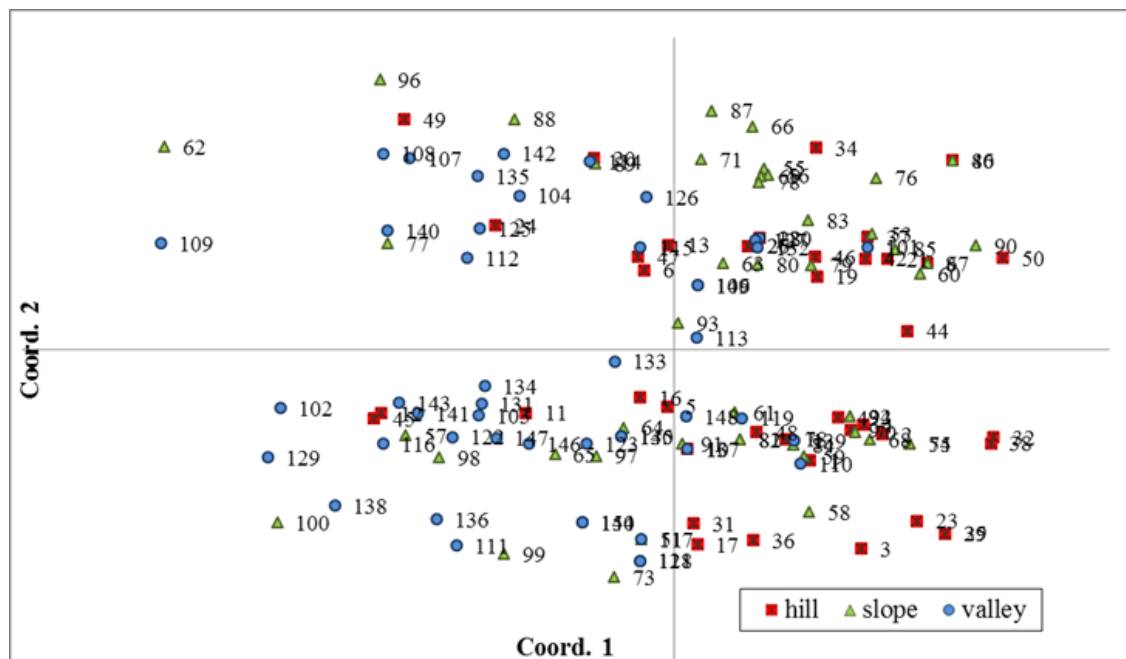


Figure 7. PCoA of SSR data from the three sub-populations

Table 2. Statistical analysis of fragment number amplified by RE markers (h – hill, sl – slope, sw- swamp).

Pop	Average fragment no.	Standard deviation	Standard error	Difference in fr. no.	Standard error of difference	Coefficient of probability	F-criteria
	Xvid	$\delta$	Sx	d(Xvh-Xvsl/sw)	Sd (h-sl/sw)	$t > 3$	F, P= 0.05, df(40) $\geq 1.51$
h	116,37	6,80	1,10	-	-	-	-
sl	109,59	6,08	0,92	6,78	1,42	4,77	1,25
sw	107,89	5,40	0,80	8,48	1,34	6,33	1,58

The difference (d) between average values, standard error of difference (Sd) and probability coefficient (t) were determined (Table 2). The difference in average fragment number between the hill and slope populations is  $6.78 \pm 1.42$ , and the difference between hill and swamp populations is  $8.48 \pm 1.34$ . These differences were significant ( $p < 0.01$ ). According to the F-criteria, the difference in the average fragment number amplified in the hill and swamp population is statistically significant, but not the difference between hill and slope populations (Table 1). While the variation within populations is high, it can be seen that a higher average number of fragments is amplified by each marker in the hill population (Figure 8). The number of fragments amplified by some markers (IR\_6, IR\_4, IR\_1) is also higher in the swamp population. The differences in amplified fragment number between the hill and swamp are significant for the markers IR\_2, IR\_7, IR\_5 and IR\_6, while differences in amplified fragment number between the hill and slope are significant only for the marker IR\_6 (Figure 9).



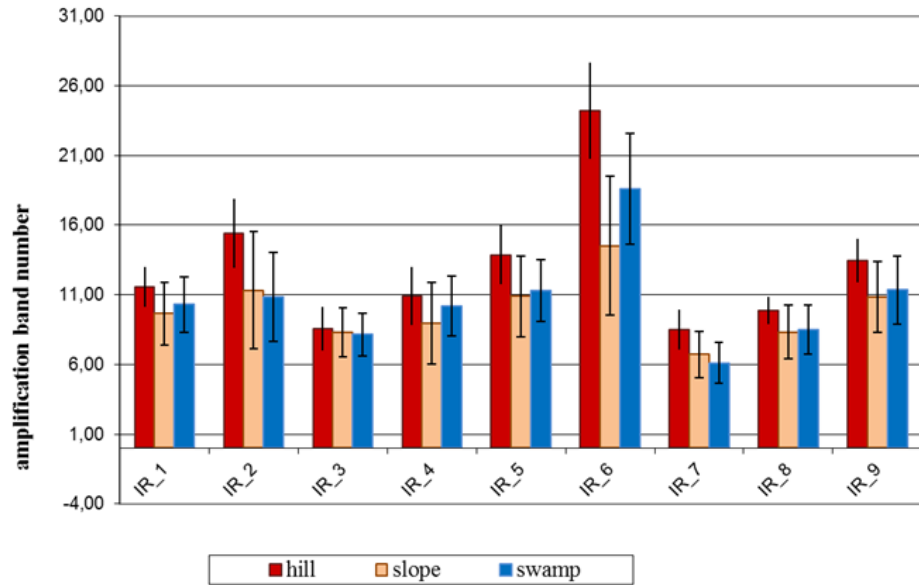


Figure 8. Average number of fragments amplified by each RE-derived marker

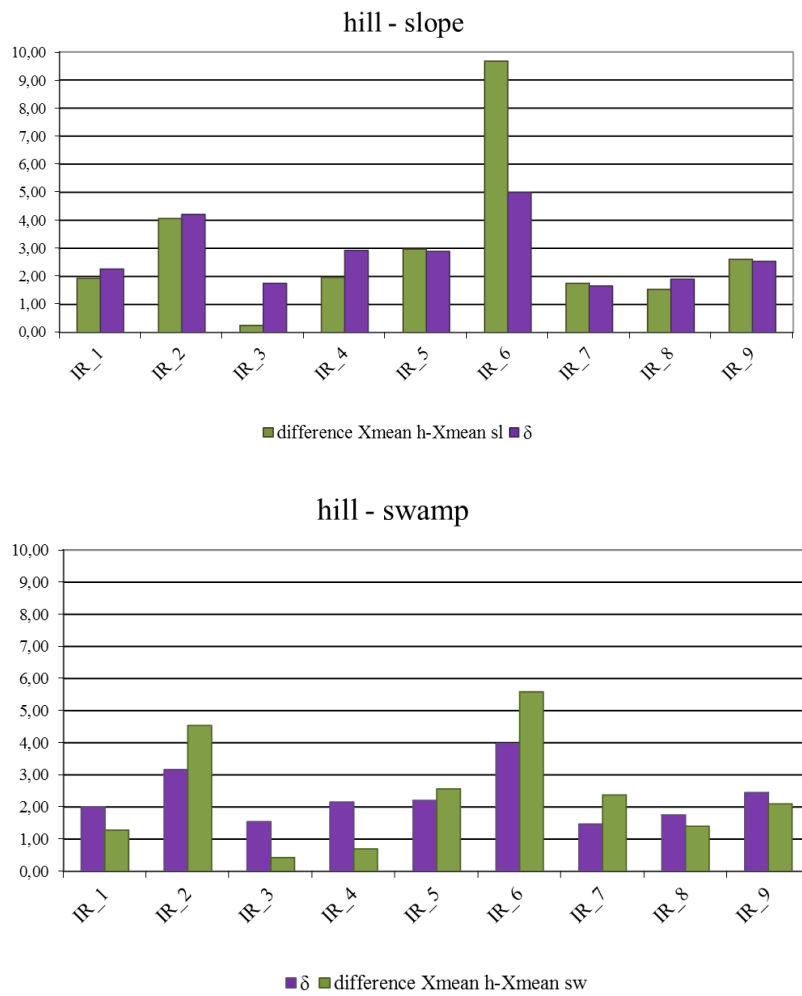


Figure 9. Comparison of differences in average number of amplified fragments per marker with standard deviation ( $\sigma$ ).

### 3.8 Assessment of clonal variation

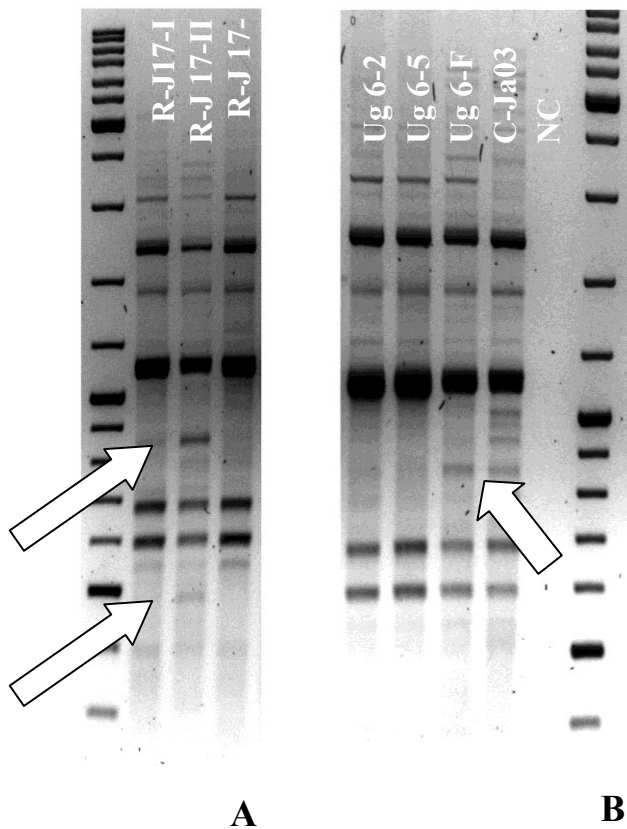


Figure 10. IR\_7h (T73, Gypsy-I) IRAP amplification product polymorphism among pine ramets (A)- R-J 17, (B)- Ug 6; C-control Ja03, NC-negative control.

Nine of the developed IRAP markers were used to assess 26 year old pine ramets representing 4 different clones, growing in two different plantations – “Dravas” and “Sāviena”. The ramets of one clone were identical according to SSR marker genotyping. One of the markers (IR\_2) amplified fragments polymorphic between the ramets of a single pine clone (Figure 10). Seven polymorphic fragments were excised, but only five were successfully cloned, 17 plasmid inserts were sequenced. The IR\_2 primer sequence was identified at the ends of each sequenced fragment. Five unique sequences were identified, 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) was partially similar to the LTR of *Copia-17\_PAb-I* from *Picea abies* (GenBank accession Nb. MA\_172979). This group contained 98% identical sequences derived from two similarly sized excised polymorphic

fragments from different clones (R-J17-II and Ug6). A homologous 5986 bp retrotransposon sequence was identified in a *P.taeda* BAC sequence, and which was designated as *Copia-17-PTa-I* (submitted to Repbase database). A third sequence group (507bp) derived from one polymorphic fragment was similar to *Copia-17\_PAb-I*, spanning the 3' region of the internal domain and the 5' region of the LTR. The internal sequence of *Copia-17-PTa-I* is 73% similar to *Copia-17\_PAb-I*, but the element from *Picea abies* contains a 404 bp deletion in the reverse transcriptase domain. The LTRs are only partially similar between these elements; only 231 bp and 79 bp of the LTRs are 70% similar. The LTRs in the *Copia-17-PTa-I* element are 947 bp long and are 92% identical to each other. The conserved plant RE promoter motif 3'-TATATA-5' (Vicent *et al.* 2005; Benachenhou *et al.* 2013) was only identified in the U3 region of the *P.sylvestris* LTR region, the polyadenylation motif (3'-AATAAA-5') was identified in the 5' LTR of *Copy-17-PTa-I*. The full-length sequence of the *Pinus sylvestris* RE has not been determined, and further sequencing is required to elucidate this.

### 3.9 Gymnosperm phylogenetic analysis with IRAP markers

DNA from individuals representing 60 different gymnosperm species was analysed with six IRAP markers. Four Scots pine individuals were also included in the analysis. A total of 241 fragments were genotyped, on average 38.8 fragments were amplified from each individual (variation 134.59 and standard deviation 11.6). Species were grouped by genus: *Abies* (7), *Picea* (14), *Pinus* (19) other *Pinales* (20) and other gymnosperms (5).

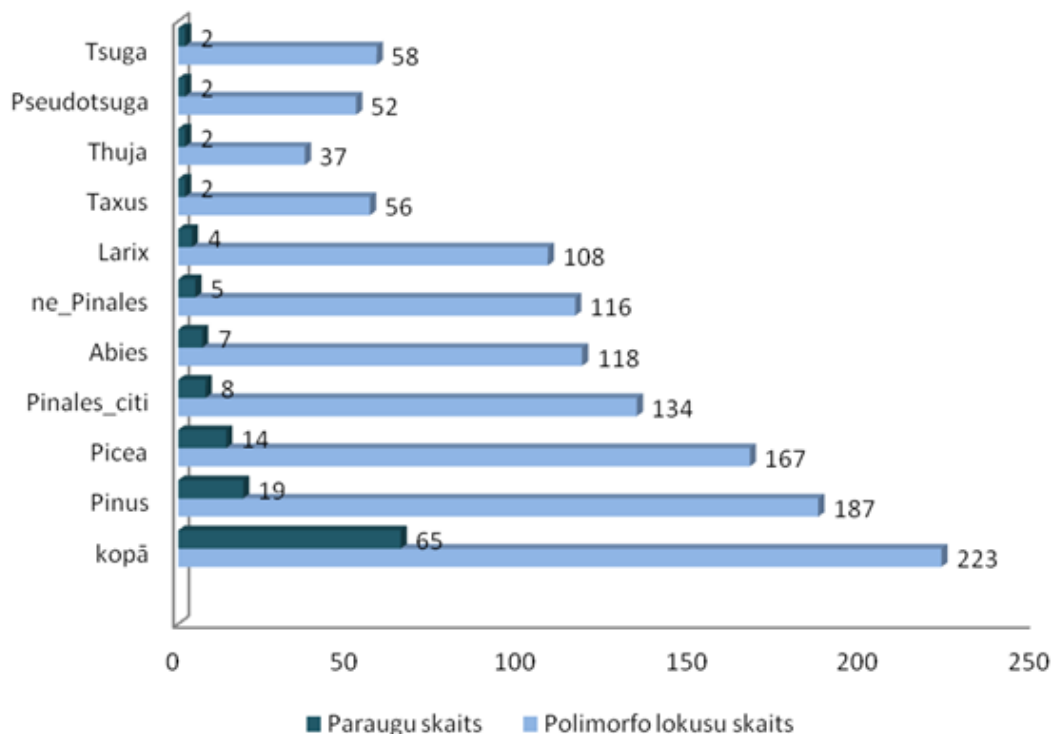


Figure 11. Number of individuals and polymorphic loci in each gymnosperm species group.

Each gymnosperm genus was not equally represented, but the number of polymorphic loci was lower in *Thuja* (37), compared to other groups of similar size: *Tsuga* (58), *Taxus* (56) and *Pseudotsuga* (52) (Figure 11). The frequency of polymorphic fragments in each group relative to the total number of amplified fragments was calculated (Figure 12).

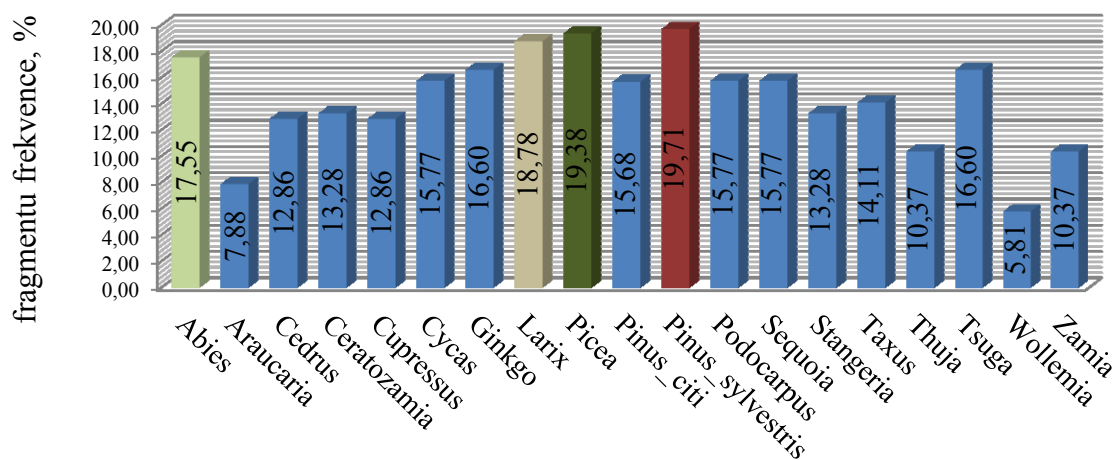


Figure 12. Representation frequency of fragments in each gymnosperm group

In groups consisting of more than one species, the average number of fragments was used. The highest frequency was found in *P. sylvestris* (19.71), the

species from which the markers were developed. A similarly high frequency was found in the related *Picea*. The lower frequency found in the other *Pinus* group can be explained by the inclusion of distantly related pine species (*P.banksiana*, *P.contorta*, *P.pumila*, *P.canariensis*). The *Thuja* group also has a low frequency in addition to a low total number of amplified fragments. Especially low frequencies were found in the exotic gymnosperms which are not native to the northern hemisphere (*Wollemia nobilis*, *Zamia integrifolia*, *Araucaria*).

Neighbour joining phylogenetic analysis of the IRAP data grouped Scots pine samples separately with high bootstrap values (Figure 13). This result was expected, as other species were represented only with one sample. According to previous phylogenetic studies of gymnosperms, different genera of conifers were similar and were grouped with low bootstrap values. *Abies*, *Picea* and *Pinus* were separated from other samples.

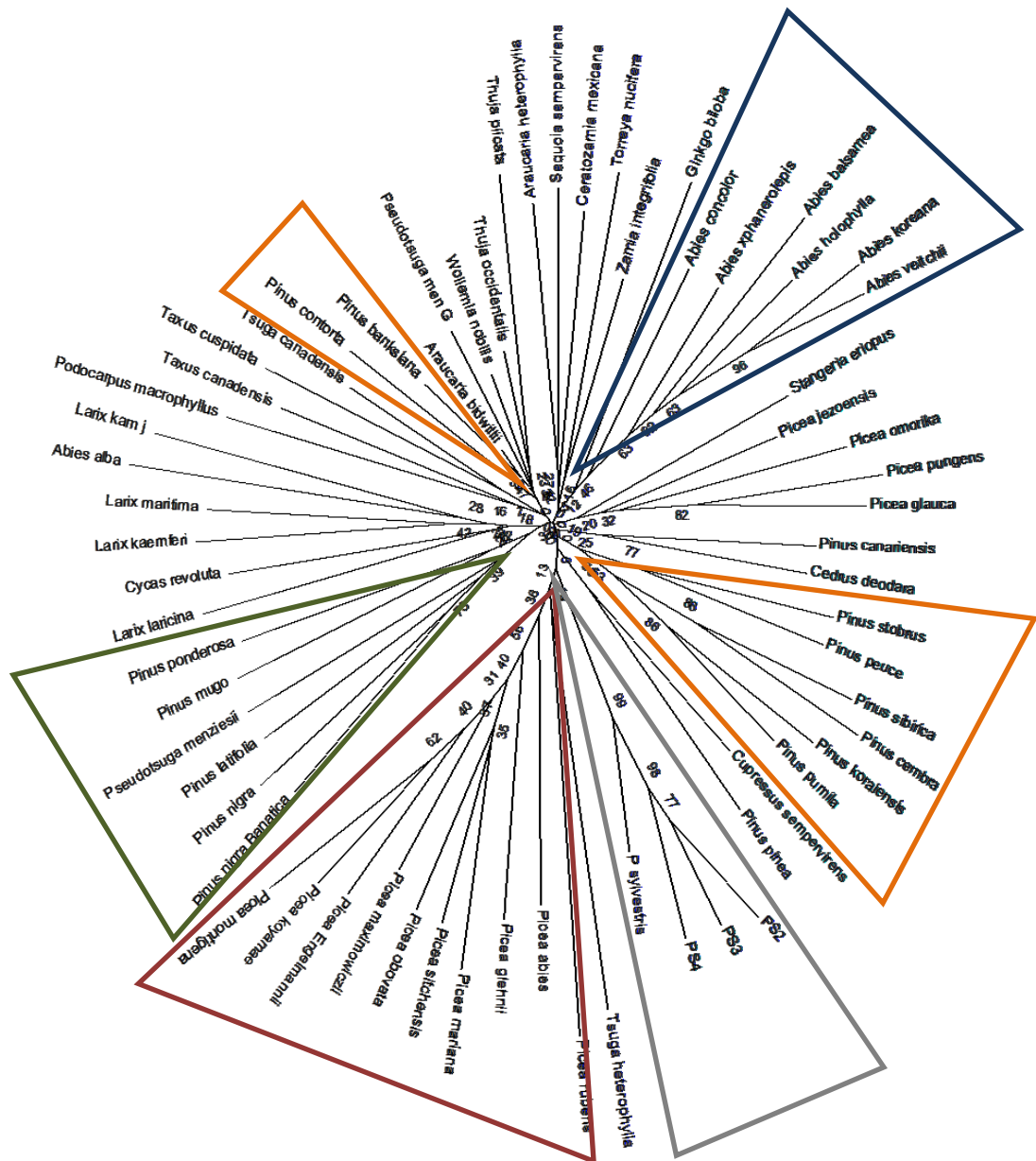


Figure 13. Gymnosperm phylogenetic analysis using pine RE specific markers. Orange triangle - Strobos, green triangle - Pinus, red triangle – Picea; blue triangle – Abies; grey triangle - Pinus sylvestris.

The distribution of identified REs in gymnosperm genomes was analysed by amplification of internal sequences. psTE283: PCR primers were designed to amplify the gag-RT domain from the RE psTE283. These amplified the largest number of fragments of the expected size range from *P. sylvestris*, as well as from five other *Pinus* species, two *Picea* species, *Tsuga*, *Cedrus* and two more distant taxa *Ginkgo biloba* and *Ceratozamia mexicana*. No amplification with these primers was observed in *Pseudotsuga menziesii*.

psTE101: PCR primers were designed to amplify part of the 5' LTR and the gag domain. Amplification was only observed in seven pine species (*P. sylvestris*, *P. nigra* var. *banatica*, *P. latifolia*, *P. mugo*, *P. ponderosa*, *P. canariensis*, *P. pinea*). A wider size range of fragments was also observed, from 500-1500bp (Figure 14).

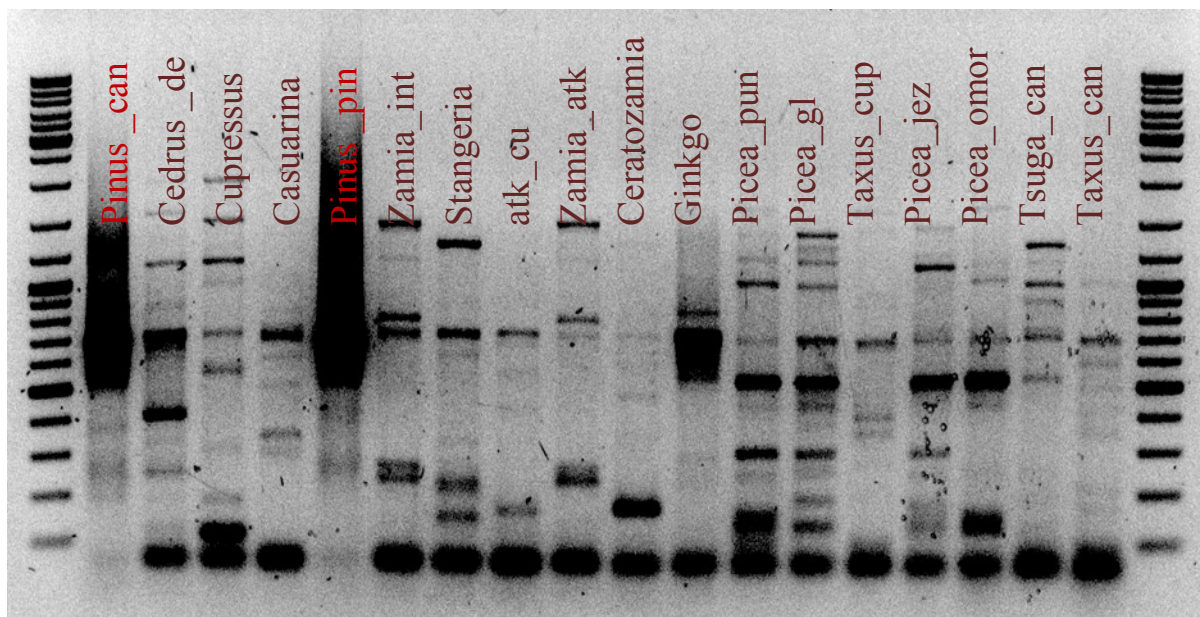


Figure 14. Amplification of the RE psTE101 LTR-gag domain. Expected fragment length in *P. sylvestris* is 700bp.

TE121: PCR primers amplifying a 1608bp LTR fragment were designed. Fragments varying in size from 1.2-2.5 kb were amplified from *P. nigra*, *P. latifolia*, *P. nigra* var. *banatica*, *P. mugo*, *P. ponderosa* and *P. sylvestris*. A specific and intense fragment of the expected size was amplified from *Picea rubens*, *P. peuce*, *P. sibirica*, *P. koraiensis*, *P. canariensis*, *P. pinea* as well as from the more distant *Ginkgo biloba* and *Araucaria heterophylla*.

SA07: Two primer pairs amplifying the *Copia-18BD*-like region as well as the chloroplast-like region were utilised. The primers targeting the chloroplast-like region amplified a single, specific fragment from all gymnosperm species. The primers targeting the *Copia-18BD*-like region amplified the expected 120bp product from 15 *Pinus* species, 10 *Picea* species, 2 *Abies* species, one *Larix* species and *Ginkgo biloba*. A number of polymorphic fragments were only amplified from *Stangeria eriopus*.

The putative *P. sylvestris* TEs ptTE101, ptTE121 and psTE283 were also present in the genomes of *P. nigra*, *P. latifolia*, *P. mugo* and *P. ponderosa*. In contrast, the RE psTE283, identified after heat stress, was found in the genomes of a much wider range of species, including *Cedrus deodara*, *Ginkgo biloba* and *Ceratozamia mexicana*. This suggests that it is an older RE that has been



maintained in *Pinus*, but lost in *Picea*. Specific amplification of the psTE283 domain was also not observed in *Abies* and *Larix*. The two species with the least amplification of *P. sylvestris* REs were *Pseudotsuga menziesii* var. *glauca* and *Wollemia nobilis*. Interestingly, according to the utilised markers, the pine species most distinct from *P. sylvestris* was *P. contorta*.

## 4. DISCUSSION

### 4.1. Identification of TE-like transcripts

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). This study investigated stress-responsive transcripts that are associated with REs or are parts of actively transcribed mobile elements. The iPBS approach utilized in this study has identified a range of coding polyproteins belonging to known mobile element families which are transcribed under stress conditions. Contrary to the type of fragments expected to be amplified from genomic DNA with PBS primers (Kalendar *et al.* 2010), only a few differentially expressed transcripts contained LTR sequences in opposite orientations at both ends. Existence of transcripts containing LTRs from different RE elements confirms that continuous transcription can occur within clusters of REs, and that the expected types of fragments are also amplified. Amplification of genomic DNA with iPBS markers results in approximately 50-70% of TE-like fragments (Kalendar *et al.* 2010). The lower proportion of identified TE-like fragments in this study is due to the lower representation of these sequences in the transcriptome, as well as the low number of annotated pine and conifer TE sequences in databases, which are divergent from angiosperm TEs (Kovach *et al.* 2010; Nystedt *et al.* 2013). Some of the differentially expressed sequences 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. Sequence analysis of the fragment with homology to the DNA transposon *EnSpm* indicated that the PBS primer sequences are present in the sequence of the *EnSpm* ORF, excluding only the first three 5' base pairs at the end of the primer. The highest proportion of identified fragments was homologous to TE coding domains, because these regions are the most conserved TE regions (Wicker *et al.* 2007). The low number of LTRs identified in this study is probably due to the fact that LTRs are the most divergent sequences between RE element families, and therefore many of them will not be identified by cross-species sequence homology. A recent computational analysis of RE elements in the poplar genome reveals that homologous RE sequences from other species could be found in data bases for only 42% of TE elements (Cossu *et al.* 2012). In the *P. abies* genome, only approximately half of the repeat sequences are homologous to previously described TEs (Nystedt *et al.* 2013). More transcripts of the ancient *Gypsy* superfamily polyproteins were identified in comparison to *Copia* type representatives. *In silico* analysis of the distribution of plant TE in maize ESTs has revealed that *Gypsy* superfamily elements are the most abundant among transcripts, but LTR *Copia* and CACTA family DNA transposon transcripts had lower transcription levels (Vincent 2010). Similarly, an abundance

of *Gypsy* superfamily transcripts was observed in potato (Yadav and Singh 2013), triticeae (Echenique *et al.* 2002) and spruce (Nystedt *et al.* 2013).

#### **4.2. Stress dependent expression of TE-like sequences**

Recently it was reported that RE activation during heat stress increases the transcriptional response of stress inducible genes by incorporation of additional RE-derived signals (Ito *et al.* 2011, Feng *et al.* 2012), suggesting that mobilization of active REs could generate novel gene-response networks, and leading to advantageous phenotypes. It was shown that naturally occurring stress conditions such as heat stress and insect damage initiate a significant increase of transcription of iPBS fragments in Scots pine, including expression of TE-like sequences. In comparison, ABA and SA plant hormone treatment resulted in a lower level of iPBS fragment expression as well as a lower proportion of identified TE-like fragments. This could be due to the activation of a wider range of stress responses by heat and insect attack in contrast to ABA and SA, which activate specific response pathways. Abscisic acid (ABA) is one of the main plant hormones involved in response to water deficit, and it is synthesized in roots and leaves. ABA receptors are located in the plasma membranes of stomatal guard cells in the epidermis (Wilkinson and Davies 2002). ABA can be stored in the symplasts, and not stimulate a reaction from the plant until a water deficit induced change in the pH of leaves and roots (Wilkinson, Davies 2002). The relatively smaller effect of ABA on transcription of TE-like sequences could be due to the fact that the pine ramets were not subjected to water deficit. It has been previously reported that the strawberry *Ty-1 copia* RE *FaRel* transcription is induced by ABA (He *et al.* 2010). Additionally, the *BARE-1* RE from barley contains ABA response elements, and water stress may induce an increase in copy number of *BARE-1* (Kalendar *et al.* 2000). In this study, the highest diversity of TE-like transcripts was identified after heat stress, including sequences with similarity to members of various TE superfamilies (DNA transposon, non-LTR LINE element, *Gypsy* and *Copia* LTR REs). In the heat stress experiment, we identified several CACTA family DNA transposon *EnSpm*-like RNA transcripts. *EnSpm*-like DNA transposons have been shown to be active in a broad range of distantly related species including maize (Peterson 1953; McClintock 1954), wheat (Wicker *et al.* 2003), sorghum (Chopra *et al.* 1999), rice (Jurka 2005) and *Poaceae* species (Altinkut *et al.* 2006). The unexpected identification of *EnSpm*-like RNA transcripts indicates the possibility that this DNA transposon is activated in *P. sylvestris* in response to heat stress. However, isolation of a full-length transcript will be required to determine whether it is functional. No similar transcripts were found between heat stress and response to plant hormones. This could be a result of activation of different stress-responsive regions or genes within the pine genome via *cis*-acting elements and promoters localized in RE LTRs (Kumar and Bennetzen 1999; Takeda *et al.* 1999; Havecker *et al.* 2004). In addition, solo LTRs can also regulate gene expression (Butteli *et al.* 2012). These regulatory sequences are dependent upon various signal molecules, and therefore can be differentially regulated in differing stress or environmental conditions (Hirochika 1993; Beguiristain *et al.* 2001; Tapia *et al.* 2005; Ito *et al.* 2011). In aphid damaged trees, a large abundance of transcripts with similarity to various domains of the *Gypsy Ogre* family were identified. *Ogre* is a large and

transcriptionally active *Gypsy* LTR RE described in the genome of *Pisum sativum* and found also in other dicots such as *Medicago truncatula* (Neumann *et al.* 2003). However, several transcripts were shared between response to insect damage and salicylic acid, which is consistent with the known functions of SA. SA is a well characterized signaling molecule that induces plant defense reactions during pathogen attack (Malamy *et al.* 1990; Malamy and Klessig 1992; Klessig *et al.* 2000; Durrant and Dong 2004). In addition, the elements *TLC1.1* (Tapia *et al.* 2005), *Tnt1* (Grandbastien *et al.* 1997), *OARE-1* from oat (Kimura *et al.* 2001) are expressed after SA treatment.

A differential expression pattern of TE-like and other transcripts was identified for each stress condition. However, comparison of short sequences can lead to misleading results. TEs have multiple copies in the pine genome, and conserved regions of different TEs can have high homology to other sequences, e.g. the chloroplast-like genomic insertions. Additionally, different members of one TE family can be transcribed in different stress conditions due to polymorphisms in LTR regulatory sequences (Kumar, Bennetzen 1999; Takeda *et al.* 1999; Havecker *et al.* 2004; Beguiristain *et al.* 2001). Homology with database sequences of coding domains from one TE family may represent various TE families in pine. For example, transcripts were frequently identified with high nucleotide homology to various *Gypsy Ogre* domains, but no specific PCR amplification of genomic DNA between these domains was observed. Possibly, these transcripts were derived from different elements and these primers amplified anonymous genomic regions between the elements. It is also possible that this element is inactive and it is no longer transcribed in full length. In this case, the scattered conservative TE regions generate chimeric transcripts, which were partially identified in this study. However, amplification products with one of the *Ogre*-like transcript primers are highly polymorphic; indicating that probably each of these families is represented in multiple copies in the pine genome.

#### **4.3. Analysis of chimeric transcripts**

The largest transcript group identified in *P. sylvestris* after ABA treatment was homologous to the *Copia-18\_BD* element from *Brachypodium*. However, the *Copia-18\_BD* TE from *Brachypodium* contains an insertion of a chloroplast sequence, and the pine transcripts were homologous to this chloroplast region, and no differentially expressed fragments were homologous to other regions of the *Copia-18\_BD* TE. The chloroplast sequences are conservative among all plants, and therefore there is possible that the pine genome does not contain a *Copia-18\_BD*-like RE and that in addition to retrotransposition, expansion of this sequence is possible by other mechanisms such as insertion of chloroplast sequences into the nuclear genome. Chloroplast and mitochondrial DNA insertions into the nuclear genome are common in plants (Lin *et al.* 1999; Stupar *et al.* 2001; Guo *et al.* 2008; Ueda *et al.* 2005), and insertion of transposable elements into chloroplast or mitochondrial genomic insertions may contribute to rearrangement of these regions (Ueda *et al.* 2005; Guo *et al.* 2008). Identification of transcripts highly similar to chloroplast sequences with variable flanking sequences suggests that such insertions are also present in the *P. sylvestris* genome. The identified chloroplast sequences are either overlapping or are adjacent to each other in the



chloroplast genome. In addition, this chloroplast-like sequence is highly conserved within the plant kingdom and in addition to *Arabidopsis lyrata* (ORF64c), it is also present in most plastid genomes of green, red and brown algae. Insertion of this sequence into the nuclear genome can be found in the genomes of legumes (*Glycine max* and *Medicago truncatula*), *Populus trichocarpa* and *Solanaceae* (potato and tomato). In monocots, the rice genome contains at least eight insertions of this sequence into the nuclear genome, whereas five nuclear genome insertions are present in *Brachypodium distachyon*, and in *Sorghum bicolor* two insertions. The *Cereba*-like B153 and SA07 *Copia-18*\_BD-like transcripts from *Pinus sylvestris* could be a result of such insertions. Increased expression of the chloroplast-like region of the transcript was determined in the aphid and heat stress conditions but expression was also detected in control samples, whereas the *Cereba* RE-like region was also expressed in stressed samples but at significantly lower levels in comparison to the chloroplast-like region. This result indicates that in addition to the chloroplast genome locus, there could be several insertions of this chloroplast region sequence into nuclear DNA, which are associated with retroelements which can activate transcription of this fragment. The situation with the SA07 fragment is different, where the RE 3' non-coding internal sequence was also homologous to chloroplast sequences. Non-coding DNA between pol and the 3'LTR in some RE elements contains a conservative region with unknown function (Havecker *et al.* 2004). The SA07 and B153 fragment chloroplast sequences were similar to over 1000 plant ESTs, indicating that this chloroplast insertion is found in a large range of species, and it is transcribed (probably due to RE element insertions) in stress conditions. EST database searches using the RE-like part of the B153 sequence results in only one highly similar EST from *Pinus banksiana*, demonstrating that this RE family could be transcribed specifically in pine species. The expression of both transcripts (B153 and SA07) increases in aphid damaged trees and after heat stress, with the SA07 transcript having on average a two-fold higher expression level.

#### **4.4. Distribution of identified TE-like fragments among gymnosperms**

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 exists in distinct plant species (Flavell *et al.* 1992; Voytas *et al.* 1992; Xiong and Eickbush 1990). In genomes of gymnosperms, REs of common origin have also been described (Friesen *et al.* 2001; Stuart-Rogers and Flavell 2001). Therefore, identification of TE homologues from grasses and legumes in Scots pine transcripts was not unexpected. Evolutionary analysis of the *Copia* superfamily TE in *Triticeae*, rice and *Arabidopsis* genomes reveals that six ancient lineages existed before the divergence of monocots and dicots (Wicker and Keller 2007). A similar study of soybean, rice and maize REs revealed seven *Copia* and five *Gypsy* evolutionary lineages (Du *et al.* 2010). Different lineages share some common features, furthermore, some RE sequences retain an unexpected degree of conservation at the DNA level in distinct species and horizontal gene transfer has been suggested (Wicker and Keller 2007). In this study, transcripts were identified with homology

to conservative polyprotein sequences belonging to the *Copia* superfamily *Maximus* lineage (*Copia-2-TA* like fragments B104 and T30), as well as to both the *Angela* and *Ivana* lineages (fragment T12). In addition, transcripts with homology to the *Gypsy* superfamily *Tat* lineage (*Ogre*-like transcripts), *CRM* lineage (*Appalachian*, fragment B193) and *Athila* lineage (*PIGY-1*, fragment B301). The *Maximus*, *Athila* and *Tat* lineages TEs contain endogenous retroviral elements that suggest their ancient origin. Representatives of these lineages are not only present in the Scots pine genome, but might be also transcriptionally active in stress conditions.

Three full-length LTR TE-like constructs were identified in *P. taeda* BAC sequences (ptTE101, ptTE121) and *P. glauca* (psTE283), with homology to differentially expressed transcripts. Two of these constructs were subsequently confirmed to be recently described conifer TEs. psTE121 is homologous to *P. taeda* *PtCumberland\_I* (Wegrzyn, *Repbase*), and psTE283 is homologous to *PtAppalachian\_I* (Wegrzyn, *Repbase*). Structurally, ptTE101 and ptTE121 belong to the *Copia* LTR RE superfamily, and psTE283 belongs to the *Gypsy* superfamily. PCR amplification of domains from these elements from *P. sylvestris* genomic DNA results in a smear of fragments of various sizes, indicating that they are present in high copy number. A similar range of fragments is amplified from other pine species, indicating that these elements are also highly represented in the genomes of other pine species. Amplification of domains from the psTE283 *PtAppalachian\_I*-like TE is observed even in such distant species such as *Ceratozamia mexicana* and *Ginkgo biloba* as well as the conifers *Cedrus deodara* and *Tsuga canadensis*, but it is not as highly represented in *Picea*, *Abies* and *Larix*. In contrast, the ptTE101 element is highly represented in 10 pine species, but no amplification is observed in two pine species (*P. contorta* and *P. siberica*). This element was also found in *Picea rubens* and *Larix kaemferi*. The element ptTE121 is wide-spread in only seven pine species, with only weak amplification from the other pines. The distribution of these elements in gymnosperm species indicates that ptTE101 and ptTE121 have undergone a more recent expansion than psTE283. The uneven distribution of TEs between related species has been reported previously (Capy *et al.* 1994; Langdon *et al.* 2000; Friesen *et al.* 2001; Du *et al.* 2010; Kejnovsky *et al.* 2012).

#### **4.5. Transcriptional and transpositional activity of identified retrotransposons in *P. sylvestris***

Relatively short differentially expressed transcripts were analyzed (less than 10 kB), therefore no conclusions regarding the activation of full-length RE elements in the Scots pine genome can be made. Potentially, degenerate or truncated transcripts could be expressed, in addition to the chimeric sequences (Vignols *et al.* 1995; Betran *et al.* 2002; Wang *et al.* 2002; Rebollo *et al.* 2012). Analysis of such short fragments can indicate that they are similar to TE regions, but the origin of them cannot be determined. Conifer genomes are highly syntenic, and sequencing of the spruce genome has shown that spruce genes can contain very large introns (Nystedt *et al.* 2013). Considering that a non-autonomous TE can be only a few hundred bp long (Kumar, Bennetzen 1999), even the isolation of a full-length TE from cDNA may not be indicative of specific activation of a TE.

One full-length LTR element sequence (psTE283) and the functional domains of two REs (ptTE101 and ptTE121) were sequenced from cDNA obtained from stressed pine ramets, indicating that these are expressed as full-length REs. REs can significantly influence gene expression and splicing, even if they are found in introns (Vignols *et al.* 1995; Marillonnet, Wessler 1997). Two ORFs were identified within psTE283, encoding a reverse transcriptase and an integrase domain. Active REs are characterised by one continuous polyprotein ORF, and recently transposed REs have identical LTRs (Kumar, Bennetzen 1999). Multiple copies of REs complicate the isolation of a single locus, therefore the LTRs of a single element were not able to be compared as multiple copies of the RE are PCR amplified and these may contain polymorphisms between copies. However, due to the presence of two ORFs in consensus sequence (rather than one polyprotein ORF) as well as other features of the transcriptional regulatory motifs, the psTE283 element has probably lost transpositional activity, while maintaining expressional activity. The transcripts of the three full-length TEs were identified after heat stress, but some were found in other stress conditions as well. psTE283 transcripts were also identified after insect infestation and 64 hours after ABA and SA treatments. ptTE101 transcripts were only found after heat stress. ptTE121 transcripts were found in all stressed and hormone treated samples. Transcription of TEs is the initial step in transposition, and transpositionally active TEs are not required to encode functional proteins. There are many examples of transposition of nonautonomous TEs utilising proteins encoded by other TEs (Berg, Howe 1989; Flavell *et al.* 1994; Feschotte *et al.* 2002; Dewannieux *et al.* 2003; Kazazian, 2004).

Previous reports on pine REs have not described transcribed and active elements. Computational analyses of pine BACs have identified some elements with conserved LTRs that imply recent possible activity of these elements (Kovach *et al.* 2010). Analysis of polymorphism between ramets of a single clone identified RE that potentially could have induced the polymorphism by transposition. The IRAP marker technique has been used to identify somaclonal variation (Asif, Othman 2005; Lightbourn *et al.* 2007; Bairu *et al.* 2011; Campbell *et al.* 2011; Everensel *et al.* 2011; Bayram *et al.* 2012). IRAP primers amplify the genomic region between two sufficiently close elements, and transposition of a RE can induce amplification of additional fragments (Kalendar *et al.* 1999). Sequencing of the polymorphic fragments identified two groups of sequences with the same putative LTR *Copia* element, at one end, which may indicate that the same element inserted into different loci in different ramets. Further isolation and characterisation of this element is required to fully elucidate its activity.

Analysis of natural pine populations with non-specific IRAP 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 accurately characterise. The selected pine population was of natural origin and grew in differing topographic conditions that indicate long-standing differences in moisture conditions. Analysis of the number of amplified IRAP fragments showed a statistically significant increase in the average fragment number in the population that grew on the hill compared to the swamp population. Principal coordinates analysis (PcoA) drew differentiated the hill and swamp populations, while the

slope population was intermediate between the two. Alleles were equally represented in all populations, and the difference between the allele frequencies at each locus did not exceed 0.3. These results are consistent with the detailed wild barley *BARE-1* RE study, where the copy number of this element increased in populations growing in drought conditions (Kalendar *et al.* 2000). Changes of the number of amplification products in relation to geographical origin is also described in other studies, such as the increased number of amplification in north-western *Aegilops tauschii* populations (Saeid *et al.* 2008). Exact determination of copy number requires alternative methods; however these results provide an insight into the relative degree of variation and diversity of natural pine population determined by TE distribution.

#### **4.6. Phylogenetic analysis of gymnosperms using the developed retrotransposon markers**

Phylogenetic relationships of gymnosperms have been studied at the molecular level using *rbcL* (Chase *et al.* 1993; Hasebe *et al.* 1992), 18S rRNA (Chaw *et al.* 1997), 28S rRNA sequences (Stefanovic *et al.* 1998), nuclear and plastid small sub-unit rRNA (Troitsky *et al.* 1991) and mitochondrial sequences (Bowe *et al.* 2000; Ran *et al.* 2010). Conifers are also highly monophyletic (Raubeson, Jansen 1992; Bowe *et al.* 2000; Kan *et al.* 2007). However, some phylogenetic relationships have not been elucidated, e.g. the relative ages of some of the clades (Raubeson, Jansen 1992; Bowe *et al.* 2000; Donohue, Doyle 2000). It is known that *Pinus*, *Picea*, *Cathaya*, *Pseudotsuga* and *Larix* are monophyletic, but *Abies*, *Tsuga*, *Keteleeria*, *Nothotsuga* and *Pseudolarix*, but *Cedrus* forms another branch (Wang *et al.* 2000). A study utilising ITS and SSR markers separated *Larix* and *Pseudotsuga* from *Pinus*, *Picea* and *Cathaya* (Kan *et al.* 2007). In this study, some of the *Larix* species clustered with *Abies*, but not with *Pseudotsuga*, which was represented by only one species.

In a more detailed study of 10 *Pinus* species using 18 nuclear gene sequences, the genus was sub-divided into *Pinus* and *Strobus*. In the sub-genus *Pinus*, *P.contorta* and *P.banksiana* clustered together in the *Contortae* division, and *P.sylvestris* and *P.nigra* also formed one cluster (Palmé *et al.* 2009). In this study, the *Strobus* sub-family *Cembrae* was represented by *P.cembra*, *P.pumila*, *P.koraiensis* and *P.sibirica*, which were clustered together using both the specific TE-like markers. The IRAP marker data also clustered these species, together with the *Strobus* species *P.strobus* and *P.peuce*.

The specific TE-like marker data only differentiated *Pinus*, as the markers were developed from *P. sylvestris* sequences and the other species were not well represented. The non-specific IRAP marker data differentiated *Picea* and *Abies*, although the bootstrap values were low. This is largely determined by the non-specificity and small number of utilised markers. Also, the IRAP markers were designed using hypothetical LTR sequences that may not reflect the most widespread pine RE LTR sequences. However, the results show that the use of TE-derived markers have potential for gymnosperm phylogenetic studies as these markers analyse the polymorphism of the non-coding part of the genome and may provide additional information about phylogenetic relationships.

## CONCLUSIONS

1. Using the non-specific iPBS method it is possible to identify retrotransposon transcripts as well as chimeric transcripts containing parts of retrotransposons without previous sequence knowledge of species specific transposable elements.
2. Retrotransposon-like sequences are differentially expressed in *Pinus sylvestris* after heat stress, insect damage and treatment with plant hormones.
3. Transcriptionally active retrotransposons in *P.sylvestris* are weakly similar to angiosperm retrotransposons. Identification of highly conserved angiosperm retrotransposon-like sequences does not indicate presence of homologous retrotransposon elements in the *Pinus sylvestris* genome.
4. Transcripts homologous to chloroplast sequences were identified (SA07, ABA03, ABA02, B153, SA01) that are derived from expression of nuclear insertions of chloroplast sequences. Chloroplast-like sequences are intensively expressed after pine wooly aphid damage and heat shock. One identified chimeric transcript B153 contains both chloroplast and retrotransposon-like sequences.
5. A full length retrotransposon *Ps\_Appalachian-I* (psTE283) was sequenced and transcripts of this element were identified after heat stress, pine wooly aphid damage and after treatment with plant hormones (ABA and SA). *Ps\_Appalachian-I* belong to the LTR *Gypsy* class of retrotransposons and is homologous to *PtAppalachian\_I* from *P.taeda*. *Ps\_Appalachian-I* transcripts are derived from different copies in the complex genome of *P.sylvestris*. The hypothetical polyprotein ORF of this element contains stop codons, which indicates a limited capacity for transposition. However, the ancient origin of the *Ps\_Appalachian-I* element is confirmed by the distribution of its internal domain in distinct species of gymnosperms from *Cycadales* and *Gingkoales*.
6. Retrotransposons homologous to LTR *Copia* ptTE101 and ptTE121 are expressed as full-length transcripts in *P.sylvestris* after heat shock. PtTE121 transcripts were also identified after pine wooly aphid damage and after salicylic acid treatment. These retrotransposons are of more recent origin as their distribution in gymnosperms is uneven and they are particularly widespread in the *Pinus* genus.
7. Nine *P.sylvestris* specific IRAP markers were developed, which identified high genetic diversity in individuals from a natural pine population. These markers were able to differentiate sub-populations growing in differing topographic conditions. Differences in the number of amplification fragments

between sub-populations is significant for the markers IR\_2, IR\_5, IR\_6 and IR\_7.

8. IRAP analysis of somaclonal variation indicates possible recent transpositional activity a *P.sylvestris* retrotransposon homologous to the *Copia-17-Pta-I* LTR element.
9. The developed IRAP markers, as well as *P.sylvestris* specific retrotransposon markers are useful for identification of novel phylogenetic relationships in gymnosperms.

## ACKNOWLEDGMENTS

I am grateful to my research supervisor Dr. biol. Dainis Ruņģis for assistance in the understanding of the experimental results and valuable advice during the development and analysis of the work. I express my gratitude to the LSFRI “Silava” staff and especially to Dr.chem. Ilze Veinberga for support and encouragement to develop this study, Dr. biol. Agnese Gailite, Mg.math. Anita Gaile, Dr. silv. Vilnis Šķipars and Mg. biol. Viktorija Beļēviča for their advice and practical assistance for the experimental work and thesis writing. I am thankful to the Forest breeding group and especially to Dr. silv. Imants Baumanis for providing pine clonal material. I am thankful to Dr. silv. Āris Jansons for discussions about experimental requirements and help in the identification of the naturalpine stand. Many thanks to Mg. biol. Mārtiņš Zeps for assistance in measurement of pine trees. I thank Dr. biol. Nils Rostoks and his team for the fast and efficient help with cloning and transformation methods, as well as the advice given during the course of work.

I thank Dr. Alan Schulman, professor of the University of Helsinki and Institute of Biotechnology MTT/BI Plant Genomics Lab group leader, and his team for the encouragement of the development of this work and for consultations about plant retrotransposons and their analysis. I thank assistant professor Ruslan Kalendar for permission to use his developed PBS primers prior to publication.

I express my gratitude to Linda Strode, manager of the Department of dendroflora in the National Botanical Garden and assistant Inese Graudiņa for help in identifying and collecting specimens of the conifer species. I thank the staff of the Botanical Garden of the University of Latvia for assistance in identification and collection of exotic gymnosperm species.

Years 2009-2012 the research was supported by funding from the European Social Fund Project “Importance of Genetic Factors in Formation of Forest Stands with High Adaptability and Qualitative Wood Properties” (Nr. 2009/0200/1DP/1.1.1.2.0/09/APIA/VIAA/146).



## IZMANTOTĀS LITERATŪRAS SARAKSTS/ REFERENCES

- Altinkut A, Raskina O, Nevo E, Belyayev A (2006) *En/Spm*-like transposons in Poaceae species: transposase sequence variability and chromosomal distribution. *Cell Mol Biol Lett* 11(2):214-30
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
- Asif JM, Othman FY (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 Molecular Biology Reporter* 23(3): 241-249
- Bairu MW, Aremu AO, Staden JV (2011) Somaclonal variation in plants: causes and detection methods. *Plant Growth Regul* 63: 147-173
- Bayram E, Yilmaz S, Hamat-Mecbur H, Kartal-Alacam G, Gozukirmizi N (2012) *Nikita* retrotransposon movements in callus cultures of barley (*Hordeum vulgare* L.) *POJ* 5(3): 211-215
- Beguiristain T, Grandbastien MA, Puigdomènech P, Casacuberta JM (2001) Three *Tnt1* Subfamilies Show Different Stress-Associated Patterns of Expression in Tobacco. Consequences for Retrotransposon Control and Evolution in Plants. *Plant Physiol* 127: 212-221
- Benachenhou F, Sperber GO, Bongcam-Rudloff E, Andersson G, Boeke JD, Blomberg J (2013) Conserved structure and inferred evolutionary history of long terminal repeats (LTRs). *Mobile DNA*. doi: 10.1186/1759-8753-4-5
- Berg DE, Howe MH (1989) *Mobile DNA* (American Society for Microbiology Press, Washington, DC).
- Betran E, Thornton K, Long M. (2002) Retroposed new genes out of the X in *Drosophila*. *Genome Res* 12: 1854–1859.
- Birnboim HC, Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* 7(6): 1513-1523
- Boeke JD, Eickbush T, Sandmeyer SB, and Voytas DF (2006) *Index of Viruses – Pseudoviridae*. ICTVdB - The Universal Virus Database, version 4. Büchen-Osmond, C (Ed), Columbia University, New York, USA.
- Boerkoel CF, Kung HJ (1992) Transcriptional interaction between retroviral long terminal repeats (LTRs): mechanism of 5' LTR suppression and 3' LTR promoter activation of c-myc in avian B-cell lymphomas. *J Virol* 66(8): 4814-23
- Bowe LM, Coat G, dePamphilis CW (2000) Phylogeny of seed plants based on all three genomic compartments: extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers. *Proc Natl Acad Sci USA* 97(8): 4092-4097
- Brandes A, Heslop-Harrison JS, Kamm A, Kubis S, Doudrick RL, 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
- Brunner S, Fengler K, Morgante M, Tingey S, Rafalski A (2005) Evolution of DNA sequence nonhomologies among maize inbreds. *Plant Cell* 17: 343-360

Butelli E, Licciardello C, Zhang Y, Liu J, Mackay S, Bailey P, Reforgiato-Recupero G, Martin C (2012) Retrotransposons Control Fruit-Specific, Cold-Dependent Accumulation of Anthocyanins in Blood Oranges. *Plant Cell* 24:1242–1255

Campbell BC, LeMare S, Piperidis G, Godwin ID (2011) IRAP, a retrotransposon-based marker system for the detection of somaclonal variation in barley. *Mol Breeding* 27:193-206

Capy P (2005) Classification and nomenclature of retrotransposable elements. *Cytogenet Genome Res* 110(1-4):457-461

Capy P, Anxolabéhère D, Langin T (1994) The strange phylogenies of transposable elements: are horizontal transfers the only explanation? *Trends Genet* 10: 7-12

Chase MW, Soltis DE, Olmstead RG *et al.* (1993) DNA sequence phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. *Ann MO Bot Gard* 80: 528-580

Chaw SM, Zharkikh A, Sung HM, Lau TC, Li WH (1997) Molecular phylogeny of extant gymnosperms and seed plant evolution: analysis of nuclear 18S rRNA sequences. *Mol Biol Evol* 14(1): 56-68

Chopra S, Brendel V, Zhang J, Axtell JD, Peterson T (1999) Molecular characterization of a mutable pigmentation phenotype and isolation of the first active transposable element from *Sorghum bicolor*. *Proc Natl Acad Sci USA* 96:15330–15335

Cossu RM, Buti M, Giordani T, Natali L, Cavallini A (2012) A computational study of the dynamics of LTR Retrotransposons in the *Populus trichocarpa* genome. *Tree Genet Genomes* 8(1):61-75

Curcio M., Derbyshire, K. (2003). The outs and ins of transposition: from mu to kangaroo. *Nature Rev. Mol. Cell Biol.* 4: 865–877

Dewannieux M, Esnault C, Heidmann T (2003) LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet* 35(1):41-8

Donoghue MJ, Doyle JA (2000) Seed Plant Phylogeny: Demise of the Anthophyte Hypothesis. *Curr Biol* 10(3): R106-R109

Doolittle WF, Sapienza C (1980) Selfish genes, the phenotype paradigm and genome evolution. *Nature* 284: 601–603

Du J, Tian Z, Hans CS, Laten HM, Cannon SB, Jackson SA, Shoemaker RC, Ma J (2010) Evolutionary conservation, diversity and specificity of LTR-retrotransposons in flowering plants: insights from genome-wide analysis and multi-specific comparison. *Plant J.* 63(4): 584-598

Durrant WE, Dong X (2004) Systemic Acquired Resistance. *Annu Rev Phytopathol* 42:185-209

Echenique V, Stamova B, Wolters P, Lazo G, Carollo VL, Dubcovsky J (2002) Frequencies of *Ty1-copia* and *Ty3-gypsy* retroelements within the Triticeae EST databases. *Theor Appl Genet* 104:840–844

Emerson JJ, Kaessmann H, Betran E, Long M (2004) Extensive gene traffic on the mammalian X chromosome. *Science* 303: 537–540

Feng G, Leem Y, Levin HL (2012) Transposon integration enhances expression of stress response genes. *Nucleic Acids Res.* doi: 10.1093/nar/gks1185

Feschotte C, Jiang N, Wessler SR (2002) Plant transposable elements: where genetics meets genomics. *Nature Reviews Genetics* 3: 329-341



- Finnegan DJ (1989) Eukaryotic transposable elements and genome evolution. *Trends Genet* 5: 103–107
- Finnegan EJ, Genger RK, Peacock WJ, Dennis ES (1998) DNA Methylation in plants. *Annu Rev Plant Physiol Plant Mol Biol* 49:223-247
- Flavell AJ, Dunbar E, Anderson R, Pearce SR, Hartley R, Kumar A (1992) *Ty1-copia* group Retrotransposons are ubiquitous and heterogeneous in higher plants. *Nucleic Acids Res* 20:3639–3644
- Flavell AJ, Pearce SR, Kumar A (1994) Plant transposable elements and the genome. *Curr Opin Genet Dev* 4: 838–844
- Friesen N, Brandes A, Heslop-Harrison JS (2001) Diversity, Origin and Distribution of Retrotransposons (*gypsy* and *copia*) in Conifers. *Mol Biol Evol* 18(7): 1176-1188
- Gabriel A, Mules EH (1999) Fidelity of Retrotransposon Replication. *Ann N Y Acad Sci* 18: 108-118
- Gama Sosa MA, Rosas DH, DeGasperi R, Morita E, Hutchison MR, Ruprecht RM (1994) Negative regulation of the 5' long terminal repeat (LTR) by the 3' LTR in the murine proviral genome. *J Virol* 68(4): 2662–2670
- Gao D, Chen J, Chen M, Meyers BC, 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 MA, Audeon C, Casacuberta JM, Grappin P, Lucas H, Moreau C, Pouteau S (1994) Functional analysis of the tobacco *Tnt1* retrotransposon. *Genetica* 93(1-3): 181-189
- Grandbastien M-A, Lucas H, Morel JB, Corinne MC, Vernhettes S, Casacuberta JM (1997) The expression of the tobacco *Tnt1* Retrotransposon is linked to plant defense responses. *Genetica* 100:241–252
- 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
- Guo X, Ruan S, Hu W, Cai D, Fan L (2008) Chloroplast DNA insertions into the nuclear genome of rice: the genes, sites and ages of insertion involved. *Funct Integr Genomics*. doi:10.1007/s10142-007-0067-2
- Hasebe M., Kofuji R, Ito M, Kato M, Iwatsuki K, Ueda K (1992) Phylogeny of gymnosperms inferred from *rbcL* gene sequences. *Bot Mag Tokyo* 105: 673-679
- Hashida S, Kitamura K, Mikami T, Kishima Y (2003) Temperature shift coordinately changes the activity and the methylation state of transposon *Tam3* in *Antirrhinum majus*. *Plant Physiol* 132(3):1207-1216
- Havecker ER, Gao X, Voytas DF (2004) The diversity of LTR retrotransposons. *Genome Biology*. doi:10.1186/gb-2004-5-6-225
- Hawkins JS, Kim H, Nason JD, Wing RA, Wendel JF (2006) Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*. *Genome Res* 16: 1252-1261
- Hawkins JS, Proulx SR, Rapp RA, Wendel JF (2009) Rapid DNA loss as a counterbalance to genome expansion through retrotransposon proliferation in plants. *Proc Natl Acad Sci USA* 106(42): 17811-17816
- He P, Ma Y, Zhao G, Dai H, Li H, Chang L, Zhang Z (2010) FaRE1: a transcriptionally active *Ty1-copia* Retrotransposon in strawberry. *J Plant Res* 123(5):707-714

- Hirochika H (1993) Activation of tobacco Retrotransposons during tissue culture. *EMBO J* 12(6):2521-2528
- 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
- Jin YK, Bennetzen JL (1994) Integration and nonrandom mutation of a plasma membrane proton ATPase gene fragment within the *BsI* retroelement of maize. *Plant Cell*. 6(8): 1177-1186
- Johns MA, Mottinger J, Freeling MA (1985) A low copy number, *copia*-like transposon in maize. *EMBO J*. 4, 1093 – 1102
- Jurka J (2005) OSHOOTER: *EnSpm*-type DNA transposon from rice. *Rebase Reports* 5(8):205-205
- Kalendar R, Antonius K, Smykal P, Schulman AH (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, Tanskanen J, Immonen S, Nevo E, Schulman AH (2000) Genome evolution of wild barley (*Hordeum spontaneum*) by *BARE-1* Retrotransposon dynamics in response to sharp microclimatic divergence. *Proc Natl Acad Sci USA* 97(12):6603-6607
- Kalendar R, Tanskanen JA, Chang W, Antonius K, Sela H, Peleg P, Schulman AH (2008) *Cassandra* Retrotransposons carry independently transcribed 5S RNA. *Proc Natl Acad Sci USA* 105:5833–5838
- Kamm A, Doudric RL, Heslop-Harrison JS, 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
- Kan XZ, Wang SS, Ding X, Wang XQ (2007) Structural evolution of nrDNA ITS in *Pinaceae* and its phylogenetic implications. *Mol Phylogen Evol* 44(2): 765-777
- Kashkush K, Feldman M, Levy AA (2003) Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nat Genet* 33(1): 102-106.
- Kato H, Sriprasertsak P, Seki H, Ichinose Y, Shiraishi T, Yamada T (1999) Functional Analysis of Retrotransposons in Pea. *Plant Cell Physiol* 40(9): 933-41
- Kazazian Jr HH (2004) Mobile elements: Drivers of genome evolution. *Science* 303: 1626–1632
- Kazazian Jr HH (2011) Mobile DNA finding treasure in junk. Pearson Education Inc. ISBN-10: 0-13-707062-4
- Kejnovsky E, Hawkins JS, Feschotte C (2012) Plant Transposons and Genome Dynamics in Evolution. Ed. Fedoroff NV, Wiley-Blackwell, US
- Kidwell MG, Lisch DR (2001) Perspective: transposable elements, parasitic DNA, and genome evolution. *Evolution* 55(1): 1-24

- Kimura Y, Tosa Y, Shimada S, Sogo R, Kusaba M, Sunaga T, Betsuyaku S, Eto Y, Nakayashiki H, Mayama S (2001) *OARE-1*, a *Ty1-copia* Retrotransposon in Oat Activated by Abiotic and Biotic Stresses. *Plant Cell Physiol* 42 (12):1345-1354
- Klaver B, Berkhout B (1994) Comparison of 5' and 3' long terminal repeat promoter function in human immunodeficiency virus. *J Virol* 68(6): 3830-3840
- Klessig DF, Durner J, Noad R, Navarre DA, Wendehenne D, Kumar D, Zhou JM, Shah J, Zhang S, Kachroo P, Trifa Y, Pontier D, Lam E, Silva H (2000) Nitric oxide and salicylic acid signaling in plant defense. *Proc Natl Acad Sci U S A* 97(16): 8849-8855
- Kobayashi S, Goto-Yamamoto N, Hirochika H (2004) Retrotransposon-Induced Mutations in Grape Skin Color. *Science* 304:982
- Kohany O, Gentles AJ, 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
- Komatsu M, Shimamoto K, Kyojuka J (2003) Two-Step Regulation and Continuous Retrotransposition of the Rice *LINE*-Type Retrotransposon Karma. *Plant Cell* 15:1934-1944
- Kossack DS, Kinlaw CS (1999) IFG, a *gypsy*-like retrotransposon in *Pinus* (*Pinaceae*), has an extensive history in pines. *Plant Mol Biol* 39: 417-426
- Kovach A, Wegrzyn JL, Parra G, Holt C, Bruening GE, Loopstra CA, Hartigan J, Yandell M, Langley CH, Korf I, Neale DB (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 JL (1999) Plant Retrotransposons. *Annu Rev Genet* 33: 479-532
- Kumekawa N, Ohtsubo H, Horiuchi T, Ohtsubo E (1999) Identification and characterization of novel retrotransposons of the *gypsy* type in rice. *Mol Gen Genet* 260: 593-602
- L'Homme Y, Seguin A, Tremblay FM (2000) Different classes of retrotransposons in coniferous spruce species. *Genome* 43: 1084-1089
- Langdon T, Seago C, Mende M, Leggett M, Thomas H, Forster JW, Jones RN, Jenkins G (2000) Retrotransposon evolution in diverse plant genomes. *Genetics* 156(1): 313-325
- Lightbourn GJ, Jelesko JG, Veilleux RE (2007) Retrotransposon-based markers from potato monoloids used in somatic hybridization. *Genome* 50(5): 492-501
- Lin X, Kaul S, Rounsley S, Shea TP, Benito MI, Town CD, Fujii CY, Mason T, Bowman CL, Barnstead M, Feldblyum TV, Buell CR, Ketchum KA, Lee J, Ronning CM, Koo HL, Moffat KS, Cronin LA, Shen M, Pai G, Van Aken S, Umayam L, Tallon LJ, Gill JE, Adams MD, Carrera AJ, Creasy TH, Goodman HM, Somerville CR, Copenhaver GP, Preuss D, Nierman WC, White O, Eisen JA, Salzberg SL, Fraser CM, Venter JC. (1999) Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature* 402(6763):761-768
- Llorens C, Futami R, Covelli L, Dominguez-Escriba L, Viu JM, Tamarit D, Aguilar-Rodriguez J, Vicente-Ripolles M, Fuster G, Bernet GP, Maumus F, Munoz-Pomer A, Sempere JM, LaTorre A, Moya A (2011) The Gypsy Database (*GyDB*) of mobile genetic elements: release 2.0 *Nucleic Acids Res*. doi:10.1093/nar/gkq1061

- Lorenz R, Bernhart SH, Siederdisen HC, Tafer H, Flamm C, Stadler PF, Hofacker IL (2011) ViennaRNA Package 2.0. Algorithms Mol Biol. doi: 10.1186/1748-7188-6-26
- Madlung A, Comai L (2004) The Effect of Stress on Genome Regulation and Structure. Annals of Botany 94: 481-495
- Mak J, Kleiman L (1997) Primer tRNAs for reverse transcription. J Virol 71(11): 8087-8095
- Malamy J, Carr JP, Klessig DF, Raskin I (1990) Salicylic acid: A likely endogenous signal in the resistance response of tobacco to viral infection. Science 250:1002–1004
- Malamy J, Klessig DF (1992) Salicylic acid and plant disease resistance. Plant J 2(5):643-654
- Marillonnet S, Wessler SR (1997) Retrotransposon insertion into the maize waxy gene results in tissue-specific RNA processing. Plant Cell 9: 967 – 978
- Mattick JS, Makunin IV (2006) Non-coding RNA. Hum Mol Genet 15:17-29
- May BP, Lippman ZB, Fang Y, Spector DL, Martienssen RA (2005) Differential regulation of strand-specific transcripts from *Arabidopsis* centromeric satellite repeats. PLoS Genet. doi: 10.1371/journal.pgen.0010079
- McClintock B (1954) Mutations in maize and chromosomal aberrations in Neurospora. Carnegie I Wash 53:254-261
- McClintock B (1984) The significance of responses of the genome to challenge. Science 226:792–801
- Melayah D, Bonnard E, Chalhoub B, Audeon C, Grandbastien M (2001) The mobility of the tobacco *Tnt1* Retrotransposon correlates with its transcriptional activation by fungal factors. Plant J 28(2):159-168
- Mieczkowski PA, Lemoine FJ, Petes TD (2006) Recombination between retrotransposons as a source of chromosome rearrangements in the yeast *Saccharomyces cerevisiae*. DNA repair 5: 1010–1020
- Miguel C, Simoes M, Oliveira MM, Rocheta M (2008) Envelope-like retrotransposons in the plant kingdom: evidence of their presence in *Gymnosperms* (*Pinus pinaster*). J Mol Evol 67: 517-525
- Morgante M, De Paoli E, Radovic S (2007) Transposable elements and the plant pan-genomes. Curr Opin Plant Biol 10:149-155
- Morse AM, Peterson DG, Islam-Faridi MN, Smith KE, Magbanua Z, Garcia SA, Kubisiak TL, Amerson HV, Carlson JE, Nelson CD, Davis JM (2009) Evolution of Genome Size and Complexity in *Pinus*. PLoSone. doi:10.1371/journal.pone.0004332
- Mourier T, and Willerslev E (2009) Retrotransposons and non-protein coding RNAs. Brief Funct Genomic Proteomic 8(6):493-501
- Murray BG (1998) Nuclear DNA Amounts in Gymnosperms. Ann Bot 82:3-15
- Murray BG (2005) When does Intraspecific C-value Variation become Taxonomically Significant? Ann Bot 95: 119-125
- Neumann P, Koblizkova A, Navratilova A, Macas J (2006) Significant expansion of *Vicia pannonica* genome size mediated by amplification of a single type of giant retroelement. Genetics 173:1047-1056
- 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

- Neumann P, Yan H, Jiang J (2007) The centromeric retrotransposons of rice are transcribed and differentially processed by RNA interference. *Genetics* 176: 749–761
- Nevo E (2001) Evolution of genome–phenome diversity under environmental stress. *Proc Natl Acad Sci USA* 98(11): 6233-6240
- Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin YC, Scofield DG, 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 JA, Sena J, Svensson T, Talavera-López C, Theißen G, Tuominen H, Vanneste K, Wu ZQ, Zhang B, Zerbe P, Arvestad L, Bhalerao R, Bohlmann J, Bousquet J, Garcia Gil R, Hvidsten TR, de Jong P, MacKay J, Morgante M, Ritland K, Sundberg B, Thompson SL, Van de Peer Y, Andersson B, Nilsson O, Ingvarsson PK, Lundeberg J, Jansson S (2013) The Norway spruce genome sequence and conifer genome evolution. *Nature*. doi:10.1038/nature12211
- Orgel LE, Crick FHC (1980) Selfish DNA—the ultimate parasite. *Nature* 284: 604–607
- Palmé AE, Pyhäjärvi T, Wachowiak W, Savolainen O (2009) Selection on nuclear genes in a *Pinus phylogeny*. *Mol Biol Evol* 26(4): 893-905
- Panstruga R, Büschges R, Piffanelli P, Schulze-Lefert P (1998) A contiguous 60 kb genomic stretch from barley reveals molecular evidence for gene islands in a monocot genome. *Nucleic Acids Res* 26(4): 1056-1062
- Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6: 288-295
- Peterson PA (1953) A mutable pale green locus in maize. *Genetics* 38:682-683
- Piegu B, Guyot R, Picault N, Roulin A, Sanyal A, Kim H, Collura K, Brar DS, Jackson S, Wing RA, Panaud O. (2006) Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. *Genome Res* 16: 1262-1269
- Piriyapongsa J, Marino-Ramirez L, Jordan IK (2007) Origin and Evolution of Human microRNAs From Transposable Elements. *Genetics* 176:1323–1337
- Ponting CP, Oliver PL, Reik W (2009) Evolution and Functions of Long Noncoding RNAs. *Cell* 136:629-641
- Pouteau S, Huttner E, Grandbastien MA, Caboche M (1991) Specific expression of the tobacco *Tnt1* retrotransposon in protoplasts. *EMBO J* 10(7): 1911-1918
- Preston BD (1996). Error-prone retrotransposition: rime of the ancient mutators. *Proc Natl Acad Sci USA* 93: 7427–7431
- Ramallo E, Kalendar R, Schulman AH, Martínez-Izquierdo JA (2008) *Remel*, a *Copia* retrotransposon in melon, is transcriptionally induced by UV light. *Plant Mol Biol* 66(1-2): 137-150
- Ran JH, Gao H, Wang XQ (2010) Fast evolution of the retroprocessed mitochondrial *rps3* gene in Conifer II and further evidence for the phylogeny of gymnosperms. *Mol Phylogenet Evol* 54(1): 136-149
- Raubeson LA, Jansen RK (1992) A rare chloroplast DNA structural mutation is shared by all conifers. *Biochem Sys. Evol* 20: 17-24
- Rebollo R, Romanish MT, Mager DL (2012) Transposable Elements: An Abundant and Natural Source of Regulatory Sequences for Host Genes. *Annu Rev Genet* 46:21-42

Rocheta M, Carvalho L, Viegas W, Morais-Cecilio L (2012) *Corky*, a gypsy-like Retrotransposon is differentially transcribed in *Quercus suber* tissues. BMC Res Notes. doi:10.1186/1756-0500-5-432

Rocheta M, Cordeiro J, Oliveira M, Miguel C (2007) *PpRT1*: the first complete gypsy-like retrotransposon isolated in *Pinus pinaster*. Planta 225: 551-562.

Rostoks N, Park Y-J, Ramakrishna W, Ma J, Druka A, *et al.* (2002) Genomic sequencing reveals gene content, genomic organization, and recombination relationships in barley. Functional & Integrative Genomics 2: 51–59

Sabot F, Schulman AH (2006) Parasitism and the retrotransposon life cycle in plants: a hitchhiker's guide to the genome. Heredity 97: 381-388

Sabot F, Schulman AH (2007) Template switching can create complex LTR retrotransposon insertions in *Triticeae* genomes. BMC Genomics. doi:10.1186/1471-2164-8-247

Saeidi H, Rahiminejad MR, Heslop-Harrison JS (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

SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z, Bennetzen JL (1996) Nested retrotransposons in the intergenic regions of the maize genome. Science 274: 737–738

Sayah DM, Sokolskaja E, Berthoux L, Luban J (2004) Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1. Nature 430: 569–573

Schlüter PM, Harris SA (2006) Analysis of multilocus fingerprinting data sets containing missing data. Mol Ecol Notes 6: 569-572

Schmittgen TD, Livak KJ (2008) Analyzing real-time PCR data by the comparative C (T) method. Nat Protoc 3:1101–1108

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(7):908-15

Stupar RM, Lilly JW, Town CD, Cheng Z, Kaul S, Buell CR, Jiang J (2001) Complex mtDNA constitutes an approximate 620-kb insertion on *Arabidopsis thaliana* chromosome 2: implication of potential sequencing errors caused by large-unit repeats. Proc Natl Acad Sci USA 98:5099-5103

Suoniemi A, Narvanto A, Schulman AH (1996) The *BARE-1* retrotransposon is transcribed in barley from an LTR promoter active in transient assays. Plant Mol Biol 31:295-306

Takano M, Kanegae H, Shinomura T, Miyao A, Hirochika H, Furuya M (2001) Isolation and characterization of rice phytochrome A mutants. Plant Cell 13: 521 – 534.

Takeda S, Sugimoto K, Otsuki H, Hirochika H (1999) A 13-bp cis-regulatory element in the LTR promoter of the tobacco Retrotransposon *Tto1* is involved in responsiveness to tissue culture, wounding, methyl jasmonate and fungal elicitors. Plant J 18(4):383-393

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Method. Molecular Biology and Evolution 28: 2731-2739

- Tapia G, Verdugo I, Yañez M, Ahumada I, Theoduloz C, Cordero C, Poblete F, González E, Ruiz-Lara S (2005) Involvement of Ethylene in Stress-Induced Expression of the *TLC1.1* Retrotransposon from *Lycopersicon chilense* Dun. *Plant Physiol* 138(4): 2075–2086
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22(22):4673-4680
- Topp CN, Zhong CX, Dawe RK (2004) Centromere-encoded RNAs are integral components of the maize kinetochore. *Proc Natl Acad Sci USA* 101: 15986–15991
- Troitsky AV, Melekhovets YuF, Rakhimova GM, Bobrova VK, Valiejo-Roman KM, Antonov AS. (1991) Angiosperm origin and early stages of seed plant evolution deduced from rRNA sequence comparison. *J Mol Evol* 32: 255-261
- Turcich MP, Bokhari-Riza A, Hamilton DA, He C, Messier W, Stewart CB, Mascarenhas JP (1996) *PREM-2*, a *copia*-type retroelement in maize is expressed preferentially in early microspores. *Sex Plant Reprod* 9: 65–74
- Ueda M, Tsutsumi N, Kadowaki K (2005) Translocation of a 190-kb mitochondrial fragment into rice chromosome 12 followed by the integration of four Retrotransposons. *Int J Biol Sci* 1(3): 110-113
- Varagona MJ, Purugganan M, Wessler SR (1992) Alternative splicing induced by insertion of retrotransposons into the maize waxy gene. *Plant Cell* 4: 811 – 820
- Vernhettes S, Grandbastien M-A, Casacuberta JM (1998) The evolutionary analysis of the *Tnt1* Retrotransposon in *Nicotiana* species reveals the high variability of its regulatory sequences. *Mol Biol Evol* 15:827–836
- Vicient CM (2010) Transcriptional activity of transposable elements in maize. *BMC Genomics*. doi:10.1186/1471-2164-11-601
- Vicient CM, Kalendar R, Schulman AH (2005) Variability, recombination and mosaic evolution of the barley *BARE-1* retrotransposon. *J Mol Evol* 61: 275–291
- Vicient CM, Suoniemi A, Anamthawat-Jónsson K, Tanskanen J, Beharav A, Nevo E, Schulman AH (1999) Retrotransposon *BARE-1* and Its Role in Genome Evolution in the Genus *Hordeum*. *Plant Cell* 11(9): 1769-1784
- Vignols F, Rigau J, Torres MA, Capellades M, Puigdomenech P (1995) The brown midrib3 (bm3) mutation in maize occurs in the gene encoding caffeic acid O-methyltransferase. *Plant Cell* 7: 407 – 416
- Voronova A, Jansons Ā, Ruņģis D (2011) Expression of retrotransposon-like sequences in Scots pine (*Pinus sylvestris*) in response to heat stress. *Environmental and Experimental Biology* 9: 121-127
- Voytas D F, Cummings MP, Konieczny A, Ausubel FM, Rodermel SR (1992) *Copia*-like Retrotransposons are ubiquitous among plants. *Proc Natl Acad Sci USA* 89:7124–7128
- Wang W, Brunet FG, Nevo E, Long M (2002) Origin of sphinx, a young chimeric RNA gene in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* 99: 4448–4453
- Wang W, Zheng H, Fan C, Li J, Shi J, Cai Z, Zhang G, Liu D, Zhang J, Vang S, Lu Z, Wong GK, Long M, Wang J (2006) High rate of chimeric gene origination by retroposition in plant genomes. *Plant Cell* 18(8): 1791-1802
- Wang XQ, Tank DC, Sang T (2000) Phylogeny and divergence times in *Pinaceae*: evidence from three genomes *Mol Biol Evol* 17: 773–781

- Wessler S (2001) Plant Transposable Elements. A Hard Act to Follow. *Plant Physiology* 125: 149–151
- Wessler SR (1996) Plant Retrotransposons: Turned on by stress. *Curr Biol* 6 (8): 959–961
- Wicker T, Guyot R, Yahiaoui N, Keller B (2003) CACTA Transposons in *Triticeae*. A Diverse Family of High-Copy Repetitive Elements. *Plant Physiol* 132:52–63
- Wicker T, Keller B (2007) Genome-wide comparative analysis of *copia* retrotransposons in *Triticeae*, rice, and *Arabidopsis* reveals conserved ancient evolutionary lineages and distinct dynamics of individual copia families. *Genome Res* 17(7):1072–1081
- Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P, Schulman AH (2007) A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* 8(12): 973–982
- Wilkinson S, Davies W (2002) ABA-based chemical signaling: the co-ordination of responses to stress in plants. *Plant Cell Environ* 25:195–210
- Xiong Y, Eickbush TH (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J* 9:3353–3362
- Yadav CB, Singh HN (2013) In-Silico Identification of LTR type Retrotransposons and Their Transcriptional Activities in *Solanum Tuberosum*. *IJSCE* 3(1):160–164
- 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
- Zhang J, Dean AM, Brunet F, Long M (2004) Evolving protein functional diversity in new genes of *Drosophila*. *Proc Natl Acad Sci USA* 101:16246–16250
- Zhang R, Zhang L, Yu W (2012) Genome-wide expression of non-coding RNA and global chromatin modification. *Acta Biochim Biophys Sin* 44(1): 40–47
- Zhang Y, Wu Y, Liu Y, Han B (2005) Computational identification of 69 retrotransposons in *Arabidopsis*. *Plant Physiol*. 138: 935–948