NERC Open Research Archive
sharing our research

UK Centre for Ecology & Hydrology

Article (refereed) - postprint

This is the peer reviewed version of the following article:

Zaborowska, Julia; Perry, Annika; Cavers, Stephen; Wachowiak, Witold M. 2023. **Evolutionary targets of gene expression divergence in a complex of closely related pine species**. *Journal of Systematics and Evolution*, 61 (1) 198-212, which has been published in final form at https://doi.org/10.1111/jse.12896

This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.

© 2022 Institute of Botany, Chinese Academy of Sciences

This version is available at https://nora.nerc.ac.uk/id/eprint/534885

Copyright and other rights for material on this site are retained by the rights owners. Users should read the terms and conditions of use of this material at https://nora.nerc.ac.uk/policies.html#access.

This document is the authors' final manuscript version of the journal article, incorporating any revisions agreed during the peer review process. There may be differences between this and the publisher's version. You are advised to consult the publisher's version if you wish to cite from this article.

The definitive version is available at https://onlinelibrary.wiley.com/

Contact UKCEH NORA team at
noraceh@ceh.ac.uk

The NERC and UKCEH trademarks and logos ('the Trademarks') are registered trademarks of NERC and UKCEH in the UK and other countries, and may not be used without the prior written consent of the Trademark owner.

Accepted Article

Check for updates

Julia Zaborowska ORCID iD: 0000-0002-9225-9129

Witold Wachowiak ORCID iD: 0000-0003-2898-3523

Evolutionary targets of gene expression divergence in a complex of closely related pine species

Gene expression in closely related pines

Julia Zaborowska [1,*], Annika Perry[2], Stephen Cavers[2], Witold M. Wachowiak[1]

1. Institute of Environmental Biology, Adam Mickiewicz University, Uniwersytetu Poznanskiego 6,
2. 61-614 Poznan, Poland
3. UK Centre for Ecology and Hydrology - Edinburgh site, Bush Estate, Penicuik, Midlothian, EH26 0QB, United Kingdom

* Correspondence: Julia Zaborowska, Department of Plant Ecology and Environmental Protection, Institute of Environmental Biology, Adam Mickiewicz University, Uniwersytetu Poznanskiego 6, 61-614 Poznan, Poland e-mail: julia.zaborowska@amu.edu.pl

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/jse.12896.

This article is protected by copyright. All rights reserved.

Accepted Article

ABSTRACT

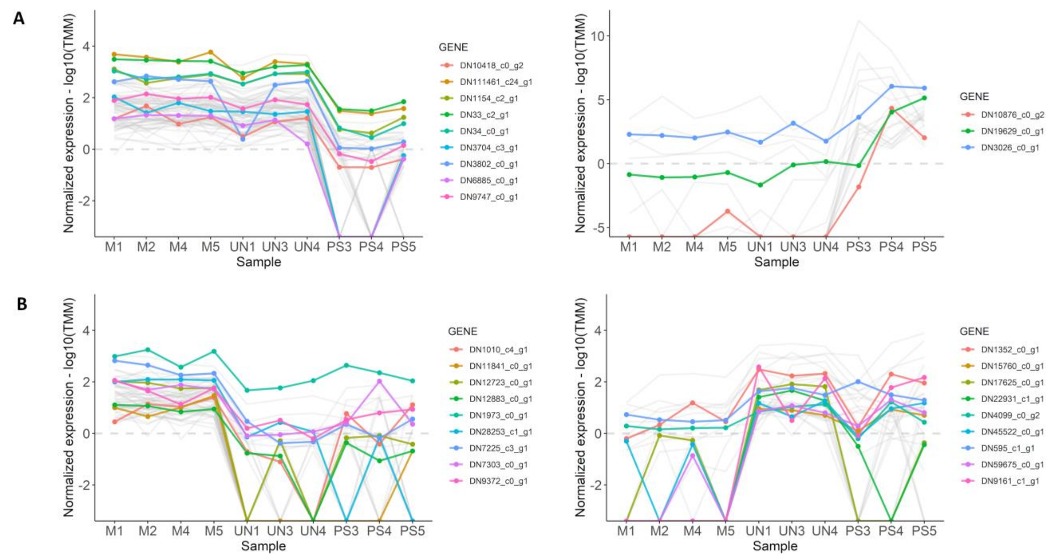
The environment is a powerful selective pressure for sessile organisms, such as plants, and adaptation to the environment is particularly important for long-lived species, like trees. Despite the importance of adaptive trait variation to the survival and success of trees, the molecular basis of adaptation is still poorly understood. Gene expression patterns in three closely related, but phenotypically and ecologically divergent, pine species were analyzed to detect differentiation that may be associated with their adaptation to distinct environments. Total RNA of *Pinus mugo*, *P. uncinata* and *P. sylvestris* samples grown under common garden conditions was used for *de novo* transcriptome assembly, providing a new reference dataset that includes species from the taxonomically challenging *Pinus mugo* complex. Gene expression profiles were found to be very similar with only 121 genes significantly diverged in any of the pairwise species comparisons. Functional annotation of these genes revealed major categories of distinctly expressed transcripts including: wood trait properties, oxidative stress response, and response to abiotic factors such as salinity, drought and temperature. We discuss putative associations between gene expression profiles and adaptation to different environments, for example: upregulation of genes involved in lignin biosynthesis in the species which have adapted to mountainous regions characterized by strong winds and thick snow cover. Our study provides valid candidates for verification of the importance of the gene expression role, in addition to evidence for selec-

This article is protected by copyright. All rights reserved.

tion within genomic regions, in the process of ecological divergence and adaptation to higher altitudes in pine taxa.

GRAPHICAL ABSTRACT

The expression profiles of annotated (in color) and not (in gray) gene models that discriminate three European pine taxa. The upper (A) panel shows markers significantly up-regulated in tests comparing two mountain pines (*Pinus mugo* and *P. uncinata*) with their close relative, Scots pine (*P. sylvestris*); the lower (B) panel presents markers with differential expression between the two mountain pines.



KEYWORDS

differential gene expression; high-altitude adaptations; mountain pines; *Pinus mugo* complex; transcriptome assembly

This article is protected by copyright. All rights reserved.

Accepted Article

Accepted Article

1 INTRODUCTION

Plant growth, fitness and survival is highly dependent on their surrounding environment. This reliance is highly pronounced in perennials such as temperate forest trees, as they need to cope with changing seasons and extreme events. Although these species are often characterized by wind pollination and high outcrossing rates, resulting in generally low neutral among-population variation and weak population structure at large geographical areas (Petit & Hampe, 2006), their quantitative traits usually express strong geographical and environmental patterns of differentiation (Aitken et al., 2008). Such intra- and interspecific variation has been quantitatively assessed for many traits related to temperature, photoperiod or water availability in numerous common garden and provenance trial experiments (Lascoux et al., 2016). This heritable differentiation is linked to the species adaptations resulting from natural selection driven by local environmental demands. At the genome level, over relatively short timescales, without time for new mutations to arise, adaptations of trees are based on the standing genetic variation of populations, operating mostly through changes in allele frequency spectra (Neale et al., 2017). However, genetic variation, even from extensive genome-wide studies, rarely seems sufficient to explain observed phenotypic diversity or species adaptations to different habitat gradients (Kremer, 2011, Plomion et al., 2016). Despite growing examples of nucleotide polymorphisms under selection in genomic regions of forest trees, less is known about variation in gene ex-

This article is protected by copyright. All rights reserved.

pression patterns that may play important role in development of plant adaptation. Studies of closely related but phenotypically and ecologically diverse species may be particularly useful to shed light on the evolution of such traits.

The three hard pines investigated here - *Pinus mugo* Turra, *P. uncinata* Ramond and *P. sylvestris* L. - are morphologically diverged and adapted to different habitats, enforced by disjunction of their geographical ranges and isolation during the Pleistocene glaciation (Christensen, 1987). *Pinus mugo*, the dwarf mountain pine, and *P. uncinata*, the Pyrenean pine, are sister taxa related to subalpine habitats of European mountains. The dwarf pine is represented by individuals of shrubby habit (multiple trunks, up to a few meters high), which inhabit central and eastern massifs – Alps, Sudetes, Carpathians and Balkan mountain chains (Hamernik & Musil, 2007). Pyrenean pine is a typical coniferous tree (grows straight and may reach up to 25 m), that occurs from the Alps westward – in the Massif Central, Jura and Vosges, Pyrenees, and a few remote populations inside Iberian Peninsula (Jalas & Suominen, 1973). Although they are distinguished by a number of other, more subtle traits, e.g., phenology, cone size and shape, needle characteristics, composition of volatiles and allozymes (please, see Table S1; Adams & Tashev, 2019; Boratynska & Boratynski, 2007; Boratynska et al., 2015; Lewandowski et al., 2000; Monteleone et al., 2006; Wachowiak et al., 2018), they are often aggregated under the *P. mugo* complex together with a few less defined groups (Christensen, 1987; Hamernik & Musil, 2007). For simplicity reasons, and

This article is protected by copyright. All rights reserved.

Accepted Article

following Businsky & Kirschner (2010), we call them here separate species. The two taxa are closely related to Scots pine (*P. sylvestris*), that is mostly monocormic and upright tree, up to 45 m high. Range of this pine spreads from the Mediterranean climate on Iberian Peninsula, through vast areas of Europe and Asia, up to the cold temperate conditions of the Siberian taiga, covering broad diversity of mainly lowland habitats. As a result of its large distribution, the species demonstrates high phenotypic differentiation with dozens of ecotypes described (www.theplantlist.org). Pines from the *P. mugo* complex and *P. sylvestris* diverged about 5 million years BP (Wachowiak et al., 2011, Labiszak & Wachowiak, 2021). However, they are genetically similar and previous studies showed little differentiation between the taxa at karyotype (Bogunic et al., 2011), mitochondrial and plastid genomes (Dzialuk et al., 2017; Heuertz et al., 2010; Sokolowska et al., 2020; Zaborowska et al., 2019), nuclear loci (Monteleone et al., 2006; Wachowiak et al., 2013; Zaborowska et al., 2021), and candidate genes that showed low (<0.1%) net genetic divergence between the taxa (Wachowiak et al., 2013). So far, only singular species diagnostic markers have been found at mitochondrial, plastid and nuclear regions (Kormutak et al., 2005; Wachowiak et al., 2000; Zaborowska et al., 2019; Zukowska & Wachowiak, 2017). This close genetic similarity between the species is usually explained by relatively recent speciation in presence of gene flow (Christensen, 1987; Jasinska et al., 2010; Monteleone et al., 2006) and segregation of ancestral variation (Wachowiak et al., 2011; Wachowiak et al., 2013). Comparative transcriptome analysis of these pines (Wachowiak et al.,

This article is protected by copyright. All rights reserved.

Accepted Article

2015) and revealed thousands of polymorphic markers that were used for development of SNPs genotyping array (Perry et al., 2021) which has since been used to identify loci significantly associated with key adaptive traits, including growth and phenology (Perry et al., 2022). Genomic studies support the phylogenetically close relationship between two mountain pine taxa as compared to Scots pine and reveal some candidate regions under selection during the species evolution (Zaborowska et al., 2021).

In the presented work we focused on the information hidden in the expression profiles of these pines to look at the genes and metabolic pathways that could influence their distinct phenotype and ecology. As the study builds on samples from a common garden experiment, it helped us exclude the impact of population-specific environmental determinants on gene transcription profiles of the samples. We were primarily interested in the relationships of the two mountain species (*P. mugo* and *P. uncinata*), as these share longer period of common history and higher proportion of common variation at the nucleotide sequence level but are highly phenotypically distinct. The Scots pine was included to provide a reference, and to evaluate the extent to which the variation in expression profiles reflects the phylogenetic relationships between the taxa. We looked specifically for transcriptome outliers potentially related to adverse environmental factors known to operate on higher altitudes, such as reduced atmospheric pressure, photo-oxidative stress, or more demanding substrates.

Accepted Article

This article is protected by copyright. All rights reserved.

Accepted Article

2 MATERIALS AND METHODS

2.1 Materials

Genetic material was obtained from needles of 11 two year-old seedlings of three pine species (*P. mugo* – 4 samples, M; *P. uncinata* – 3 samples, UN; *P. sylvestris* L. – 4 samples, PS) grown in a glasshouse facility of the UK Centre of Ecology and Hydrology in Edinburgh. The seedlings were derived from open-pollinated seeds collected in natural populations in Europe (Fig. 1, Table 1; for details, see Wachowiak et al., 2018). Raw paired-end Illumina reads from teams' previous RNA-sequencing attempts were used, data is deposited in the European Nucleotide Archive under accession number PRJEB6877 (for precise sample identifiers, please see Table 1). Details regarding RNA isolation, library preparation and sequencing are specified in Wachowiak et al. (2015).

2.2 Transcriptome assembly and expression analysis

The raw reads of all 11 samples were quality checked in FASTQC v0.11.9 (www.github.com/s-andrews/FastQC) and used for *de novo* transcriptome construction by TRINITY v2.11.0 (Haas et al., 2013). The paired-end assembly procedure adapted default parameter values, except it was run with simultaneous cleaning and trimming of reads provided by TRIMMOMATIC v0.39 (Bolger et al., 2014) and contigs below 200 bp were rejected. Afterwards reads were verified with FASTQC again.

This article is protected by copyright. All rights reserved.

Accepted Article

The resulting transcriptome, hereafter *MUS assembly*, was adopted as the reference in all subsequent examination. Sample PS2, used in earlier studies for Scots pine reference transcript construction and sequenced at much higher coverage than the remaining samples (Wachowiak et al., 2015), was excluded from further analyses.

For the quality assessment of the resulting *MUS assembly*, basic statistics from TRINITY run were investigated, and the transcripts were compared to the known 'core' gene sequences present in plants as deposited in the BUSCO v5.0.0 database (Manni et al., 2021), both, the viridiplante_odb10 and embryophyta_odb10 clusters were used.

Analyses of expression profiles were run at isoform and gene levels, using the raw *MUS assembly* transcripts and the quality-filtered reads. Three methods were exploited for read count estimations. First, two fast pseudo-aligner software, KALLISTO v0.46.2 (Bray et al., 2016) and SALMON v1.4.0 (Patro et al., 2017), were used adapting suggestions from TRINITY abundance estimation protocol (www.github.com/trinityrnaseq/trinityrnaseq/wiki). Default settings were run in SALMON, whereas KALLISTO was executed with additional 100 bootstrap samples. The third approach included the classical alignment-based method implemented in RSEM, it was run with scripts from the DETONATE v1.11 package (Li et al., 2014). The required read mapping was performed simultaneously using BOWTIE2 v2.3.4.3 (Langmead & Salzberg, 2012), default program options were used except the value of

This article is protected by copyright. All rights reserved.

Accepted Article

the parameter '–k' changed from 200 to 100 and the addition of '–phred33-quals' for proper read quality scoring. The expression estimations were normalized to FPKMs (fragments per kilobase transcript length per million fragments mapped), TPMs (transcripts per million), and to cross-sample normalized TMMs (trimmed M-means). The general ExN50 statistics were recorded. They are calculated like the standard N50 (or more broadly Nx, which indicate the length of contig for which the collection of all contigs of equal or longer length produces 50% of the total bases in the transcriptome), however these are limited to the x% of the total normalized expression data - the most highly expressed transcripts. They are therefore recommended as more reliable indicators of transcriptome quality as they consider the read support (www.github.com/trinityrnaseq/trinityrnaseq/wiki). Transcripts that did not pass further expression filtering (FPKM > 0.5 for average across samples) were discarded.

2.4 Differential expression of gene models

Differential expression analysis was conducted using two BIOCONDUCTOR (Gentleman et al., 2004) packages based on the negative binomial distribution model - EDGER v3.32.0 (Robinson et al., 2010) and DESEQ. 2 v1.30.0 (Love et al., 2014). The expression estimates were compared between the species at the gene level, and each individual sample was treated as biological replicate. We tested patterns of change in expression among four pairs of taxa: *P. mugo vs P. uncinata* (hereafter M *vs* UN), *P. mugo vs P. sylvestris* (M *vs* PS), *P. uncinata vs P. sylvestris* (UN *vs* PS)

This article is protected by copyright. All rights reserved.

Accepted Article

and *P. mugo* and *P. uncinata* (as sister, high-altitude taxa) that were jointly opposed to *P. sylvestris* (M-UN *vs* PS). Both packages are included in the TRINITY kit so were run using scripts provided therein. Default parameter values were adopted, i.e. minimum 4-fold change in gene expression and *p*-value cutoff for false discovery rate (FDR) set at 0.001 were thresholds for significant result. Outputs from different software were compared and only intersection of the sets was considered to contain differentially expressed sequences. For M-UN *vs* PS analyses we distinguished two sets of results further described as 'broad' and 'narrow'. These corresponded to direct outcome from M-UN *vs* PS test ('board'), and its subset significant in 3 tests - confirmed by M *vs* PS and UN *vs* PS comparisons ('narrow'). Overlaps between marker sets were identified and depicted with Venn diagrams produced with the VENNDIAGRAM v1.6.20 package (Chen & Boutros, 2011).

2.5 Functional annotation of transcriptome

For recognition of the gene specific functions, we chose the ENTAP v0.10.8-beta program (Hart et al., 2020) designed for improved accuracy and speed in non-model organisms. Particular steps of transcriptome annotation were run with the following software, data and specific parameters: 1) Expression analysis and filtering step were omitted, previous RSEM results were taken and only isoforms of genes passing the FPKM > 0.5 cutoff were utilized; 2) Identification of protein coding regions was conducted with TRANSDECODER v5.5.0

This article is protected by copyright. All rights reserved.

(www.github.com/TransDecoder/TransDecoder/wiki) using the following arguments: 'runP=true', 'transdecoder-m=100', 'complete=false' and 'transdecod-er-no-refine-starts=false'; 3) For similarity search and identification of contaminants we used DIAMOND v2.0.11 (Buchfink et al., 2015), two UniProt KB databases - manually annotated Swiss-Prot and computationally analyzed TrEMBL (release 2021_03 of 02-Jun-2021; The UniProt Consortium, 2021), and the collection of plant proteins stored in NCBI RefSeq database (release207 of 12-Jul-2021; O'Leary et al., 2016). Thresholds of 50 bp for minimum query and target coverages along with *e*-value cutoff of 10^{-5} for hits were set. The Pinidae lineage (NCBI txid3313) was preferred to be the reported hit ('taxon=pinidae'), also two lists of keywords were used for filtration of extraneous or uncertain subjects (including 'con-tam=bacteria,opisthokonta' and 'uninforma-tive=conserved,predicted,unknown,unnamed,hypothetical,putative,unidentified,uncharacterized,uncultured,uninformative'); 4) Sequence functional analysis, the assign-ments of protein domains and gene ontology (GO) terms were performed with the EGGNOG-MAPPER v2.1.5 and eggNOG database version v5.0 (Cantalapiedra et al., 2021; Huerta-Cepas et al., 2019). Single isoform was chosen to represent predicted gene and the longest contig was chosen if there was more than one annotated. In case of the differentially expressed markers, we took the most reliably annotated sequence that was most often the longest one. Additionally, the GO term enrichment analysis was performed on the terms assigned to the differentially expressed markers. The

This article is protected by copyright. All rights reserved.

Accepted Article

Accepted Article

script run_GOseq.pl from TRINITY package was employed for that purpose and results were further reduced and visualized by REVIGO (accessed 18 May 2022; Supek et al., 2011).

3 RESULTS

3.1 Characteristic and quality assessment of de novo transcriptome assembly

Out of 259,291,524 input read pairs that passed quality filtering and were trimmed (Table S2), we built the new transcriptome *MUS assembly*. In total, 371,779 unique contigs for 241,804 putative gene models were generated, summing up to about 266.5 Mb total length and 42.17% of GC nucleotide pairs (Table S3). The newly generated transcriptome was characterized by a mean contig length of 716.77 bp. The N50 value based on all transcripts was 1290 bp, while N50 based on the longest isoforms per gene model was 674 bp, indicating that first value might be exaggerated due to generation of surplus isoforms during assembly, especially in longer transcripts. The new *MUS assembly* (submitted to the Dryad database under accession number XXX), which was compared to the expected plant genes of two BUSCO collections, represented nearly complete sets of those sequences. In case of the 425 'core' genes of Viridiplante only 0.1% were missing, while of the complete sequences 37.6% were single-copy and 61.6% were duplicated BUSCOs. Representation of Embryophyta genes was slightly lower - 4.2% of 1614 were missing, however there were fewer duplicated

This article is protected by copyright. All rights reserved.

Accepted Article

sequences among complete BUSCOs – 39.2% compared to 54.7% single-copy representations. The observed significant proportion of duplicated sequences is expected in non-filtered and non-clustered transcriptomes (Madritsch et al., 2021).

The analysis of isoform expression performed using three methods yielded quite consistent results. According to RSEM read abundance estimates, the *MUS assembly* E50N50 (N50 for the top 50% of genes with the highest expression), calculated here on 1,939 gene models, was 1,577 bp, while E90N50, measured on 17,830 gene models, was 2,173 bp. Three outputs of the transcript expression estimation methods showed similar ExN50 maximum, however those peaks were reached at slightly different Ex values (Fig. S1). 56,077 gene models met the minimal mRNA volume criteria for further analyses. For details on the success of mapping reads to the reference *MUS assembly* that was performed to enable RSEM estimations (alignment-based method), please see Table S2.

3.2 Functional annotation of the transcriptome

After filtering to remove markers with low expression, the number of isoforms dropped from 371,779 to 159,919 (Table S3). In about half of these sequences - 73,339 isoforms (23,815 gene models) an open reading frame (ORF) could be predicted (Table S4). They represented mainly complete genes (44,874 isoforms), but partial sequences were recognized too (5’-fragments: 11,112, 3’-fragments: 7,979,

This article is protected by copyright. All rights reserved.

internal sequences: 9,374). All the isoforms with at least partial ORF detected were subjected to similarity searches through local alignments to three protein databases. Of the 77,339 isoforms checked, 41,300 were successfully aligned to plant reference sequences in RefSeq, 32,664 to Swiss-Prot records, and 40,825 to sequences deposited in TrEMBL. In total 59,369 contigs, corresponding to 19,731 gene models, had at minimum one significant hit in any of the collections (Table S5). Depending on the database screened (and its specificity), diverse plants dominated the significant best hits: RefSeq – *Amborella trichopoda* (13.6%), Swiss-Prot – *Arabidopsis thaliana* (61.3%), TrEMBL – *Picea sitchensis* (32.3%); the last species was also the top one in global consideration (29.3%). Discarded contaminant sequences, including those from two most commonly found taxa - *Orchasella cincta* and *Photinus pyralis* (107 and 46 hits, respectively) - summed up to 1.4% of all significant alignments (details in Table S6). In the BUSCO analysis performed at this stage - on the set of gene models with detected ORFs but free of contaminants - 96.5% of the Viridiplantae and 89.4% of Embryophyta near universal single-copy orthologues were covered (Fig. 2; the longest contig per gene model was included). About one quarter (25.9%) of the aligned sequences were further rejected if their DIAMOND tags were uninformative (e.g., ‘predicted’, ‘hypothetical’). Considering only queries with informative, non-contaminant best-hit alignments, total number of 43,573 isoforms, corresponding to 15,084 coding gene models (hereafter named genes), was left for final annotation steps in EGGNOG.

This article is protected by copyright. All rights reserved.

Accepted Article

Accepted Article

For 14,666 genes at least one isoform was successfully annotated, providing function and/or structure information based on assigned GO term, PFAM or SMART domain content, or other data from EGGNOG mapping. Screening PFAM collection resulted in total of 27,523 recognized (3214 unique) domains in 14,022 sequences, while identification of compositionally biased structure with SMART was possible for 10,768 sequences in which 20,631 domains (607 unique) were found (Table S7). Together, 13,293 genes were annotated with at least one GO tag of all 803,761 recorded (considering terms of level 1 or higher, i.e. more specific). In that set 9,816 were unique, inclusive of 6,163 in the Biological Process aspect, 2,662 Molecular Functions and 991 Cellular Components. Level 4 GO terms, that might be considered as general function description, are listed in Fig. 3 (details in Table S8). In the category of Biological Processes metabolic processes prevailed, particularly macromolecule biosynthesis and modifications - protein modifications, phosphorylation and regulation of gene expression. Amongst the most frequently annotated Molecular Functions we found activities such as binding of nucleic acid, nucleotides or nucleosides, ion binding, also hydrolase and kinase activities. Plastids and nucleus, and more generally cytoplasm, intracellular membrane-bounded organelles and integral components of membrane, were the most common Cellular Components.

This article is protected by copyright. All rights reserved.

Accepted Article

3.3 Differentially expressed genes and their ontology

Identification of distinguished patterns was simultaneously run on all gene models that passed expression filtering; it was conducted in six combinations: two differential expression analysis tools on results from three transcript estimation software. A gene model was considered differentially expressed if it differed significantly in results of all these combinations, in total 393 adequate markers were recorded from all taxa pairs analyzed. There were 86 markers pointed in M-UN *vs* PS test: 68 up regulated in first plus 18 in latter group; 90 in M vs UN: 56 + 34; 308 in M *vs* PS: 246 + 62; and 74 found in UN *vs* PS: 64 + 10 (Figs. 4 and 5). Nearly half of the gene models discriminating both mountain pines from Scots pine were confirmed in the 'narrow' subset (Fig. 6). Open reading frames were found in 121 sequences (M-UN *vs* PS: 26; M *vs* UN: 28, M *vs* PS: 90; UN *vs* PS: 26; Table 2 and Table S9), and none was indicated as a contaminant, so this set of coding genes is further treated as the final group of differentially expressed genes – DEGs. However, successful annotation with predicted function description or specific gene recognition was possible only for 74 of them (12 in M-UN *vs* PS, 18 in M *vs* UN, 54 in M *vs* PS, and 13 in UN *vs* PS comparison; see Table 2 and Tables S10-S11 for detailed lists of genes, and Fig. 5 for comparison of their expression levels). In total 135 PFAM and 78 SMART protein domains were recognized (respectively: 85 and 24 unique) in this subset, and 6,433 GO terms were assigned: 3,882 (1,179 unique) in category of Biological Processes,

This article is protected by copyright. All rights reserved.

Accepted Article

1,127 (244 unique) Molecular Functions and 1,424 (206 unique) Cellular Components. Terms representative for the clusters of functions enriched in comparisons of both high-altitude taxa and Scots pine, and between the first two are given in Figs. S2-S5 and Table S12. No terms were found to be depleted in any of the sets of GO terms assigned to DE markers. Please, notice that GO-term enrichment is highly dependent on the success of sequence annotation, so the outcome should be treated with caution in case of non-model taxa.

In the group of proteins with larger production in two mountain pines compared to Scots pine, we identified three different O-metylotransferases (COMT, CCoAOMT1 and unspecified one), oxidoreductese from 2OG-FE(II) oxygenase family, cis-zetain O-glucosylotransferase, protein disulfide isomerase, elongation factor, cullin 1 and photosystem II 10kDa polypeptide. All of these were similarly up-regulated in comparison of dwarf mountain pine *vs* Scots pine, while in case of *P. uncinata*, the expression of the three last molecules did not differ from *P. sylvestris* levels. In the long list of DEGs differentiating dwarf from Scots pine, besides the few mentioned above, 30 other could be characterized (Table S9 and S11). There was one more O-metylotransferase, two additional peptides from 2OG-Fe(II) oxygenase family, another component of photosystem II and elongation factor. Some distinct functions occurred too, among the more frequently represented were lipoxygenase activity (three genes), ribulose bisphosphate carboxylase oxygenase activity (two genes), and

This article is protected by copyright. All rights reserved.

3-beta hydroxysteroid dehydrogenase/ isomerase activity (two genes; for full list, please see Table S11). In addition to the products of six mentioned DEGs discriminating both mountain pines and Pyrenean pine alone from Scots pine simultaneously, there were six other proteins: a reverse transcriptase, ribosomal protein RPL23, RING finger and CHF zinc finger domain-containing protein, heat-shock protein, 26S proteasome non-ATPase regulatory subunit, and a stem-specific protein.

Regarding markers with enlarged expression in Scots pine, albeit there were some genes concurrently and significantly up-regulated in all three tests involving this taxon, none was annotated. Two aldehyde dehydrogenase were found to have significantly increased expression in Scots pine compared to both high-altitude taxa, but not in tests on individual species – M *vs* PS or UN *vs* PS. Single transcript discriminating M-UN *vs* PS was also found to differ between Scots pine and dwarf mountain pine, it encodes an abietadienol/abietadienal oxidize. 14 more annotated DEGs distinguished Scots pine from the shrubby species (Table S11); beside a few protein functions represented by solitary genes, we found two more common - transcription factor and aldehyde dehydrogenase, with five and two occurrences respectively. The *tufA* gene, encoding an elongation factor, was the only defined DEG found to have higher expression in *P. sylvestris* compared to *P. uncinata*.

Amongst genes with satisfying annotation and variant expression between two focal mountain plants, we found nine sequences up-regulated in dwarf pine and other nine

This article is protected by copyright. All rights reserved.

Accepted Article

with higher expression in Pyrenean pine. In the first group one function occurred twice - both transcripts coded for alpha subunits of elongation factor 1 that act in ribosomes during translation. Moreover, there were two agents engaged in terpenoid synthesis: diphosphate synthase and terpene synthase, and singular representatives of aldehyde dehydrogenases, peroxidases, fatty acid desaturases and plasma membrane H^+-ATPases, together with one receptor-like serine threonine-protein like kinase. The very last function was also found among predicted activities of genes with increased expression in Pyrenean pine. Except that one, there occurred two heat-shock proteins, two cysteine-rich repeat secretory proteins and another cysteine-rich receptor-like protein kinase, a 60S ribosomal protein L18, one stem-specific protein containing DUF3700 domain, and, most probably, an aldose 1-epimerase.

Taking into consideration only the part of markers that distinguished individual species – significant and exhibiting the same direction of adjustment in two species-wise tests – we found five DEGs unique for dwarf and two for Pyrenean pine. In the former, genes coding for elongation factor, fatty acid desaturase and terpene synthase were up-regulated, at the same time, two coding for aldose 1-epimerase and for cysteine-rich receptor-like protein kinase were down-regulated in *P. mugo* compared to other two taxa. Pyrenean pine stood out with its higher expression of *HSP90-1* and a gene coding for some stem-specific protein.

This article is protected by copyright. All rights reserved.

Accepted Article

Accepted Article

4 DISCUSSION

4.1 New reference transcriptome

We examined differentiation in the gene expression between two mountain species from *Pinus mugo* complex and *P. sylvestris* to look at the genetic relationships between the taxa and to identify potential drivers of their ecological and phenotypic divergence. The newly generated reference transcriptome sequence (*MUS assembly*) builds on published transcriptome data (Wachowiak et al., 2015) but extends the available reference by including the two mountain pines. In the new assembly, the number of raw gene models reconstructed is considerably higher as compared to the original *P. sylvestris* raw assembly and to single species studies in other pines (Duran et al., 2019; Parchman et al., 2010; Pinosio et al., 2014). On the other hand, at 14,666, the number successfully annotated protein coding genes is lower than previously reported from *P. sylvestris* (19,659; Wachowiak et al., 2015) or from better studied relatives like *P. lambertiana* (26,568; Gonzales-Ibeas et al., 2016) or *Picea abies* (28,354; Nystedt et al., 2013), and the model species of *Arabidopsis thaliana* (about 27,500; Cheng et al., 2016; www.bioinformatics.psb.ugent.be/plaza/versions/plaza_v3_dicots). However, the good representation of the BUSCOs indicates that our sequence collection is a fair representation of the plants' basal genes, and the number is higher than found in other expression studies in pine seedling needles (Canas et al., 2017). Since the *MUS as-*

This article is protected by copyright. All rights reserved.

sembly incorporates samples from other taxa, it covers new, previously unidentified transcripts. Most of these represent species-specific isoforms of known genes, however several new genes were also detected, mostly expressed in *P. uncinata* or *P. mugo* but not in *P. sylvestris*. These were reported from ORF-containing, significantly differentiated gene models (DEGs), though, some more could have been dropped based on their low expression.

4.2 Interspecific gene expression patterns

Our data provide a gene expression perspective on the mutual relationships between the studied pines. We assumed that controlled glasshouse environment and uniform setting of growth (Wachowiak et al., 2018) equalized expression between the species, leading to rather conservative measures of differentiation as compared to natural, *in situ* variation. Therefore, the observed patterns were expected to reflect variation resulting from fixed and heritable determinants that mirror the species evolutionary history. As compared to the total number of protein-coding genes, we found relatively low numbers of differentially expressed markers (121). The adopted expression-based filtering criteria should be regarded as a mild cutoff, enough to exclude transcripts with the lowest signal. Additionally, the conservative approach of limiting the reported markers to the intersection of results from concurrent methods of differential expression detection impacted these counts moderately (the respective unions were just 3-5 times larger). Counts of DEGs exhibited interesting pattern showing similar

This article is protected by copyright. All rights reserved.

Accepted Article

Accepted Article

numbers of genes in comparisons of *P. mugo* and *P. uncinata vs P. sylvestris* (26), *P mugo vs P. uncinata* (28) and *P. uncinata vs P. sylvestris* (26), whilst triple that number were observed in comparison of *P. mugo* and *P. sylvestris* (90). The pattern denotes asymmetric relations between Scots pine and two mountain pines, consistent with most available data (e.g., Wachowiak et al., 2015; Zaborowska et al., 2021). The number of markers that unite the mountain pines and discriminate them from Scots pine ((M *vs* PS) □ (U *vs* PS): 14) was not much larger than the numbers of DEGs specific to each of them ((M *vs* PS) □ (M *vs* UN): 8 and (U *vs* PS) □ (U *vs* M): 2). Moreover, markers found between two mountain taxa and Scots pine, those present in the 'broad' M-UN *vs* PS set, were dominated by differences between dwarf pine and Scots pine, which disappeared when filtering to include those which were also differentiated in UN *vs* PS comparison – were absent from the 'narrow' intersection. We observed generally lower intraspecific variation in expression of markers in individuals of dwarf mountain pine and the other two species occasionally showed bipolar spread of the transcription estimates. This results contrast with the pattern observed in a broad set of genome-wide SNPs markers (Zaborowska et al., 2021), where *P. mugo* showed the highest intraspecific diversity among these three taxa. However, due to the limited sample sizes used here, much of the variation among the pine species remains unexplored. Overall, the analysis exposed a limited set of genes with diverged patterns of expression between species as compared to the number of transcripts analyzed.

This article is protected by copyright. All rights reserved.

Accepted Article

4.3 Putative signatures of mountain pine adaptations

The group of genes most differentiated in the expression between the *P. mugo* complex and *P. sylvestris* could be grouped into several major categories related to environmental gradients and conditions of the species occurrence. Three genes encoding O-methyotransferases (OMTs) were discovered in a group of sequences highly up-regulated in the *P. mugo* complex taxa (Table 2), and the O-methyltransferase activity was as one of the main GO terms enriched in these group of DEGs (Table S12). OMTs are a large family of enzymes that add methyl groups to target oxygen atoms of a variety of secondary metabolites such as phenylpropanoids, flavonoids and some alkaloids, playing important roles in lignin biosynthesis, defense and stress resistance (Lam et al., 2007). Two were annotated as caffeoyl CoA O-methyltransferase (*CCoAOMT*) and catechol-O-methyltransferase (*COMT*), and third was also grouped with *COMT*-like sequences or type II OMTs from the genus *Pinus*. The best recognized function of *CCoAOMT* is its role in lignin synthesis, and interestingly the patterns of nucleotide polymorphism at the gene showed signatures of natural selection in *Pinus taeda* and *P. sylvestris* (Gonzalez-Martinez et al., 2006; Wachowiak et al., in press). Finding the *COMT* gene was unexpected, since to our knowledge, such sequences have not been noted earlier in Scots pine, despite extensive research related to the wood production process in this species (Lim et al., 2021; Paasela et al., 2017). Another OMT was highly differentiated among *P. mugo* and *P. sylvestris* but showed

This article is protected by copyright. All rights reserved.

Accepted Article

intermediate levels of expression in *P. uncinata*. The results suggest that the pine species adapted to mountain regions characterized by strong winds and often thick snow cover, adjusted their wood properties via the lignin synthesis pathway. On the other hand, group of phenylpropanoid metabolic processes was indicated to also be enriched in genes up-regulated in Scots pine (Figs. S2-S3). Furthermore, the mentioned OMTs are vital in plant responses to environmental stressors too, mainly salinity, drought and high ozone concentrations (Chiron et al., 2000; Chun et al., 2021). The last aspect seems important for subalpine flora (Matyssek & Sandermann, 2003), and *P. mugo* and *P. uncinata* specifically, as both are very sensitive to ozone (Bicarova et al., 2019; Diaz-de-Quijano et al., 2019).

Other gene up-regulated in mountain pines as compared to Scots pine encodes oxidoreductase from the 2-oxogluterate (2OG) Fe(II)-dependent oxygenase superfamily that exhibit great diversity of metabolite biosynthesis, biodegradation, regulatory or structural roles (Herr & Hausinger, 2018). Another transcript represented nuclear *PsbR* gene for plastid photosystem II (PSII) 10 kDa polypeptide. The polypeptide is essential for stable assembly of proteins in the oxygen-evolving complex including water splitting and electron transport in PSII. It was shown to have enhanced activity under low light conditions (Suorsa et al., 2006), be dependent on ultraviolet-B radiation (Peng et al., 2021), and has expression negatively correlated with sun, temperature and wind, but positively with precipitation and humidity (Sjodin et al., 2008). As

This article is protected by copyright. All rights reserved.

Accepted Article

the light intensity and UV-B radiation are greater at higher altitudes, up-regulation in both mountain pines could reflect their adaptation and reaction to changed glasshouse conditions. Enrichment analysis indicated the regulation of circadian rhythm as over-represented term in *P. mugo* and *P. uncinata*, meaning that these plants might evolved to better synchronize with light cycle, or other diurnal cycles altered in their environment. Another DEG encoded protein disulfide isomerase (*PDI*), which catalyzes conversion of thiol-disulfide by formation and breakage of SS-bonds between cysteine residues, being responsible for proper folding of proteins (protein folding in endoplasmic reticulum was also an enriched GO term). PDIs in plants are involved in responses to biotic and abiotic stress (Feldeverd et al., 2020; Zhang et al., 2018), and may play a role in redox signaling (Wittenberg & Danon, 2008). Also, an elongation factor was distinctly up-regulated in the subalpine taxa (better specified below).

Also strongly differentiated, but exhibiting the opposite direction of regulation (i.e. greater expression in Scots pine) was abietadienol/abiedienal oxigenase, a cytochrome P450 monooxygenase unique to conifers. It is specialized in synthesis of tricyclic diterpene resin acids (DRAs) (Bathe & Tissier, 2019, Ro et al., 2005), essential components of conifers' oleoresin - the defense blend against pathogens, pests and herbivores, and a chemotaxonomic marker effective in *Pinus* (Mitic et al., 2017). Potentially, this could be a consequence of the extensive range of Scots pine, and its need to defend against a broader range of pests and pathogens.

This article is protected by copyright. All rights reserved.

Accepted Article

4.4 Gene expression divergence between the mountain pine species

Diverse genes were found to distinguish the two mountain pines. A fatty acid desaturase was one of the best markers, these enzymes modify properties of fatty acids (FAs) chains by forming double bonds between adjacent carbon atoms. They are responsible for appropriate structure and fluidity of plasma membranes, helping to deal with temperature changes (Kates et al., 1984; Makarenko et al., 2014) or ozone effects (Matyssek & Sandermann, 2003). The composition of FAs was also successfully used as chemotaxonomic discrimination tool between the main families and within some species groups of conifers (Wolff et al., 2001). Our result indirectly supports the utility of this tool, and further investigation of FA variation in these species is merited. Furthermore, peroxidase PRX20, had greatly increased expression in *P. mugo* compared to *P. uncinata*. This protein belongs to a plant-specific family of class III peroxidases, involved in diverse processes such us cell elongation, lignification, seed germination and stress responses (Shigeto & Tsutsumi, 2016). Although we lack data from gymnosperms, the PRX20 of *Populus* trees exhibited strong activity towards coniferyl alcohol, a monolignol participating in lignin synthesis (Ren et al., 2014). These data suggest that *P. uncinata* might be less vulnerable to H_2O_2 toxicity or that wood properties differ in species of the *P. mugo* complex. Similar bias in expression modes was observed in two elongation factors, annotated as EF-1α enzymes. Essentially, various elongation factors (EFs) act in ribosomes during translation, among

This article is protected by copyright. All rights reserved.

Accepted Article

others facilitating the elongation of synthesized peptide (Sasikumar et al., 2012). These specific two are related to regulation of growth, as indicated by GO terms assigned (and enriched in the discussed group of DEGs, Table S12 and Fig. S4), so they represent the candidates for determinants of the shrubby habit of *P. mugo*. Another possible agent might be seen in the diphosphate synthase, it is related to the gibberellin metabolic processes (Li et al., 2017), and these hormones are known to effect dwarfism in plants (Ford et al., 2018).

In the list of GO terms over-represented in transcripts that were more abundant in Pyrenean pine, many represented diverse responses to exogenous stimuli, like defense responses, response to water deprivation or responses to inorganic substances (Table S12, Fig. S5). Among the more significant, response to heat and heat acclimation were found. These were mainly represented by two heat-shock proteins of family 90 (HSP90) up-regulated in *P. uncinata*. In plants, this gene family functions as a molecular chaperone, acting in stress signal transduction and influencing responses to different abiotic stresses, but also participates in plant development and resistance to pests and diseases (Mozharovskaya, 2018; Xu et al., 2012). That observation might speak for adaptation of Pyrenean pine to warmer climate of the western European mountain ranges. Furthermore, transcripts encoding three proteins sharing a plant-specific stress-antifungal domain (two cysteine-rich repeat secretory proteins, one cysteine-rich receptor-like protein kinase) were also found up-regulated in *P.*

This article is protected by copyright. All rights reserved.

unicnata. They belong to a big group of signaling, transmembrane proteins which respond to diverse developmental and environmental prompts (Mou et al., 2021; Vaattovaara et al., 2019; Wrzaczek et al., 2010). Another differentially expressed transcript encoded a stem-specific protein that putative orthologues act as regulators of the growth of *Ricinus communis* internodes, and of the development and differentiation of *Eucalyptus* callus (Hu et al., 2016; Zhang et al., 2022).

All the genes described above, and listed in tables, reveal the expression component of interspecific variation that may play a role in the species ecological divergence and adaptation. As such, they well deserve testing their phenotypic effects. First, validation of the gene expression profiles in form of qRT-PCR or other relevant analysis (Dallas et al., 2005) should be performed. Furthermore, for finer resolution of the interactions between expression, phenotype and environment, robust statistical modeling of transcriptome data and environmental factors, conducted on larger sample of populations, would be supportive. Finally, complementary test, that would confirm how actual the reveled expression patterns are, could be governed on individuals collected in their native conditions, across natural populations of the three pine taxa.

This article is protected by copyright. All rights reserved.

Accepted Article

Accepted Article

4.5 Conclusions

This study reveals an expression component of interspecific variation that may play a role in the ecological divergence and differential adaptation among the study species. As such, they merit further examination to assess their association with phenotype among standing populations of the species. Furthermore, a new genomic reference for the pine transcriptome, with particular relevance for the taxonomically challenging *P. mugo* complex (including *P. mugo* and *P. uncinata*) will facilitate research into these species. We identified several genes that exhibited good discrimination ability between the sister mountain pines and their close relative, Scots pine. Those genes could be grouped into several functional categories including: wood trait properties, oxidative stress response, other abiotic factors related to salinity, drought and temperature, as well as some biotic stressors. Although the molecular basis for adaptation to different environments is likely to be highly complex and difficult to validate, the identified markers are excellent candidates for further investigation given their putative function and corresponding expression patterns among the three taxa. The relatively low number of differentially expressed genes discovered in the study is in line with previous reports showing high molecular and genetic similarity between the species, suggesting that additional determinants of their phenotypic and ecological diversity may involve variation in the noncoding part of the genome, and epigenetic interactions.

This article is protected by copyright. All rights reserved.

Accepted Article

ACKNOWLEDGEMENTS

The authors would like to express their gratitude to institutions which funded the study: the Polish National Science Centre (UMO-2017/27/B/NZ9/00159) which supported analyses of the expression data and publication of the study, and the Natural Environment Research Council (NE/K012177/1) for financing data acquisition. There is no conflict of interest to declare.

REFERENCES

Adams RP, Tashev AN. 2019. Composition of the leaf volatile terpenoids of *Pinus mugo* Turra from Bulgaria. *Phytologia* 101: 74-80.

Aitken SN, Yeaman S, Holliday JA, Wang T, Curtis-McLane S. 2008. Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications* 1: 95-111.

Bathe U, Tissier A. 2019. Cytochrome P450 enzymes: A driving force of plant diterpene diversity. *Phytochemistry* 161: 149-162.

Bicarova S, Sitkova Z, Pavlendova H, Fleischer P (Jr), Fleischer P (Sr), Andrzej Bytnerowicz. 2019. The role of environmental factors in ozone uptake of *Pinus mugo* Turra. *Atmospheric Pollution Research* 10: 283-293.

This article is protected by copyright. All rights reserved.

Accepted Article

Bogunic F, Siljak-Yakovlev S, Muratovic E, Pustahija F, Medjedovic S. 2011. Molecular cytogenetics and flow cytometry reveal conserved genome organization in *Pinus mugo* and *P. uncinata*. *Annals of Forest Science* 68: 179-187.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics* 30: 2114-2120.

Boratynska K, Boratynksi A. 2007. Taxonomic differences among closely related pines *Pinus sylvestris*, *P. mugo*, *P. uncinata*, *P. rotundata* and *P. uliginosa* as revealed in needle sclerenchyma cells. *Flora* 202: 555-569.

Boratynska K, Jasinska AK, Boratynski A. 2015. Taxonomic and geographic differentiation of *Pinus mugo* complex on the needle characteristics. *Systematics and Biodiversity* 13: 901-915.

Bray NL, Pimentel H, Melsted P and Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology* 34: 525–527.

Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nature Methods* 12: 59-60.

Businsky R, Kirschner J. 2010. *Pinus mugo* and *P. uncinata* as parents of hybrids: a taxonomic and nomenclatural survey. *Phyton* 50: 27-57.

This article is protected by copyright. All rights reserved.

Accepted Article

Canas RA, Li Z, Pascual MB, Castro-Rodriguez V, Avila C, Sterck L, Van de Peer Y, Canovas FM. 2017. The gene expression landscape of pine seedling tissues. *The Plant Journal* 91: 1064-1087.

Cantalapiedra CP, Hernandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021. eggNOG-mapper v2: functional annotation, orthology assignments, and prediction at the metagenomic scale. *bioRxiv* 2021.06.03.446934.

Chen H, Boutros PC. 2011. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics* 12: 35.

Cheng C-Y, Krishnakumar V, Chan AP, Thibaud-Nissen F, Schobel S, Town CD. 2017. Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *The Plant Journal* 89: 789-804.

Chiron H, Drouet A, Claudot AC, Eckerskorn C, Trost M, Heller W, Ernst D, Sandermann H Jr. 2000. Molecular cloning and functional expression of a stress-induced multifunctional O-methyltransferase with pinosylvin methyltransferase activity from Scots pine (*Pinus sylvestris* L.). *Plant Molecular Biology* 44: 733-45.

Christensen KI. 1987. A morphometric study of the *Pinus mugo* Turra complex and its natural hybridization with *P. sylvestris* L. (Pinaceae). *Feddes Repertorium* 98: 623-635.

This article is protected by copyright. All rights reserved.

Chun HJ, Lim LH, Cheong MS, Baek D, Park MS, Cho HM, Lee SH, Jin BJ, No DH, Cha YJ, Lee YB, Hong JC, Yun D-J, Kim MC. 2021. *Arabidopsis* CCoA-OMT1 plays a role in drought stress response via ROS- and ABA-dependent manners. *Plants* 10: 831.

Dallas PB, Gottardo NG, Firth MJ, Beesley AH, Hoffmann K, Terry PA, Freitas JR, Boag JM, Cummings AJ, Kees UR. 2005. Gene expression levels assessed by oligonucleotide microarray analysis and quantitative real-time RT-PCR – how well do they correlate? *BMC Genomics* 6: 59.

Diaz-de-Quijano M, Kefauver S, Ogaya R, Vollenweider P, Ribas A, Penuelas J. 2019. Visible ozone-like injury, defoliation, and mortality in two *Pinus uncinata* stands in the Catalan Pyrenees (NE Spain). *European Journal of Forest Research* 135: 687-696.

Duran R, Rodriguez V, Carrasco A., Neale D, Balocchi C, Valenzuela S. 2019. SNP discovery in radiata pine using a *de novo* transcriptome assembly. *Trees* 33: 1505-1511.

Dzialuk A, Boratynska K, Romo A, Boratynski A. 2017. Taxonomic and geographic variation of the *Pinus mugo* complex on chloroplast microsatellite markers. *Systematics and Biodiversity* 15: 464-479.

This article is protected by copyright. All rights reserved.

Accepted Article

Accepted Article

European Forest Genetic Resources Programme (EUFORGEN) [online]. Available from www.euforgen.org [accessed 22 Steptember 2009].

FastQC: A quality control tool for high throughput sequence data [online]. Available from www.github.com/s-andrews/FastQC [accessed 27 January 2021].

Feldeverd E, Porter BW, Yuen CYL, Iwai K, Carrillo R, Smith T, Barela C, Wong K, Wang Pe, Kang B-H, Matsumoto K, Christopher DA. 2020. The *Arabidopsis* protein disulfide isomerase subfamily M isoform, PDI9, localizes to the endoplasmic reticulum and influences pollen viability and proper formation of the pollen exine during heat stress. *Frontiers in Plant Science* 11.

Ford BA, Foo E, Sharwood R, Karafiatova M, Vrana J, MacMillan C, Nichols DS, Steuernagel B, Uauy C, Dolezel J, Chandler PM, Spielmeyera W. 2018. *Rht18* semidwarfism in wheat is due to increased GA2-oxidaseA9 expression and reduced GA content. *Plant Physiology* 177: 168-180.

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JYH, Zhang J. 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology* 5: R80.

This article is protected by copyright. All rights reserved.

Accepted Article

Gonzalez-Martinez SC, Ersoz E, Brown GR, Wheeler NC, Neale DB. 2006. DNA sequence variation and selection of tag SNPs at candidate genes for drought-stress response in *Pinus taeda* L. *Genetics* 172: 1915-1926.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, Macmanes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, Leduc RD, Friedman N, Regev A. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* 8: 1494-512.

Hamernik J, Musil I. 2007. *Pinus mugo* complex - its structuring and general overview of the used nomenclature. *Journal of Forest Science* 53: 253-266.

Hart AJ, Ginzburg S, Xu MS, Fisher CR, Rahmatpour N, Mitton JB, Paul R, Wegrzyn JL. 2020. EnTAP: Bringing faster and smarter functional annotation to non-model eukaryotic transcriptomes. *Molecular Ecology Resources* 20: 591-604.

Herr CQ, Hausinger RP. 2018. Amazing diversity in biochemical roles of Fe(II)/2-oxoglutarate oxygenases. *Trends in Biochemical Sciences* 43: 517-532.

This article is protected by copyright. All rights reserved.

Accepted Article

Heuertz M, Teufel J, Gonzalez-Martinez SC, Soto A, Fady B, Alia R, Vendramin GG. 2010. Geography determines genetic relationships between species of mountain pine (*Pinus mugo* complex) in western Europe. *Journal of Biogeography* 37: 541-556.

Hu W, Chen L, Qiu X, Lu H, Wei J, Bai Y, He N, Hu R, Sun L, Zhang H, Shen G. 2016. Morphological, physiological and proteomic analyses provide insights into the improvement of castor bean productivity of a dwarf variety in comparing with a high-stalk variety. *Frontiers in Plant Science* 7.

Huerta-Cepas J, Szklarczyk D, Heller D, Hernandez-Plaza A, Forslund SK, Cook H, MendeDR, Letunic I, Rattei T, Jensen LJ, von Mering C, Bork P. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research* 47: D309–D314.

Jalas J, Suominen J. 1973. *Atlas Florae Europaeae: Gymnospermae (Pinaceae to Ephedraceae).* Helsinki: Committee for Mapping the Flora of Europe & Suomen Biologian Seura Vanamo.

Jasinska AK, Wachowiak W, Muchewicz E, Boratynska K, Montserrat JM, Boratynski A. 2010. Cryptic hybrids between *Pinus uncinata* and *P. sylvestris*. *Botanical Journal of the Linnean Society* 163: 473-485.

This article is protected by copyright. All rights reserved.

Accepted Article

Kates M, Pugh EL, Ferrante G. 1984. Regulation of membrane fluidity by lipid desaturases. In: Kates M., Manson L.A. eds. *Membrane Fluidity*. Biomembranes vol 12. Boston: Springer.

Kormutak A, Ostrolucka M, Vookova B, Pretova A, Feckova M. 2005. Artificial hybridization of *Pinus sylvestris* L. and *Pinus mugo* Turra. *Acta Biologica Cracoviensia s. Botanica* 47: 129-134.

Kremer A. 2011. Missing heritability and missing Fst of candidate genes: why does gene variation differ from trait variation in trees? *BMC Proceedings* 5: I1.

Labiszak B, Wachowiak W. 2021. Molecular signatures of reticulate evolution within the complex of European pine taxa. *Forests* 12: 489.

Lam KC, Ibrahim RK, Behdad B, Dayanandan S. 2007. Structure, function, and evolution of plant O-methyltransferases. *Genome* 50: 1001-1013.

Langmead B, Salzberg S. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9: 357-359.

Lascoux M, Glemin S, Savolainen O. 2016. Local adaptation in plants. In: *eLS*. Chichester: John Wiley & Sons, Ltd.

Lewandowski A, Boratynski A, Mejnartowicz L. 2000. Allozyme investigations on the genetic differentiation between closely related pines - *Pinus sylvestris*, *P.*

This article is protected by copyright. All rights reserved.

Accepted Article

mugo, *P. uncinata*, and *P. uliginosa* (Pinaceae). *Plant Systematics and Evolution* 221: 15-24.

Li B, Fillmore N, Bai Y, Collins M, Thomson JA, Stewart R, Dewey CN. 2014. Evaluation of *de novo* transcriptome assemblies from RNA-Seq data. *Genome Biology* 15: 553.

Li C, Lu Y, Shen WH, Yu Y, Dong A. 2018. Chromatin-remodeling factor OsINO80 is involved in regulation of gibberellin biosynthesis and is crucial for rice plant growth and development. *Journal of Integrative Plant Biology* 60: 144– 159.

Lim KJ, Paasela T, Harju A, Venalainen M, Paulin L, Auvinen P, Karkkainen K, Teeri TH. 2021. A transcriptomic view to wounding response in young Scots pine stems. *Scientific Reports* 11: 3778.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq. 2. *Genome Biology* 15: 550.

Makarenko SP, Konenkina TA, Suvorova GG, Oskorbina MV. 2014. Seasonal changes in the fatty acid composition of *Pinus sylvestris* needle lipids. *Russian Journal of Plant Physiology* 61: 119-123.

This article is protected by copyright. All rights reserved.

Accepted Article

Madritsch S, Burg A, Sehr EM. 2021. Comparing de novo transcriptome assembly tools in di- and autotetraploid non-model plant species. *BMC Bioinformatics* 22: 146.

Manni M, Berkeley MR, Seppey M, Simao FA, Zdobnov EM. 2021. BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Molecular Biology and Evolution* 38: 4647-4654.

Matyssek R, Sandermann H. 2003. Impact of ozone on trees: an ecophysiological perspective. In: Esser K, Lüttge U, Beyschlag W, Hellwig F eds. *Progress in Botany*, vol. 64. Heidelberg: Springer.

Mitic ZS, Jovanovic SC, Zlatkovic BK, Nikolic BM, Stojanovic GS, Marin PD. 2017. Needle terpenes as chemotaxonomic markers in *Pinus*: subsections *Pinus* and *Pinaster*. *Chemistry & Biodiversity* 14: e1600453.

Monteleone I, Ferrazzini D, Belletti P. 2006. Effectiveness of neutral RAPD markers to detect genetic divergence between the subspecies *uncinata* and *mugo* of *Pinus mugo* Turra. *Silva Fennica* 40: 391-406.

Mou S, Meng Q, Gao F, Zhang T, He W, Guan D, He S. 2021. A cysteine-rich receptor-like protein kinase CaCKR5 modulates immune response against *Ralstonia solanacearum* infection in pepper. *BMC Plant Biology* 21: 382.

This article is protected by copyright. All rights reserved.

Accepted Article

Mozharovskaya LV. 2018. Functional annotation of genes of *Pinus sylvestris* associated with resistance to phytopathogenic micromycetes. *Journal of the Belarusian State University. Biology* 2: 78-84. [in Russian]

Neale DB, Martinez-Garcia PJ, De La Torre AR, Montanari S, Wei XX. 2017. Novel insights into tree biology and genome evolution as revealed through genomics. *Annual Review of Plant Biology* 68: 457-483.

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, Hallman J, Keech O, Klasson L, Koriabine M, Kucukoglu M, Kaller M, Luthman J, Lysholm F, Niittyla T, Olson A, Rilakovic N, Ritland C, Rossello JA, Sena J, Svensson T, Talavera-Lopez C, Theißen G, Tuominen H, Vanneste K, Wu Z-Q, Zhang B, Zerbe P, Arvestad L, Bhalerao R, Bohlmann J, Bousquet J, Garcia Gil R, Hvidsten TR, de Jong P, MacKay J, Morgante M, Ritland K, Sundberg B, Lee Thompson S, Van de Peer Y, Andersson B, Nilsson O, Ingvarsson PK, Lundeberg J, Jansson S. 2013. The Norway spruce genome sequence and conifer genome evolution. *Nature* 497: 579-584.

O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y,

This article is protected by copyright. All rights reserved.

Accepted Article

Blinkova O, Brover V, Chetvernin V, Choi J, Cox E, Ermolaeva O, Farrell CM, Goldfarb T, Gupta T, Haft D, Hatcher E, Hlavina W, Joardar VS, Kodali VK, Li W, Maglott D, Masterson P, McGarvey KM, Murphy MR, O'Neill K, Pujar S, Rangwala SH, Rausch D, Riddick LD, Schoch C, Shkeda A, Storz SS, Sun H, Thibaud-Nissen F, Tolstoy I, Tully RE, Vatsan AR, Wallin C, Webb D, Wu W, Landrum MJ, Kimchi A, Tatusova T, DiCuccio M, Kitts P, Murphy TD, Pruitt KD. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research* 44: D733-745.

Paasela T, Lim KJ, Pietiainen M, Teeri TH. 2017. The O-methyltransferase PMT2 mediates methylation of pinosylvin in Scots pine. *New Phytologist* 214: 1537-1550.

Parchman TL, Geist KS, Grahnen JA, Benkman CW, Buerkle A. 2010. Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery. *BMC Genomics* 11: 180.

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods* 14:417-419.

This article is protected by copyright. All rights reserved.

Accepted Article

Peng T, Wang Y, Yang T, Wang F, Luo J, Zhang Y. 2021. Physiological and bio-chemical responses, and comparative transcriptome profiling of two *Angelica sinensis* cultivars under enhanced ultraviolet-B radiation. *Frontiers in Plant Science* 12.

Perry A, Wachowiak W, Downing A, Talbot R, Cavers S. 2020. Development of a SNP array for population genomic studies in four European pine species. *Molecular Ecology Resources* 20: 1697-1705.

Petit RJ, Hampe A. 2006. Some evolutionary consequences of being a tree. *Annual Review of Ecology Evolution and Systematics* 37: 187-214.

Pinosio S, Gonzalez-Martinez SC, Bagnoli F, Cattonaro F, Grivet D, Marroni F, Lorenzo Z, Pausas JG, Verdu M, Vendramin GG. 2014. First insights into the transcriptome and development of new genomic tools of a widespread circum-Mediterranean tree species, *Pinus halepensis* Mill. *Molecular Ecology Resources* 14: 846-856.

PLAZA 3.0: *Arabidopsis thaliana* gene overview [online]. Available from www.bioinformatics.psb.ugent.be/plaza/versions/plaza_v3_dicots [accessed 31 March 2022].

Plomion C, Bastien C, Bogeat-Triboulot MB, Bouffier L, Dejardin A, Duplessis S, Fady B, Heuertz M, Le Gac AL, Le Provost G, Legue V, Lelu-Walter MA,

This article is protected by copyright. All rights reserved.

Accepted Article

Leple JC, Maury S, Morel A, Oddou-Muratorio S, Pilate G, Sanchez L, Scotti I, Scotti-Saintagne C, Segura V, Trontin JF, Vacher C. 2016. Forest tree genomics: 10 achievements from the past 10 years and future prospects. *Annals of Forest Science* 73: 77-103.

Ren L-L, Liu Y-J, Liu H-J, Qian T-T, Qi L-W, Wang X-R, Zeng Q-Y. 2014. Subcellular relocalization and positive selection play key roles in the retention of duplicate genes of *Populus* class III peroxidase family. *The Plant Cell* 26: 2404–2419.

Ro D-K, Arimura G-I, Lau SYW, Piers E, Bohlmann J. 2005. Loblolly pine abietadienol/abietadienal oxidase *PtAO* (CYP720B1) is a multifunctional, multisubstrate cytochrome P450 monooxygenase. *Proceedings of the National Academy of Sciences* 102: 8060-8065.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26: 139-140.

Sasikumar AN, Perez WB, Kinzy TG. 2012. The many roles of the eukaryotic elongation factor 1 complex. *Wiley Interdisciplinary Reviews RNA* 3: 543-555.

Shigeto J, Tsutsumi Y. 2016. Diverse functions and reactions of class III peroxidases. *New Phytologist* 209: 1395-1402.

This article is protected by copyright. All rights reserved.

Accepted Article

Sjodin A, Wissel K, Bylesjo M, Trygg J, Jansson S. 2008. Global expression profiling in leaves of free-growing aspen. *BMC Plant Biology* 8: 61.

Sokolowska J, Fuchs H, Celinski K. 2021. New insight into taxonomy of European mountain pines, *Pinus mugo* complex, based on complete chloroplast genomes sequencing. *Plants* 10: 1331.

Suorsa M, Sirpio S, Allahverdiyeva Y, Paakkarinen V, Mamedov F, Styring S, Aro EM. PsbR, a missing link in the assembly of the oxygen-evolving complex of plant photosystem II. *Journal of Biological Chemistry* 281: 145-50.

Supek F, Bosnjak M, Skunca N, Smuc T. 2011. REVIGO summarizes and visualizes long lists of Gene Ontology terms. *PLOS ONE* 6: e21800.

The Plant List: *Pinus sylvestris*. Version 1.1 (2013) published by Royal Botanic Garden Kew and Missouri Botanic Garden [online]. Available from www.theplantlist.org [accessed 10 March 2021].

The UniProt Consortium. 2021. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research* 49: D480–D489.

TransDecoder by Brian Haas [online]. Available from www.github.com/TransDecoder/TransDecoder/wiki [accessed 22 October 2018].

This article is protected by copyright. All rights reserved.

Accepted Article

Trinity Transcript Quantification by Brian Haas [online]. Available from www.github.com/trinityrnaseq/trinityrnaseq/wiki [accessed 20 April 2021].

Vaattovaara A, Brandt B, Rajaraman S, Safronov O, Veidenberg A, Luklova M, Kangasjarvi J, Loytynoja A, Hothorn M, Salojarvi J, Wrzaczek M. 2019. Mechanistic insights into the evolution of DUF26-containing proteins in land plants. *Communications Biology* 2: 56.

Wachowiak W, Boratynska K, Cavers S. 2013. Geographical patterns of nucleotide diversity and population differentiation in three closely related European pine species in the *Pinus mugo* complex. *Botanical Journal of the Linnean Society* 172: 225-238.

Wachowiak W, Lesniewicz K, Odrzykoski IJ, Augustyniak H, Prus-Glowacki W. 2000. Species specific cpDNA markers useful for studies on the hybridization between *Pinus mugo* and *P. sylvestris*. *Acta Societatis Botanicorum Poloniae* 69: 273-276.

Wachowiak W, Palme AE, Savolainen O. 2011. Speciation history of three closely related pines *Pinus mugo* (T.), *P. uliginosa* (N.) and *P. sylvestris* (L.). *Molecular Ecology* 20: 1729-1743.

This article is protected by copyright. All rights reserved.

Accepted Article

Wachowiak W, Perry A, Donnelly K, Cavers S. 2018. Early phenology and growth trait variation in closely related European pine species. *Ecology and Evolution* 8: 655-666.

Wachowiak W, Perry A, Zaborowska J, Gonzalez-Martinez SC Cavers S. 2022. Admixture and selection patterns across the European distribution of Scots pine, *Pinus sylvestris* (Pinaceae) [in press, accepted by the *Botanical Journal of the Linnean Society*].

Wachowiak W, Trivedi U, Perry A, Cavers S. 2015. Comparative transcriptomics of a complex of four European pine species. *BMC Genomics* 16: 234.

Wikimedia Commons, the free media repository [online]. Available from www.commons.wikimedia.org [accessed 4 October 2018].

Wittenberg G, Danon A. 2008. Disulfide bond formation in chloroplasts: Formation of disulfide bonds in signaling chloroplast proteins. *Plant Science* 175: 459-466.

Wolff RL, Lavialle O, Pedrono F, Pasquier E, Deluc LG, Marpeau AM, Aitzetmüller K. 2001. Fatty acid composition of Pinaceae as taxonomic markers. *Lipids* 36: 439-451.

This article is protected by copyright. All rights reserved.

Accepted Article

Wrzaczek M, Brosche M, Salojarvi J, Kangasjarvi S, Idanheimo N, Mersmann S, Robatzek S, Karpinski S, Karpinska B, Kangasjarvi J. 2010. Transcriptional regulation of the CRK/DUF26 group of receptor-like protein kinases by ozone and plant hormones in *Arabidopsis*. *BMC Plant Biology* 10: 95.

Xu Z-S, Li Z-Y, Chen Y, Chen M, Li L-C, Ma Y-Z. 2012. Heat shock protein 90 in plants: Molecular mechanisms and roles in stress responses. *International Journal of Molecular Sciences* 13: 15706-15723.

Zaborowska J, Labiszak B, Perry A, Cavers S, Wachowiak W. 2021. Candidate genes for the high-altitude adaptations of two mountain pine taxa. *International Journal of Molecular Sciences* 22: 3477.

Zaborowska J, Labiszak B, Wachowiak W. 2019. Population history of European mountain pines *Pinus mugo* and *Pinus uncinata* revealed by mitochondrial DNA markers. *Journal of Systematics and Evolution* 58: 474-486.

Zhang Y, Li J, Li C, Chen S, Tang Q, Xiao Y, Zhong L, Chen Y, Chen B. 2022. Gene expression programs during callus development in tissue culture of two *Euca-lyptus* species. *BMC Plant Biology* 22: 1.

Zhang Z, Liu X, Li R, Yuan L, Dai Y, Wang X. 2018. Identification and functional analysis of a protein disulfide isomerase (*At*PDI1) in *Arabidopsis thaliana*. *Frontiers in Plant Science* 9.

This article is protected by copyright. All rights reserved.

Zukowska WB, Wachowiak W. 2017. Cross-amplification and multiplexing of cpSSRs and nSSRs in two closely related pine species (*Pinus sylvestris* L. and *P. mugo* Turra). *Dendrobiology* 77: 59-64.

Table 1. Location of pine populations sampled and glasshouse used for experiment.

Sp.[†]	**Sample, ENA Acc.**[‡]	**Country code, population**[§]	**Latitude**	**Longitude**	**Alt.**[¶]
P. mugo	**M1**, SAMEA2672716	RO, Busteni	45°25′55″ N	25°27′06″ E	2070
	M2, SAMEA2672717	BA, Bjelasnica Mts	43°45′00″ N	18°13′08″ E	2120
	M4, SAMEA2672719	AT, Scharnitz	47°22′42″ N	11°17′45″ E	1400
	M5, SAMEA2672720	PL, Slaskie Kamienie	50°46′35″ N	15°36′08″ E	1400
P. uncinata	**UN1**, SAMEA2672721	FR, Col de la Croix de Morand	45°35′58″ N	2°50′44″ E	1200
	UN3, SAMEA2672723	AD, Vall de Ransol	42°35′02″ N	1°38′21″ E	2025

This article is protected by copyright. All rights reserved.

	UN4, SAMEA2672724	ES, Sierra de Gudar	40°28′49″ N	0°41′51″ W	2000
P. sylvestris	**PS2**, SAMEA2672712[#]	UK, Glen Tanar	57°02′60″ N	2°51′36″ W	334
	PS3, SAMEA2672713	FI, Punkaharju	61°45′33″ N	29°23′21″ E	80
	PS4, SAMEA2672714	PL, Jarocin	51°58′20″ N	17°28′40″ E	120
	PS5, SAMEA2672715	ES, Trevenque	37°05′47″ N	3°32′51″ E	1170
Glasshouse facility, UK CEH		**Edinburgh, UK**	55°57′00″ N	3°11′56″ W	189

Footnote: [†] Sp. – species name; [‡] ENA Acc. – European Nucleotide Archive Sample Accession; [§] Country codes: RO – Romania, BA – Bosnia and Herzegovina, AT – Austria, PL – Poland, FR – France, AD – Andorra, ES – Spain, UK – United Kingdom, FI – Finland; [¶] Alt. – Altitude in meters above sea level; [#] Reference sample used only for the transcriptome assembly.

Accepted Article

This article is protected by copyright. All rights reserved.

Table 2. Successfully annotated DEGs from comparisons of Scots pine against both mountain pines and between the latter. Listed sequences correspond to color-marked genes from Fig. 5.

Gene, query isoform	**Best hit sequence, species of origin**	**Predicted gene and/or function description**
Up-regulated in both mountain pines in relation to *P. sylvestris*		
DN10418_c0_g2 i1	tr|A0A223PIL1_PICGL *Picea glauca*	cis-zeatin O-glucosyltransferase
DN111461_c24_g1 i1	tr|A9NVS6_PICSI *Picea sitchensis*	oxidoreductase, 2OG-Fe(II) oxygenase family protein
DN1154_c2_g1 i6	tr|A0A0A7E9L1_PINRA *Pinus radiata*	*COMT*, caffeic acid 3-O-methyltransferase
DN33_c2_g1 i1	sp|CAMT_PINTA *Pinus taeda*	*CCoAOMT1*, caffeoyl-CoA O-methyltransferase 1
DN34_	XP_024928927.1	O-methyltransferase

This article is protected by copyright. All rights reserved.

Accepted Article

c0_g1 i1	*Ziziphus jujube*	
DN3704_c3_g1 i1	tr|A0A0D3CES4_BRAOL *Brassica oleracea*	elongation factor
DN3802_c0_g1 i1	tr|A9NK29_PICSI *Picea sitchensis*	*PSBR*, photosystem II 10 kDa polypeptide
DN6885_c0_g1 i1	XP_028103945.1 *Camellia sinensis*	*CUL1*, cullin 1
DN9747_c0_g1 i1	tr|V4T2T7_CITCL *Citrus clementina*	*PDI*, protein disulfide isomerase
Up-regulated in *P. sylvestris* in relation to *P. mugo* and *P. uncinata*		
DN10876_c0_g2 i1	tr|A9NV57_PICSI *Picea sitchensis*	*ALDH1A2*, aldehyde dehydrogenase
DN19629_c0_g1	XP_023522297.1 *Cucurbita pepo*	*ALDH11A3*, aldehyde dehydrogenase

This article is protected by copyright. All rights reserved.

Accepted Article

i1		
DN3026_c0_g1 i2	sp|C72B1_PINTA *Pinus taeda*	abietadienol/abietadienal
Up-regulated in *P. mugo* in relation to *P. uncinata*		
DN1010_c4_g1 i2	tr|B8LLS6_PICSI *Picea sitchensis*	*OSI_37124*, receptor-like serine threonine-protein kinase
DN11841_c0_g1 i2	XP_023921427.1 *Quercus suber*	*ALDH2C4*, aldehyde dehydrogenase family 2 member C4-like
DN12723_c0_g1 i1	tr|A0A5B9T6I9_ARAAG *Araucaria angustifolia*	*EEF1A1*, elongation factor-1 alpha
DN12883_c0_g1 i3	sp|3CAR1_PICAB *Picea abies*	terpene synthase, N-terminal domain
DN1973_c0_g1 i8	tr|A9NU81_PICSI *Picea sitchensis*	*PRX20*, peroxidase

This article is protected by copyright. All rights reserved.

Accepted Article

DN28253_c1_g1 i1	tr|A0A7J7M4R7_9MAGN *Kingdonia uniflora*	fatty acid desaturase
DN7225_c3_g1 i1	tr|Q5ME93_PSEMZ *Pseudotsuga menziesii*	*EEF1A1*, elongation factor-1 alpha
DN7303_c0_g1 i3	tr|A0A2K3K6A4_TRIPR *Trifolium pratense*	*HA1*, plasma membrane H^+-ATPase
DN9372_c0_g1 i10	sp|TPSD1_PINBN *Pinus banksiana*	*CPS1*, diphosphate synthase
Up-regulated in *P. uncinata* in relation to *P. mugo*		
DN1352_c0_g1 i10	tr|A9P2M9_PICSI *Picea sitchensis*	cysteine-rich repeat secretory protein
DN15760_c0_g1 i1	tr|A9NT90_PICSI *Picea sitchensis*	cysteine-rich repeat secretory protein

This article is protected by copyright. All rights reserved.

Accepted Article

DN17625_c0_g1 i1	tr|A7Y7E4_STYHA *Stylosanthes hamata*	*HSP90-1*, heat shock protein
DN22931_c1_g1 i1	tr|A0A443PX63_9MAGN*Cinnamomum micranthum*	stem-specific protein
DN4099_c0_g2 i1	tr|A0A0K0M729_PINTB *Pinus tabuliformis*	*OSI_37124*, receptor-like serine threonine-protein kinase
DN45522_c0_g1 i4	tr|A9NME8_PICSI *Picea sitchensis*	*RPL18*, 60S ribosomal protein L18
DN595_c1_g1 i1	XP_030442509.1 *Syzygium oleosum*	converts alpha-aldose to the beta-anomer, active on D-glucose, L-arabinose, D-xylose, D-galactose, maltose and lactose (by similarity)
DN59675_c0_g1 i5	XP_024931526.1 *Ziziphus jujube*	cysteine-rich receptor-like protein kinase
DN9161_c1_g1	tr| A0A7J7C285_TRIWF *Tripterygium wil-*	*HSP90-1*, heat shock protein

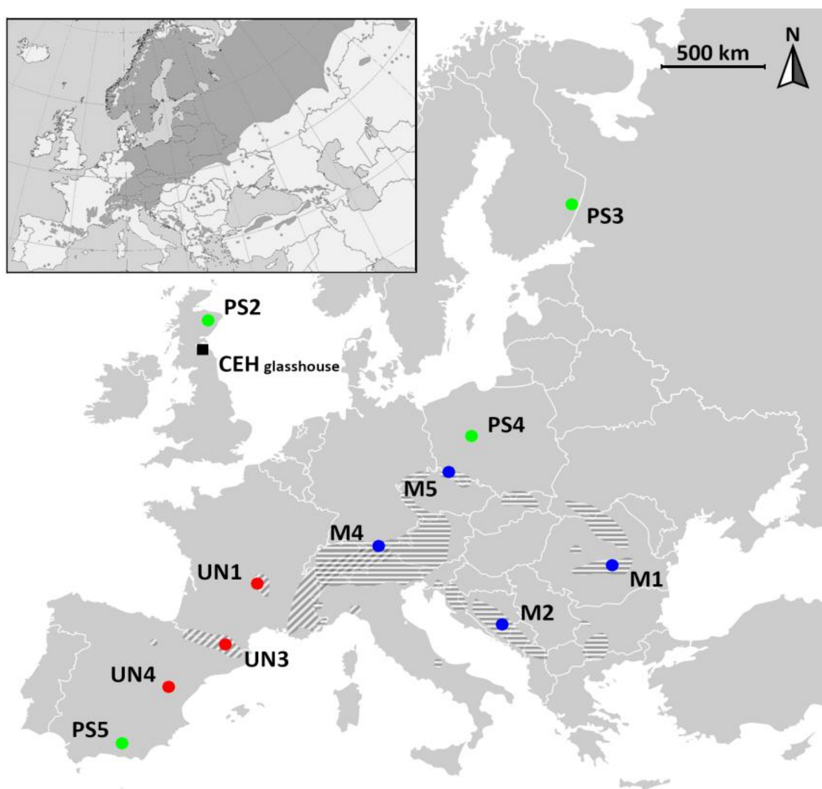
This article is protected by copyright. All rights reserved.

Accepted Article

i1	*fordii*

FIGURES

Fig. 1. Geographic distribution and sample collection sites of three investigated pine species. Blue dots and horizontal shading represent *P. mugo* (M) sites and range; red dots and diagonal shading correspond to *P. uncinata* (UN) locations; green points and dark shading on the inset represent *P. sylvestris* (PS) sample locations and European part of its range. Black square marks localization of the glasshouse where the seedlings were grown. The mountain pines' distribution map was created by the authors based on the Empty Political Map of Europe iso3166-1 (www.commons.wikimedia.org) and ranges taken from Jalas and Suominen (1973). The inset map has been prepared by EUFORGEN (www.euforgen.org) and adapted by authors.



This article is protected by copyright. All rights reserved.

Accepted Article

Fig. 2. Representation of 23,425 protein coding genes from *MUS assembly* transcriptome compared to two sets of the BUSCOs - present in green plants (upper panel) and in land plants (lower panel). Only one isoform, the longest, per gene was involved in the test.

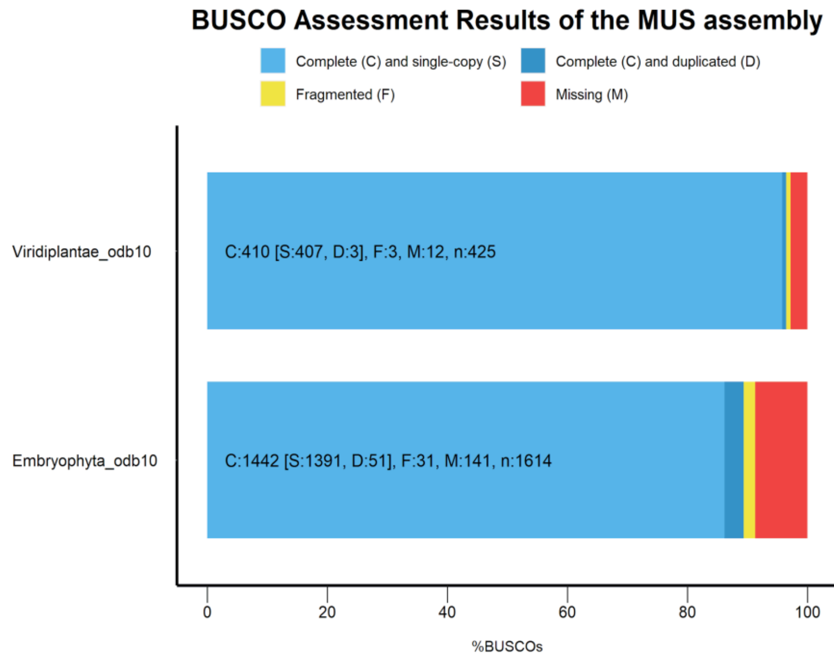
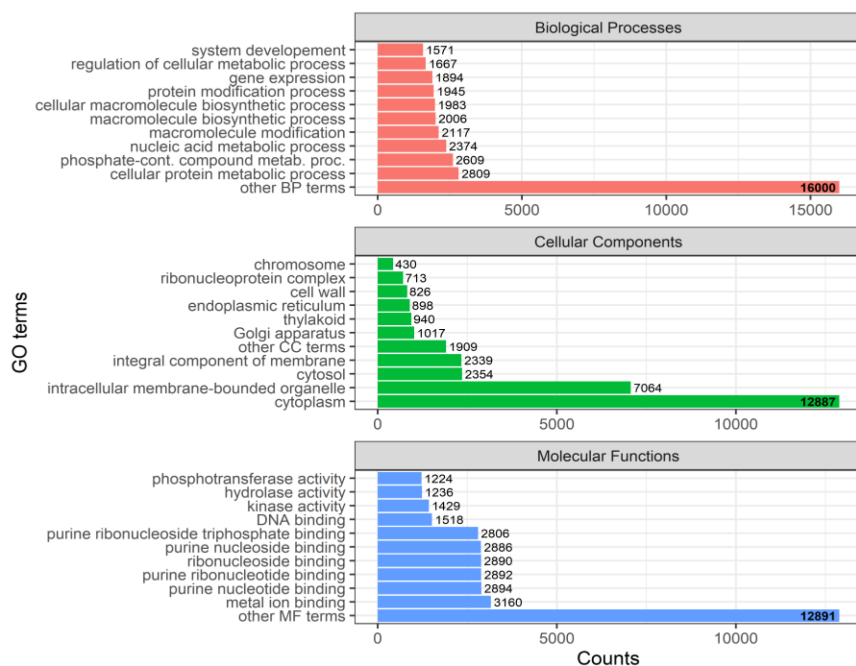


Fig. 3. The most frequently reported terms of level 4 in each main gene ontology (GO) aspect, as found in the annotated isoforms of the *MUS assembly*.



This article is protected by copyright. All rights reserved.

Accepted Article

Fig. 4. Venn diagrams of results from different analysis tools used in search of differentially expressed gene models in the *MUS assembly*. **A**) Discriminating markers found between two mountain pines and Scots pine ('broad' M-UN *vs* PS comparison); **B**) Discriminating markers found between dwarf mountain pine and Pyrenean pine (M *vs* UN).

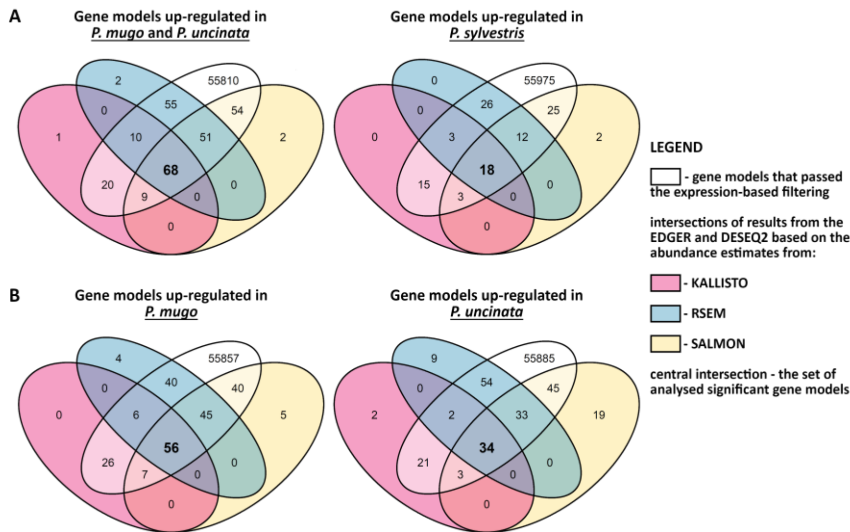
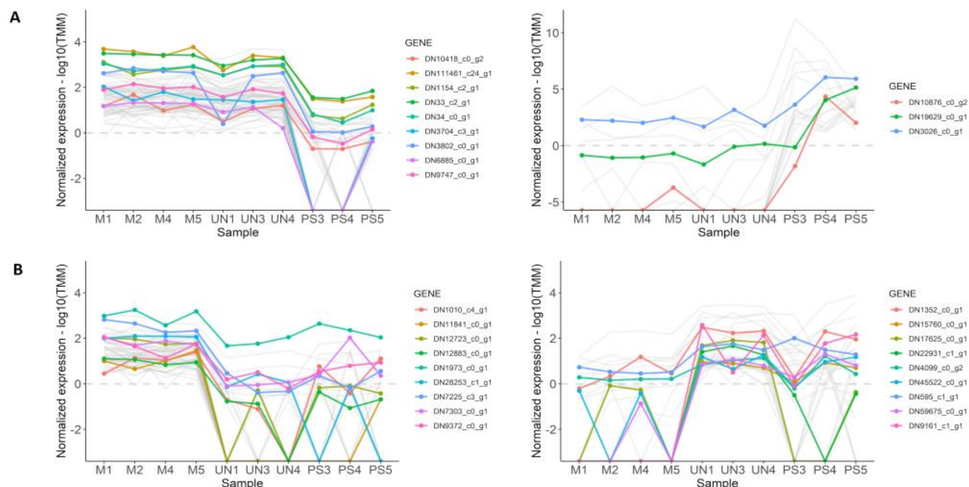


Fig. 5. Expression profiles of differentially expressed gene models, the successfully annotated ones are marked with colors. **A**) Significant markers found in M-UN vs PS pair, up-regulated in both mountain pines (left) or in Scots pine (right); **B**) Markers significantly differentiating two mountain pines, up-regulated in dwarf pine or in Pyrenean pine (left and right, respectively).

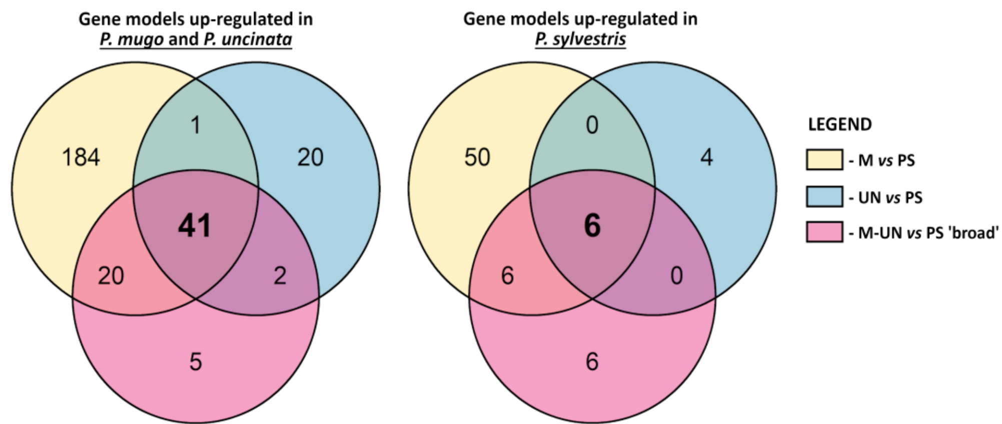


This article is protected by copyright. All rights reserved.

Accepted Article

Accepted Article

Fig. 6. Venn diagrams of results from three complementary comparisons of differentially expressed genes between two mountain pine taxa and Scots pine.



This article is protected by copyright. All rights reserved.