Target Journal: *Plant, Cell and Environment*

Title: Evolution of the biosynthesis of two hydroxyacetophenones in plants

Authors: Geneviève J. Parent[a,b,c*], Isabelle Giguère[a,b], Melissa Mageroy[d,e], Joerg Bohlmann[d], John J. MacKay[a,b,c]

[a]Centre d'étude de la forêt, Département des sciences du bois et de la forêt, Université Laval, Québec, Qc, Canada G1V 0A6

[b]Institut de biologie intégrative et des systèmes, Université Laval, Québec, Qc, Canada G1V 0A6

[c]Department of Plant Sciences, University of Oxford, Oxford, UK OX1 3RB

[d]Michael Smith Laboratories, University of British Columbia, Vancouver, BC, Canada V6T 1Z4

[e]Norwegian Institute for Bioeconomy Research, Ås, Norway, 1433

*Corresponding author: genevieve.parent@plants.ox.ac.uk

**Abstract**

Acetophenones are phenolic metabolites of plant species. A metabolic route for the biosynthesis and release of two defense-related hydroxyacetophenones in white spruce (*Picea glauca*) was recently proposed to involve three phases: (i) biosynthesis of the acetophenone aglycons catalyzed by a currently unknown set of enzymes, (ii) formation and accumulation of the corresponding glycosides catalyzed by a glucosyltransferase, and (iii) release of the aglycons catalyzed by a glucosylhydrolase (PgβGLU-1). We tested if this biosynthetic model is conserved across Pinaceae and land plant species. We assayed and surveyed the literature and sequence databases for possible patterns of the presence of the acetophenone aglycons piceol and pungenol and their glucosides, as well as sequences and expression of *Pgβglu-1* orthologues. In the Pinaceae, the three phases of the biosynthetic model are present and differences in expression of *Pgβglu-1* gene orthologues explain some of the interspecific variation in hydroxyacetophenones. The phylogenetic signal in the metabolite phenotypes was low across species of six plant divisions. Putative orthologues of PgβGLU-1 do not form a monophyletic group in species producing hydroxyacetophenones. The biosynthetic model for acetophenones appears to be conserved across Pinaceae, while convergent evolution has led to the production of acetophenone glucosides across land plants.

**Keyword index:** biosynthesis, evolution, gene expression, enzyme, convergence, phenolics, secondary metabolism.

## Introduction

Acetophenones are a group of phenolic metabolites (Fig. 1a) present in many gymnosperms including white spruce (*Picea glauca*) (Delvas et al. 2011), Scots pine (*Pinus sylvestris*) (Härtling & Schulz 1998), and various angiosperms (Jeon et al. 2008, Hacibekiroglu & Kolak 2011, Wang et al. 1999). In shoots of white spruce, acetophenones are the most abundant soluble phenolics (70%) and their glucosides are the main constituents (90%) of the glucosidic fraction (Kraus & Spiteller 1997). Acetophenones have been described in a context of plant defense in several plants species. For example, the 3-hydroxyacetophenone, piceol, (Fig. 1b) appears to contribute fungitoxic properties in Norway spruce (*P. abies*) (Osswal et al. 1989, Boufalis & Pellissier 1987) and the glucoside, picein, is associated with a lack of mycorrhizas in fine roots of European larch (*Larix decidua*) (Fig. 2) (Münzenberger et al. 1995). The constitutive accumulation of piceol and pungenol in white spruce foliage (Fig. 2) underlies defense and resistance against spruce budworm (*Choristoneura fumiferana*) (Strunz et al. 1986; Delvas et al. 2011; Parent et al. 2017). Levels of phenolic glucosides dominated by picein reduce vole feeding in the bark of willow (*Salix myrsinifolia*) (Heiska et al. 2007). In carnation (*Dianthus caryophyllus*), piceol is associated with complete resistance against the fungal parasite, *Fusarium oxysporum* f. sp. *dianthi* (Curir et al. 1996).

A biosynthetic pathway for acetophenones in *Nicotiana tabacum* has been suggested to use feruloyl-CoA as a precursor and involve an as-of-yet unknown set of enzymes to produce acetovanillone and acetosyringone (Fig. 2, Negrel & Javelle 2010). The formation of acetophenone glucosides may allow the accumulation of these metabolites without the putative cytotoxic effects of the aglycons (Hoque et al. 1995). For example, the acetophenone glucosides picein and pungenin (Fig. 2) accumulate in the foliage of white spruce (Mageroy et al. 2015, Parent et al. 2017). The formation of pungenin from pungenol in white spruce is catalyzed by a

64 UDP-sugar dependent glucosyltransferase encoded by the *PgUGT5b* gene (Mageroy et al.
65 2017b). Release of the aglycons piceol and pungenol from their corresponding glucosides results
66 from activity of the β-glucosidase PgβGLU-1, encoded by the *Pgβglu-1* gene (Mageroy et al.
67 2015; 2017a). Variation of *Pgβglu-1* expression in different genotypes of white spruce, or over
68 the development of foliage, appears to effect the phenotypic variation of acetophenone aglycon
69 accumulation (Mageroy et al. 2015, Parent et al. 2017). Levels of piceol and pungenol in white
70 spruce trees are under genetic control and high levels are linked to greater survival in natural
71 population (Parent et al. 2017). Orthologues of *Pgβglu-1* may play a similar role in other
72 members of the Pinaceae.

73 Based on the prior work in tobacco (Negrel & Javelle 2010) and white spruce (Mageroy et
74 al. 2015, 2017a; 2017b) three phases has been proposed for the biosynthesis of
75 hydroxyacetophenones: (i) biosynthesis of the acetophenone aglycons, (ii) formation and
76 accumulation of the corresponding glycosides catalyzed by glucosyltransferase activity, and (iii)
77 release of the aglycons catalyzed by glucosylhydrolase activity (Fig. 2).

78 Here, we tested if the biosynthetic model is conserved in the Pinaceae and a larger group
79 of taxonomically more diverse plant species. We analyzed the occurrence of the
80 hydroxyacetophenones piceol and pungenol and their corresponding glucosides, as well as the
81 sequences and expression of *Pgβglu-1* gene orthologues in different members of the Pinaceae.
82 We also compiled data on the occurrence of these acetophenones in species that represent six
83 different plant divisions and assessed the phylogenetic relationships between protein orthologues
84 of PgβGLU-1 in plants that produce both the aglycon and glycosylated acetophenones.

85 **Materials and methods**

86 *Literature survey of plant acetophenones*

We searched the literature for reports on plant species producing either piceol, pungenol, picein, and pungenin with these names, their IUPAC names and some of their synonyms (Table S1). We used only studies that used mass spectrometry (MS) or nuclear magnetic resonance (NMR) to confirm identity of acetophenones unless stated otherwise, and studies measuring molecular mass to ensure a reliable identification of compounds across the survey. We cited data from one study that reported the presence of at least one of the four surveyed hydroxyacetophenones per species; therefore, we did not form an exhaustive list of studies on acetophenones in each species. We then searched in each order (i.e. taxonomic rank) for at least three studies of phenolic compounds in three different species to test the effectiveness of identifying species with compounds of interest. No other species containing one of the four acetophenones were identified in this validation step of the literature survey. Names were corrected based on accepted identification (theplantlist.org) for all names of species listed in Table S2. We also used species habits to qualitatively review if these acetophenones were specific to any particular plant features in general. Species habits were mostly obtained from tropicos.org. For species without species habit information in tropicos.org, we classified as climber any plant that needed support from another structure to grow away from the ground. Herbs are defined as plants without any abundant lignified structures lower than 2 m. Shrubs are defined as low height bushes with abundant lignin. Trees are defined as higher than 4 m with a stem diameter greater than 10 cm.

In silico *identification and phylogeny of* Pgβglu-1*orthologues*

Putative orthologues of the *P. glauca* gene *Pgβglu-1* (Mageroy et al. 2015) were identified by means of sequence similarity searches (min. 42%) by using BLAST analysis (Altschul et al. 1990). The protein sequences of the putative orthologues were obtained from the National Center for Biotechnology Information (NCBI) (blast.ncbi.nlm.nih.gov) and Congenie (congenie.org).

Amino acid sequences were aligned using MUSCLE with default settings implemented in MEGA v7 (Kumar et al. 2016). Alignments were inspected to identify artefacts in annotation or alignments and to manually remove amino acid positions not shared by all of the species. Single-gene trees were built with RAxML using the best evolutionary model (LG) inferred in MEGA v7.

*Plant material and sampling*

Foliage of *P. glauca* (N = 30), *P. mariana* (N = 10), *P. rubens* (N = 10), and *Abies balsamea* (N = 8) was sampled from May to September (except *A. balsamea* until August) in 2014 (see Table S3 for dates, Table S4 for origins for all species, except *P. glauca* see Parent et al. 2017). Samples of *P. glauca*, *P. mariana*, and *P. rubens* were taken from mature trees located in a common garden established in 1999, 1972, and 1959, respectively, in Valcartier, Quebec, Canada (46°56'N, 71°29'W). A natural population of *A. balsamea* young trees (1.15 to 8.75 cm in diameters, 4.33 ± 0.78 cm) was sampled on the Université Laval campus (46.78$^{o}$N, 71.14$^{o}$W). *P. abies* trees (established in 1969, N = 29) were sampled in Valcartier once on September 12$^{th}$ 2014, (see Table S4 for origins). Samples were comprised of current-year foliage from the north side of the midcrown. In addition for the analysis of interspecific variation of acetophenones, foliage of mature spruce trees (*P. asperata*, *P. breweria*, *P. crassifolia*, *P. jezoensis*, *P. koraiensis*, *P. pungens*, *P. sitchesis*, *P. smithiana*, *P. wilsonii*) were sampled in spring from the UBC Botanical Garden (see Table S5 for species and sampling period). All samples were frozen in liquid nitrogen immediately after removal from the trees and stored at -80°C to avoid RNA degradation and putative hydrolysis of glycosylated hydroxyacetophenones. Foliage was ground to a fine powder using a MixerMill 300 (Retsch) and stored at -80°C until further analyses.

*Acetophenone extraction and metabolite analysis*

Samples of 50–100 mg of fine needle powder were extracted with HPLC grade aqueous methanol as described in Parent et al. (2017) using 70% methanol a set of samples that was analysed for temporal variation and 100% methanol for a set of samples that was analysed for interspecific variation. Hydrolysis of glycosylated hydroxyacetophenone do not occur using this method or prolonged extraction time. Benzoic acid at 1 mg $ml^{-1}$ was used as an internal standard. Metabolite analyses were performed on a liquid chromatography (LC) (Agilent 1200 series) coupled to a MS detector (Agilent 6210 TOF). Acetophenones were separated on a 4.6 mm guard column (Polaris MetaGuard) and a column Polaris 250 x 4.6 mm C18-A (Agilent Technologies Inc.). The solvent and solvent gradient were as described in Mageroy et al. (2015) with a solvent flow rate of 1.5 ml $min^{-1}$. Five microlitres of extract were injected. Quantification was done using calibration curves for piceol (Sigma-Aldrich), pungenol (TCI America), and picein (Toronto Research Chemicals Inc.) using authentic standards. A pungenin standard was not available.

*DNA extraction,* Pgβglu-1 *sequence amplification and analyses*

Genomic DNA was extracted using DNeasy Plant Mini Kit (Qiagen). Primers matching the 5' and 3' ends of the full length (FL) DNA sequence *Pgβglu-1* were used to amplify orthologues in *P. mariana*, *P. rubens*, *P. abies*, and *A. balsamea* by PCR using Platinum Taq DNA Polymerase High Fidelity (LifeTechnologies, lifetechnologies.com). The PCR program was: 5 min at 94°C followed by 35 cycles of 15 sec at 94°C and 1 min at 62°C and a final elongation step of 1 min at 68°C. Amplicons were purified with ExoSAP-IT® for PCR Product Cleanup (Affymetrix, affymetrix.com) and then analysed by using niched Sanger sequencing (using primers pairing at interval of ca. 300 bp. See table S4 of Mageroy et al. 2015 for primer pairs). The sequence assembly was performed using the CHROMASPRO 1.5 software (technelysium.com.au). The amplicons yielded poor quality results for *P. rubens*. Therefore we isolated FL DNA sequence

using a TOPO TA Cloning® Kit for sequencing (invitrogen.com) (details given in Mageroy et al. 2015) and then analysed the DNA by using niched Sanger sequencing. Sequences were aligned and analysed (for pairwise distance using default parameters) with MEGAv7.

*RNA extraction and analysis of transcripts abundance*

Total RNA was extracted as in Chang et al. (1993) with modifications as in Pavy et. al. (2008) and stored at -80°C. The total RNA concentration was determined using a NanoDrop 1000 (Thermo Scientific) and assessed for quality with an Agilent 2100 Bioanalyzer and RNA 6000 Nano Kit LabChips (Agilent Technologies Inc.). Reverse transcriptase-qPCR with gene-specific primers was used to quantify *Pgβglu-1* transcripts as described in Mageroy et al. (2015). cDNA synthesis used 500 ng of total RNA and the Superscript First-Strand cDNA synthesis system for RT-PCR (Invitrogen). PCR reactions were performed using the QuantiFast SYBR Green PCR kit (Qiagen) as follows: 1x master mix, 300 nM of 50 and 30 primers and 5 µl of cDNA in a final volume of 15 µl. PCR reactions were carried out in a LightCycler 480 (Roche) as described in Boyle et al. (2009). The linear regression of efficiency (LRE) method (Rutledge & Stewart 2008) was used to calculate the number of transcript molecules. Transcripts quantified with two different pairs of primers ($N = 30$, $r = 0.97$, $P < 0.0001$, Fig. S1) were highly correlated for *P. abies* as for *P. glauca* (Mageroy et al. 2015). Similar results are expected in other species assayed for gene expression as sequences are highly similar across the Pinaceae (see results).

*Statistical analyses*

Correlations, t-tests, and analyses of variance (ANOVAs) were conducted with SAS 9.4 (SAS Institute Inc).

## Results

### Presence of acetophenones in species of the Pinaceae

The literature survey and our own analyses for the four acetophenones (piceol, picein, pungenol, and pungenin) (Table S5) covered 26 different Pinaceae species in four genera, of which 12 (46%) contained at least one of the four acetophenones (Fig. 3). Ten of these 12 species accumulated a glycosylated acetophenone (Fig. 3), four of which (three *Picea* and one *Pinus* species) accumulated both a glycosylated acetophenone and the corresponding aglycon (Fig. 3). *P. abies* and *P. glauca*, which represent two different *Picea* clades, contain both picein and pungenin. *P. glauca* was the only species of our survey to accumulate both of the aglycons piceol and pungenol in its foliage (Fig. 3). More than half of the *Picea* species accumulated glycosylated acetophenones in their foliage. Out of the eight *Pinus* species surveyed, low levels of picein were observed in the foliage of mature individuals of *P. sylvestris* (Fig. 3). In *Larix decidua*, picein accumulated in the roots (Table S5), but not in the foliage (Niemann & Bass 1978). In general, glycosylated hydroxyacetophenones were present in higher amounts than aglycons in the tissues that have been surveyed. In two of the *Picea* species, foliar concentrations of picein and piceol measured in mature individuals sampled during the summer were of 0.63 and 0.13% of dry weight (average) in *P. glauca* (Parent et al. 2017) and 0.33 and 0.13% in *P. abies*, respectively. In *P. sylvestris*, the concentrations were 0.1% for picein and 0.03% for piceol (Härtling & Schulz 1998).

### Levels of *βglu-1* transcripts in species of the Pinaceae.

We identified and sequenced the putative *βglu-1* orthologues in three *Picea* species (*P. abies, P. glauca, P. mariana*) that accumulate aglycon acetophenones and in two species that do not (*P.*

*rubens, A. balsamea*). The gene is highly conserved in the five species tested with a maximum of 2 % of sequence divergence observed between the five species over the whole DNA sequence (2952 bp) including 13 exons (1508 bp) and 12 introns (1374 bp) (Fig. S2, GenBank accession numbers MG049733, MG049734, MG049735, MG049736, MG049737). The *P. rubens* gene had an additional insertion of 299 bp in the 8$^{th}$ intron (Fig. S2).

Analysis of transcript levels of *βglu-1* orthologues showed that temporal variation over the course of a growing season was not conserved across the five species tested (*P. abies, P. glauca, P. mariana, P. rubens, A. balsamea*) (Fig. 4, Table S4). As described previously (Parent et al. 2017), *Pgβglu-1* transcripts levels increased sharply in June in *P. glauca* and remain stable until the end of the sampling in September (Fig. 4a). The transcript levels were also low in *P. mariana* at the beginning of the growing season (June) and increased up to 32-fold by September (Fig. 4b). Transcript levels in *P. rubens* and *A. balsamea* were consistently low through the sampling period (Fig. 4c,d). Interspecific differences in transcript levels at the end of the growth season (i.e. September 12$^{th}$) were observed for *Picea* species located in the common garden (ANOVA type IV; $df = 3$, $F = 12.8$, $P < 0.0001$; multiple comparisons adjusted with Tukey $\alpha = 0.05$). Transcript levels were higher in *P. glauca* ($3.8 \pm 0.2$ $\log_{10}$ ng$^{-1}$ RNA) than in *P. rubens* ($1.7 \pm 0.1$ $\log_{10}$ ng$^{-1}$ RNA, $P < 0.0001$) and in *P. mariana* ($2.8 \pm 0.4$ $\log_{10}$ ng$^{-1}$ RNA, $P = 0.05$). *P. abies* transcript levels ($3.5 \pm 0.2$ $\log_{10}$ ng$^{-1}$ RNA) were higher than in *P. rubens* ($P < 0.0001$).

Significant correlations exist between the levels of transcripts of *βglu-1* orthologues and the levels of piceol or pungenol in *P. glauca* and piceol in *P. abies* (Table 1, Fig. S3). In contrast, pungenin is present in *P. abies*, but lack of detectable levels of pungenol (Table 1). In *P. mariana*, transcript levels of *Pmβglu-1* do not correlate with levels of pungenol (Table 1, Fig. S3).

**Presence of acetophenones in plant species of other taxonomic groups**

We surveyed literature data of 238 plant species which have been analyzed for phenolic compounds to obtain information on the presence of piceol, pungenol, picein, and pungenin, covering 103 families across 55 orders and six divisions of plants (Fig. 5, Table S2). The habits of the species surveyed include climbers, herbs, shrubs, and trees (Table S6). In addition to species of the Coniferophyta division, 61 species across 18 orders in the Bryophyta, Monilophyta, and Magnoliophyta divisions accumulate at least one of the four acetophenones surveyed (Fig. 5a, Table S2). In general, accumulation the four acetophenones appears to be sporadic across land plants and not linked to a specific habit (Fig. 5b, Tables S2, S6).

Ancestral lineages of plants appear to accumulate hydroxyacetophenones only in the aglycon form, whereas glycosylated hydroxyacetophenones appear in plants that evolved later (Fig. 5a). Piceol is the only hydroxyacetophenone observed in Bryophyta and Monilophyta. It is reported to accumulate at low levels (i.e. 0.006%) in the leaves of *Sphagnum papillosum* (Mellegård et al. 2009). In Coniferophyta, only the Pinaceae and the Podocarpaceae, both from the Pinales order, contain glycosylated hydroxyacetophenones (Fig. 5b, Table S2). In the Magnoliophyta, species from nine different orders (of 48 orders surveyed) are reported to accumulate picein or pungenin (Fig. 5b, Table S2). Orders in the Coniferophyta and Magnoliophyta that accumulate the picein or pungenin do not form a monophyletic group (Fig. 5b).

A co-occurrence of both aglycons and the glycosides of hydroxyacetophenones (Fig. 2) is observed in species of the Pinales, as well as in three species of Magnoliophyta (Fig. 5b, Table S2). In monocots, piceol is reported in four species and pungenin is reported in one species (Table S2). In eudicots, species of 14 orders contain at least one of the four acetophenones. Pungenin is reported only in the Lamiales, whereas picein or piceol is present in species of 13

orders (Fig. 5b). Three species outside of the Pinaceae, namely *Salix hultenii* (Malpighiales) (Jeon et al. 2008), *Fabiana imbricata* (Solonales) (Quispe et al. 2012) and *Rhodiola litwinowii* (Saxifragales) (Krasnov et al. 1975) are reported to contain both glycosylated and corresponding aglycon acetophenones (Table S2). Note that orders of Rosales, Asterales and Lamiales contain some species that accumulated either picein or piceol.

**Phylogenetic relationship of proteins encoded by putative orthologues of *Pgβglu-1***

We investigated the phylogenetic relatedness of predicted sequences that may be orthologues of the gene encoding PgβGLU-1 in *P. glauca* across taxonomic groups. We hypothesized that species in the orders that contain both glycosylated acetophenones and the corresponding aglycon may contain a glucosylhydrolase of common ancestry that function in the release of the acetophenone aglycons. Alternatively, such hydrolases may have evolved convergently.

Using BLAST homology searches for putative orthologues of PgβGLU-1 we identified 23 sequences in the Magnoliophyta across 14 species in the Malpighiales, Solanales, and Rosales (Table S6). Full-length amino acid sequences were retrieved and had at least 94% coverage and 51% identity score when aligned with PgβGLU-1. In cases where a species contained multiple highly similar sequences, one representative sequence was selected randomly to be included in the phylogeny (Fig. 5c), while all sequences are presented in Table S7. Sequences obtained by PCR in the Pinaceae (N = 5, this study) or identified in the Congenie database (congenie.org) were also included in the phylogeny (N = 4) along with sequences from *Arabidopsis thaliana* (N = 6, Brassicales), *Amborella trichopoda* (N = 1, Amborellales), and *Selaginella moelledorffii* (N = 1, Lycophyta) to identify a putative ancestral sequence (min. coverage 94%, min. identity 42%).

The phylogeny shows three groups of sequences similar to PgβGLU-1 (Fig. 5c). One group is common to all of the Coniferophyta whereas the two other groups are found among the Magnoliophyta. Sequences in the Coniferophyta appear to have common ancestry with an *Arabidopsis thaliana* sequence, AtβGLU40. Across the Coniferophyta and Magnoliophyta, intraspecific diversification appears to be common for these sequences (Fig. 5c, Table S7). *Arabidopsis thaliana* has four closely related sequences with 3 to 11% amino acid substitutions among them. For *P. abies* and *Prunus persica*, two and four intraspecific sequence variants were identified with 3 to 21% and 1 to 44% of amino acid substitutions, respectively (Table S7).

## Discussion

The hydroxyacetophenones piceol, pungenol, picein, or pungenin are produced in a wide variety of plant species, although with a sporadic pattern of distribution across plant taxa. We used metabolite occurrence data, as well as gene sequences and expression data of *Pgβglu-1* to investigate how widely the biosynthesis and accumulation of hydroxyacetophenones as recently described in *P. glauca* (Fig. 2; Mageroy et al. 2015, 2017a) may be conserved in other plants. The complete biosynthesis model appears to be conserved across the Pinaceae, but the formation and accumulation of glucosides seems to have evolved independently in the Coniferophyta and Magnoliophyta divisions.

### Broadly conserved biosynthetic model across the Pinaceae

Three observations support that the biosynthetic model of hydroxyacetophenones proposed for *P. glauca* (Fig. 2) is broadly conserved amongst the Pinaceae. First, the different species tested accumulated glycosylated forms of hydroxyacetophenones alone or in higher concentration than

aglycons. Accumulation of picein or pungenin was found in eight out of 15 *Picea* species, in one species of *Pinus* and in one species of *Larix*. In *P. glauca*, higher concentrations of picein than piceol were observed in most plants (Mageroy et al. 2015, Parent et al. 2017). These results indicate that accumulation in a glycosylated form of hydroxyacetophenone is a common feature in the Pinaceae, suggesting that phase I and II are present and conserved. Second, the general pattern of expression of *Pgβglu-1* orthologues is related to levels of aglycon hydroxyacetophenones in *Picea*. For instance, *P. abies* accumulated high levels of piceol and had high transcript levels of *Paβglu-1,* whereas *P. rubens* did not accumulate pungenol and had low transcript levels of *Prβglu-1*. These results suggest the active role of the βGLU-1 enzyme in the synthesis of aglycon from accumulated glycosylated hydroxyacetophenones in *Picea*. Third, although variation in gene expression was not tested across all of the genera in the Pinaceae, the highly conserved sequences (2% divergence) of *Pgβglu-1* orthologues in distantly related taxa such as *Picea* and *Abies* suggest that phase III first appeared in their common ancestor and was conserved across the family. Such highly conserved sequences are on average unusual between plant species. *Picea glauca* sequence divergence was reported as 3.0% on average with the congeneric *P. abies* and as 17.8% on average with *Pinus taeda* (Warren et al. 2015). In angiosperms, coding sequences that share a high degree of identity may diverge by 10-30% (Ober & Kaltenegger 2009). Well-conserved gene sequence may also be observed for enzymes involved in secondary metabolism such as cinnamyl alcohol dehydrogenase, which is 94% similar for amino acid sequences between *Pinus taeda* and *Picea abies* (MacKay et al. 1995). In our study, high sequence similarity across exons and introns between orthologs suggests that the whole sequence is under strong purifying selection, which may indicate a crucial role in survival for this defense mechanism. *βglu-1* gene sequences are also likely to be orthologues that have the same

function across the Pinaceae species. These observations support that the three phases of the model for hydroxyacetophenone production in *P. glauca* (Fig. 2) are conserved across the Pinaceae.

The role of hydroxyacetophenones in defense may influence the conservation of its biosynthetic pathway across the Pinaceae. Hydroxyacetophenones can reduce insect herbivory or fungal growth in *Picea* and *Larix* (Strunz et al. 1986; Münzenberger et al.1990; Münzenberger et al.1995; Delvas et al. 2011; Parent et al. 2017) and have been reported to have allelopathic effects (Ruan et al. 2011; Fernandez et al. 2016). In *P. glauca*, the foliar accumulation of piceol and pungenol from late June is synchronized with the development of the most damaging stage of spruce budworm and is linked with resistance (Parent et al. 2017). High levels of piceol and pungenol are detected in *P. glauca* from areas where historical spruce budworm damage was the greatest in their eastern distribution (Parent et al. 2017). Such pattern in distribution is likely caused by a greater survival of trees with high levels of hydroxyacetophenones to spruce budworm attack. Interestingly, the species *P. mariana* and *P. rubens* as well as their hybrids are defoliated by spruce budworm with contrasted damage levels (Manley & Fowler 1969) that may be related to differences in their hydroxyacetophenone accumulation phenotypes (resistance *P. mariana* > hybrids > *P. rubens*). The synchronized development of spruce budworm larvae and the accumulation of pungenol in *P. mariana* foliage also indicate that pungenol may affect the insect. *P. glauca* and *P. mariana* have distributions that overlap widely with spruce budworm distribution, unlike *P. rubens* which is more southerly distributed. In these two species, hydroxyacetophenones accumulation may be conserved and necessary to survive to spruce budworm attack. It is worth mentioning that *A. balsamea* sustains the highest levels of spruce budworm damage (Nealis & Régnière 2004) and contained no hydroxyacetophenones in the present study. Hydroxyacetophenones have also been linked with cold tolerance in boreal areas.

In *P. glauca* and in *P. abies*, concentrations of hydroxyacetophenones have been observed to increase with latitude (Rummukainen et al. 2007; Parent et al. 2017). Increased resistance to spruce budworm and to cold suggest that the biosynthesis of hydroxyacetophenones may increase the fitness of the Pinaceae producing them.

High levels of hydroxyacetophenones may also decrease fitness due to allelopathy. The presence pungenol in the needle leachate decreased seed germination for *P. schrenkiana* (Ruan et al. 2011) suggesting that high levels of hydroxyacetophenones have negative effects on fitness. The greater fitness of trees with high levels of hydroxyacetophenones may therefore only be observed in areas where spruce budworm attacks may be frequent. Elsewhere, trees with low hydroxyacetophenones may have greater fitness due the greater survival of their offspring. Contrasting fitness under different environmental conditions may explain the wide variety of hydroxyacetophenones levels observed in the eastern range of *P. glauca* (Mageroy et al. 2015). In addition, contrasted fitness of trees with high or low levels of hydroxyacetophenones in different environment may avoid the fixation of one phenotype within the tree population. Insects or pathogens are therefore confront to diversity of phenotypes that avoids the fixation of a counter response to the tree defense. Such hypotheses needs to be tested.

Our study shows that variations in gene expression of *Pgβglu-1* orthologues do not explain the extent phenotypic variability in aglycon hydroxyacetophenone content in closely related taxa of *Picea*. In *P. glauca*, levels of both aglycon acetophenones and *Pgβglu-1* transcripts are highly correlated (Mageroy et al. 2015; Parent et al. 2017). In contrast, *P. abies* foliage contains pungenin but no pungenol although high *Paβglu-1* transcript levels were observed. Transcript levels of the *Pmβglu-1* were also not correlated with levels of pungenol in *P. mariana*. The lack of correlation between levels of pungenol and transcripts of *Pgβglu-1*orthologues suggests that a different beta-glucosidase is required to hydrolyse the glucose

moiety of pungenin in *P. abies* and *P. mariana*. For *P. abies*, it can also suggest that picein and pungenin are not both colocalized with the PgβGLU-1 orthologue, and thus may accumulate in different cell types or cellular compartments. It may also suggest that pungenol is rapidly being converted and prevented from accumulation at detectable levels in these two species.

Variation in gene expression has been proposed as the mechanism to explain the sporadic presence of compounds within plant clades (Wink 2003, Agrawal 2006). In contrast, convergent evolution explains most of the sporadic occurrence of the same secondary metabolites in distantly related clades (reviewed in Wink 2003; Agrawal 2007; Pichercky & Lewinsohn 2011). While convergent evolution of secondary metabolism has been shown in several systems, the role of variation in gene expression as a factor causing sporadic patterns of secondary metabolites is less well documented. Our results show that the relationship between the occurrence of secondary metabolites and gene expression is not consistent within a clade as predicted.

High levels of intraspecific diversification in *Pgβglu-1* orthologues are present across the land plants (Table S6). In *Prunus serotina*, one the putative orthologue has been linked to different functions such as amygdalin hydrolase, which is also associated with defence against herbivores (Zheng & Poulton 1995). The presence of numerous slightly different forms of *Pgβglu-1* in *Picea* species supports the notion that a different enzyme catalyses phase III for pungenin in *P. abies* and *P. mariana*. Following sequence duplications within each of the species, these conserved *βglu-1* sequences may have undergone subfunctionalization to hydrolyse glucose from pungenin, but this hypothesis needs to be tested.

Our experimental test within the Pinaceae also showed that variation in acetophenone content was large at the temporal scale in the Pinaceae species. The temporal variation in acetophenones could explain inconsistent patterns of aglycon occurrence across *Picea* due to sampling times for some species. For instance, the only three species that we identified with a

combination of glycosylated and aglycon forms were those sampled over the complete growth season (Table S2). In contrast, the four species with only glycosylated acetophenones were sampled early in the growth season (Table S2). *Picea schrenkiana* accumulated only the acetophenone aglycon and was sampled at the end of growth season (Table S2). In this report and in a previous study (Parent et al. 2017), there is clear temporal variation in acetophenone content in foliage in *Picea*. Here, we show that although they are rarely presented, detailed accounts of temporal variation can be highly informative to study the evolution of metabolic pathways.

**Conservation and divergence in the biosynthetic model of hydroxyacetophenones across plant divisions**

The accumulation of piceol is reported in the Bryophyta, Coniferophyta, and Magnoliophyta. The most parsimonious interpretation of the pattern of occurrence of piceol across plant phylogeny is that the phase I of the biosynthetic model (Fig. 2) is the same as assumed for other compounds in land plants (e.g. Qualley et al. 2012), although the enzymes involved in the synthesis of the skeleton of piceol are unknown. The glycosylation (phase II, Fig. 2) is likely not required or may be optional for the accumulation of piceol in the early land plant divisions of the Bryophyta and Monilophyta, based on the detection of low levels of piceol but not picein (Rasmussen et al. 1995, Mellegård et al. 2009). Similarly, a major lineage of the Magnoliophyta, the monocots, mainly accumulate piceol and not picein.

The phase II involving the addition of a glucose moiety diverges across plant divisions. The accumulation of picein or pungenin in Pinales and other lineages of Magnoliophyta has evolved convergently and thus results from the evolution multiple glycosyltransferases. The broad patterns of occurrence of picein are consistent with the evolution of two different glycosyltransferases for their synthesis in Pinales and in the common ancestor of superosids and

superasterasids. For pungenin, three different glycosyltransferases may have evolved independently in the Pinales, Poales, and Lamiales. Furthemore, the biosynthesis of glycosylated acetophenones as proposed in the *P. glauca* model involves two different glycosyltransferases (i.e. one for picein and one for pungenin) that may be conserved across Pinales, and this is supported by the recent identification of a glycosyltransferase (PgUGT5b, Fig. 2) that specifically acts on pungenol for the formation of pungenin (Mageroy et al. 2017a).

The deglycosylation phase III does not appear to be present in land plants other than Pinaceae. The phylogeny reported here does not support a monophyletic origin for the βGLU-1 enzyme for the release of aglycons among the Coniferophyta and Magnoliophyta. These observations indicate the terminal phase in the biosynthesis model proposed in *P. glauca* (Fig. 2) may apply only to the Pinaceae and not to the superosids and superasterids. Thus, convergent evolution most likely explains the simultaneous occurrence of piceol and picein in both the Coniferophyta and the Magnoliophyta. Convergence is frequent in the evolution of secondary metabolites (Wink 2003; Pichersky & Lewinsohn 2011).

We have interpreted reports of acetophenone content with caution based on a well-established phylogeny and this allowed us to estimate the minimum number of times the traits evolved in plants. However, we do not have a complete picture of the evolution of these traits in plants because our survey of the presence of specific acetophenones is broadly based but includes on a small fraction of all plant taxa. Here, trait occurrence in only one species within an order was not interpreted as convergent evolution since data in sister taxa are required to draw such a conclusion.

## Conclusion

Research on acetophenones is in its infancy unlike research on other phenolic compounds such as flavonoids and monolignols. Our study provides a first broad overview of the occurrence of some acetophenones and the evolution of their production in land plants. Although the four hydroxyacetophenones surveyed here are found in a wide variety of plants, the model first described for their production in *P. glauca* appears to be conserved only in the Pinales. In other land plants such as the Bryophyta, Monilophyta and Magnoliophyta, their production may result from a conserved phase I but glycosylation and deglycolysation phases II and III are not conserved (Fig. 2).

**Acknowledgements**

This work was possible thanks to the collaboration of the Valcartier Forest Research station of Natural Resources Canada. The authors thank Jean Bousquet, Jean Beaulieu, Martin Perron, Gaby Germanos, Cyril Vanghelder, Dapeng Wang, Gabriel Piette-Lauzière, Jeanne Piette, David Vigneault, Karen Guay, Claudia Mendez-Espinoza, and Kimmy Beaupré-Boivin for valuable discussions, for field, laboratory, or bioinformatics assistance. GJP would like to thank the plant taxonomy team at Kew Garden and NSERC (UK) for their fellowship, which greatly enhanced her understanding of plant diversity and evolution. Funding was received from the Natural Resources and Engineering Council of Canada as Discovery and Strategic Grants to JJM and JB, and Fonds de recherche du Québec - Nature et technologies to GJP.

461 **References.**


462 Agrawal A.A. (2007) Macroevolution of plant defense strategies. *Trends in Ecology and*
463 *Evolution* **22**, 103–109.

464 Altschul S.F., Gish W., Miller W., Myers E.W. & Lipman D.J. (1990) Basic local alignment
465 search tool. *Journal of Molecular Biology* **215**, 403–410.

466 Boufalis A. & Pellissier F. (1987) Allelopathic effects of phenolic mixtures on respiration of two
467 spruce mycorrhizal fungi. *Journal of Chemical Ecology* **20**, 2283–2289.

468 Boyle B., Dallaire N. & MacKay J. (2009) Evaluation of the impact of single nucleotide
469 polymorphisms and primer mismatches on quantitative PCR. *BMC Biotechnology* **9**, 75.

470 Chang S., Puryear J. & Cairney J. (1993) A simple and efficient method for isolating RNA from
471 pine trees. *Plant Molecular Biology Reporter* **11**, 113–116.

472 Chase M.W., Christenhusz M.J.M., Fay M.F., Byng J.W., Judd W.S., Soltis D.E., … Weber A.
473 (2016) An update of the Angiosperm Phylogeny Group classification for the orders and
474 families of flowering plants: APG IV. *Botanical Journal of the Linnean Society* **181**, 1–20.

475 Curir P., Marchesini A., Danielit~ B. & Mariani F. (1996) 3-hydroxyacetophenone in carnations
476 is a phytoanticipin active against *Fusarium oxysporum* f. sp. *dianthi*. *Ecological*
477 *Biochemistry* **41**, 447–450.

478 Delvas N., Bauce É., Labbé C., Ollevier T. & Bélanger R. (2011) Phenolic compounds that
479 confer resistance to spruce budworm. *Entomologia Experimentalis et Applicata* **141**, 35–44.

Fernandez C., Monnier Y., Santonja M., Gallet C., Weston L.A., Prévosto B., … Bousquet-Mélou A. (2016) The impact of competition and allelopathy on the trade-off between plant defense and growth in two contrasting tree species. *Frontiers in Plant Science* **7**, 594.

Gernandt D.S., Geada López G., Ortiz García S. & Liston A. (2005) Phylogeny and classification of *Pinus*. *Taxon* **54**, 29–42.

Hacibekiroglu I. & Kolak U. (2011) Antioxidant and anticholinesterase constituents from the petroleum ether and chloroform extracts of *Iris suaveolens*. *Phytotherapy Research* **25**, 522–529.

Härtling S. & Schulz H. (1998) Biochemical parameters as biomarkers for the early recognition of environmental pollution in Scots pine trees. I. Phenolic compounds. *Verlag der Zeitschrift für Naturforschung* **53**, 331–340.

Heiska S., Tikkanen O.P., Rousi M. & Julkunen-Tiitto R. (2007) Bark salicylates and condensed tannins reduce vole browsing amongst cultivated dark-leaved willows (*Salix myrsinifolia*). *Chemoecology* **17**, 245–253.

Hoque, E. (1985) Norway spruce dieback: occurrence, isolation and biological activity of p-hydroxy acetophenone and p-hydroxy acetophenone-O-glucoside and their possible roles during stress phenomena. *Forest Pathoogy* **15,** 129–145.

Jeon S.H., Chun W., Choi Y.J. & Kwon Y.S. (2008) Cytotoxic constituents from the bark of *Salix hulteni*. *Archives of Pharmacal Research* **31**, 978–982.

499 Krasnov E.A., Demidenko L.A. & Zeman L.P. (1975) p-hydroxyacetophenone and picein from
500 *Rhodiola litvinovii*. *Chemistry of Natural Compounds* **9**, 388–389.

501 Kraus C. & Spiteller G. (1997) Comparison of phenolic compounds from galls and shoots of
502 *Picea glauca*. *Phytochemistry* **44**, 59–67.

503 Kumar S., Stecher S. & Tamura K. (2016) MEGA7: Molecular evolutionary genetics analysis
504 version 7.0. *Molecular Biology and Evolution* **33**, 1870-1974.

505 Lockwood J.D., Aleksić J.M., Zou J., Wang J., Liu J. & Renner S.S. (2013) A new phylogeny for
506 the genus *Picea* from plastid, mitochondrial, and nuclear sequences. *Molecular*
507 *Phylogenetics and Evolution* **69**, 717–727.

508 MacKay, J. J., Liu, W., Whetten, R., Sederoff, R. R. & O'Malley, D. M. (1995) Genetic analysis
509 of cinnamyl alcohol dehydrogenase in loblolly pine: single gene inheritance, molecular
510 characterization and evolution. *Molecular and General Genetics* **247**, 537–545.

511 Mageroy M.H., Parent G., Germanos G., Giguère I., Delvas N., Maaroufi H., … Mackay J.J.
512 (2015) Expression of the β-glucosidase gene *Pgβglu-1* underpins natural resistance of white
513 spruce against spruce budworm. *Plant Journal* **81**, 68–80.

514 Mageroy M.H., Jancsik S., Yuen M.M., Fischer M., Withers S.G., Paetz C., … Bohlmann J.
515 (2017a) A conifer UDP-sugar dependent glycosyltransferase contributes to acetophenone
516 metabolism and defense against insects. *Plant Physiology* **Published Online Aug 9**. DOI:
517 https://doi.org/10.1104/pp.17.00611

518 Mageroy M. H., Lachance D., Jancsik S., Parent G., Seguin A., Mackay J., Bohlmann, J.
519 (2017b). In vivo function of *Pgβglu-1* in the release of acetophenones in white spruce.
520 *PeerJ* **5**, e3535-e3535.

521 Manley S.A.M. & Fowler D.P. (1969) Spruce budworm defoliation in relation to introgression in
522 red and black spruce. *Forest science* **15**, 365-366.

523 Mellegård H., Stalheim T., Hormazabal V., Granum P.E. & Hardy S.P. (2009) Antibacterial
524 activity of sphagnum acid and other phenolic compounds found in *Sphagnum papillosum*
525 against food-borne bacteria. *Letters in Applied Microbiology* **49**, 85–90.

526 Münzenberger B., Kottke I. & Oberwinkler F. (1995) Reduction of phenolics in mycorrhizas of
527 *Larix decidua* Mill. *Tree Physiology* **15**, 191–196.

528 Münzenberger B., Heilemann J., Strack D., Kottke I. & Oberwinkler F. (1990) Phenolics of
529 mycorrhizas and non-mycorrhizal roots of Norway spruce. *Planta* **182**, 142–148.

530 Nealis V.G. & Régnière J. (2004) Insect-host relationships influencing disturbance by the spruce
531 budworm in a boreal mixedwood forest. *Canadian Journal of Forest Research* **34**, 1870–
532 1882.

533 Negrel J. & Javelle F. (2010) The biosynthesis of acetovanillone in tobacco cell-suspension
534 cultures. *Phytochemistry* **71**, 751–759.

535 Neimann G.J. & Baas W.J. (1978) Phenolics from *Larix* needles XIV. Flavonoids and phenolic
536 glucosides and ester of *L. decidua*. *Zeitschrift für Naturforschung* **33**, 780–782.

537 Ober D. & Kaltenegger E. (2009) Pyrrolizidine alkaloid biosynthesis, evolution of a pathway in
538 plant secondary metabolism. *Phytochemistry* **70**, 1687–1695.

539 Osswald W.F. & Benz B. (1989) p-hydroxyacetophenone and picein contents of healthy and
540 damaged spruce needles from different locations in Bavaria. *Forest Pathology* **19**, 323–334.

541 Parent G.J., Giguère I., Germanos G., Lamara M., Bauce É. & MacKay J.J. (2017) Insect
542 herbivory (*Choristoneura fumiferana*, Tortricidea) underlies tree population structure (*Picea*
543 *glauca*, Pinaceae). *Scientific Reports* **199**, 379–398.

544 Pavy N., Boyle B., Nelson C., Paule C., Giguère I., Caron S., … Mackay J. (2008) Identification
545 of conserved core xylem gene sets: conifer cDNA microarray development, transcript
546 profiling and computational analyses. *New Phytologist* **180**, 766–786.

547 Pichersky E. & Lewinsohn E. (2011) Convergent evolution in plant specialized metabolism.
548 *Annual Review of Plant Biology* **62**, 549–566.

549 Qualley, a. V., Widhalm, J. R., Adebesin, F., Kish, C. M. & Dudareva, N. (2012) Completion of
550 the core β-oxidative pathway of benzoic acid biosynthesis in plants. *Proceedings of the*
551 *National Academy of Sciences of the United States of America* **109**, 16383–16388.

552 Quispe C., Viveros-Valdez E. & Schmeda-Hirschmann G. (2012) Phenolic constituents of the
553 Chilean herbal tea *Fabiana imbricata* R. et P. *Plant Foods for Human Nutrition* **67**, 242–
554 246.

555 Rasmussen S., Wolff C. & Rudolph H. (1995) Compartmentalization of phenolic constituents in
556 sphagnum. *Phytochemistry* **38**, 35–39.

557 Ruan X., Li Z.-H., Wang Q., Pan C.-D., Jiang D.-A. & Wang G.G. (2011) Autotoxicity and
558 allelopathy of 3,4-dihydroxyacetophenone isolated from *Picea schrenkiana* needles.
559 *Molecules* **16**, 8874–8893.

560 Ruhfel B.R., Gitzendanner M.A., Soltis P.S., Soltis D.E., Burleigh J., Govaerts R., … Stamatakis
561 A. (2014) From algae to angiosperms–inferring the phylogeny of green plants
562 (Viridiplantae) from 360 plastid genomes. *BMC Evolutionary Biology* **14**, 23.

563 Rummukainen A., Julkunen-Tiitto R., Räisänen M. & Lehto T. (2007) Phenolic compounds in
564 Norway spruce as affected by boron nutrition at the end of the growing season. *Plant and*
565 *Soil* **292**, 13–23.

566 Rutledge R.G. & Stewart D. (2008) A kinetic-based sigmoidal model for the polymerase chain
567 reaction and its application to high-capacity absolute quantitative real-time PCR. *BMC*
568 *Biotechnology* **8**, 47.

569 Strunz G.M., Giguère P. & Thomas A.W. (1986) Synthesis of pungenin, a foliar constituent of
570 some spruce species, and investigation of its efficacy as a feeding deterrent for spruce
571 budworm [*Choristoneura fumiferana* (Clem.)]. *Journal of Chemical Ecology* **12**, 251–260.

572 Väisänen E.E., Smeds A.I., Fagerstedt K. V., Teeri T.H., Willför S.M. & Kärkönen A. (2015)
573 Coniferyl alcohol hinders the growth of tobacco BY-2 cells and *Nicotiana benthamiana*
574 seedlings. *Planta* **242**, 747–760.

575 Wang M., Kikuzaki H., Lin C.C., Kahyaoglu A., Huang M.T., Nakatani N. & Ho C.T. (1999)
576 Acetophenone glycosides from thyme (*Thymus vulgaris* L.). *Journal of Agricultural and*
577 *Food Chemistry* **47**, 1911–1914.

578 Warren R.L., Keeling C.I., Yuen M.M., Raymond A., Taylor G.A., Vandervalk B.P., …
579 Bohlmann J. (2015) Improved white spruce (*Picea glauca*) genome assemblies and
580 annotation of large gene families of conifer terpenoid and phenolic defense metabolism.
581 *Plant Journal* **83**, 189–212.

582 Wink M. (2003) Evolution of secondary metabolites from an ecological and molecular
583 phylogenetic perspective. *Phytochemistry* **64**, 3–19.

584 Xiang Q.-P., Xiang Q.-Y., Guo Y.-Y. & Zhang X.-C. (2009) Phylogeny of *Abies* (Pinaceae)
585 inferred from nrITS sequence data. *Taxon* **58**, 141–152.

586 Zheng L. & Poulton J.E. (1995) Temporal and spatial expression of amygdalin hydrolase and
587 (R)-(+)-mandelonitrile lyase in black cherry seeds. *Plant Physiology* **109**, 31–39.

588

589

590 **Tables**

591

592 Table 1. Correlation between levels of hydroxyacetophenones and transcripts of *βglu-1*

593 orthologues. The number of measurements during the temporal series (or punctual sampling for

594 *P. abies*), correlation coefficient, and P-value are indicated by N, r, and P, respectively.

| Species | Trait 1 | Trait 2 | N | r | P |
|---|---|---|---|---|---|
| *P. glauca* | pungenol | exp | 267 | 0.33 | <0.0001 |
| *P. glauca* | piceol | exp | 267 | 0.3 | <0.0001 |
| *P. glauca* | picein | exp | 265 | 0.05 | 0.39 |
| *P. glauca* | picein | piceol | 463 | 0.33 | <0.0001 |
| *P. glauca* | picein | pungenol | 463 | 0.082 | 0.08 |
| *P. glauca* | piceol | pungenol | 463 | 0.75 | <0.0001 |
| *P. mariana* | pungenol | exp | 49 | 0.2 | 0.18 |
| *P. abies* | piceol | picein | 28 | -0.06 | 0.77 |
| *P. abies* | piceol | exp | 28 | 0.45 | 0.014 |
| *P. abies* | picein | exp | 28 | -0.26 | 0.19 |

595

596

597

## Figure legends

**Fig. 1** a) acetophenone, b) 3-hydroxyacetophenone

**Fig. 2** Proposed acetophenone biosynthesis routes in tobacco (*Nicotiana tabacum*) and white spruce (*Picea glauca*). In tobacco, the synthesis of acetovanillone has been suggested to involve four enzymatic reactions (phase I) for formation of the acetophenone skeleton (Negrel & Javelle 2010; Mageroy et al. 2017a) and the production androsin (glycosylated acetovanillone) has suggested the presence of phase II (Väisänen et al. 2015) but none of the enzyme are known. In white spruce, a model biosynthetic route was proposed based on the tobacco phase I, the hydroxyacetophenone accumulation patterns in spruce, the characterizations of a UDP-sugar glucosyltransferase (Mageroy et al. 2017a) and a glucosidase enzyme PgβGLU-1 (Mageroy et al. 2015). Picein and pungenin accumulate to high levels in spruce (phase II) but the corresponding aglycons are only detected when *Pgβglu-1* is expressed, suggesting that phase III is required for aglycon accumulation (Mageroy et al. 2015). PgβGLU-1 was shown to be active on both picein and pungenin. Additionally, *Pgβglu-1* expression was well correlated with the accumulation of both of the aglycon acetophenones in *P. glauca* (Mageroy et al. 2015, Parent et al. 2017).

**Fig. 3** Interspecific variation in hydroxyacetophenone content in Pinaceae. The phylogeny presented on the left side is based on Lockwood et al. 2013 (*Picea* and overall phylogeny), Gernandt et al. 2005 (*Pinus*), and Xiang et al. 2009 (*Abies*); the three shades of grey in the background indicate the three clades of *Picea* (Lockwood et al. 2013). Filled and empty squares indicate presence and presumed absence, respectively, of the compounds in tissue of at least one individual. All of the species were analyzed for their foliage, except *Larix decidua* (roots). Details on source (literature or experimental) are given in Table S2 and methods. The four

hydroxyacetophenones surveyed were: picein (PiGl), piceol (Pi), pungenin (PuGl), and pungenol (Pu).

**Fig. 4** Temporal variation in transcripts abundance of *Pgβglu-1* orthologues in five Pinaceae species. Foliage was sampled in 2014 in Valcartier research station for *Picea glauca, P. mariana, P. rubens, P. abies* and in 25 km south-east site for *Abies balsamea* (Table S3, see methods for details). The first data points report transcript levels in previous year foliage (formed in 2013) whereas all the other dates report that of current year (2014) foliage. We tested for temporal variation with an ANOVA type IV for *P. glauca* (white circles, N = 30, data reformatted from Parent et al. 2017; $df = 11$ $F = 28.2$ $P < 0.0001$), *P. mariana* (black circles, N = 10; $df = 5$ $F = 5.9$ $P = 0.0002$), *P. rubens* (red circles, N = 10; $df = 5$ $F = 2.13$ $P = 0.08$), *P. abies* (blue circle, N = 30), and *Abies balsamea* (yellow circles, N = 10; $df = 5$ $F = 1.9$ $P = 0.10$). Different letters indicate difference of the variable across the time interval (multiple comparisons adjusted with Tukey $\alpha = 0.05$). Mean and SEM are presented.

**Fig. 5** Hydroxyacetophenones content and phylogeny of PgβGLU-1 putative orthologues in plant orders. Presence of picein (PiGl), piceol (Pi), pungenin (PuGl), and pungenol (Pu) are reported in (a) an overview of the six investigated plant divisions and (b) a detailed summary for the Coniferophyta and Magnoliophyta orders. For panels (a) and (b), the phylogeny presented on the left side is based on Chase et al. 2016 and Ruhfel et al. 2014. Empty and filled squares indicate presence or presumed absence of the compounds, respectively, in any tissue of at least one individual (for tissues , see Table S2). We considered studies using mass spectrometry (MS) or nuclear magnetic resonance (NMR) to confirm compound identity, except when otherwise

644 indicated Table S2. Species habits are summarised in Table S6. (c) Phylogeny of PgβGLU-1
645 putative orthologues in plants. The phylogenetic tree was produced by using a maximum
646 likelihood approach (RAxML) based on 31 sequences of 449 amino acids (Table S7). The
647 phylogeny only includes one putative orthologue per species of Magnoliophyta for clarity, but a
648 larger number of sequences found through blast analyses are indicated in squared brackets next to
649 the accession numbers in regular brackets (Fig. 5c, Table S7 for supplementary sequences
650 information). Protein functions are identified in purple font. βGLU: betaglucosidase, PsAH-1:
651 *Prunus serotina* amygdalin hydrolase-1. All bootstrap support values are indicated near the node.
652 The scale indicates rate of substitution per nucleotide.
653
654

Fig. 1

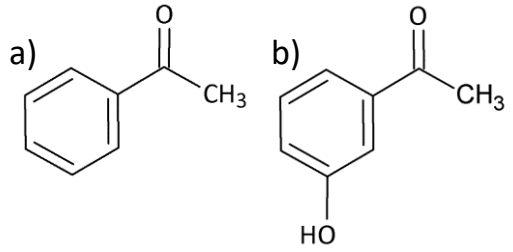


Fig. 2

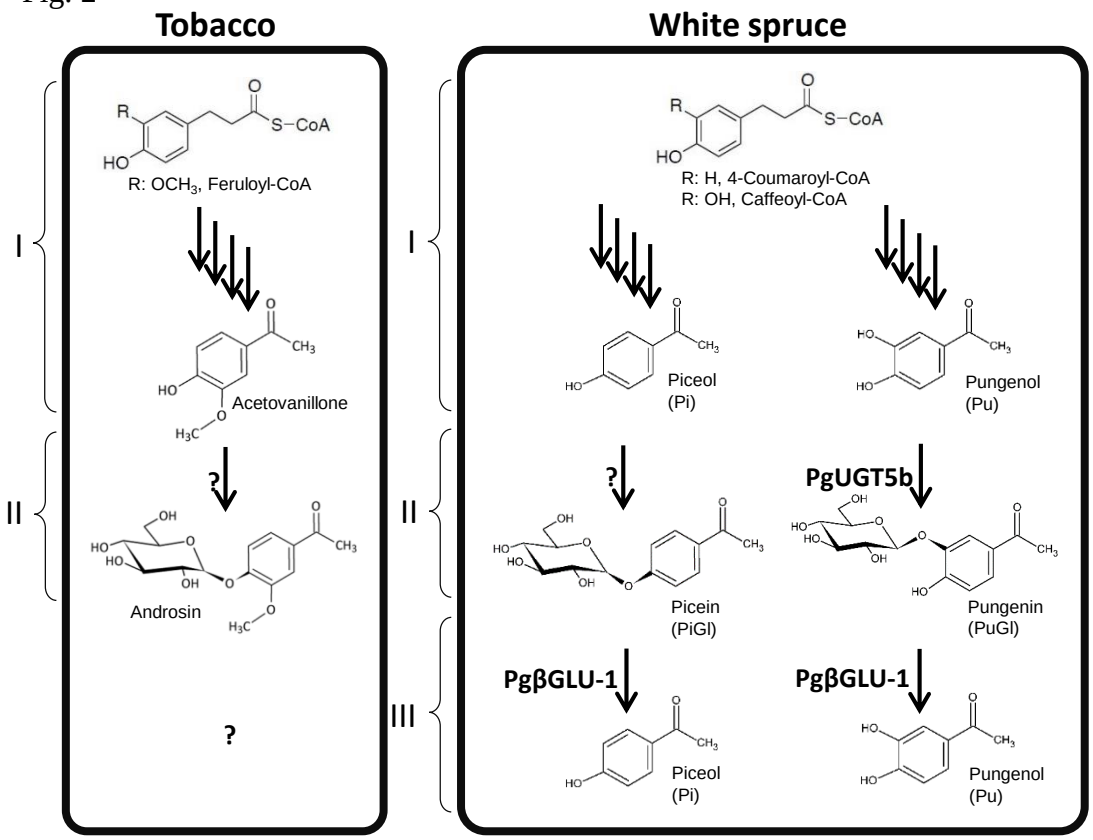

Tobacco
White spruce
R: OCH3, Feruloyl-CoA
Acetovanillone
Androsin
?
I
II
III
R: H, 4-Coumaroyl-CoA
R: OH, Caffeoyl-CoA
Piceol
(Pi)
Pungenol
(Pu)
Picein
(PiGl)
Pungenin
(PuGl)
PgUGT5b
PgβGLU-1
PgβGLU-1

Fig. 3

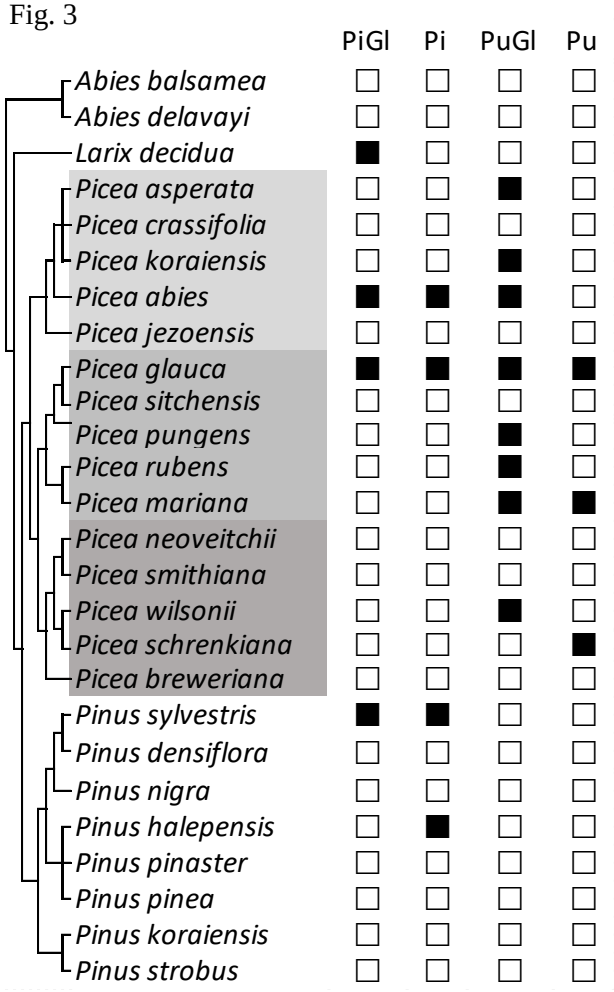

PiGl
Pi
PuGl
Pu
Abies balsamea
Abies delavayi
Larix decidua
Picea asperata
Picea crassifolia
Picea koraiensis
Picea abies
Picea jezoensis
Picea glauca
Picea sitchensis
Picea pungens
Picea rubens
Picea mariana
Picea neoveitchii
Picea smithiana
Picea wilsonii
Picea schrenkiana
Picea breweriana
Pinus sylvestris
Pinus densiflora
Pinus nigra
Pinus halepensis
Pinus pinaster
Pinus pinea
Pinus koraiensis
Pinus strobus

Fig.4

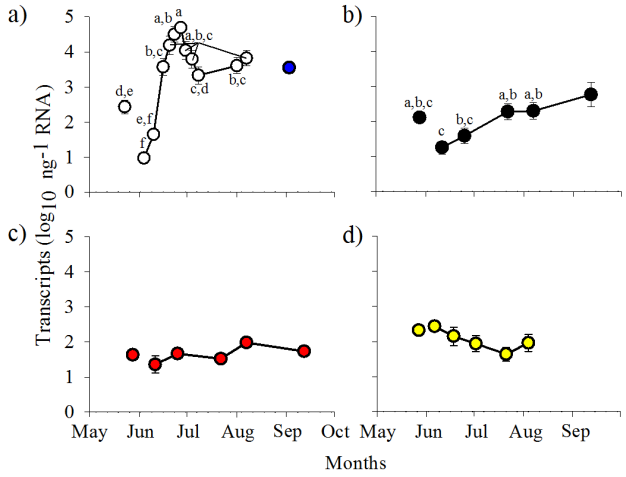
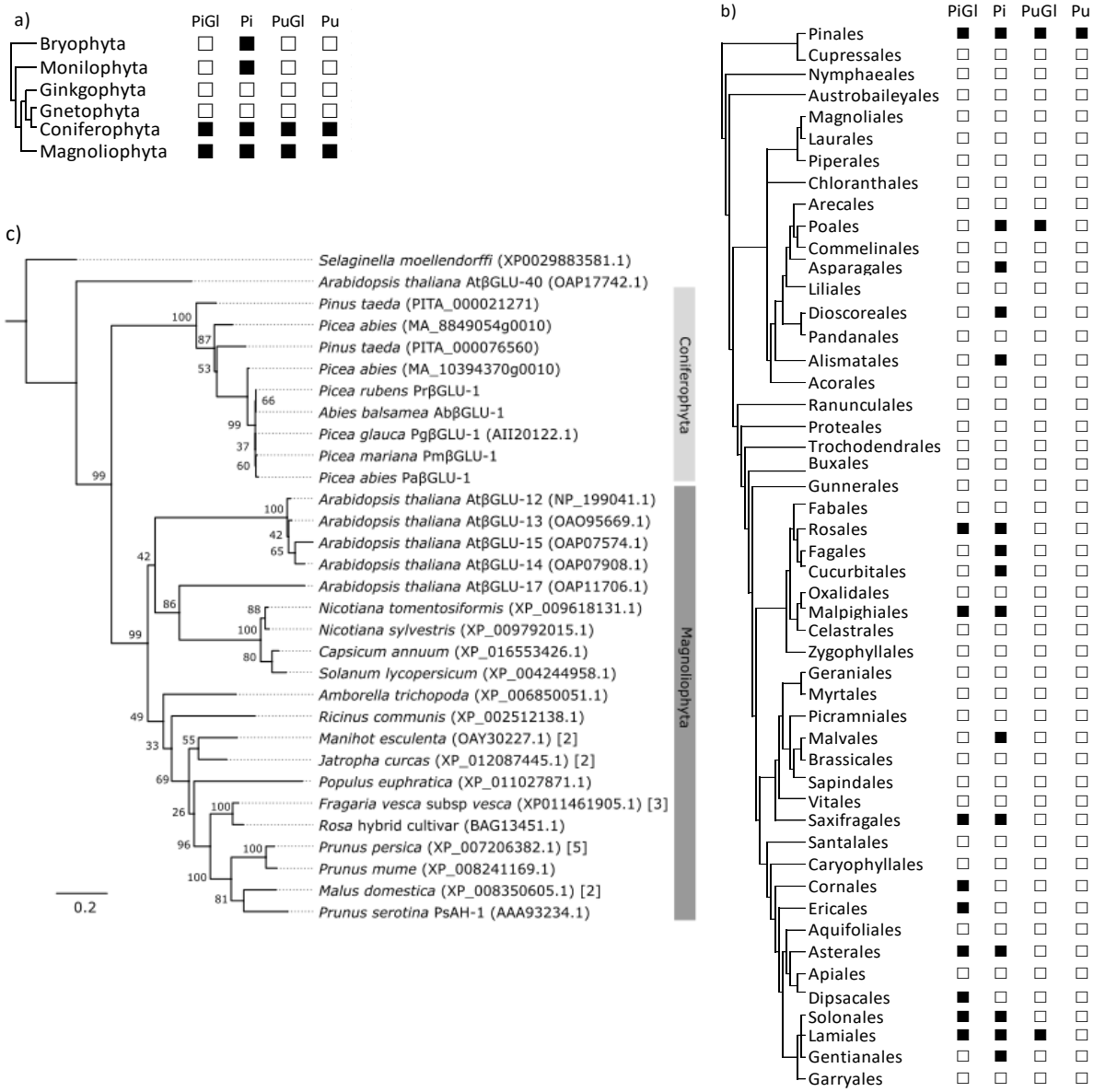

a)
b)
c)
d)
Transcripts (log10 ng-1 RNA)
Months
May
Jun
Jul
Aug
Sep
Oct
a,b
a
a,b,c
b,c
d,e
e,f
f
c,d
b,c
a,b,c
c
b,c
a,b
a,b

Fig. 5


a)
PiGl
Pi
PuGl
Pu
Bryophyta
Monilophyta
Ginkgophyta
Gnetophyta
Coniferophyta
Magnoliophyta
b)
PiGl
Pi
PuGl
Pu
Pinales
Cupressales
Nymphaeales
Austrobaileyales
Magnoliales
Laurales
Piperales
Chloranthales
Arecales
Poales
Commelinales
Asparagales
Liliales
Dioscoreales
Pandanales
Alismatales
Acorales
Ranunculales
Proteales
Trochodendrales
Buxales
Gunnerales
Fabales
Rosales
Fagales
Cucurbitales
Oxalidales
Malpighiales
Celastrales
Zygophyllales
Geraniales
Myrtales
Picramniales
Malvales
Brassicales
Sapindales
Vitales
Saxifragales
Santalales
Caryophyllales
Cornales
Ericales
Aquifoliales
Asterales
Apiales
Dipsacales
Solonales
Lamiales
Gentianales
Garryales
c)
Selaginella moellendorffi (XP0029883581.1)
Arabidopsis thaliana AtβGLU-40 (OAP17742.1)
Pinus taeda (PITA_000021271)
Picea abies (MA_8849054g0010)
Pinus taeda (PITA_000076560)
Picea abies (MA_10394370g0010)
Picea rubens PrβGLU-1
Abies balsamea AbβGLU-1
Picea glauca PgβGLU-1 (AII20122.1)
Picea mariana PmβGLU-1
Picea abies PaβGLU-1
Arabidopsis thaliana AtβGLU-12 (NP_199041.1)
Arabidopsis thaliana AtβGLU-13 (OAO95669.1)
Arabidopsis thaliana AtβGLU-15 (OAP07574.1)
Arabidopsis thaliana AtβGLU-14 (OAP07908.1)
Arabidopsis thaliana AtβGLU-17 (OAP11706.1)
Nicotiana tomentosiformis (XP_009618131.1)
Nicotiana sylvestris (XP_009792015.1)
Capsicum annuum (XP_016553426.1)
Solanum lycopersicum (XP_004244958.1)
Amborella trichopoda (XP_006850051.1)
Ricinus communis (XP_002512138.1)
Manihot esculenta (OAY30227.1) [2]
Jatropha curcas (XP_012087445.1) [2]
Populus euphratica (XP_011027871.1)
Fragaria vesca subsp vesca (XP011461905.1) [3]
Rosa hybrid cultivar (BAG13451.1)
Prunus persica (XP_007206382.1) [5]
Prunus mume (XP_008241169.1)
Malus domestica (XP_008350605.1) [2]
Prunus serotina PsAH-1 (AAA93234.1)
Coniferophyta
Magnoliophyta
0.2