

Thaumatin-like protein gene copy number variation in Scots pine (*Pinus sylvestris*)

Vilnis Šķipars, Baiba Krivmane, Dainis Ruņģis*

Latvian State Forest Research Institute “Silava” Genetic Resource Centre, Rigas 111, Salaspils LV-2169, Latvia

*Corresponding author, E-mail: dainis.rungis@silava.lv

Abstract

Copy number variations (CNVs) are structural variations in genomes, resulting from the duplication of a section of DNA that is usually from one kilobase to several megabases in size. The aim of this research was to investigate gene copy number variation in Scots pine (*Pinus sylvestris*). Twenty nine Scots pine trees from a research pine plantation were tested for CNVs of a thaumatin-like protein (TLP) gene, by utilising real-time PCR with TaqMan probes and a relative quantification approach. DNA quantitation results indicated that CNVs of the *TLP* gene were detected in the analysed samples. The results were confirmed by repeating the quantitation analysis using a different endogenous control. In addition, all except one members of an open-pollinated half sib family had increased copy numbers of the *TLP* gene, indicating the genetic basis of these results. The results presented in this paper are to our knowledge the first reported results of CNV polymorphism in a conifer species.

Key words: copy number variations, *Pinus sylvestris* L., real-time PCR, thaumatin-like protein.

Abbreviations: CNV, copy number variation; PsBBs, pinosylvin synthase; *PsR*, *Pinus sylvestris* resistance gene (also known as *PsACRE*); TLP, thaumatin-like protein.

Introduction

Scots pine (*Pinus sylvestris*) is an important forest tree species in northern Europe. The pathogenic fungus *Heterobasidion annosum*, the main agent of root rot in pine, causes large economic losses for pine forest owners. In Norway spruce stands in Latvia butt rot related losses account for 800 to 4790 EUR ha⁻¹ (Gaitnieks et al. 2007) indicating that considerable economic losses are most probably also inflicted by *H. annosum* root rot to Scots pine stands. Root rot in pines is mainly localised in the root system, and the infection leads to increased risk of wind throw, reduced growth and, ultimately, to death of the tree. The infection spreads mainly via root contacts and the growth rate is up to 2 m year⁻¹ (Risbeth 1962). The host can remain asymptomatic, even if half of the root system is already colonised by the pathogen, leading to delayed discovery of the problem in forest stands and in individual trees. The only method to detect *H. annosum* infection by visual inspection of a living tree is by presence of fruiting bodies of *H. annosum*, but the formation of fruiting bodies is dependent on climatic conditions; if the summer is dry, there will be less fruiting bodies. Fruiting bodies are often very small and covered with moss, and located at the very base of the stem, making them hard to find. Even if disease symptoms are present, they are not clearly distinguishable from those inflicted by other root pathogens on visual examination of living trees. Additionally, if a forest stand

has been infected with *H. annosum*, it is virtually impossible to decontaminate it and other, less susceptible tree species need to be planted instead, but many local tree species are susceptible to *H. annosum*, which makes the choice difficult. These problems raise the importance of identification and breeding of pines with increased resistance against root rot.

While tree lines or trees with differing levels of resistance to *H. annosum* have been described in *Picea abies* and *P. sylvestris* (Fossdal et al. 2006; Korshikov, Demkovich 2008), no absolutely resistant clone has been identified, due to the quantitative nature of resistance to root rot, which is most probably not dependent on a single gene of large impact; rather, it depends of a number of genes of smaller effect. Many antifungal proteins have been described in various species in virtually all taxa, including thaumatin-like proteins, which are described as potent antimicrobial agents against a wide range of plant and some human pathogens (Selitrennikoff 2001). Furthermore, thaumatin-like proteins (TLP) are of interest in practical applications in plant disease control and other fields, as the structure of thaumatin-like proteins includes eight disulphide bonds, which make these proteins very stable (Roberts, Selitrennikoff 1990). Genes encoding TLP have been also identified in conifers, including *Pinus taeda* and *P. sylvestris*. It has been clarified that the expression of these genes increases in *P. sylvestris* after inoculation with *H. annosum* (Li, Asiegbu 2004b), but the mode of action of TLP is not fully understood. These proteins have not been extracted

from Scots pine and their effect on growth of *H. annosum* needs to be tested *in vitro*.

Copy number variations (CNVs) are structural variations in genomes, resulting from the duplication of a section of DNA that is usually from one kilobase to several megabases in size (Stankiewicz, Lupski 2010). If the duplicated regions contain functional genes, then this can result in variation in gene copy number between individuals, leading to differing levels of gene expression. CNVs were initially studied in detail in the human genome (Iafra et al. 2004; Sebat et al. 2004), but recently they have begun to be studied in plants. Although the majority of studies examining CNV have been performed using high throughput arrays (Lucito et al. 2001; Pollack et al. 2002; Järvinen et al. 2008; Myllykangas et al. 2008), real-time PCR is often used to confirm these results (Aarhus et al. 2010; Zhang et al. 2011) as well as in independent studies (Chen et al. 2006; Rose-Zerilli et al. 2009).

There have been no published studies investigating CNV in conifers. Studies in other species have indicated that CNV can be linked to variation in the gene expression level (Bradeen et al. 2009; Orozco et al. 2009; Ortiz-Estevéz et al. 2011), and that CNV is a major component of genetic variation within the human genome (Li et al. 2009). The copy number variation of the *TLP* gene was chosen for study, as increased copy number of an antifungal gene might lead to an increased amount of its product in tree cells prior to infection, compared to other trees, or to stronger induction of gene expression after infection. Increased levels of antifungal proteins might make a tree less susceptible to pathogen infection, or lead to stronger defence response after infection. Existence of CNV of a particular gene does not imply differences in gene expression, and expression analyses are required to detect expression differences as a result of CNV. However, gene expression analyses are biologically and technically more demanding than analysis of genomic DNA, particularly in mature individuals of conifer tree species. Creating controlled experimental conditions when studying mature pine trees growing in forest stands is difficult, particularly with regard to environmental fluctuation and availability of genetically identical individuals (clones). In addition, gene expression analyses require careful consideration about the type of tissue and time points to be collected. Even if no correlation between gene copy number and mRNA levels is detected, it is possible that at different developmental stages, genes can be regulated differently. Investigation of variation of gene copy number can identify potential genes for further gene expression analyses, which are more experimentally demanding.

Thaumatin-like proteins are encoded by a family of related genes in *Pinus monticola* (Liu et al. 2010) and possibly in other conifer species as well. However, only one *TLP* gene sequence from *P. taeda* is available from the NCBI database, and this sequence was used for primer

and probe construction for CNV analyses conducted in this study. *TLP* genes of *P. monticola* have been reported to be involved in constitutive and induced defense against pathogens, wounding, and cold-hardiness, they have disparate expression profiles in different tissues, and DNA and deduced amino acid sequences of these genes are highly diverse (Liu et al. 2010).

In this study, twenty nine Scots pine trees from a research pine plantation were tested for CNV of a *TLP* gene, by utilising real-time PCR with TaqMan probes and a relative quantification approach.

Materials and methods

Plant material

Samples were collected in a 29 year old experimental pine (*Pinus sylvestris* L.) forest plantation near Kalsnava, Latvia (experiment Nr. 235, register of long-term research sites). Pine needles were collected from 29 trees from 15 different open-pollinated families for DNA extraction.

DNA extraction

DNA was extracted from fresh needles using the Genomic DNA isolation kit (Fermentas, Lithuania). Needles were cut into small pieces, put into 2 mL centrifuge tubes containing a 5 mm diameter steel ball each, frozen in liquid nitrogen, and homogenised in a ball mill (Retch, model MM400). After that, the procedure described in the DNA isolation kit manual was performed with the following modifications: polyvinylpyrrolidone was added to the lysis buffer provided in the kit to a final concentration of 4%, and centrifugation speed and time in all steps were increased to 16 100 g and 15 min. Incubation with RNase A was added at the NaCl precipitation step by addition of 4 µg of RNase A to each sample together with the NaCl solution. Incubation was carried out at 37 °C for 30 min. After DNA extraction, additional DNA purification was performed using the OneStep™ PCR Inhibitor Removal kit from Zymo Research (USA) following the manufacturer's instructions. DNA concentration was measured with the Qubit system (Invitrogen) using the Quant-iT™ dsDNA BR Assay Kit (Invitrogen, USA).

Real-time PCR for gene copy number determination

The DNA sequence of the *Pinus taeda* thaumatin-like protein (NCBI accession No. EF532603) was utilised for the design of real-time PCR primers and probe. The *P. sylvestris* GAPDH gene was chosen as the endogenous control (NCBI accession No. L07501). Primers and probe were designed for the *P. sylvestris* resistance gene (also called *PsAcre*) (NCBI accession No. AY423270), used as an alternative endogenous control. Primers and probes were designed using PrimerExpress software (Applied Biosystems). Sequences of primers and probes used for CNV analyses are listed in Table 1.

Table 1. Sequences of primers and probes used for CNV analyses

Primer / probe name	Sequence (5' → 3') and fluorescent labels
TLP-F	CAGTGCCACAGGCATACAG
TLP-R	CCACCAGGGCAGGTGAAG
TLP-Z	6-FAM-TATGCCAAGGACGATGCCACCAGC-TAMRA
GAPDH-F	ACGGTTTTGGTCGAATTGGA
GAPDH-R	CCCCACGAGCTCGATATCAT
GAPDH-Z	VIC-CTCGTCGCCCCGTGTGGCTCTG-TAMRA
PsR-F	CATCATTACTTCCCCACACATATTCT
PsR-R	TGGGCTCTTCCTTGTCTTCAA
PsR-Z	6-FAM-CACTTTCATGCATTCCGCATCA-TAMRA

The Maxima™ Probe/ROX qPCR Master Mix (2×) (Fermentas, Lithuania) was used to perform real-time PCR; reaction setup per sample was as follows (reaction volume 30 µL): 15 µL of master mix, 27 pMol of each primer (final concentration 900 nM), 7.5 pMol of probe (final concentration 250 nM), 10 ng DNA, water to 30 µL.

Real-time PCR conditions were 2 min at 50 °C, 10 min 95 °C, followed by 40 cycles of denaturation for 15 s at 95 °C and annealing and extension for 1 min at 60 °C. PCR efficiency was checked using three serial (2×) DNA dilutions (40 ng, 20 ng, and 10 ng per reaction); for all targets PCR efficiency was nearly 100%.

Interpretation of results

The copy number determination procedure described by D'haene was followed for interpretation of results (D'haene et al. 2010). Biogazelle qbasePLUS software and calculations in Excel were used. D'haene's protocol describes the calculation of rescaling factors. Dividing the normalised relative copy number (normalised relative quantity) with the rescaling factor produces a value for absolute gene copy number. We were not able to fully apply the described procedure because of the lack of reference samples with a known copy number. Samples with similar CNV trends in *TLP* and *PsBB* (pinosylvin synthase gene) (data not shown) real-time PCR analyses (in total 14 samples) were used as reference samples for determination of rescaling factors for gene copy number calculation for CNV analysis of the *TLP* gene.

$$RF_j = \sqrt[n]{\left(\frac{NRQ_1}{p}\right) \times \left(\frac{NRQ_2}{p}\right) \times \left(\frac{NRQ_n}{p}\right)},$$

where RF_j is rescaling factor; NRQ – normalised relative quantity; p – ploidy of analysed organism (in this case $p = 2$); indexes 1, 2 and n – sample numbers.

Z scores calculated according to D'haene et al. (2010) were used for estimation of the credibility of absolute copy number data. If the absolute value of the Z score for a sample was less than 2, the result was considered as significant.

The role of standard samples used to calculate the rescaling factor is to indicate the possible amplitude of real-

time PCR Ct values for samples with identical copy numbers for a particular gene. The boundaries between gene copy number classes were calculated using the geometric mean between each adjacent copy number class.

Results

Given the possibility that the *TLP* gene in *P. sylvestris* is part of a larger gene family, this might lead to speculation about possible co-amplification of various members of the *TLP* gene family. However, previously conducted real-time PCR experiments using the same *TLP* primers with SybrGreen detection did not show any variation in the melting curves – a single peak was detected (data not shown). In addition, sequencing of the analysed *P. sylvestris* *TLP* gene did not reveal the presence of other gene family members (data not shown), indicating that no co-amplification of a different *TLP* gene family member occurred. These results indicate that the PCR primers utilised in this study amplified one specific sequence from the *P. sylvestris* genome. Previously published study of the *P. monticola* *TLP* gene family indicated a high degree of diversity at the nucleotide sequence level, with identities between family members varying from 4 to 96% (Liu et al. 2010), which suggested that it is possible to develop oligonucleotides specific only to one of the *TLP* gene family members.

Copy number polymorphism of the *P. sylvestris* *TLP* gene was detected using real-time PCR and relative quantitation. The relative quantities of the *TLP* gene ranged from 1.005 (sample GE21/5) to 3.421 (sample GE27/6). The average of the relative quantities was 1.772. The endogenous control, against which the quantity of the *TLP* genomic sequence was normalised in these experiments, was the *GAPDH* gene. The *GAPDH* gene has been utilised as an endogenous control in many gene expression studies in numerous species. However, as CNV can affect large portions of the genome, it is possible that the *GAPDH* gene also had various copy numbers in the experimental samples. To test for this, we compared average C_t values for the *TLP* gene and *GAPDH*. The PCR efficiency for both primer pairs was previously determined to be 100% (data not shown). The average C_t values of technical replicates of one sample

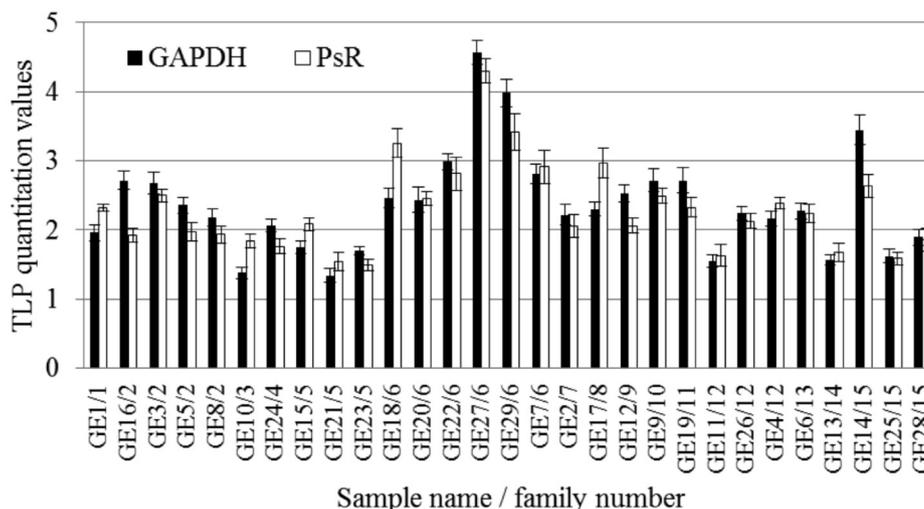


Fig. 1. Comparison of rescaled quantitation values of the *TLP* gene using *GAPDH* and *PsR* genes as endogenous controls.

for the *TLP* gene ranged from 24.76 to 30.09, while for *GAPDH* they ranged from 28.75 to 31.08. Given that the same amount of DNA was utilised in all assays, and that the PCR efficiencies for both primer pairs were similar, this suggests that there was larger variation in the quantity of the *TLP* gene sequence than in the *GAPDH* gene sequence. Data regarding amplitude of C_t values for *TLP* and *GAPDH* genes should be considered as additional information supporting the existence of CNV for the *TLP* gene, but not as direct evidence. To confirm the results of the *TLP* CNV analysis, the CNV experiment was repeated with a different endogenous control – *Pinus sylvestris* resistance gene (*PsR*). In previous experiments in our laboratory, no CNV of the *PsR* gene was found in these samples (data not shown). The overall tendencies of the *TLP* genomic sequence quantitation were similar when using the *GAPDH* or the *PsR* genes as endogenous controls (Fig. 1). The Spearman's rank order coefficient between the two sets of data was 0.812, indicating a high correlation between the ranks of the two values. Furthermore, analysing the same samples with primers for various candidate genes, CNV were not identified in all of the analysed genes, indicating that these results were specific for the *TLP* gene, and not dependent on other factors such as DNA quality or PCR efficiency differences. Considering that the calculated relative quantity of the *TLP* gene sequence in a particular sample was similar after normalisation with both the *GAPDH* gene and the *PsR* gene, data from the experiment with *GAPDH* as the endogenous control was used in further analyses.

In order to assign these relative quantity values to gene copy number classes, they must be rescaled using a rescaling factor calculated from a set of reference samples with known gene copy number. Unfortunately, no data on gene CNV is available for the *P. sylvestris* genome. In order to calculate the rescaling factor, we utilised three different sets of reference samples, which were chosen according to

different criteria. The first set (RF_a) was chosen by taking samples with similar CNV trends in *TLP* and *PsBBs* (pinosylvin synthase gene) (data not shown) (14 samples). The second set (RF_b) contained samples with relative quantities close to the average (10 samples). The third set (RF_c) contained all samples that were below the average relative quantity (17 samples). We made the assumption that the majority of samples will have two copies of the *TLP* gene, given that the *P. sylvestris* genome is diploid.

Using the reference sample set RF_a , the estimated copy number was five for one sample, four for one sample and three for nine samples. Two copies of the *TLP* gene were detected in genomes of 16 samples, and one for two samples (Fig. 2). In several samples, the rescaled values are close to the boundaries of different gene copy number groups. The effect of using different reference sample sets is to shift the quantitation values of entire data set up or down. Using the reference sample set RF_b , which contained samples with quantitation values close to the average, the rescaled quantitation values were lower. This was due to the fact that samples with lower quantitation values were not assumed to have two copies of the *TLP* gene. As a result, some samples that were close to the copy number class boundaries were reclassified into the lower copy number class. Use of the reference sample set RF_c (containing all samples with quantitation values below the average) resulted in the rescaled quantitation values being shifted to higher values, as all samples with lower quantitation levels were assumed to have two copies of the *TLP* gene (Fig. 2).

Regardless of the classification of the samples into absolute gene copy number classes, the results strongly suggest the existence of copy number variation for this gene. Furthermore, samples GE18, GE29, GE27, GE22, GE7, and GE20 (family 6) are descendants of one mother tree. Using RF_a , in four of these samples the copy number of the *TLP* gene was either three or four, and one sample

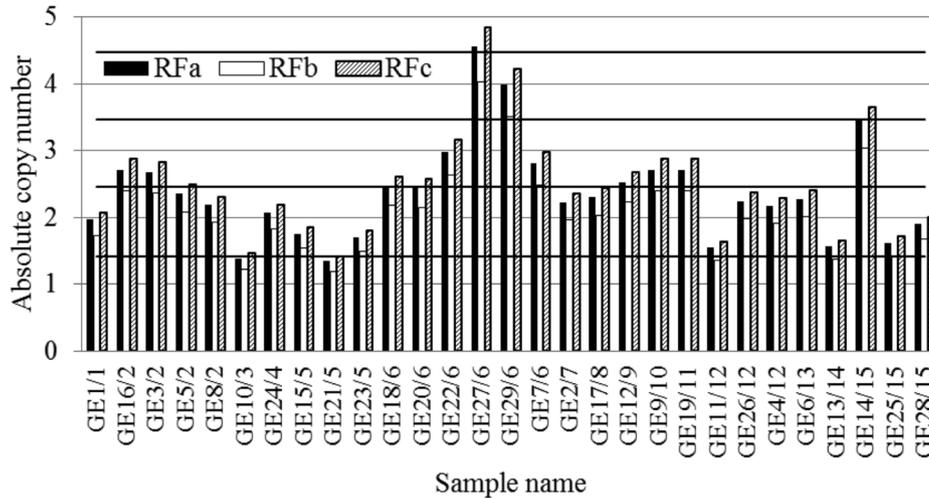


Fig. 2. Rescaled gene copy number value calculated using three different rescaling factors. Absolute gene copy number classes are delineated on the y axis.

slightly exceeded the border value to a gene copy number of five. Sample GE20 was extremely close to the delimiting border of the sample group with three copies of the *TLP* gene. Differences in CNV of the pollen donors might provide explanation for the above results – if both gametes originated from parents with four copies of the *TLP* gene, the result would be four copies in the offspring but if one parent contained four but the other contained two copies of this gene, it could lead to an offspring with three copies of this gene.

Following the procedure of D’haene (D’haene et al. 2010), Z scores were used as an indicator of significance of the results. As the absolute values of Z scores for all samples were less than 2, the results can be considered as significant. Z scores for *TLP* gene CNV analyses using both endogenous controls are given in Fig. 3.

Discussion

Although CNV has been detected in plant genomes (Springer et al. 2009; DeBolt 2010; Swanson-Wagner et al. 2010), to our knowledge this is the first report of CNV in a conifer species – *P. sylvestris*. The existence of gene families consisting of a gene with related, but not identical, sequences within conifer genomes has been reported (Kinlaw, Neale 1997; Preisig-Müller et al. 1999; Li, Asiegbu 2004; Liu et al. 2010) but no CNV analyses of representatives of these gene families have been performed. Various methods, differing in their throughput and technology platforms, can be used for CNV studies: microarray-based comparative genome hybridisation (Springer et al. 2009; DeBolt 2010; Swanson-Wagner et al. 2010), CNV-seq – a recently developed sequencing-based method (Xie, Tammi 2009), and real-

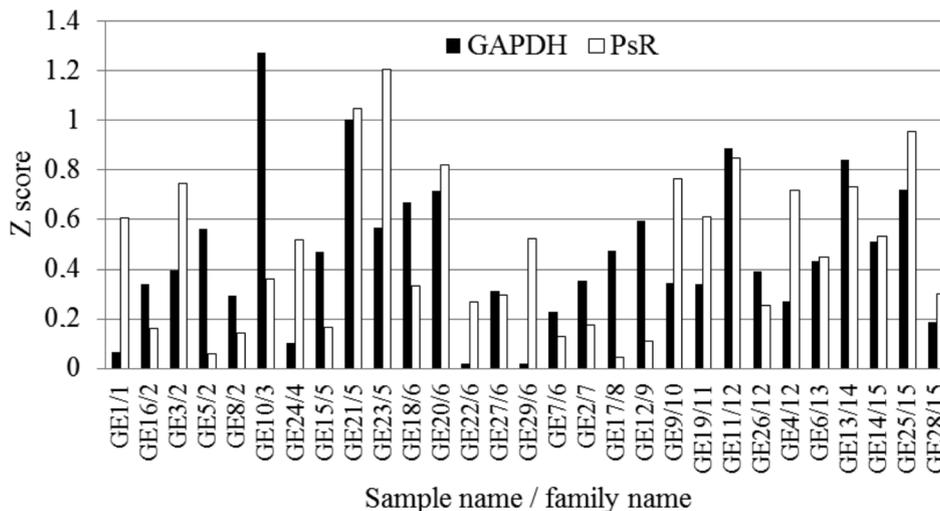


Fig. 3. Z values of *TLP* gene CNV analyses using *GAPDH* and *PsR* as endogenous controls.

time PCR (D'haene et al., 2010). In this research the major hurdle in the analysis of the results was the lack of reference samples with a defined copy number for the studied gene. Further CNV studies of the *P. sylvestris* *TLP* gene, as well as other genes, will allow for more robust selection of these reference samples, as well as further information on the extent and frequency of CNV within the *P. sylvestris* genome. One possibility would be to utilise an alternative method, such as microarray-based comparative genome hybridisation, in order to examine the extent of CNV on a larger scale. However, the lack of a genome-wide microarray for *P. sylvestris* means that these results would have to be confirmed for individual genes using the real-time PCR approach described in this study. Confirmation of CNV polymorphism could also be made by using other methods such as Southern blotting, sequencing of megagametophyte DNA, high resolution melting curve analysis or FISH, providing that these methods offered a high enough resolution to detect these CNVs. Additional real-time PCR experiments with other endogenous controls than *GAPDH* (and *PsR*) might also serve to confirm the results presented in this paper. In addition, the detection of an elevated copy number of the *TLP* gene in all members of family 6 except one, indicates the genetic basis of this variation, and confirms the results of the real-time CNV analysis. In the absence of a high resolution microarray platform and the lack of genome sequencing data for *P. sylvestris*, which would allow for a more general overview of the patterns of CNV in Scots pine, we utilised a real-time PCR analysis technique to investigate CNV of specific genes of interest. Gene copy number variation analysis using real-time PCR is a useful tool for detection of CNV, even when analysing such complicated and large genomes as those of *P. sylvestris*.

The results obtained indicate that the utilised PCR primers amplify a unique sequence from the *P. sylvestris* genome. We detected comparatively recent duplications of the *TLP* gene, which have yet to accumulate a significant amount of mutations. Gene families are presumed to have evolved from duplicated genes that accumulate mutations over time. Two widely used genome analysis techniques – analysis of EST databases and the use of shotgun genome sequencing techniques are not conducive for the detection of CNV. The widespread use of both of these methods for the analysis of a large range of species has led to CNV being undetected or underreported (Schrider, Hahn 2010).

In this report we show that CNV of the *TLP* gene is present in the *P. sylvestris* genome. One effect of an increased copy number of a gene could be higher levels of transcription of this gene. In the case of CNV of the *TLP* gene, this could lead to an increase of the constitutive levels of this gene product, and to higher levels of the protein after induction. The effectiveness of *TLP* against *H. annosum* has not been investigated *in vitro*; however this group of proteins is described as effective against a range of plant pathogens (Selitrennikoff 2001). Future research

will include determination of the efficacy of the *TLP* gene product against *H. annosum in vitro*, as well as gene expression studies to determine the effect of increased copy number of the *TLP* gene on mRNA levels. CNV analysis using the real-time PCR detection method described in this study is a robust and relatively fast tool for the analysis of CNV of candidate genes within natural and breeding populations. Detection of copy number polymorphism is a potentially useful tool for enhancing breeding programs for creation of tree lines more resistant to *H. annosum* as well as to other bacterial or fungal infections.

In conclusion, variation in the gene copy number of the thaumatin-like protein gene has been identified in the *P. sylvestris* genome. Quantitation results were similar when two different endogenous controls (*GAPDH* and *PsR*) were used. Evidence of the effect of family structure on the presence of CNV was identified, confirming the results obtained using the real-time PCR method.

Acknowledgements

This research was funded by ESF project number 2009/0228/1DP/1.1.1.2.0/09/APIA/VIAA/035.

References

- Aarhus M., Helland C.A., Lund-Johansen M., Wester K., Knappskog P.M. 2010. Microarray-based gene expression profiling and DNA copy number variation analysis of temporal fossa arachnoid cysts. *Cerebrospinal Fluid Res.* 7: 1–8.
- Bradeen J.M., Iorizzo M., Molloy D.S., Raasch J., Colton Kramer L., Millett B.P., Austin-Phillips S., Jiang J., Carputo D. 2009. Higher copy numbers of the potato *RB* transgene correspond to enhanced transcript and late blight resistance levels. *Mol. Plant Microbe Interact.* 22: 437–446.
- Chen Q., Book M., Fang X., Hoeft A., Stuber F. 2006. Screening of copy number polymorphisms in human beta-defensin genes using modified real-time quantitative PCR. *J. Immunol. Methods* 308: 231–240.
- DeBolt S. 2010. Copy number variation shapes genome diversity in *Arabidopsis* over immediate family generational scales. *Genome Biol. Evol.* 2: 441–453.
- D'haene B., Vandesompele J., Hellems J. 2010. Accurate and objective copy number profiling using real-time quantitative PCR. *Methods* 50: 262–270.
- Fossdal C.G., Hietala A.M., Kvaalen H., Solheim H. 2006. Changes in host chitinase isoforms in relation to wounding and colonization by *Heterobasidion annosum*: early and strong defense response in 33-year-old resistant Norway spruce clone. *Tree Physiol.* 26: 169–177.
- Gaitnieks T., Arhipova N., Donis J., Stenlid J., Vasaitis R. 2007. Butt rot incidence and related losses in Latvian *Picea abies* (L.) Karst. stands. In: Garbelotto M., Gonthier M. (eds) *Proceedings of the 12th IUFRO International Conference on Root and Butt Rots*. Berkeley, California, Medford, Oregon/USA, pp. 177–179.
- Iafrate A.J., Feuk L., Rivera M.N., Listewnik M.L., Donahoe P.K., Qi Y., Scherer S.W., Lee C. 2004. Detection of large-scale variation in the human genome. *Nat. Genet.* 36: 949–951.

- Järvinen A.K., Autio R., Kilpinen S., Saarela M., Leivo I., Grénman R., Mäkitie A.A., Monni O. 2008. High-resolution copy number and gene expression microarray analyses of head and neck squamous cell carcinoma cell lines of tongue and larynx. *Gene Chromosome Canc.* 47: 500–509.
- Kinlaw C.S., Neale D.B. 1997. Complex gene families in pine genomes. *Trends Plant Sci.* 2: 356–359.
- Korshikov I.I., Demkovich A.E. 2008. Genotypic peculiarities of resistance to root fungus of Scotch pine plants in artificial stands of the Ukrainian steppe-zone. *Tsitol. Genet.* 42: 41–46. [in Russian]
- Li G., Asiegbo F.O. 2004a. Induction of *Pinus sylvestris* PsACRE, a homology of *Avr9/Cf-9* rapidly elicited defense-related gene following infection with root rot fungus *Heterobasidion annosum*. *Plant Sci.* 167: 535–540.
- Li G., Asiegbo F.O. 2004b. Use of Scots pine seedling roots as an experimental model to investigate gene expression during interaction with the conifer pathogen *Heterobasidion annosum* (P-type). *J. Plant Res.* 117: 155–162.
- Li J., Yang T., Wang L., Yan H., Zhang Y., Guo Y., Pan F., Zhang Z., Peng Y., Zhou Q., He L., Zhu X., Deng H., Levy S., Papasian C.J., Drees B.M., Hamilton J.J., Recker R.R., Cheng J., Deng H.-W. 2009. Whole genome distribution and ethnic differentiation of copy number variation in Caucasian and Asian populations. *PLoS ONE* 4: e7958.
- Liu J.J., Zamani A., Ekramoddoullah A.K. 2010. Expression profiling of a complex thaumatococcus-like protein family in western white pine. *Planta* 231: 637–651.
- Lucito R., Alexander J., Rostan S., West J., Wigler M. 2001. Detection of gene copy number fluctuations with microarrays using genomic representations. *Genetics* 27: 70.
- Myllykangas S., Junnila S., Kokkola A., Autio R., Scheinin I., Kiviluoto T., Karjalainen-Lindsberg M.L., Hollmén J., Knuutila S., Puolakkainen P., Monni O. 2008. Integrated gene copy number and expression microarray analysis of gastric cancer highlights potential target genes. *Int. J. Cancer.* 123: 817–825.
- Orozco L.D., Cokus S.J., Ghazalpour A., Ingram-Drake L., Wang S., van Nas A., Che N., Araujo J.A., Pellegrini M., Lusis A.J. 2009. Copy number variation influences gene expression and metabolic traits in mice. *Hum. Mol. Genet.* 18: 4118–4129.
- Ortiz-Estevéz M., De Las Rivas J., Fontanillo C., Rubio A. 2011. Segmentation of genomic and transcriptomic microarrays data reveals major correlation between DNA copy number aberrations and gene-loci expression. *Genomics* 97: 86–93.
- Pollack J.R., Sorlie T., Perou C.M., Rees C.A., Jeffrey S.S., Lonning P.E., Tibshirani R., Botstein D., Børresen-Dale A.-L., Brown P.O. 2002. Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proc. Natl. Acad. Sci. USA* 99: 12963–12968.
- Preisig-Müller R., Schwekendiek A., Brehm I., Reif H.-J., Kindl H. 1999. Characterization of a pine multigene family containing elicitor-responsive stilbene synthase genes. *Plant Mol. Biol.* 39: 221–229.
- Rishbeth J. 1962. Rate of spread of *Fomes annosus*. *Trans. Br. Mycol. Soc.* 45: 280.
- Roberts W., Selitrennikoff C.P. 1990. Zeamatin, an antifungal protein from maize with membrane-permeabilizing activity. *J. Gen. Microbiol.* 136: 1771–1778.
- Rose-Zerilli M.J., Barton S.J., Henderson A.J., Shaheen S.O., Holloway J.W. 2009. Copy-number variation genotyping of *GSTT1* and *GSTM1* gene deletions by real-time PCR. *Clin. Chem.* 55: 1680–1685.
- Sebat J., Lakshmi B., Troge J., Alexander J., Young J., Lundin P., Månér S., Massa H., Walker M., Chi M., Navin N., Lucito R., Healy J., Hicks J., Ye K., Reiner A., Gilliam T.C., Trask B., Patterson N., Zetterberg A., Wigler M. 2004. Large-scale copy number polymorphism in the human genome. *Science* 305: 525–528.
- Schrider D.R., Hahn M.W. 2010. Gene copy-number polymorphism in nature. *Proc. Royal Soc. B* 277: 3213–3221.
- Selitrennikoff C.P. 2001. Antifungal proteins. *Appl. Environ. Microbiol.* 67: 2883–2894.
- Springer N.M., Ying K., Fu Y., Ji T., Yeh C.-T., Jia Y., Wu W., Richmond T., Kitzman J., Rosenbaum H., Iniguez A.L., Barbazuk W.B., Jeddloh J.A., Nettleton D., Schnable P.S. 2009. Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. *PLoS Genet.* 5: e1000734.
- Stankiewicz P., Lupski J.R. 2010. Structural variation in the human genome and its role in disease. *Annu. Rev. Med.* 61: 437–455.
- Swanson-Wagner R.A., Eichten S.R., Kumari S., Tiffin P., Stein J.C., Ware D., Springer N.M. 2010. Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. *Genome Res.* 20: 1689–1699.
- Xie C., Tammi M. 2009. CNV-seq, a new method to detect copy number variation using high-throughput sequencing. *BMC Bioinformatics* 10: 80.
- Zhang D., Qian Y., Akula N., Alliey-Rodriguez N., Tang J., Gershon E.S., Liu C. 2011. Accuracy of CNV detection from GWAS data. *PLoS ONE* 6: e14511.