

## RESEARCH ARTICLE

## Open Access



# Association genetics of acetophenone defence against spruce budworm in mature white spruce

Mebarek Lamara<sup>1,3\*</sup>, Geneviève J. Parent<sup>2</sup>, Isabelle Giguère<sup>1</sup>, Jean Beaulieu<sup>1,3</sup>, Jean Bousquet<sup>1,3</sup> and John J. MacKay<sup>1,2,3</sup>

## Abstract

**Background:** Outbreaks of spruce budworm (SBW, *Choristoneura fumiferana* Clem.) cause major recurrent damage in boreal conifers such as white spruce (*Picea glauca* [Moench] Voss) and large losses of forest biomass in North America. Although defensive phenolic compounds have recently been linked to chemical resistance against SBW, their genetic basis remains poorly understood in forest trees, especially in conifers. Here, we used diverse association genetics approaches to discover genes and their variants that may control the accumulation of acetophenones, and dissect the genetic architecture of these defence compounds against SBW in white spruce mature trees.

**Results:** Out of 4747 single nucleotide polymorphisms (SNPs) from 2312 genes genotyped in a population of 211 unrelated individuals, genetic association analyses identified 35 SNPs in 33 different genes that were significantly associated with the defence traits by using single-locus, multi-locus and multi-trait approaches. The multi-locus approach was particularly effective at detecting SNP–trait associations that explained a large fraction of the phenotypic variance (from 20 to 43%). Significant genes were regulatory including the NAC transcription factor, or they were involved in carbohydrate metabolism, falling into the binding, catalytic or transporter activity functional classes. Most of them were highly expressed in foliage. Weak positive phenotypic correlations were observed between defence and growth traits, indicating little or no evidence of defence-growth trade-offs.

**Conclusions:** This study provides new insights on the genetic architecture of tree defence traits, contributing to our understanding of the physiology of resistance mechanisms to biotic factors and providing a basis for the genetic improvement of the constitutive defence of white spruce against SBW.

**Keywords:** Association genetics, Phenolic compounds, *Pgβglu-1* expression, Spruce budworm, White spruce, Metabolic trade-offs

## Background

Trees use a battery of constitutive and inducible defence strategies to limit the damage of herbivory from insects over their long life span [1, 2]. Constitutive chemical defence barriers are particularly well developed in conifers, which produce a wide range of secondary metabolites

such as oleoresin terpenoids and phenolic compounds to reduce herbivore attacks [3–6]. The arsenal of constitutive and inducible terpenes that are produced by conifers such as spruce, pine or fir have become some of the best studied secondary metabolites in trees, particularly in regard to the mechanisms of synthesis and the molecular bases of their regulation [7–11]. However, the molecular basis of heritable variation in chemical defences is only partially understood.

The spruce budworm (SBW) *Choristoneura fumiferana* Clemens (Lepidoptera: Tortricidae) is one of the most destructive native insect pests in coniferous and

\* Correspondence: [mebarek.lamara.1@ulaval.ca](mailto:mebarek.lamara.1@ulaval.ca)

<sup>1</sup>Forest Research Centre and Institute for Systems and Integrative Biology, Département des sciences du bois et de la forêt, Université Laval, Qc, Québec G1V 0A6, Canada

<sup>3</sup>Canada Research Chair in Forest Genomics, Université Laval, Qc, Québec G1V 0A6, Canada

Full list of author information is available at the end of the article



mixed forests of North America, particularly in the East [12–15]. In the last decade, recurrent outbreaks of SBW in Canada have caused high levels of tree mortality in fir and spruce trees through intensive leaf herbivory [16]. The outbreaks have spread over millions of hectares [17] and caused losses varying from 3 to 68 m<sup>3</sup>/ha of wood [16] depending on stand and region, with substantial damages occurring in both natural and plantation forests [18]. SBW larvae preferentially feed on the new foliage of conifers, which include, in decreasing order of susceptibility, balsam fir (*Abies balsamea* [L.] Mill.), white spruce (*Picea glauca* [Moench] Voss), red spruce (*P. rubens* Sarg.) and black spruce (*P. mariana* [Mill.] BSP) [18–20]. Despite the ecological and economic importance of these spruce and fir trees in North American forests, little is known of naturally-occurring defence mechanisms against SBW.

Quantitative genetics studies of chemical defence compounds such as monoterpenes in conifers have reported considerable intraspecific variation at the phenotypic level and relatively high estimates of heritability, which indicates the strong genetic control underlying these traits [5, 21–23]. Variability in some phenolic compounds and other secondary metabolites that may accumulate in leaf tissues is also under genetic control in trees [24–26]. For instance, the accumulation of sideroxylonal was shown to be highly heritable in eucalypt species [24, 27]; however, compared to terpenes, the biosynthesis and genetic control of phenolic compounds involved in defence against insects is less well understood in conifers. The shikimic acid and phenylpropanoid pathways form the core biosynthetic route leading to the production of both defensive and structural phenolic compounds such as lignin [28], but the mechanisms by which the former may accumulate are largely unknown in conifers.

In white spruce trees, constitutive chemical defence against SBW has been linked to the accumulation of piceol and pungenol, which belong to a class of phenolic compounds known as acetophenones [19, 29]. Considerable variation was observed in this naturally-occurring resistance mechanism and it was shown to be linked to the expression of the *Pgβglu-1* gene [29]. Population [30], sib and clonal analyses, showed that variation in acetophenone concentrations was highly heritable and positively impacted white spruce fitness. Laboratory feeding experiments also showed a decrease in SBW larvae survival [19]. In contrast, the glycosylated conjugates picein and pungenin accumulated in both resistant and non-resistant trees and were not biologically active against SBW [19]. These findings have raised questions regarding the molecular basis of the genetic control underlying these defence traits and the considerable natural genetic variation observed at the population level.

Association genetics approaches are often used to dissect complex traits in forest trees including wood quality [31–34] and defence against insect herbivory [5, 6, 35]. Single-locus association studies have identified associations between SNPs in candidate genes and defence against insect herbivory in *Eucalyptus* [35] and *Pinus* [5, 6]. However, most of the marker-trait associations only explained a small proportion of the phenotypic variation. This is due to the fact that variation in complex traits appears to be based on many loci with small effects [36]. Alternative approaches may be more effective at uncovering the network of gene effects which underpins phenotypic variations. For example, multi-locus analyses have been developed for more effective capture of combined effects [37] and multi-trait models have also been used to account for trait interactions [38].

The high constitutive levels of acetophenones reported for white spruce [39] also raise questions regarding possible trade-offs between defence and growth. The production cost of phenolic compounds could be relatively high and may compete with the formation of new tissues or the accumulation of reserves. Different hypotheses related to the balance of energy between growth and defence have been proposed [40–43], but little evidence of trade-offs has been observed in trees to date. Such trade-offs would be detected as negative phenotypic correlations between defence and growth traits and could have consequences on breeding strategies for improved defence against SBW in white spruce [44].

This study pursued two major objectives: (1) to identify genes and SNPs associated with variation in acetophenone concentrations. To date, the level of acetophenone aglycons has been explained in part by *Pgβglu-1* expression but it accounted for less than half of the variation [29]. Moreover, variation in the acetophenone glucoside picein remains unexplained. We thus used three different association study approaches to identify genes associated with phenolic compounds and *Pgβglu-1* expression as quantitative defence traits. (2) to examine potential trade-offs between acetophenone defences and growth given that phenolic compounds such as picein accumulate to very high concentrations in the foliage of some trees but not in others.

## Methods

The four defence traits assessed in this study were the acetophenones piceol and pungenol, the expression levels of the *Pgβglu-1* gene, which is responsible for their release, and the acetophenone glucoside picein. These traits were first described in [19] and in [29]. Sampling and laboratory analyses are summarized in the following sections, but more details are available from [39]. In a separate analysis, the phenotypic data of three growth traits (tree height in m (Ht), stem diameter at breast

height in cm (DBH), and growth ring width averaged from pith to bark in cm (RW)) from [45] were used to assess defence-growth trade-offs.

#### Plant materials

Foliage of 211 unrelated 38-year-old mature white spruce (*Picea glauca*) trees, each from a distinct open-pollinated family and representing 42 geographic origins (provenances), were sampled in a provenance-progeny test established by the Canadian Forest Service in the field with three-year old trees in 1979 at the site of Mastigouche, Québec, Canada (46°38'N, 73°13'W) (described in [45]). Only current-year foliage was sampled from the north side of the mid-crown on 24 July 2014 (trees aged 38), frozen immediately in liquid nitrogen after removal from the trees and stored at  $-80^{\circ}\text{C}$ . The foliage was ground to a fine powder using a MixerMill 300 (Retsch) and steel grinding balls cooled in nitrogen. Powdered tissue was stored at  $-80^{\circ}\text{C}$  until further analyses.

The sampling was non-destructive and the trees were part of an experimental plantation established for research on land of the government of Québec. A collaborative research agreement between the organizations as part of the Arborea II project gave permission for the sampling, which followed guidelines of the institutions involved in the research and in force in Québec (Canada).

#### RNA extraction and transcript determination assays

Total RNA was extracted as in [46] with modifications as in [47] and stored at  $-80^{\circ}\text{C}$ . The total RNA concentration was determined using a NanoDrop 1000 (Thermo Fisher Scientific, Wilmington, DE, USA) and assessed for quality with an Agilent 2100 Bioanalyzer and RNA 6000 Nano Kit LabChips (Agilent Technologies Inc.). Only RNA isolates with an integrity score (RIN) of 7.0 or more were used for analyses. Reverse transcriptase-quantitative PCR (RT-qPCR) with gene-specific primers was used to quantify transcript accumulation levels of the *Pgβglu-1* gene (see [29] for more details).

#### Acetophenone extractions and determinations

The hydroxyl-acetophenones piceol and pungenol and the hydroxyl-acetophenone glucoside picein were extracted as described in [39]. Assays were conducted on a LC (Agilent 1200 series) coupled to a MS detector (Agilent 6210 TOF). Acetophenones were separated in a pre-column Polaris MetaGuard 4.6 mm and a column Polaris 250 mm  $\times$  4.6 mm C18-A, particular size 5  $\mu\text{m}$  (Agilent Technologies Inc.). The solvent and solvent gradient were as described in [39]. The column flow rate was  $1.5\text{ ml min}^{-1}$  and ten microlitres of extract were injected. Quantification was done using external calibration curves for picein, piceol and pungenol. No pungenin was commercially available.

#### Genotypic data

High-quality genotyping data based on single nucleotide polymorphisms (SNPs) were obtained using an Infinium iSelect genotyping chip (Illumina, San Diego, CA) and were previously described [45]. In the current study, a cut-off of 0.10 for minor allele frequency (MAF) was used, resulting in a set of 4767 high-quality SNPs in 2312 genes without any missing genotypes from a starting dataset of 6385 SNPs in 2652 genes. The gene sequences are described in the white spruce gene catalogue [48] and genes were selected based on multiple criteria as described in [49] and the Supporting Information in [34]. Briefly, these criteria were related to 1) predicted functions relevant for wood formation, growth, phenology, and adaptation to biotic and abiotic factors as indicated by database searching and scientific literature from *Arabidopsis* and poplar (e.g. [50–52]; 2) expressional candidate genes related to phenology [53] and vascular tissue differentiation [47]; 3) overexpression of R2R3-MYB genes, HD-zips and other transcription factors in spruce trees [54–57]; 4) co-localization with QTLs for bud flush, bud set and height growth [58]; and 5) genes harbouring SNPs implicated in local adaptation [59]. The genes were well distributed across the 12 linkage groups of white spruce [60].

#### Simulations

Simulations were used to assess the effectiveness of the multi-locus mixed model (MLMM) and single-locus mixed model (SLMM) in detecting associations under different genetic architectures of the complex traits in the present population. Simulated phenotypic data sets were generated by simulating genetic effects based on real genotype data (4767 SNPs) drawn from this study using the R-package BGLR [61]. A theoretical normally distributed phenotypic trait was simulated for the 211 trees under two different scenarios differing in the number of SNPs controlling the phenotype; scenario I, 10 SNPs and scenario II, 50 SNPs. For both scenarios, we tested three different heritability levels, i.e. 0.50, 0.75 and 1.

#### Association analyses

Data for the four investigated defence traits were normalized using the rank-based inverse normal transformation, implemented as the *rntransform* function in the GenABEL R Library [62] in order to comply with assumptions of association genetics testing that residuals be normally distributed. Principal component analysis (PCA) and a pairwise kinship matrix were used to assess for the presence of population structure in the set of 211 trees using the 4767 SNPs. The association analyses between SNPs and traits were performed using the three following approaches.

The single-locus mixed model (SLMM) implemented in TASSEL v5.2.1 [63] as described by [64] was used to take into account potential relatedness among the 211 trees as well as a weak population structure previously noted [32] so to remove any spurious association effects. We set a uniform threshold  $P < 2 \times 10^{-4}$  (calculated according to  $P = 1/n$ ;  $n$  = total number of SNPs used in the analysis), which is roughly equivalent to a Bonferroni correction [65, 66], to determine if the SNP markers were significantly associated with the four defence traits for the different analyses.

The modified version of multi-locus mixed model (MLMM), as developed by [37] where PCA scores and kinship coefficients are defined as cofactors, was used to further identify SNPs potentially associated with the four defence traits. The approach relies on a simple, stepwise mixed-model regression with forward inclusion and backward elimination while re-estimating the genetic and error variances at each step of the regression. This method may well lead to higher detection power and a lower FDR when compared with traditional single-locus approaches [37]. For each phenotype, the percentage of phenotypic variation explained (PVE) by markers was determined at the optimal step. The multi-trait mixed model (MTMM) [38] was used to analyse pairs of correlated traits. This approach is based on the principle that measurements taken for the correlated traits may be combined to increase the power to detect common SNPs in genetic association with both traits [38, 67, 68].

#### Trade-offs between defence and growth traits

We investigated whether there may be trade-off relationships between the constitutive defence and three growth traits, tree height in m (Ht), stem diameter at breast height in cm (DBH), and growth ring width averaged from pith to bark in cm (RW) as reported previously in [45]. First, pairwise Pearson correlation coefficients were determined between all traits to estimate the magnitude of trade-offs using the transformed data. Second, a principal component analysis (PCA) was conducted using the *prcomp* function implemented in R [69] to graphically illustrate the relationship between SBW defence traits represented by acetophenone compounds and *Pgβglu-1* transcripts on one hand, and growth traits on the other hand by examining the biplot graphics. Third, association analyses were performed between SNPs and all of the growth and defence traits by using permissive statistical test conditions (SLMM method, threshold of  $P < 0.05$  without correction for multiple testing [34]) in order to uniquely determine the extent of overlap among the sets of genes that may be linked to the different traits.

## Results

### Phenotypic variation

Table 1 shows the summary statistics for the four defence traits determined in 211 unrelated trees each representing a different open-pollinated family from 42 natural populations [45]. A broad range of variation was observed for each trait. In particular, acetophenone defence compounds accumulated to high levels in some trees and were undetected in others (Table 1); in addition, the data were not normally distributed and skewed toward low values (Fig. 1a, c, e). Similar observations were made for the *Pgβglu-1* transcript levels though the distribution was skewed toward high levels (Fig. 1g). Data were transformed using the rank-based inverse normal transformation such that residuals were normalized (Fig. 1b, d, f, h) prior to conducting the association genetics analyses that follow.

### Simulations

We used genotyping data for 4767 high-quality SNPs from 2312 candidate genes [45] to search for and analyse SNP-trait associations potentially controlling the defence phenotypes against SBW. The simulations used to assess the potential to detect SNPs in the study population indicated that the multi-locus mixed model (MLMM) detected a larger number of SNP-trait associations compared to the single-locus mixed model (SLMM) (Table 2). The power to detect SNP-trait associations declined when SNPs controlling the trait increased from 10 to 50, especially at a moderate heritability level.

### Identification of SNPs and genes associated with defence traits

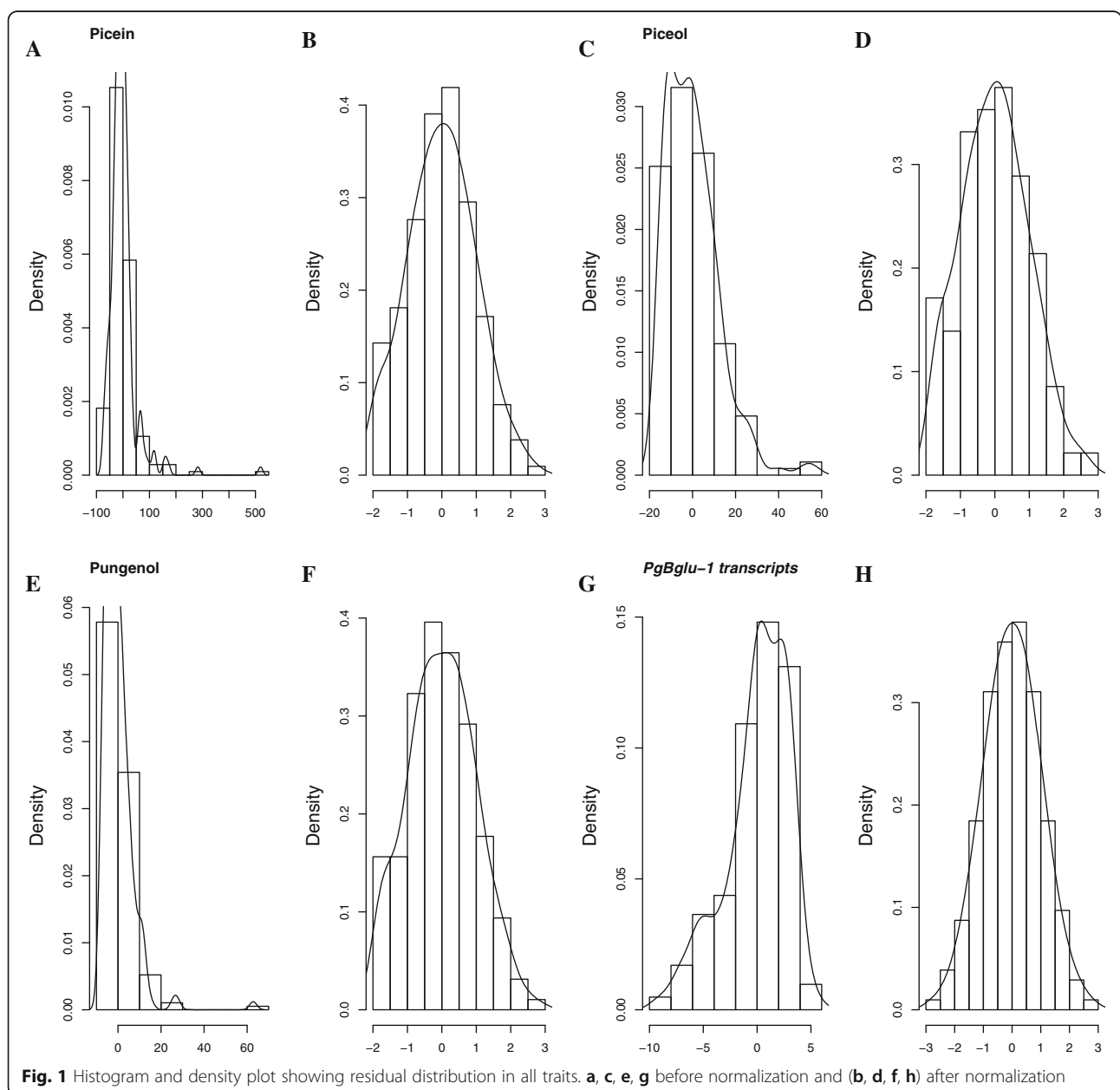
In a first step, we used the SLMM and MLMM approaches and identified a total of 31 SNPs in 29 genes that were significantly associated with variation in at least one of the acetophenone compounds and *Pgβglu-1* transcripts levels at the Bonferroni-corrected statistical threshold ( $-\log P > 3.68$ ,  $\alpha = 1$ ) (Table 3). The SLMM method identified eight significant associations involving seven different SNP (Table 3). The proportion of the phenotypic variation explained (PVE) by all significant SNPs varied from as little as 2.3% for piceol to as high as 11.2% for picein (Table 3). In contrast, significant associations were obtained for 26 SNPs with MLMM. Three of the SNPs were associated with the glucosylated phenolic compound picein and explained 20% of phenotypic variation; two of them were also identified with the SLMM approach. The acetophenone piceol was significantly associated with nine SNPs with a PVE of 43%, and pungenol was significantly associated with six SNPs with a PVE of 27%. A total of eight SNPs were significantly associated with *Pgβglu-1* transcripts and explained 23% of phenotypic variation. Our results indicate that

**Table 1** Summary statistics of constitutive defence traits in the white spruce association population

Defence traits	Number of trees	Minimum	Maximum	Median	Mean
Picein (mg/g)	210	0	590.2	53.3	62.4
Piceol (mg/g)	211	0	71.7	11.2	12.9
Pungenol (mg/g)	211	0	70.0	5.3	6.4
<i>Pgβglu-1</i> transcripts (ng/RNA)	206	5	180,500	4197	11,320

the analysis carried out with the MLM method by using the same SNPs genotyped in the population was more effective for detecting significant SNPs compared to the traditional SLMM approach (Table 3) as intended by its developers [37]. In total, five

SNPs representative of 5 distinct genes were detected by SLMM, and 23 SNPs representative of 23 distinct genes were detected by MLM, with three SNPs representative of three distinct genes in common between SLMM and MLM, thus resulting in a total



**Table 2** Simulation results of detecting significant<sup>a</sup> SNP-trait associations using SLMM and MLMM approaches

Association approaches <sup>b</sup>	10 SNPs			50 SNPs		
	Heritability			Heritability		
	0.50	0.75	1.0	0.50	0.75	1.0
SLMM	0	3	6	0	0	3
MLMM	6	8	10	1	3	8

<sup>a</sup>The significant threshold used was  $P < 2 \times 10^{-4}$ <sup>b</sup>SLMM, single-locus mixed model; and MLMM, multi-locus mixed model

of 31 distinct SNPs representative of 29 genes detected by the two methods.

Piceol, pungenol and *Pgβglu-1* expression were previously reported to be moderately correlated [29]. Thus, the MTMM approach was used to search for significant SNPs in common between each pair of the traits used in this study. Three SNPs (from three distinct genes) were shared between piceol and pungenol, and four times one SNP (from distinct genes) were shared in other pairs of traits (Table 4). We thus identified a total of six significant SNPs (from as many distinct genes) associated with the combined traits, including four new SNPs identified only with the MTMM approach. The two other SNPs were also detected by using SLMM and/or MLMM approaches. In total using the three methods (SLMM, MLMM and MTMM) applied to the four defence traits, we identified 35 different SNPs representative of 33 distinct genes (Fig. 2).

#### Functional annotations and expression of genes associated with defence traits against spruce budworm

We began the characterization of the 33 genes containing the 35 SNPs significantly associated with the defence traits by conducting an analysis of the gene ontology (GO) terms associated to these functionally annotated genes. We found that the genes belonged essentially to three molecular functions: binding (GO:0005488; 6 genes), catalytic activity (GO:0003824; 19 genes) and transporter activity (GO:0005215; 2 genes). The catalytic activity category harboured the largest number of genes and involved several different enzymatic functions.

**Table 3** SNPs significantly associated with defence traits in white spruce using SLMM and MLMM approaches<sup>a</sup>, and their combined percentage of phenotypic variation explained (PVE)

Defence traits	SLMM		MLMM	
	Nb. of SNPs <sup>b</sup>	PVE (%)	Nb. of SNPs <sup>b</sup>	PVE (%)
Picein	2 (2) <sup>c</sup>	11.2	3 (3)	20
Piceol	2 (2)	2.3	9 (9)	43
Pungenol	2 (2)	4.0	6 (6)	27
<i>Pgβglu-1</i> transcripts	2 (2)	8.2	8 (8)	23
Total number of distinct SNPs	8 (7)	–	26 (26)	–

<sup>a</sup>SLMM, single-locus mixed model; MLMM, multi-locus mixed model<sup>b</sup>Number of significant SNPs associated with the trait variation<sup>c</sup>In parentheses, number of significant genes**Table 4** Number of significant SNPs associated with defence traits in white spruce using the multi-trait mixed model (MTMM) approach

Defence traits	Piceol	Pungenol	<i>Pgβglu-1</i> transcripts
Picein	1 (1) <sup>a</sup>	1 (1)	1 (1)
Piceol	–	3 (3)	1 (1)
Pungenol	–	–	0

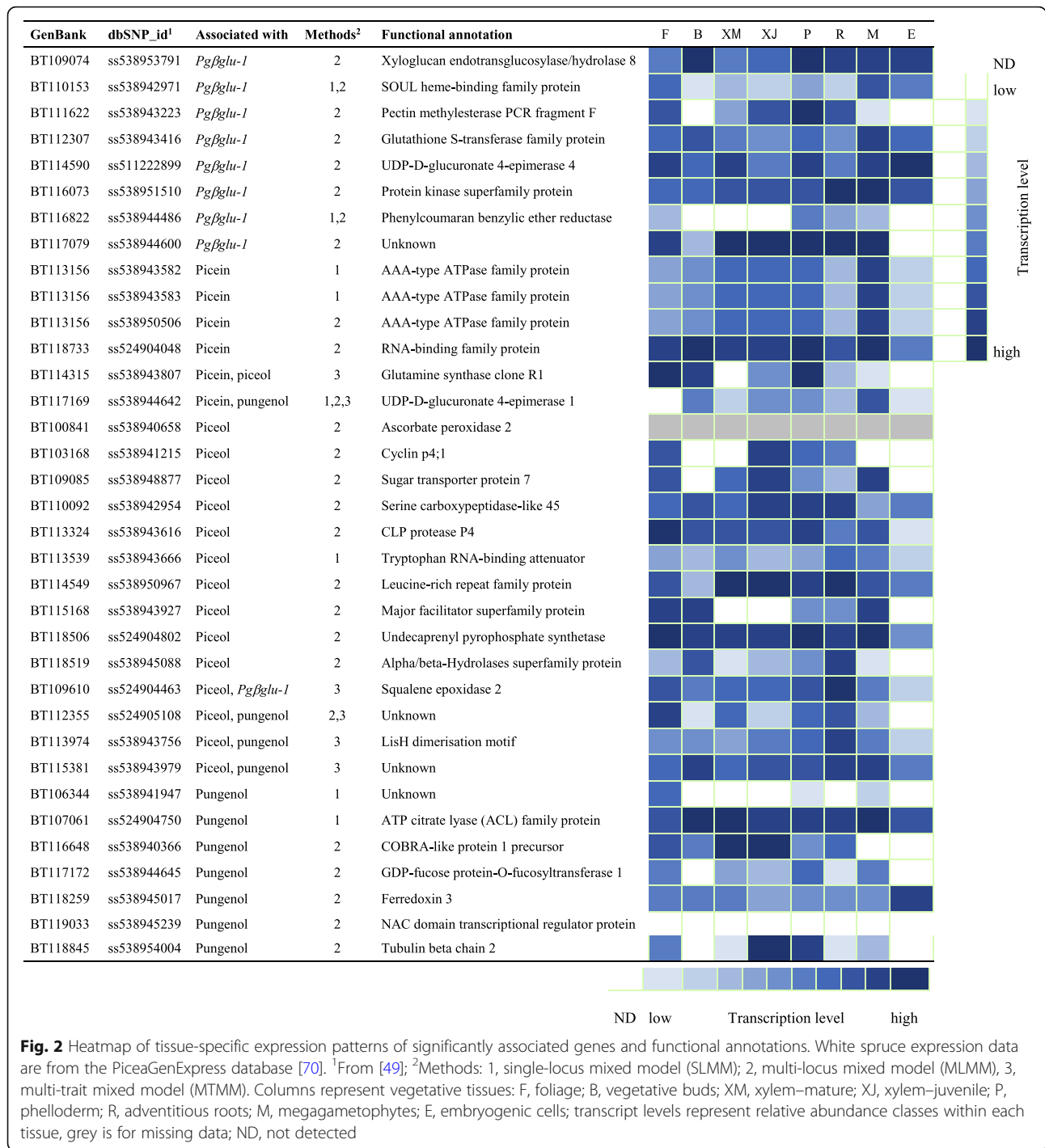
<sup>a</sup>Number of significant genes in brackets

None of the genes were annotated as encoding enzymes of the shikimic or general phenylpropanoid pathways that are responsible for the synthesis of phenolic compounds used in the formation of acetophenones, despite the fact that most of the corresponding genes were represented on the SNP array (Additional file 1: Figure S1). In contrast, the four defence traits were associated with genes involved in carbohydrate metabolism, and they were annotated as xyloglucan endotransglucosylase/hydrolase 8 (XTH8), sugar transporter protein 7 (STP7), UDP-D-glucuronate 4-epimerase 1 (GAE1) and UDP-D-glucuronate 4-epimerase 4 (GAE4). Also, genes that bear regulatory functions, including the NAC transcriptional factor, suppressor of gamma response 1 (SOG1), and genes that are involved in response to different stimulus and stress, including ascorbate peroxidase (APX), glutathione S-transferases (GST) and phenylcoumaran benzylic ether reductase1 (PCBER1) were observed as carrying significant SNPs, as well as other genes of unknown functions.

Next, we examined gene expression profiles for the 33 genes identified to carry the 35 significant SNPs by using data from the PiceaGenExpress database comprised of microarray RNA profiles [70], which indicated variable expression across tissues. The expression data indicated that most of these genes were highly expressed in foliage and also expressed at variable levels in one or several other tissues (Fig. 2).

#### Defence-growth trade-offs

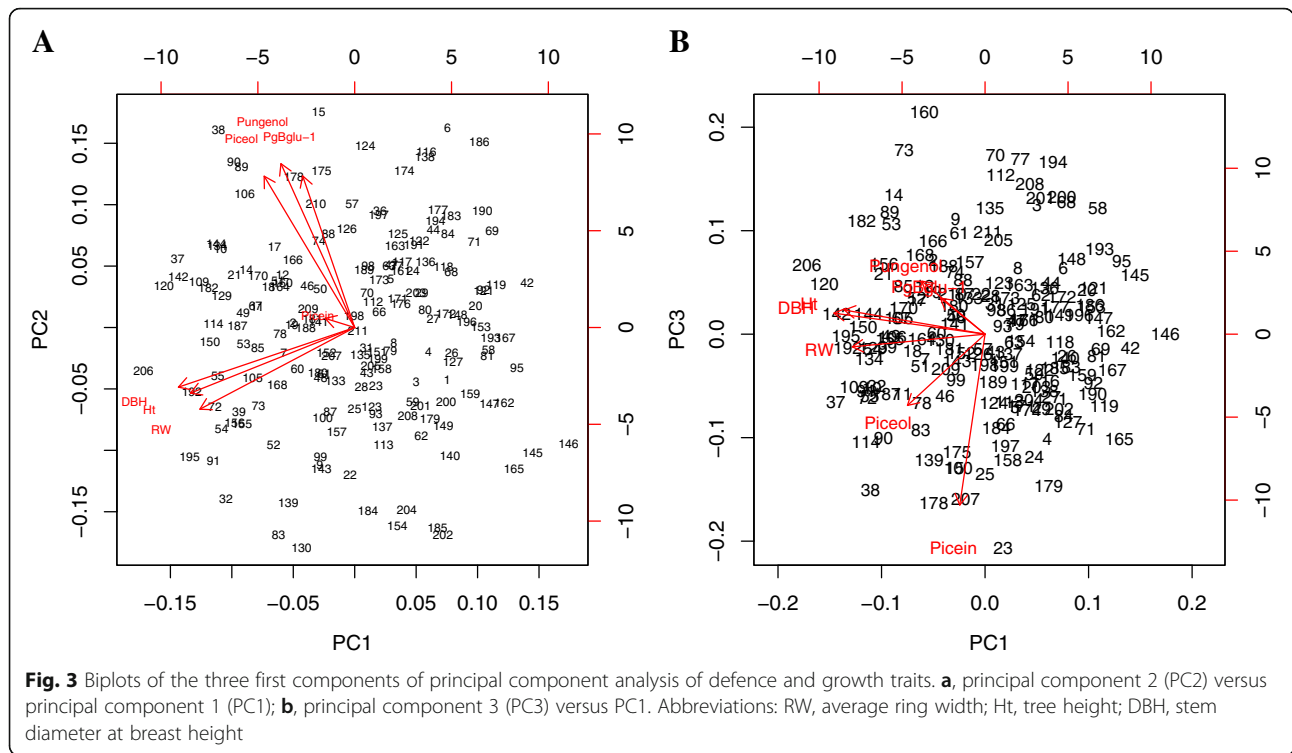
No trade-offs were identified between levels of the key defensive compounds piceol and pungenol and growth



**Table 5** Phenotypic correlations between defence traits and between defence and growth traits in white spruce

Defence traits	Piceol	Pungenol	<i>Pgβglu-1</i> transcripts	Average ring width	Total tree height	Stem diameter at breast height
Picein	0.38 (0.06) <sup>a</sup>	−0.15 (0.07)	−0.10 (0.06)	0.12 (0.07)	0.002 (0.06)	0.01 (0.06)
Piceol		0.64 (0.05)	0.43 (0.06)	0.05 (0.07)	0.06 (0.07)	0.10 (0.07)
Pungenol			0.57 (0.06)	−0.04 (0.07)	0.06 (0.07)	0.07 (0.07)
<i>Pgβglu-1</i> transcripts				−0.08 (0.06)	−0.01 (0.06)	−0.04 (0.06)

<sup>a</sup>In parentheses, standard errors



traits. First, we calculated phenotypic correlations between three different growth traits, i.e. total tree height, stem diameter at breast height and growth ring width averaged from pith to bark [45]. Phenotypic correlations were generally low between piceol or pungenol and the three growth traits (Table 5). In fact, the largest coefficient of correlation (0.12) was observed between picein and average ring width and between piceol and stem diameter, which indicates no possible trade-off. Second, a PCA analysis was carried out considering all of the traits related to defence and growth and similar results were obtained (Fig. 3). The first principal component (PC1) explained 33% of the total variation (Fig. 3), while the second (PC2) and third one (PC3) explained 29% and 18% of the variation, respectively. PC1 was largely determined by growth traits, and variation of PC2 was

controlled mostly by piceol, pungenol and the level of *Pgβglu-1* transcripts, whereas most of the variation of PC3 was controlled by picein (Table 6).

Comparisons of significantly associated genes for the different traits showed a small overlap between defence and growth traits (Fig. 4). Using a relaxed significance threshold of  $P < 0.05$ , the analyses identified close to 200 significant genes for the traits tested, and the proportion of shared significant genes ranged from 4% between piceol and stem diameter to 7% between picein and growth ring width (Fig. 4). In comparison, the overlap was two to three times higher among the defence traits and ranged from 10 to 15%. This low level of observed overlap is consistent with the weak phenotypic correlations that were observed between defence and growth traits (Table 5), thus suggesting little or no trade-off.

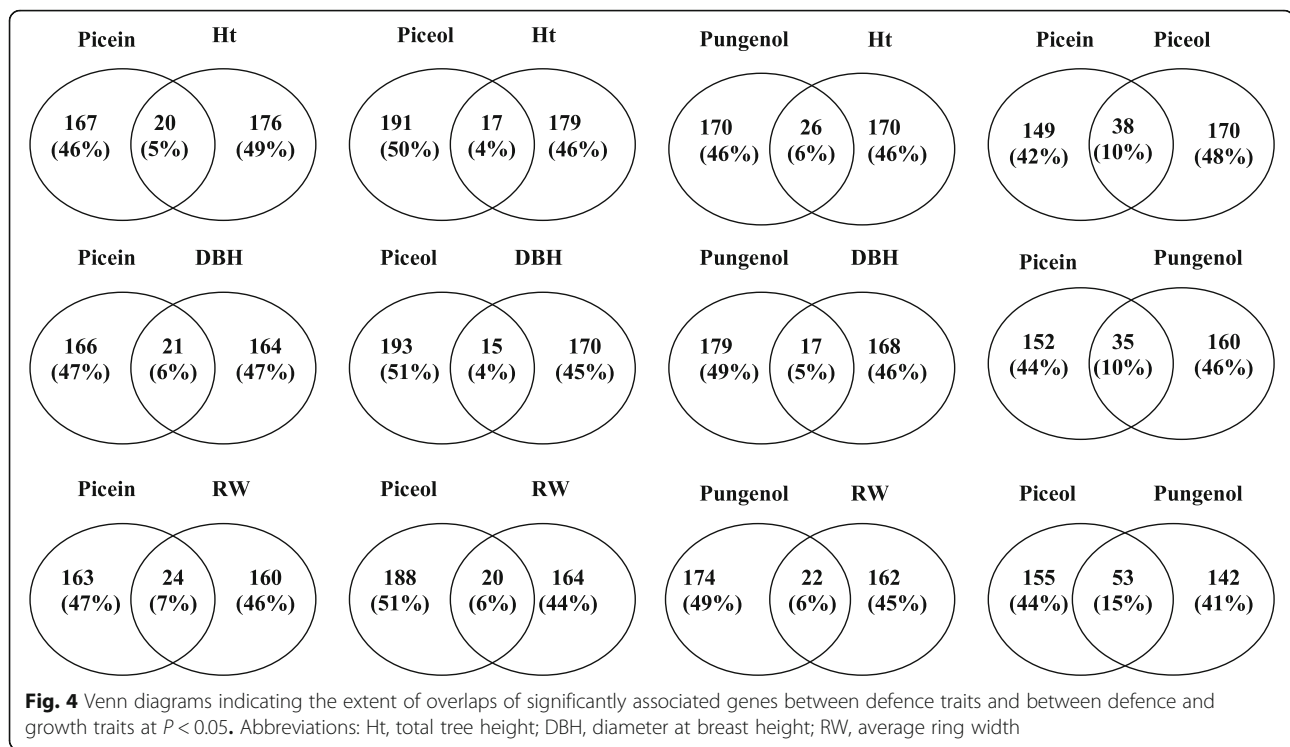
**Table 6** Factor loadings of the three first principal components (PC) for all defence and growth traits analysed in this study

Traits	PC1	PC2	PC3
Picein	-0.09	0.03	-0.86
Piceol	-0.28	0.51	-0.35
Pungenol	-0.23	0.55	0.26
<i>Pgβglu-1</i> transcripts	-0.16	0.51	0.18
Average ring width	-0.52	-0.22	0.12
Total tree height	-0.56	-0.20	0.10
Stem diameter at breast height	-0.49	-0.27	-0.06

### Discussion

Despite the economic and ecological importance of white spruce and other conifers that are attacked by the spruce budworm in North American forests, very little is known of their naturally-occurring defence mechanisms. This study aimed to contribute to the understanding of the molecular basis of SBW defence traits described by [19, 29]. Previous work has linked SBW resistance to the foliar accumulation of the acetophenones piceol and pungenol [19] and *Pgβglu-1* gene transcripts [29] based on the analysis of 20 selected white spruce trees. The glycosylated acetophenone conjugates picein and





pungenin were not linked to resistance although they accumulated to high levels in several trees [19, 29]. A recent study of full-sib families and clonal lines in white spruce found that these same chemical defences traits were moderately to highly heritable in field grown trees of six to 14 years of age [30]. Here, we looked at genetic variation and studied the molecular basis of these traits in a sample of 211 trees from as many open-pollinated families representing 42 natural populations, which were gathered and raised in a common garden experiment [32]. We identified 33 genes carrying a total of 35 SNPs significantly associated with one or more of the traits, and found that most of the genes were strongly expressed in the foliage. Acetophenones and their glycosylated conjugates accumulated to high levels in some individuals but no trade-offs were observed between defence and growth traits. We discuss the insights that are gained from these molecular analyses into the genetic control of SBW resistance.

#### Candidate genes associated with defence traits

The association genetics results presented above support a few major findings. First, several of the significantly associated genes with known predicted functions were linked to defence or included genes that have been recently found to be indirectly implicated in the biosynthesis of phenolic compounds [71, 72]. Second, the multi-locus approach allowed to identify the largest number of significant SNPs that explained a larger

proportion of the phenotypic variance. Here, SNPs identified with the MLM approach explained 20% to 43% of the phenotypic variation. In contrast, the single-locus approach only identified two significant SNPs at most for each trait, and each SNP explained only a small proportion of the phenotypic variance, as observed in several other studies in forest trees [31–33, 73–75].

Robertson [76] proposed an exponential distribution model for quantitative traits in which there are few genes with large effects and many additional genes with small effects. The exponential model represents an alternative to the infinitesimal model where a large number of loci with individual small effects contribute to the quantitative genetic variation of the trait [73, 77, 78]. Our association genetics results suggest a genetic architecture that may be closer to the exponential model for acetophenone compounds and *Pgβglu-1* expression in white spruce trees. This interpretation is supported by MLM results, which showed that a few significant genes collectively explain a large proportion of the phenotypic variation (e.g. 43% for piceol), suggesting a genetic architecture involving a moderate number of genes. It has been shown that traits involved in resistance to biotic stress may favour fixation of large-effect QTLs, and these QTLs are more common than predicted by the infinitesimal model of genetic adaptation [79]. This interpretation is also consistent with the report of moderate to high heritability for defence compounds against SBW in white spruce [30], and with

other studies on secondary metabolites that contribute to biotic resistance in plants. Research on the variation in concentration of different terpenes also suggested that they are under the control of a few major genes in conifer trees [80], eucalyptus trees [81, 82], and crop plants [83, 84].

We also observed that a large proportion of the phenotypic variation for acetophenone metabolites remained unexplained in the present association genetics study [39]. This was expected, given that genotyping data were obtained for around 10% of the transcribed genes according to conservative estimates of the gene content for spruces [48, 85].

#### Molecular basis of acetophenone accumulation

We observed that acetophenones and their glucoside conjugates reach high levels of accumulation. For instance, picein accumulated to 62.4 mg/g on a dry weight basis on average and reached much higher levels in some trees. The larvae of SBW feed primarily on newly formed foliage of spruce and fir trees in late Spring and early Summer [19, 39] and the temporal accumulation of acetophenones in the foliage is tightly linked to SBW resistance in white spruce [29, 39]. Acetophenones are thought to be derived from the phenylpropanoid pathway. However, most of the steps leading to their biosynthesis have only been proposed [86] and two genes have been shown to be directly involved in their accumulation and are involved in their glycosylation (*PgUGT5b*) [87] and deglycosylation (*Pgβglu-1*) [29]. The high levels of accumulation of acetophenones suggest that they represent a significant sink involving both phenolic and carbohydrate metabolisms. The predicted functions of the genes that we identified by association genetic approaches shed a first light onto the network of genes that may influence their synthesis and accumulation. The potential contribution of the genes is supported by data showing that nearly all of them are strongly expressed in white spruce foliage based on the transcript accumulation profiles of Raheison et al. [70]. In the following sections, we discuss our findings in light of the putative functions of the genes identified by genetic association analyses and of their potential involvement in plant metabolism.

#### Phenolic metabolism

In this study, glutamine synthetase (GS) was associated with piceol and its glycosylated form picein with the MTMM approach. In conifers, GS has been shown to be responsible for the re-assimilation of ammonium provided by the deamination of phenylalanine, the precursor for phenylpropanoid biosynthesis, in the reaction catalysed by the enzyme phenylalanine ammonia-lyase (PAL) [88, 89]. This is an efficient nitrogen recycling system

that was hypothesized to be responsible for the lack of trade-off between the accumulation of phenolic compounds and the growth of leaves or long shoots in birch [90]. Defoliation by herbivores alters the balance between nitrogen (N) sources and sinks [91] and to avoid severe N deficiency, plants have evolved an efficient N-recycling mechanism, which involves the GS enzyme system [92, 93].

Conifers produce diverse phenolic compounds that are involved in chemical defence against natural enemies [94]. The acetophenones picein and pungenin accumulate constitutively in white spruce and are believed to be synthesized via the phenylpropanoid pathway, which is also central to the synthesis of many chemical defences as well as lignin [19, 71]. However, much less is known about the genes that may be involved in the molecular regulation of the acetophenone specific branch [86]. The genes putatively involved in the phenylpropanoid pathway have been characterized in white spruce [28] and many of them were up-regulated in response to fungal infection or herbivory attack, suggesting a role in conifer defence [3]. Here, none of the core phenylpropanoid pathway genes were significantly associated genetically with the accumulation of acetophenones, although they were represented on the genotyping array. This may be explained based on two major considerations. First, most of the phenylpropanoid pathway genes are part of superfamilies in spruce [3, 28, 71] and those which are key to foliar defence may not have been adequately represented on the genotyping chip and thus may have been untested (Additional file 1: Figure S1). For example, 37 OMT /COMTL genes were identified in white spruce [71] and several of them were differentially expressed between tissues and in response to stress factors [3]. Secondly, most of the phenylpropanoid genes tested had low expression levels in foliage tissue (Additional file 1: Figure S1). This pattern suggests that other phenylpropanoid genes should be tested and selected based on the recently developed understanding of gene families [28] and expanded expression data in white spruce [95].

#### Carbohydrate metabolism

In this study, two UDP-D-Glucuronate 4-epimerases (GAEs) (GAE1 and GAE4) were significantly associated with acetophenones and *Pgβglu-1* expression. GAE1 was associated with picein by the SLMM and MLMM approaches and both picein and pungenol by the MTMM approach; GAE4 was only associated with *Pgβglu-1* transcript levels. Glycosylation consists of the attachment of a sugar moiety to phenolic compounds and is important to enhance their stability and solubility and reduce their toxicity [96]. Glycosylation by glycosyltransferases (GT) involves the transfer of sugar from its activated nucleotide sugar donor to specific acceptor molecules. One of the

common glycosyl donors in plants is UDP-glucuronate [97]. UDP-glucuronic acid (UDP-GlcA) is made from UDP-Glc via the UDP-Glc dehydrogenase activity [97]. UDP-D-Glucuronate 4-epimerases (GAEs) catalyse the reversible interconversion of UDP-D-GlcA and UDP-D-GalA [98]. One other GT gene (O-fucosyltransferase family protein) was also associated with pungenol. Plant O-fucosyltransferases (O-FuTs) are a type of GTs that catalyses the transfer of the nucleotide sugar fucose from the donor, guanosine diphosphate fucose (GDP-Fuc), to various acceptor molecules. Taken together, these observations indicate that GT enzymes influence the accumulation of at least one of the acetophenone glucoside and one of the aglycons. Functional experimentation could establish whether any of these sequences may act directly on piceol and pungenol to form the corresponding glucosides.

#### Oxidative stress control

Several of the genes identified here by association testing were potentially involved in detoxifying systems to protect cells from oxidative damage. The physiological link between acetophenones or the level of *Pgβglu-1* gene transcripts on one hand, and the oxidative stress related genes on the other hand, is unknown. But such a link is suggested by indications that piceol is cytotoxic for plant cells [99] in addition to some insects [19] and fungal pathogens [100].

The genes identified here by association genetics testing included an ascorbate peroxidase (APX) enzyme, which controls the hydrogen peroxide (H<sub>2</sub>O<sub>2</sub>) concentration in cells by catalysing its conversion to water using ascorbate as an electron donor [101, 102]. An increase in the cellular H<sub>2</sub>O<sub>2</sub> concentration in *Arabidopsis thaliana* is known to trigger DNA damage [103]. The genes SOG1, a NAC transcription factor that regulates DNA damage response [104, 105] and CYCP4;1, a cyclin, are reported to prevent oxidative damage and were both associated with pungenol and piceol. The SOG1 transcription factor regulates cyclin-dependent kinases (CDK) inhibitor genes *SMR4*, *SMR5*, and *SMR7* (belonging to the *SIAMESE/SIAMESE-RELATED* class), which are transcriptionally activated by DNA damage [103]. Cyclins are regulatory proteins that interact with CDKs to control progression through the cell cycle [106].

One of the genes significantly associated with *Pgβglu-1* transcripts encoded a phenylcoumaran benzylic ether reductase (*PCBER1*) [71], which has been reported to participate in the biosynthesis of important plant defence compounds [71, 107, 108]. Niculaes et al. [72] showed that *PCBER1* may protect against oxidative damage by chemically reducing phenylpropanoid dimers in poplar xylem. Another gene associated with *Pgβglu-1* transcripts encoded a glutathione S-transferases (GST),

which is involved in detoxification in plants [109]. In poplar trees, expression of GST increased in leaves following herbivory by forest tent caterpillars [3].

#### Trade-offs between defence and growth traits

We considered both the phenotypic correlations between traits and the list of genes significantly associated with the various traits analysed, and found little evidence for trade-offs between defence traits against SBW and growth in white spruce. These results suggest that the cost of constitutive production of piceol and pungenol as secondary metabolites does not affect primary metabolism needed to sustain growth, which is consistent with previous results in other forest trees. Similar findings were reported in a previous investigation on white spruce resistance to SBW based on the analysis of full-sib families and clonal lines where low and non-significant genetic correlations were observed between defence and growth traits [30]. In Douglas-fir and Scots pine, the accumulation of constitutive phenolic compounds in bark was also not correlated with growth [110, 111]. In terms of biosynthetic costs, phenolic compounds have also been suggested to be less costly to produce than other compounds such as alkaloids, which have a higher energy requirement to make inorganic nitrogen bioavailable (reviewed in [112]).

Among the three acetophenones studied here, the glycosylated acetophenone picein was the most abundant, and by far, as it made up 6% of the total needle dry mass on average. High foliar concentrations of phenolic compounds were also observed in other tree species and were correlated with resistance to herbivores. For instance, condensed tannins represented over 10% of the dry mass in birch leaves [90, 113] and phenolic glycosides constituted up to 4% of leaf dry weight in aspen (*Populus tremuloides*) [114]. Picein production may function as a reservoir for storing sugar (carbohydrates) in white spruce foliage. In silver birch, it has been shown that some phenolic compounds may act as a reservoir for the synthesis of other phenolic compounds when the phenylpropanoid metabolism is activated, and storing surplus carbon as cinnamoylquinates would be a better defence against herbivory than the accumulation of storage carbohydrates such as starch, thus potentially allowing a more rapid response to environmental threats [115, 116].

#### Conclusions

The present study represents a first step in understanding and dissecting the genetic architecture of defence traits against SBW in white spruce. We explored three different association genetics testing approaches and, taking advantage of the genomic resources developed for white spruce, we detected 33 genes carrying SNPs

significantly involved in the observed variation for defence traits. Our results indicate that the multi-locus association genetic approach is more powerful than the single-locus approach for identifying candidate genes implicated in the constitutive defence against SBW. We further showed that these traits are likely to be under the mixed control of minor and major genes with no significant trade-offs with growth traits. The present results should open up new opportunities for functional studies to determine the molecular roles of these genes in influencing SBW resistance. In addition, these genes and a more complete determination of their polymorphisms should allow to develop molecular tools to help identify and breed trees that are more resistant to SBW, which have been lacking to date. These tools may thus represent a means to shorten the long periods of time that tree breeders need to assess defence against SBW in the field.

## Additional file

**Additional file 1: Figure S1.** Heatmap of tissue-specific expression pattern of candidate genes involved in phenylpropanoid pathway used in this study and their functional annotations. Expression data are from the PiceaGenExpress database [70]. Columns represent vegetative tissues: F, foliage; B, vegetative buds; XM, xylem–mature; XJ, xylem–juvenile; P, phellogen; R, adventitious roots; M, megagametophytes; E, embryogenic cells; transcript levels represent relative abundance classes within each tissue, grey is for missing data; ND, not detected. (DOCX 128 kb)

## Abbreviations

DBH: Stem diameter at breast height; FDR: False discovery rate; GO: Gene ontology; GWAS: Genome-wide association studies; Ht: Tree height; LD: Linkage disequilibrium; MLM: Multi-locus mixed model; MTMM: Multi-trait mixed model; PCA: Principal component analysis; Pgluc1: *Picea glauca*  $\beta$ -glucosidase-1; PVE: Phenotypic variation explained; QTL: Quantitative trait locus; RW: Ring width; SBW: Spruce budworm; SLM: Single-locus mixed model; SNP: Single-nucleotide polymorphism

## Acknowledgements

We thank the Canadian Forest Service (Laurentian Forestry Centre) for the establishment to the white spruce field test and the Ministère des Forêts, de la Faune et des Parcs du Québec (MFFPQ) for accessing the Mastigouche arboretum. We thank F. Gagnon and S. Blais (Univ. Laval), and M. Deslauriers and S. Clément (Natural Resources Canada) for assistance with handling genotyping data, and A. Rainville (MFFPQ) for providing information on the field test. We also acknowledge C. Mendez Espinoza, G. Piette-Lauzière, J. Piette, D. Vigneault, K. Guay, and K. Beaupré-Boivin (Forest Research Centre, Univ. Laval) for field and laboratory assistance.

## Funding

Funding for the project was received from Genome Canada and Génome Québec for the large-scale spruce genomics projects SmarTForests and Spruce-Up (JJM, JBo).

## Availability of data and materials

All data generated or analysed during this study are included in the manuscript and its supplementary files. The gene expression data are from: <https://bmcbgenomics.biomedcentral.com/articles/10.1186/1471-2164-13-434#Declarations>

## Authors' contributions

M.L. J.J.M. J.Bo. and J.Be planned and designed the study. M.L. performed data analyses and drafted the manuscript. G.J.P. designed and oversaw the sampling and the acetophenone determinations, I.G. conducted the gene expression assays, J.J.M. J.Bo. and J.Be. supervised the study and revised the manuscript. All authors read and approved the final manuscript.

## Ethics approval and consent to participate

The sampling was non-destructive and the trees were part of an experimental plantation established for research on land of the government of Québec. A collaborative research agreement between the organizations as part of the Arborea II project gave permission for the sampling, which followed guidelines of the institutions involved in the research and in force in Québec (Canada). We thank the Canadian Forest Service (Laurentian Forestry Centre) for allowing the white spruce field test and the Ministère des Forêts, de la Faune et des Parcs du Québec (MFFPQ) for accessing the Mastigouche arboretum where plant samples were sourced.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Author details

<sup>1</sup>Forest Research Centre and Institute for Systems and Integrative Biology, Département des sciences du bois et de la forêt, Université Laval, Qc, Québec G1V 0A6, Canada. <sup>2</sup>Department of Plant Sciences, University of Oxford, Oxford OX1 3RB, UK. <sup>3</sup>Canada Research Chair in Forest Genomics, Université Laval, Qc, Québec G1V 0A6, Canada.

Received: 13 September 2017 Accepted: 23 September 2018

Published online: 12 October 2018

## References

- Clancy KM. Mechanisms of resistance in trees to defoliators. In: Wagner MR, Clancy KM, Lieutier F, Paine TD, editors. Mechanisms and deployment of resistance in trees to insects. The Netherlands: Kluwer Academic Publishers; 2002. p. 79–103.
- Keeling CI, Bohlmann J. Diterpene resin acids in conifers. *Phytochemistry*. 2006;67:2415–23.
- Ralph SG, Yueh H, Friedmann M, Aeschliman D, Zelnik JA, Nelson CC, Butterfield YSN, Kirkpatrick R, Liu J, Jones SJM, Marra MA, Douglas CJ, Ritland K, Bohlmann J. Conifer defense against insects: microarray gene expression profiling of Sitka spruce (*Picea sitchensis*) induced by mechanical wounding or feeding by spruce budworms (*Choristoneura occidentalis*) or white pine weevils (*Pissodes strobi*) reveals large-scale changes of the host transcriptome. *Plant Cell Environ*. 2006;29:1545–70.
- Kovalchuk A, Keriö S, Oghenekaro AO, Jaber E, Raffaello T, Asiegbu FO. Antimicrobial defenses and resistance in forest trees: challenges and perspectives in a genomic era. *Annu Rev Phytopathol*. 2013;51:221–44.
- Westbrook JW, Resende MFR, Munoz P, Wegrzyn JL, Nelson CD, Neale DB, Kirst M, Huber DA, Gezan SA, Peter GF, Davis JM. Association genetics of oleoresin flow in loblolly pine: discovering genes and predicting phenotype for improved resistance to bark beetles and bioenergy potential. *New Phytol*. 2013;199:89–100.
- Westbrook JW, Walker AR, Neves LG, Munoz P, Jr MFR, Neale DB, Wegrzyn JL, Huber DA, Kirst M, Davis JM, Peter GF. Discovering candidate genes that regulate resin canal number in *Pinus taeda* stems by integrating genetic analysis across environments, ages, and populations. *New Phytol*. 2015;205:627–41.
- Keeling CI, Bohlmann J. Genes, enzymes and chemicals of terpenoid diversity in the constitutive and induced defence of conifers against insects and pathogens. *New Phytol*. 2006;170:657–75.
- Keeling CI, Dullat HK, Yuen M, Ralph SG, Jancsik S, Bohlmann J. Identification and functional characterization of monofunctional ent-copalyl diphosphate and ent-kaurene synthases in white spruce reveal different patterns for

- diterpene synthase evolution for primary and secondary metabolism in gymnosperms. *Plant Physiol.* 2010;152:1197–208.
9. Keeling CI, Madilao LL, Zerbe P, Dullat HK, Bohlmann J. The primary diterpene synthase products of *Picea abies* levopimaradiene/abietadiene synthase (PaLAS) are epimers of a thermally unstable diterpenol. *J Biol Chem.* 2011;286:21145–53.
  10. Keeling CI, Weisshaar S, Ralph SG, Jancsik S, Hamberger B, Dullat HK. Transcriptome mining, functional characterization, and phylogeny of a large terpene synthase gene family in spruce (*Picea* spp.). *BMC Plant Biol.* 2011;11:43.
  11. Hamberger B, Ohnishi T, Hamberger B, Séguin A, Bohlmann J. Evolution of diterpene metabolism: Sitka spruce CYP720B4 catalyzes multiple oxidations in resin acid biosynthesis of conifer defense against insects. *Plant Physiol.* 2011;157:1677–95.
  12. Blais J. Trends in the frequency, extent, and severity of spruce budworm outbreaks in eastern Canada. *Can J For Res.* 1983;13:539–47.
  13. Sanders CJ. Biology of north American spruce budworm. In: Van der Geest LPS, Evenhuis HHE, editors. *Tortricid pests their biology, natural enemies and control.* Amsterdam: Elsevier; 1991. p. 579–620.
  14. Fleming RA. Climate change and insect disturbance regimes in Canada's boreal forests. *World Resour Rev.* 2000;12:521–55.
  15. Kneeshaw D, Bergeron Y. Forest ecosystem dynamics across the circumboreal forest. In: Millington AC, Blumler MA, MacDonald G, Shickhoff U, Kuuluvainen T, editors. *Handbook of biogeography.* Washington: Sage; 2011. p. 261–8.
  16. Gray DR, MacKinnon WE. Outbreak patterns of the spruce budworm and their impacts in Canada. *For Chron.* 2006;82:550–61.
  17. Rauchfuss J, Ziegler SS. The geography of spruce budworm in eastern North America. *Geogr Compass.* 2011;5:564–80.
  18. Hennigar CR, MacLean DA, Quiring DT, Kershaw JA. Differences in spruce budworm defoliation among balsam fir and white, red, and black spruce. *For Sci.* 2008;54:158–66.
  19. Delvas N, Bauce É, Labbé C, Ollevier T, Bélanger R. Phenolic compounds that confer resistance to spruce budworm. *Entomol Exp Appl.* 2011;141:35–44.
  20. Fuentealba A, Bauce É. Interspecific variation in resistance of two host tree species to spruce budworm. *Acta Oecol.* 2016;70:10–20.
  21. Aspinwall MJ, King JS, Booker FL, McKeand SE. Genetic effects on total phenolics, condensed tannins and non-structural carbohydrates in loblolly pine (*Pinus taeda* L.) needles. *Tree Physiol.* 2011;31:831–42.
  22. Sampedro L, Moreira X, Llusia J, Peñuelas J, Zas R. Genetics, phosphorus availability, and herbivore-derived induction as sources of phenotypic variation of leaf volatile terpenes in a pine species. *J Exp Bot.* 2010;61:4437–47.
  23. Ott DS, Yanchuk AD, Huber DPW, Wallin KF. Genetic variation of lodgepole pine, *Pinus contorta* var. *latifolia*, chemical and physical defenses that affect mountain pine beetle, *Dendroctonus ponderosae*, attack and tree mortality. *J Chem Ecol.* 2011;37:1002–12.
  24. Andrew RL, Wallis IR, Harwood CE, Foley WJ. Genetic and environmental contributions to variation and population divergence in a broad-spectrum foliar defence of *Eucalyptus tricarpa*. *Ann Bot.* 2010;105:707–17.
  25. Robinson KM, Ingvarsson PK, Jansson S, Albrechtsen BR. Genetic variation in functional traits influences arthropod community composition in aspen (*Populus tremula* L.). *PLoS One.* 2012;7:e37679.
  26. O'Reilly-Wapstra JM, Miller AM, Hamilton MG, Williams D, Glancy-Dean N, Potts BM. Chemical variation in a dominant tree species: population divergence, selection and genetic stability across environments. *PLoS One.* 2013;8:e58416.
  27. Andrew RL, Wallis IR, Harwood CE, Henson M, Foley WJ. Heritable variation in the foliar secondary metabolite sideroxylonal in *Eucalyptus* confers cross-resistance to herbivores. *Oecologia.* 2007;153:891–901.
  28. Warren RL, Keeling CI, Yuen MM, Raymond A, Taylor GA, Vandervalk BP, Mohamadi H, Paulino D, Chiu R, Jackman SD, et al. Improved white spruce (*Picea glauca*) genome assemblies and annotation of large gene families of conifer terpenoid and phenolic defense metabolism. *Plant J.* 2015;83:189–212.
  29. Mageroy MH, Parent G, Germanos G, Giguère I, Delvas N, Maaroufi H, Bauce É, Bohlmann J, MacKay J. Expression of the  $\beta$ -glucosidase gene *Pg $\beta$ glu-1* underpins natural resistance of white spruce against spruce budworm. *Plant J.* 2015;81:68–80.
  30. Méndez-Espinoza C, Parent GJ, Lenz P, Rainville A, Tremblay L, Adams G, McCartney A, Bauce É, MacKay J. Genetic control and evolutionary potential of a constitutive resistance mechanism against the spruce budworm (*Choristoneura fumiferana*) in white spruce (*Picea glauca*). *Heredity.* 2018;121:142–54.
  31. Dillon SK, Nolan M, Li W, Bell C, Wu HX, Southerton SG. Allelic variation in cell wall candidate genes affecting solid wood properties in natural populations and land races of *Pinus radiata*. *Genetics.* 2010;185:1477–87.
  32. Beaulieu J, Doerksen T, Boyle B, Clément S, Deslauriers M, Beauseigle S, Blais S, Poulin PL, Lez P, Caron S, Rigault P, Bicho P, Bousquet J, MacKay J, Kirst M. Association genetics of wood physical traits in the conifer white spruce and relationships with gene expression. *Genetics.* 2011;188:197–214.
  33. Guerra FP, Wegrzyn JL, Sykes R, Davis MF, Stanton BJ, Neale DB. Association genetics of chemical wood properties in black poplar (*Populus nigra*). *New Phytol.* 2012;197:162–76.
  34. Lamara M, Raheison E, Lenz P, Beaulieu J, Bousquet J, Mackay J. Genetic architecture of wood properties based on association analysis and co-expression networks in white spruce. *New Phytol.* 2016;210:240–55.
  35. Külheim C, Yeoh SH, Wallis IR, Laffan S, Moran GF, Foley WJ. The molecular basis of quantitative variation in foliar secondary metabolites in *Eucalyptus globulus*. *New Phytol.* 2011;191:1041–53.
  36. Neale DB, Savolainen O. Association genetics of complex traits in conifers. *Trends Plant Sci.* 2004;9:325–30.
  37. Segura V, Vilhjálmsson BJ, Platt A, Korte A, Seren Ü, Long Q, Nordborg M. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat Genet.* 2012;44:825–30.
  38. Korte A, Vilhjálmsson BJ, Segura V, Platt A, Long Q, Nordborg M. A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nat Genet.* 2012;44:1066–71.
  39. Parent G, Giguère I, Germanos G, Lamara M, Bauce É, Mackay J. Insect herbivory (*Choristoneura fumiferana*, Tortricidae) underlies tree population structure (*Picea glauca*, Pinaceae). *Sci Rep.* 2017;7:42273.
  40. Herms DA, Mattson WJ. The dilemma of plants: to grow or defend. *Q Rev Biol.* 1992;67:283–35.
  41. Koricheva J, Larsson S, Haukioja E, Keinänen M. Regulation of woody plant secondary metabolism by resource availability: hypothesis testing by means of meta-analysis. *Oikos.* 2012;83:212–26.
  42. Stamp N. Out of the quagmire of plant defense hypotheses. *Q Rev Biol.* 2003;78:23–55.
  43. Agrawal AA. Macroevolution of plant defense strategies. *Trends Ecol Evol.* 2006;22:103–9.
  44. Sampedro L. Physiological trade-offs in the complexity of pine tree defensive chemistry. *Tree Physiol.* 2014;34:915–8.
  45. Beaulieu J, Doerksen T, Clément S, MacKay J, Bousquet J. Accuracy of genomic selection models in a large population of open-pollinated families in white spruce. *Heredity.* 2014;113:343–52.
  46. Chang S, Puryear J, Cairney J. A simple and efficient method for isolating RNA from pine trees. *Plant Mol Biol Report.* 1993;11:113–6.
  47. Pavy N, Boyle B, Nelson C, Paule C, Giguère I, Caron S, Parsons LS, Dallaire N, Bedon F, Bérubé H, Cooke J, MacKay J. Identification of conserved core xylem gene sets: conifer cDNA microarray development, transcript profiling and computational analyses. *New Phytol.* 2008;180:766–86.
  48. Rigault P, Boyle B, Lepage P, Cooke J, Bousquet J, MacKay J. A White spruce gene catalog for conifer. *Plant Physiol.* 2011;157:14–28.
  49. Pavy N, Gagnon F, Rigault P, Blais S, Boyle B, Pelgas B, Deslauriers M, Clément S, Lavigne P, Lamothe M, Cooke J, Jaramillo Correa JP, Beaulieu J, Isabel N, MacKay J, Bousquet J. Development of high-density SNP genotyping arrays for white spruce (*Picea glauca*) and transferability to subtropical and nordic congeners. *Mol Ecol Resour.* 2013;13:324–36.
  50. Groover AT. What genes make a tree a tree? *Trends Plant Sci.* 2005;10:210–4.
  51. Demura T, Fukuda H. Transcriptional regulation in wood formation. *Trends Plant Sci.* 2006;12:64–70.
  52. Zhang J, Elo A, Helariutta Y. *Arabidopsis* as a model for wood formation. *Curr Opin Biotech.* 2011;22:293–9.
  53. EL Kayal W, Allen CCG, Ju CJT, Adams E, King-Jones S, Zaharia LI, Abrams SR, Cooke J. Molecular events of apical bud formation in white spruce, *Picea glauca*. *Plant Cell Environ.* 2011;34:480–500.
  54. Guillet-Claude C, Isabel N, Pelgas B, Bousquet J. The evolutionary implications of *knox-1* gene duplications in conifers: correlated evidence from phylogeny, gene mapping, and analysis of functional divergence. *Mol Biol Evol.* 2004;21:2232–45.
  55. Bomal C, Bedon F, Caron S, Mansfield SD, Levasseur C, Cooke JE, Blais S, Tremblay L, Morency MJ, Pavy N, Grima-Pettenati J, Séguin A, MacKay J. Involvement of *Pinus taeda* *MYB1* and *MYB8* in phenylpropanoid metabolism and secondary cell wall biogenesis: a comparative in planta analysis. *J Exp Bot.* 2008;59:3925–39.

56. Bedon F, Bomal C, Caron S, Levasseur C, Boyle B, Mansfield SD, Schmidt A, Gershenzon J, Grima-Pettenati J, Séguin A, MacKay J. Subgroup 4 R2R3-MYBs in conifer trees: gene family expansion and contribution to the isoprenoid- and flavonoid- oriented responses. *J Exp Bot.* 2010;61:3847–64.
57. Côté CL, Boileau F, Roy V, Ouellet M, Levasseur C, Morency M-J, Cooke JEK, Séguin A, Mackay JJ. Gene family structure, expression and functional analysis of HD-zip III genes in angiosperm and gymnosperm forest trees. *BMC Plant Biol.* 2010;10:273.
58. Pelgas B, Bousquet J, Meirmans PG, Ritland K, Isabel N. QTL mapping in white spruce: gene maps and genomic regions underlying adaptive traits across pedigrees, years and environments. *BMC Genomics.* 2011;12:495.
59. Namroud M-C, Beaulieu J, Juge N, Laroche J, Bousquet J. Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. *Mol Ecol.* 2008;17:3599–613.
60. Pavy N, Lamothe M, Pelgas B, Gagnon F, Birol I, Bohlmann J, Mackay J, Isabel N, Bousquet J. A high-resolution reference genetic map positioning 8 K genes for the conifer white spruce: structural genomics implications and correspondence with physical distance. *Plant J.* 2017;90:189–203.
61. Pérez P, de los Campos G. Genome-wide regression & prediction with the BGLR statistical package. *Genetics.* 2014;198:482–95.
62. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. *Bioinformatics.* 2007;23:1294–6.
63. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics.* 2007;23:2633–5.
64. Yu J, Pressoir G, Briggs WH, Vroh BI, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler E. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet.* 2006;38:203–8.
65. Wen W, Li D, Li X, Gao Y, Li W, Li H, Liu J, Liu H, Chen W, Luo J, Yan J. Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. *Nat Commun.* 2014;5:3438.
66. Ding J, Ali F, Chen G, Li H, Mahuku G, Yang N, Narro L, Magorokosho C, Makumbi D, Yan J. Genome-wide association mapping reveals novel sources of resistance to northern corn leaf blight in maize. *BMC Plant Biol.* 2015;15:206.
67. Zhou X, Stephens M. Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat Methods.* 2014;11:407–9.
68. van Heerwaarden J, van Zanten M, Kruijer W. Genome-wide association analysis of adaptation using environmentally predicted traits. *PLoS Genet.* 2015;11:e1005594.
69. Development Core Team R. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2016. <http://www.r-project.org/>.
70. Raheison E, Rigault P, Caron S, Poulin P, Boyle B, Verta J, Giguère I, Bomal C, Bohlmann J, MacKay J. Transcriptome profiling in conifers and the PiceaGenExpress database show patterns of diversification within gene families and interspecific conservation in vascular gene expression. *BMC Genomics.* 2012;13:434.
71. Porth I, Hamberger B, White R, Ritland K. Defense mechanisms against herbivory in *Picea*: sequence evolution and expression regulation of gene family members in the phenylpropanoid pathway. *BMC Genomics.* 2011;12:608.
72. Niculac C, Morreel K, Kim H, Lu F, McKee LS, Ivens B, Hastraete J, Vanholme B, De Rycke R, Hertzberg M, Fromm J, Bulone V, Polle A, Ralph J, Boerjan W. Phenylcoumaran benzylic ether reductase prevents accumulation of compounds formed under oxidative conditions in poplar xylem. *Plant Cell.* 2014;26:3775–91.
73. Quesada T, Gopal V, Cumbie WP, Eckert AJ, Wegrzyn JL, Neale DB, Goldfarb B, Huber DA, Casella G, Davis JM. Association mapping of quantitative disease resistance in a natural population of loblolly pine (*Pinus taeda* L.). *Genetics.* 2010;186:677–86.
74. Prunier J, Pelgas B, Gagnon F, Despons M, Isabel N, Beaulieu J, Bousquet J. The genomic architecture and association genetics of adaptive characters using a candidate SNP approach in boreal black spruce. *BMC Genomics.* 2013;14:368.
75. Lind M, Källman T, Chen J, Ma XF, Bousquet J, Morgante M, Zaina G, Karlsson B, Elfstrand M, Lascoux M, Stenlid J. A *Picea abies* linkage map based on SNP markers identifies QTLs for four aspects of resistance to *Heterobasidion parviporum* infection. *PLoS One.* 2014;9:e101049.
76. Robertson A. The nature of quantitative genetic variation. In: Brink RA, Styles ED, editors. *Heritage from Mendel*. Madison, WI: University of Wisconsin; 1967. p. 265–80.
77. Falconer DS, Mackay TFC, Frankham R. Introduction to quantitative genetics (4<sup>th</sup> edn). *Trends Genet.* 1996;12:280.
78. Lynch M, Walsh B. Genetics and analysis of quantitative traits. Sunderland, Massachusetts: Sinauer Associates; 1998.
79. Louthan AM, Kay KM. Comparing the adaptive landscape across trait types: larger QTL effect size in traits under biotic selection. *BMC Evol Biol.* 2011;11:60.
80. Phillips MA, Croteau RB. Resin-based defenses in conifers. *Trends Plant Sci.* 1999;4:184–90.
81. Henery ML, Moran GF, Wallis IR, Foley WJ. Identification of quantitative trait loci influencing foliar concentrations of terpenes and formylated phloroglucinol compounds in *Eucalyptus nitens*. *New Phytol.* 2007;176:82–95.
82. O'Reilly-Wapstra JM, Freeman JS, Davies NW, Vaillancourt RE, Fitzgerald H, Potts BM. Quantitative trait loci for foliar terpenes in a global eucalypt species. *Tree Genet Genomes.* 2011;7:485–98.
83. Battilana J, Costantini L, Emanuelli F, Sevini F, Segala C, Moser S, Velasco R, Versini G, Grando MS. The 1-deoxy-d-xylulose 5-phosphate synthase gene co-localizes with a major QTL affecting monoterpene content in grapevine. *Theor Appl Genet.* 2009;118:653–69.
84. Nieuwenhuizen NJ, Wang MY, Matich AJ, Green SA, Chen X, Yauk Y, Beuning LL, Nagegowda DA, Dudareva N, Atkinson RG. Two terpene synthases are responsible for the major sesquiterpenes emitted from the flowers of kiwifruit (*Actinidia deliciosa*). *J Exp Bot.* 2009;60:3203–19.
85. Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin Y-C, 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, Rilakovic N, Ritland C, Sena J, Niittylä T, et al. The Norway spruce genome sequence and conifer genome evolution. *Nature.* 2013;497:579–84.
86. Negrel J, Javelle F. The biosynthesis of acetovanillone in tobacco cell-suspension cultures. *Phytochemistry.* 2010;71:751–9.
87. Mageroy MH, Jancsik S, Saint YMM, Fischer M, Withers SG, Paetz C, Schneider B, Mackay J, Bohlmann J. A conifer UDP-sugar dependent glycosyltransferase contributes to acetophenone metabolism and defense against insects. *Plant Physiol.* 2017. <https://doi.org/10.1104/pp.17.00611>.
88. Cánovas FM, Avila C, Cantón FR, Cañas RA, de la Torre F. Ammonium assimilation and amino acid metabolism in conifers. *J Exp Bot.* 2007;58:2307–18.
89. Craven-Bartle B, Pascual MB, Cánovas FM, Ávila C. A Myb transcription factor regulates genes of the phenylalanine pathway in maritime pine. *Plant J.* 2013;74:755–66.
90. Riipi M, Ossipov V, Lempa K, Haukioja E, Koricheva J, Ossipova S, Pihlaja K. Seasonal changes in birch leaf chemistry: are there trade-offs between leaf growth and accumulation of phenolics? *Oecologia.* 2002;130:380–90.
91. Palacio S, Hernández R, Maestro-Martínez M, Camarero JJ. Fast replenishment of initial carbon stores after defoliation by the pine processionary moth and its relationship to the re-growth ability of trees. *Trees.* 2012;26:1627–40.
92. Razal R, Ellis S, Singh S, Lewis NG, Towers GHN. Nitrogen recycling in phenylpropanoid metabolism. *Phytochemistry.* 1996;41:31–5.
93. Bernards MA, Båstrup-Spöhr L. Phenylpropanoid metabolism induced by wounding and insect herbivory. In: Schaller A, editor. *Induced plant resistance to herbivory*. New York: Springer; 2008. p. 189–213.
94. Danielsson M, Lundén K, Elfstrand M, Hu J, Zhao T, Arnerup J, Ihrmark K, Swedjemark G, Borg-Karlson A-K, Stenlid J. Chemical and transcriptional responses of Norway spruce genotypes with different susceptibility to *Heterobasidion* spp. infection. *BMC Plant Biol.* 2011;11:154.
95. Raheison ESM, Giguère I, Caron S, Lamara M, Mackay JJ. Modular organization of the white spruce (*Picea glauca*) transcriptome reveals functional organization and evolutionary signatures. *New Phytol.* 2015; 207:172–87.
96. Dima O, Morreel K, Vanholme B, Kim H, Ralph J, Boerjan W. Small glycosylated lignin oligomers are stored in *Arabidopsis* leaf vacuoles. *Plant Cell.* 2015;27:695–710.
97. Sawada S, Suzuki H, Ichimaida F, Yamaguchi M, Iwashita T, Fukui Y, Hemmi H, Nishino T, Nakayama T. UDP-glucuronic acid: anthocyanin glucuronosyltransferase from red daisy (*Bellis perennis*) flowers. *J Biol Chem.* 2005;280:899–906.
98. Seifert GJ. Nucleotide sugar interconversions and cell wall biosynthesis: how to bring the inside to the outside. *Curr Opin Plant Biol.* 2004;7:277–84.
99. Curir P, Marchesini A, Danieli B, MARIANI F. 3-Hydroxyacetophenone in carnations is a phytoanticipin active against *Fusarium oxysporum* f. sp. *dianthi*. *Phytochemistry.* 1996;41:447–50.

100. Løkke H. Picein and piceol concentrations in Norway spruce. *Ecotoxicol Environ Saf.* 1990;19:301–9.
101. Asada K. The water-water cycle in chloroplasts: scavenging of active oxygens and dissipation of excess photons. *Annu Rev Plant Physiol Plant Mol Biol.* 1999;50:601–39.
102. Ramírez L, Bartoli CG, Lamattina L. Glutathione and ascorbic acid protect *Arabidopsis* plants against detrimental effects of iron deficiency. *J Exp Bot.* 2013;64:3169–78.
103. Yi D, Lessa C, Kamei A, Cools T, Vanderauwera S, Takahashi N, Okushima Y, Eekhout T, Yoshiyama KO, Larkin J, Van den Daele H, Conklin P, Britt A, Umeda M, De Veylder L. The *Arabidopsis* SIAMESE-RELATED cyclin-dependent kinase inhibitors SMR5 and SMR7 regulate the DNA damage checkpoint in response to reactive oxygen species. *Plant Cell.* 2014;26:296–309.
104. Yoshiyama KO. SOG1: a master regulator of the DNA damage response in plants. *Genes Genet Syst.* 2015;90:209–16.
105. Yoshiyama K, Conklin PA, Huefner ND, Britt AB. Suppressor of gamma response 1 (SOG1) encodes a putative transcription factor governing multiple responses to DNA damage. *Proc Natl Acad Sci U S A.* 2009;106:12843–8.
106. Torres Acosta JA, de Almeida Engler J, Raes J, Magyar Z, De Groot R, Inze D, De Veylder L. Molecular characterization of *Arabidopsis* PHO80-like proteins, a novel class of CDKA1-interacting cyclins. *Cell Mol Life Sci.* 2004;61:1485–97.
107. Gang DR, Kasahara H, Xia Z, Vander MK, Bauw G, Boerjan W, Montagu M, Davin LB, Lewis NG. Evolution of plant defense mechanisms relationships of phenylcoumaran benzylic ether reductases to pinoresinol-hariciresinol and isoflavone reductases. *J Biol Chem.* 1999;274:7516–27.
108. Vander Mijsbrugge K, Beeckman H, De Rycke R, Van Montagu M, Engler G, Boerjan W. Phenylcoumaran benzylic ether reductase, a prominent poplar xylem protein, is strongly associated with phenylpropanoid biosynthesis in lignifying cells. *Planta.* 2000;211:502–9.
109. Reymond P, Bodenhausen N, RMP VP, Krishnamurthy V, Dicke M, Farmer EE. A conserved transcript pattern in response to a specialist and a generalist herbivore. *Plant Cell.* 2004;16:3132–47.
110. Myszewski JH, Fins L, Moore JA, Rust M, Mika PG. Variation in the root bark phenolics/sugar ratio of Douglas-fir grown in two plantations in northern Idaho. *Can J For Res.* 2002;32:556–60.
111. Villari C, Faccoli M, Battisti A, Bonello P, Marini L. Testing phenotypic trade-offs in the chemical defence strategy of scots pine under growth-limiting field conditions. *Tree Physiol.* 2014;34:919–30.
112. Mithöfer A, Boland W. Plant defense against herbivores: chemical aspects. *Ann Rev Plant Biol.* 2012;63:431–50.
113. Witzell J, Ja M. Phenolic metabolites in the resistance of northern forest trees to pathogens—past experiences and future prospects. *Can J For Res.* 2008;38:2711–27.
114. Constabel CP, Yip L, Patton JJ, Christopher ME. Polyphenol oxidase from hybrid poplar. Cloning and expression in response to wounding and herbivory. *Plant Physiol.* 2000;124:285–95.
115. Yao K, De Luca V, Brisson N. Creation of a metabolic sink for tryptophan alters the phenylpropanoid pathway and the susceptibility of potato to *Phytophthora infestans*. *Plant Cell.* 1995;7:1787–99.
116. Keinänen M, Julkunen-Tiitto R, Mutikainen P, Walls M, Ovaska J, Vapaavuori E. Trade-offs in phenolic metabolism of silver birch: effects of fertilization, defoliation, and genotype. *Ecology.* 1999;80:1970–86.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

