UNIWERSYTET IM. ADAMA MICKIEWICZA W POZNANIU

WYDZIAŁ BIOLOGII

**Julia Zaborowska**

# POPULATION HISTORY AND SELECTION IN CLOSELY RELATED MOUNTAIN PINE SPECIES (GENUS *Pinus*) IN EUROPE

*Historia populacji i procesy selekcyjne blisko spokrewnionych gatunków sosen górskich (rodzaj Pinus) w Europie*

**Rozprawa doktorska**

**Poznań, lipiec 2022**

**PROMOTOR**

**Prof. dr hab. Witold Wachowiak**

Zakład Ekologii Roślin i Ochrony Środowiska

Instytut Biologii Środowiska, Wydział Biologii

Uniwersytet im. Adama Mickiewicza w Poznaniu

**RECENZENCI**

**Dr hab. Ada Wróblewska, prof. nadzw. UwB**

Katedra Biologii i Ekologii Roślin, Wydział Biologii

Uniwersytet w Białymstoku

**Prof. dr hab. Jakub Sawicki**

Katedra Botaniki i Ochrony Przyrody, Wydział Biologii i Biotechnologii

Uniwersytet Warmińsko-Mazurski w Olsztynie

Praca zastała wykonana

w **Pracowni Biologii Ewolucyjnej**

oraz **Zakładzie Ekologii Roślin i Ochrony Środowiska**

Instytutu Biologii Środowiska, na Wydziale Biologii

Uniwersytetu im. Adama Mickiewicza w Poznaniu

## PODZIĘKOWANIA

Pragnę podziękować wszystkim osobom, z którymi bezpośrednio lub pośrednio współpracowałam przy badaniach wchodzących w skład niniejszej rozprawy oraz wszystkim tym, którzy w wieloletnim okresie trwania moich studiów udzielali mi różnego rodzaju pomocy i wsparcia.

Przede wszystkim chciałabym w tym miejscu wyrazić swą wdzięczność dla

promotora pracy,

**Prof. dr. hab. Witolda Wachowiaka**,

współautorów artykułów wchodzących w skład pracy

**Dr Anniki Perry** i **Dr. Stephena Caversa**,

przyjaciół i współpracowników

z **Pracowni Biologii Ewolucyjnej**, **Zakładu Ekologii Roślin i Ochrony Środowiska** i innych jednostek **Wydziału Biologii UAM**, spośród nich w szczególności

**Dr. Bartosza Łabiszaka** i **Mgr. Jarosława Raubica**,

oraz dla moich najbliższych,

**Rodziny** i **Ukochanego**.

# FINANSOWANIE

Wykonanie niniejszej pracy odbyło się w ramach realizacji projektu badawczego nr **2015/19/B/NZ9/00024**, pt. „*Genomika środowiskowa wybranych gatunków sosen (rodzaj Pinus) w Europie*", przyznanego prof. dr. hab. Witoldowi Wachowiakowi w ramach konkursu OPUS 10, finansowanego ze środków **Narodowego Centrum Nauki**, w którym autorka była zatrudniona w roli wykonawcy na stanowisku doktoranta-stypendysty. Autorka uzyskała również środki finansowe na ukończenie rozprawy doktorskiej w ramach projektu „*Paszport do przyszłości – Interdyscyplinarne studia doktoranckie na Wydziale Biologii UAM*" (**POWR.03.02.00-00-I006/17**); stypendium to było współfinansowane ze środków **Uniwersytetu im. Adama Mickiewicza w Poznaniu** oraz **Europejskiego Funduszu Społecznego**.

W badaniach przedstawionych w niniejszej pracy wykorzystano również materiał badawczy i/ lub dane uzyskane przy wsparciu następujących projektów:

- **2017/27/B/NZ9/00159** pt. *Rewizja wzorów migracji-selekcji sosny zwyczajnej (Pinus sylvestris L.) z wykorzystaniem nowatorskich podejść genomicznych*; projekt finansowany ze środków **Narodowego Centrum Nauki**

- **NE/K012177/1** pt. *Population genomics and evolution of adaptive traits in Pines*; projekt finansowany ze środków **Natural Environment Research Council, UK**

- **BB/L012243/1** pt. *Promoting resilience of UK tree species to novel pests and pathogens: ecological and evolutionary solutions*; projekt finansowany ze środków **Biotechnology and Biological Sciences Research Council**, **UK**

## ABSTRACT

The dwarf mountain pine (*Pinus mugo* Turra) and the Pyrenean pine (*P. uncinata* Ramond ex DC.) form a pair of sister coniferous taxa inhabiting stands scattered over subalpine environments of European mountain ranges that have poorly resolved evolutionary history and affinity. Together with another closely related, rather lowland reference species, the Scots pine (*Pinus sylvestris* L.), they constitute an interesting system for comparative examination of the adaptations developed by trees in high-altitude environments. European mountain plants are challenged, among other demanding stimuli and factors, by increased solar radiation, decrease in atmospheric pressure and ambient temperatures, vegetation time contractions and often dynamic weather conditions. As long-lived phanerophytes, the two pines must have adapted to balance needs of growth, reproduction, survival, and regeneration. However, knowledge of the molecular basis of their adaptations to higher altitudes remain scarce. The dissertation represents a trial attempting to identify such corresponding genomic regions of high importance in evolution of the two European subalpine pines. The first aim of the study was to unfold their phylogeography by analysis of mutual relationships between and within a probe of 27 natural populations. Reconstructions based on a set of newly developed mitochondrial DNA markers shed a light on their current population genetic structure and possible transmission routes in the past. The next steps of the study focused on the nuclear genome and its putative adaptive variation. First, 438 samples of three species were genotyped with thousands of single nucleotide polymorphism (SNP) markers, and the genetic diversity and differentiation of species and populations were tested, followed by outlier SNP detection and ontology annotations of genes. Markers clearly differentiated the species and uncovered patterns of their population structure. Comparisons with Scots pine, allowed identification of 35 candidates for altitude-dependent selection in mountain pines, including genes encoding proteins responsible for photosynthesis, photorespiration, cell redox homeostasis, regulation of transcription, and mRNA processing. Slightly more, 75 outlier SNPs, differentiated the two sister pines; they were found mainly among genes related to gene expression and metabolism. The third part of the research focused on the relevance of the gene expression regulation in process of adaptation to mountain conditions. *De novo* transcriptome was assembled for

corresponding representatives of the taxa, each from different natural population but grown under uniform glasshouse conditions. The examined gene expression profiles appeared generally similar, allowing identification of 121 significantly diverged markers in all of the pairwise species comparisons. Functional annotation revealed major categories of distinctly expressed transcripts, including: wood trait properties and response to abiotic factors such as oxidative stress, salinity, drought, and temperature. Many of the genes constitute strong candidates for adaptation of *P. mugo* and *P. uncinata* to their demanding habitats. For instance, up-regulation of genes involved in lignin biosynthesis might be linked to strong winds and thick snow cover experienced by these pines. Overall, the presented research provides new knowledge about the drivers of background and adaptive genetic variation in the investigated pine species and advance our understanding of demographic and evolutionary processes that underlay they ecological divergence.

**KEYWORDS**

candidate genes, high-altitude adaptations, phylogeography, *Pinus mugo*, *Pinus uncinata*

## STRESZCZENIE

Sosna górska (*Pinus mugo* Turra) i sosna hakowata (*P. uncinata* Ramond ex DC.) stanowią parę taksonów siostrzanych o słabo rozpoznanej historii ewolucyjnej i relacjach pokrewieństwa. Te dwa gatunki roślin iglastych zasiedlają podobne stanowiska rozrzucone w subalpejskim paśmie europejskich masywów górskich. Sosny te, razem z innym blisko spokrewnionym, przeważnie nizinnym gatunkiem referencyjnym – sosną zwyczajną (*Pinus sylvestris* L.), stanowią dogodny układ porównawczy w badaniach genetycznych podstaw przystosowań drzew leśnych do warunków wysokogórskich. Dla roślinności górskiej wyzwaniami środowiskowymi jest m.in. wzmożone promieniowanie słoneczne, niższe ciśnienie atmosferyczne i obniżona temperatura otoczenia, a w warunkach europejskich także silne skrócenie okresu wegetacji i często dynamiczne zmiany warunków pogodowych. Jako długowieczne fanerofity, dwie omawiane sosny górskie musiały przystosować się w sposób umożliwiający im zrównoważenie nakładów potrzebnych na wzrost, reprodukcję, regenerację i przetrwanie niekorzystnych warunków. Jednakże molekularne podstawy przystosowania tych, i innych, roślin do warunków górskich pozostają wciąż słabo poznane. Niniejsza praca doktorska stanowi próbę identyfikacji regionów genomowych mogących odpowiadać za ewolucję i przystosowanie sosny górskiej oraz sosny hakowatej do ich subalpejskich siedlisk. Pierwszym etapem badań były analizy filogeograficzne dotyczące wzajemnych relacji genetycznych pomiędzy i w obrębie 27 naturalnych populacji tych taksonów. Rekonstrukcje oparte na polimorfizmie nowo opracowanych markerów mitochondrialnego DNA rzuciły światło na obecną strukturę genetyczną populacji oraz na możliwe historyczne drogi ich migracji. Kolejne etapy pracy koncentrowały się na zmienności genomu jądrowego i potencjalnej zmienności adaptacyjnej badanych gatunków. W badaniach wykorzystano dane genotypowe dla 438 prób, pozyskane z użyciem macierzy zawierającej kilkadziesiąt tysięcy markerów typu SNP (pojedynczych miejsc zmiennych) oraz przetestowano ich zmienność pomiędzy poszczególnymi gatunkami i populacjami. Przeszukano zestaw tych markerów pod kątem istotnych statystycznie odchyleń częstości alleli od częstości obserwowanych dla tła genetycznego, i dokonano przypisania prawdopodobnych funkcji genów, z których pochodziły te markery. Uzyskane dane pozwoliły na jednoznaczne rozróżnienie poszczególnych taksonów

oraz ujawniły istnienie struktury populacji u dwóch gatunków. Analizy porównawcze z sosną zwyczajną, pozwoliły na identyfikację 35 genów kandydackich, odróżniających *P. mugo* i *P. uncinata* od taksonu referencyjnego i potencjalnie związanych z kształtowaniem ich przystosowań do warunków górskich. Wśród nich znalazły się geny kodujących białka odpowiedzialne za fotosyntezę, fotorespirację, gospodarkę reaktywnymi formami tlenu (homeostazę redoks), a także regulację transkrypcji i edycję mRNA. Nieco więcej, bo 75 markerów SNP, różnicowało dwa taksony siostrzane. Te polimorfizmy obecne były głównie wśród genów związanych z ekspresją i metabolizmem. W ostatnim etapie badań weryfikacji poddano wpływ regulacji transkrypcji na proces adaptacji tych gatunków do warunków górskich. Na podstawie analizy całkowitego RNA osobników uzyskanych z nasion zebranych w naturalnych populacjach i hodowanych w jednakowych warunkach szklarniowych, dokonano składania *de novo* transkryptomu referencyjnego, a następnie analizy ekspresji poszczególnych genów. Profile ekspresji okazały się zasadniczo podobne – zidentyfikowano łącznie zaledwie 121 transkryptów o istotnym zróżnicowaniu pomiędzy którąkolwiek z par taksonów. Przypisane na podstawie oryginalnych sekwencji funkcje tych genów należały do takich kategorii jak: modyfikacja drewna oraz odpowiedź na stres wywołany przez różne czynniki biotyczne i abiotyczne, w tym stres oksydacyjny. Wiele z wykrytych genów jest dobrymi kandydatami potencjalnie związanymi z adaptacją badanych sosen górskich do wymagań ich siedliska. Przykładowo, regulacja ekspresji genów zaangażowanych w biosyntezę ligniny może być powiązana z silnymi wiatrami i grubą pokrywą śnieżną, której doświadczają te sosny w swoim środowisku. Podsumowując, przedstawione badania dostarczają nowej wiedzy na temat procesów kształtujących neutralną i adaptacyjną zmienność genetyczną badanych gatunków sosen oraz pogłębiają naszą wiedzę na temat procesów demograficznych i ewolucyjnych, które leżą u podstaw ich dywergencji ekologicznej.

**SŁOWA KLUCZOWE**

adaptacje wysokogórskie, filogeografia, geny kandydackie, *Pinus mugo*, *Pinus uncinata*

**SPIS TREŚCI**

# 1. WSTĘP

## Tło badawcze

Zmienność tła genetycznego populacji oraz zmienność adaptacyjna związana z dostosowaniem organizmów do określonych warunków środowiska jest wypadkową procesów związanych z historią populacji i ich ewolucją. Drzewa stanowią istotny element ekosystemów lądowych, oddziałując na nie poprzez interakcje biotyczne oraz procesy związane z regulacją klimatu, obiegu wody czy struktury gleby. Jednakże, genetyczne podstawy kształtowania się zasięgów i charakterystyk populacji drzew leśnych w wyniku selekcji do lokalnych warunków środowiska oraz zwykle wysokiej i odziedziczalnej zmienności fenotypowej, to procesy wciąż słabo poznane. Powstawanie i utrzymywanie się lokalnych adaptacji u tych organizmów jest szczególnie ciekawe ze względu na ich specyficzne cechy biologiczne, takie jak duża efektywna wielkości populacji, rozmnażanie krzyżowe, czy znaczne możliwości dyspersyjne pyłku i nasion (Petit i Hampe 2006). Wszystkie te cechy przyczyniają się do szybkiej homogenizacji tła genetycznego między populacjami. Jednocześnie drzewa cechuje zwykle wysoka rozrodczość i bardzo silna selekcja pomiędzy rekrutowanymi osobnikami, co ułatwia im dostosowywanie się do miejscowych warunków środowiska i utrzymywania zmienności ważnych cech adaptacyjnych pomimo możliwego intensywnego przepływu genów w zasięgu ich występowania (Sork 2016).

Procesy kształtujące zmiany zasięgów i relacji między populacjami roślin drzewiastych, a także ich zwykle wysokiej, odziedziczalnej zmienności fenotypowej są wciąż słabo poznane. Z jednej strony, drzewa są dalece niepodobne do tzw. organizmów modelowych, ich czas trwania pokolenia jest bardzo wydłużony, charakteryzuje je też wolny rozwój osobniczy, co utrudnia ich badania w tradycyjnych eksperymentach hodowlanych. Z drugiej strony, drzewa leśne tworzą zazwyczaj duże, krzyżujące się losowo populacje, w których zmienność genetyczna utrwalona przez pokolenia jest wypadkową procesów neutralnych, związanych z historią populacji, i selekcyjnych w zasięgu ich występowania. Tym samym tworzą dogodny układ eksperymentalny w analizach genetycznych podstaw ich zmienności ekotypowej, w tym poszukiwania sygnatur działania doboru na poziomie

molekularnym. Los generowanego i segregującego polimorfizmu DNA zależy od procesów neutralnych takich jak dryf genetyczny i migracje, oddziałujących na całość genomu jednakowo, oraz od procesów selekcyjnych, zazwyczaj działających na genom punktowo – w odniesieniu do miejsc funkcjonalnych oraz ich najbliższego otoczenia. Co istotne, ewolucja adaptacyjna nie jest wynikiem jedynie występowania i segregacji różnych form allelicznych genów, ale może być kształtowana także m.in. przez różnice w poziomach ekspresji czy odmienne sposoby składania (ang. *alternative splicing*) poszczególnych transkryptów (Jones i in. 2012; Smith i in. 2018).

W przypadku drzew leśnych, interesującym polem badawczym do zgłębiania różnego typu procesów mikroewolucyjnych, np. genezy i trwałości adaptacji, kształtowania się zasięgów populacji i postępującej przy tym specjacji bądź homogenizacji ich pul genowych, są gatunki przystosowane do warunków górskich. Ze względu na, zazwyczaj, wyspowy charakter ich występowania, który utrudnia wymianę materiału genetycznego, u roślin tych obserwuje się ograniczoną zmienność wewnątrz i duże zróżnicowanie pomiędzy populacjami (Ohsawa i in. 2008). Z drugiej strony, ze względu na specyfikę siedliska, populacje drzew górskich poddawane są gwałtownym zmianom czynników środowiskowych w stosunkowo niewielkim gradiencie wysokościowym, co przyczynia się do różnicowania osobników nawet w zasięgu jednej populacji. Wzrost wysokości bezwzględnej wiąże się ze spadkiem ciśnienia atmosferycznego (w tym ciśnienia cząsteczkowego $CO_2$) i średniej temperatury otoczenia, a także ze wzrostem siły promieniowania słonecznego. Taki gradient wysokościowy powoduje występowanie w licznych pasmach górskich wyraźnych pięter klimatycznych. W warunkach Europejskich, z zarysowanymi sezonami, okres wegetacji jest wyraźnie skrócony w górach, co związane jest z niższa średnią temperatur i dłużej zalegająca pokrywą śnieżną. Tereny te charakteryzują się także częstszymi i silniejszymi porywami wiatru, zwiększającymi presję mechaniczną, ale także wpływającymi na rozmnażanie roślin, zwłaszcza gatunków wiatropylnych. Dodatkowo, na wzniesieniach występują różnice siedliskowe pomiędzy stokami południowymi i północnymi, co w przypadku roślin światłolubnych istotnie ogranicza dostępność miejsc wzrostu. Na wzrost i rozwój populacji rośli wpływa również rzeźba terenu. Każdy z tego typu czynników środowiskowych silnie oddziałuje na osobniki i populacje, regulując ich śmiertelność i rozrodczość, a zatem wpływając na ich zdolności adaptacyjne.

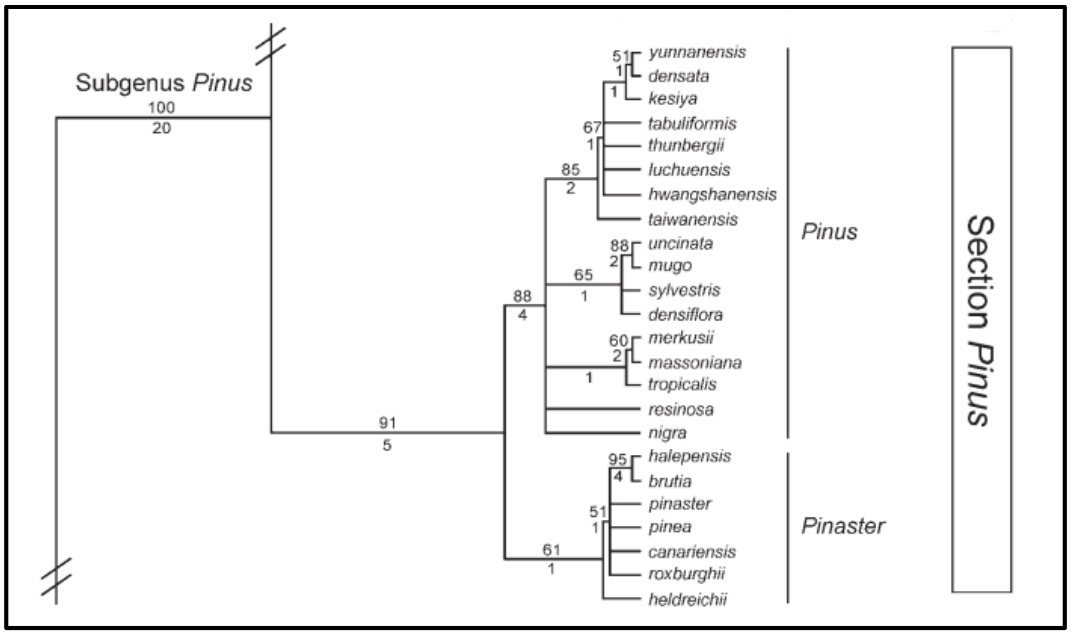
Jak dotąd jednak wciąż niewiele wiadomo na temat genetycznego podłoża przystosowań roślin, w tym drzew leśnych, do tych specyficznych wymagań środowiskowych.

**Przedmiot badań**

Prezentowana praca doktorska została poświęcona dwóm siostrzanym taksonom europejskich sosen. Pierwsza z nich – *Pinus mugo* Turra (1764), widnieje w polskiej nomenklaturze pod oficjalną nazwą sosny górskiej, a w literaturze także pod synonimami takimi jak kosodrzewina, kosodrzew, czy kosówka właściwa. W literaturze anglojęzycznej najczęściej spotykana jest pod nazwami: dwarf pine, dwarf mountain pine, mugo pine, czy scrub mountain pine. Drugi takson – *Pinus uncinata* Ramond ex DC. (1805), w polskiej nomenklaturze botanicznej funkcjonuje jako sosna hakowata, w anglojęzycznej pod synonimami: Pyrenean pine, mountain pine albo Swiss mountain pine. Mimo, że gatunki te różnią się dość wyraźnie i pod wieloma względami zostały już dobrze zbadane, ich pozycja taksonomiczna jest nadal niejasna. Zwykle zaliczane są wspólnie do szerszego kompleksu *Pinus mugo* i traktowane albo jako dwa gatunki – *P. mugo* (Turra) i *P. uncinata* (Ramond), albo ujmowane w randze podgatunku wewnątrz gatunku *Pinus mugo* jako odpowiednio *P. mugo s.s.* i *P. mugo* subsp. *uncinata*. To drugie, późniejsze ujęcie odwołuje się do biologicznej koncepcji gatunku Ernsta Mayra (1942), uwzględniając możliwość krzyżowania się tych taksonów w populacjach naturalnych. Istnieją też inne propozycje usystematyzowania tych taksonów, przykładowo: *Pinus mugo sensu stricto* i *P. uncinata* subsp. *uncinata* (*sensu* Businsky i Kirschner 2010). Niemniej, zróżnicowanie fenotypowe i ekologiczne oraz zasadniczo allopatryczny zasięg występowania, przemawiają za traktowaniem ich jako dwa osobne gatunki. Należy w tym miejscu zaznaczyć, że inny europejski takson z rodzaju *Pinus* – sosna limba (*P. cembra* L.), który również związany jest ze środowiskiem wysokogórskim, nie wchodził zakres badań niniejszego doktoratu.

Sam kompleks *Pinus mugo*, do którego powszechnie zalicza się dwie omawiane sosny, to niewielka grupa taksonomiczna, w literaturze anglojęzycznej znajdowana pod nazwami '*Pinus mugo* complex' lub '*Pinus mugo* aggregate'. Uproszczona systematyka klasyfikuje ten kompleks wśród roślin nagonasiennych, gdzie zaliczany jest do klasy iglastych (Pinopsida), rzędu sosnowców (Pinales), rodziny sosnowatych (Pinaceae) i rodzaju sosna (*Pinus*).

W ramach rodzaju, który liczy blisko 109 gatunków (Farjon 2001), dość dobrze rozpoznanych i uporządkowanych pod względem filogenezy, kompleks sytuowany jest w podrodzaju *Pinus* (*Diploxylon*, w lit. angielskiej także jako 'hard pines'), w sekcji *Pinus* i podsekcji *Pinus* (Gernandt i in. 2005; Ryc.1). Według dotychczasowych badań biometrycznych, biochemicznych i molekularnych (Gernandt i in. 2005; Grotkopp i in. 2004;) najbliżej spokrewniona z kosodrzewiną i sosną hakowatą jest sosna zwyczajna (*P. sylvestris* L.), a dalej sosna gęstokwiatowa (*P. densiflora* Siebold & Zucc.). Początek rozchodzenia się linii ewolucyjnych sosny zwyczajnej i omawianych taksonów szacowany jest na ok. 5 mln lat p.n.e. (Monteleone i in. 2006; Wachowiak i in. 2011).



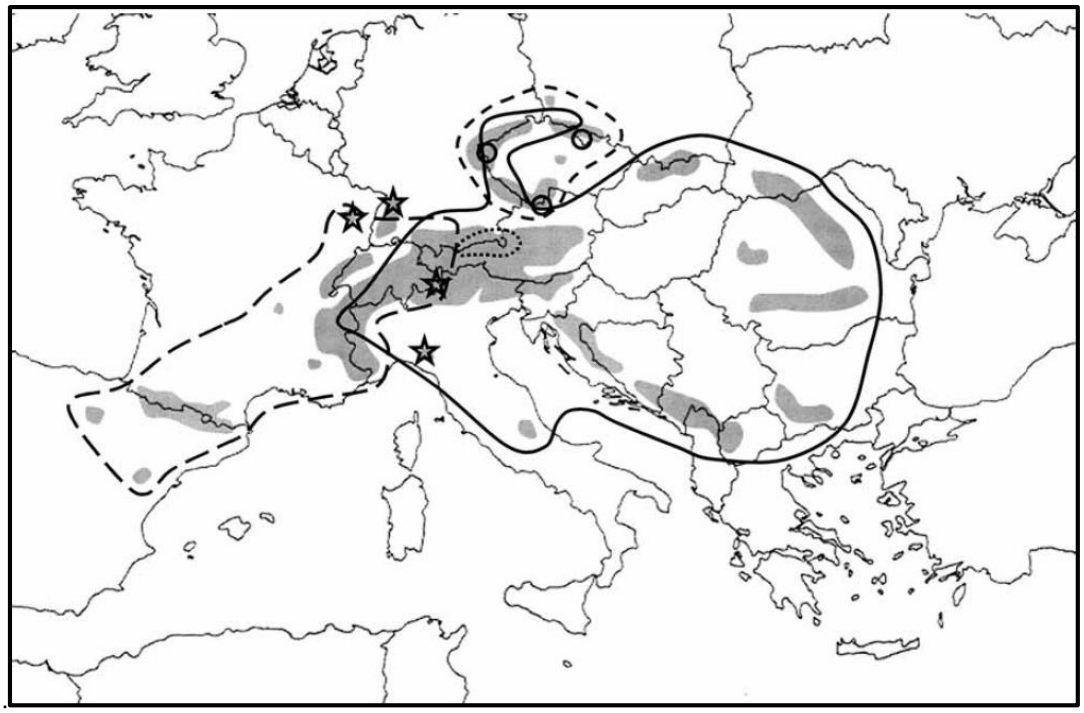

**Ryc. 1.** Pokrewieństwo gatunków z sekcji *Pinus* podgatunku *Pinus*, ukazujące bliskie relacje między omawianymi taksonami górskimi oraz sosną zwyczajną. Drzewo filogenetyczne zostało zrekonstruowane na podstawie drzewa konsensusowego otrzymanego z loci *rbcL* i *matK* chloroplastowego DNA. Przedstawiona rycina jest fragmentem większego dendrogramu i pochodzi z publikacji Gernandt i in. (2005).

Do najbardziej aktualnych, wnikliwych i systematycznych przeglądów kompleksu *P. mugo* należy publikacja Hamernika i Musila z 2007 r., która jest podsumowaniem literatury dotyczącej nomenklatury stosowanej do jej opisu w ostatnim stuleciu, oraz Businsky-iego i Kirschnera (2010), która podsumowuje wieloletnie obserwacje i badania prowadzone przez nich na blisko 150 populacjach z tego kompleksu. Warto wspomnieć, że oprócz dość dobrze wyodrębnionych taksonów *P. mugo* i *P. uncinata*, do kompleksu zalicza się również szereg

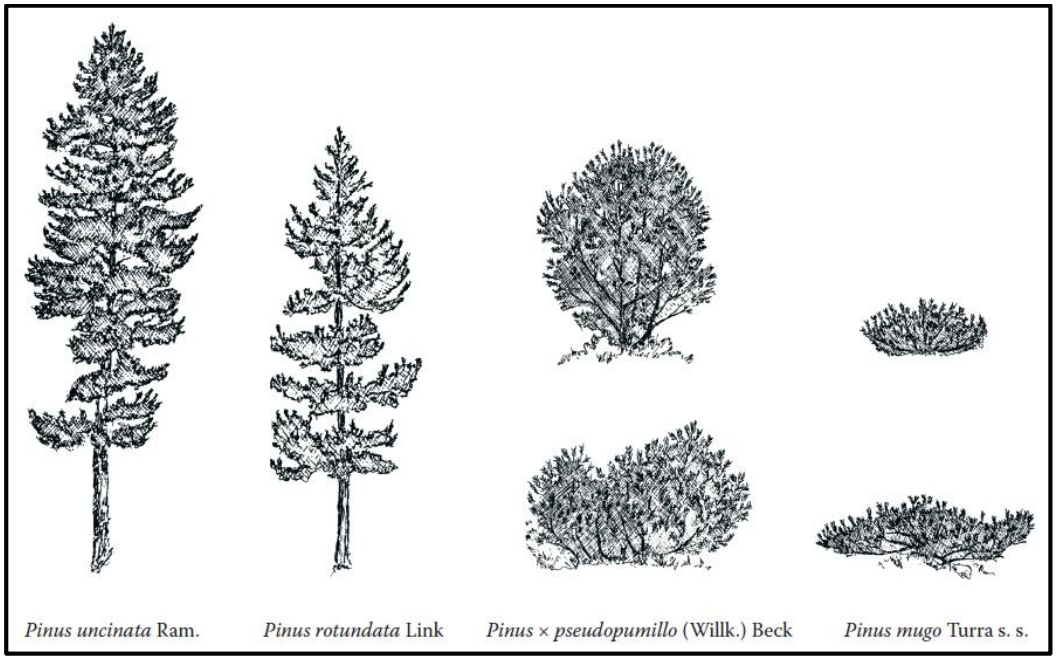
rzadszych i bardziej problematycznych grup, takich jak np. *P. uliginosa*, *P. rotundata*, czy *P.* × *pseudopumillo* (Hamernik i Musil 2007; Businsky i Kirschner 2010), które zajmują inne siedliska. Nie stanowiły one przedmiotu prezentowanych badań i nie zostały tu bliżej przedstawione.

Oba badane gatunki reprezentują europejską roślinność subalpejską, a w swoim zasięgu występowania nie przekraczają wysokości ok. 2700 m n.p.m. (Businsky i Kirschner, 2010). Na niższych wysokościach sosna hakowata tworzy mniej lub bardziej otwarte lasy, w wyższych położeniach jej zasięg wyznacza górną granicę lasu (ang. *timberline*), a także górną granicę obecności drzew (ang. *treeline*). Na podobnych wysokościach kosodrzewina tworzy zwykle grube, krzewiaste dywany. Oba gatunki dobrze radzą sobie na trudnych, często inicjalnych podłożach. Zasiedlają skaliste zbocza i rumowiska o bardzo zróżnicowanej genezie i właściwościach – zarówno podłoża magmowe, metamorficzne, jak i osadowe (m.in. granity, porfiry, gnejsy, wapienie, gipsy, kwarce, serpentyny, czy łupki; Businsky i Kirschner 2010; Monteleone i in. 2006). Taksony te są światłożądne i preferują wystawy południowe, jednocześnie mają ciemniejszą barwę igieł (np. w porównaniu z sosną zwyczajną), co świadczyć może o ich przystosowaniu do silniejszego oddziaływania promieniowania UV w wyższych położeniach.

Znanych jest wiele cech różnicujących sosnę górską i hakowatą. Pierwotnie podkreślano osobne występowanie taksonów, ze stosunkowo niewielkim obszarem nakładania się ich zasięgów (Ryc. 2 na następnej stronie) oraz widoczne różnice fenotypowe. Z czasem opisywano bardziej subtelne różnice anatomiczne i morfologiczne, wreszcie także molekularne i fenologiczne. Businsky i Kirschner (2010) wymieniają dwie zasadnicze cechy, po których należy je rozróżniać – pokrój rośliny i budowę szyszek, które to cechy jednak nie zawsze są w pełni skorelowane. U sosny górskiej pokrój jest karłowaty, przez to odmienny niż u większości przedstawicieli rodzaju. Wykształca ona liczne płożące pędy, wzniesione zwykle nie wyżej niż na 2-3 m nad ziemię. Z tego względu, zgodnie z systemem form życiowych roślin, zaliczana powinna być raczej do nano- lub mikrofanerofitów niż drzew, jednak w literaturze takie jej traktowanie jest powszechne (Alexandrov i in. 2019; Ballian i in. 2016; Willis i in. 2004).



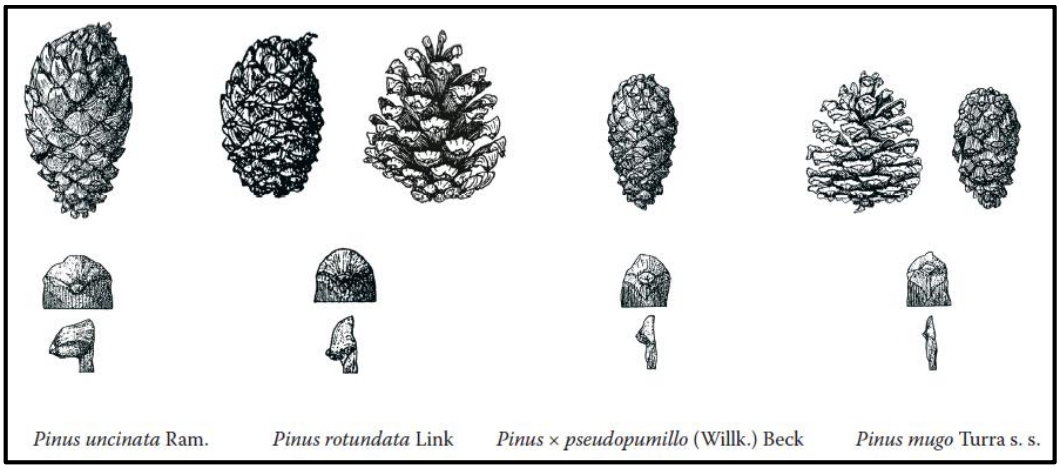
**Ryc. 2.** Zasięgi występowania omawianych sosen górskich – linią ciągłą otoczone są populacje *Pinus mugo*, linią przerywaną, dłuższą populacje *P. uncinata*, linią przerywaną, krótszą populacje *P. uliginosa*; gwiazdkami oznaczono przypuszczalne miejsca hybrydyzacji sosny górskiej i sosny hakowatej. Rycina została zaczerpnięta z publikacji Businsky i Kirschner (2010).




**Ryc. 3.** Formy wzrostu przedstawicieli kompleksu *Pinus mugo*; sosna hakowata skrajnie po lewej, sosna górska skrajnie po prawej. Rycina zapożyczona z artykułu Hamernika i Musila (2007), a oryginalnie z publikacji Kindel-a (1995).

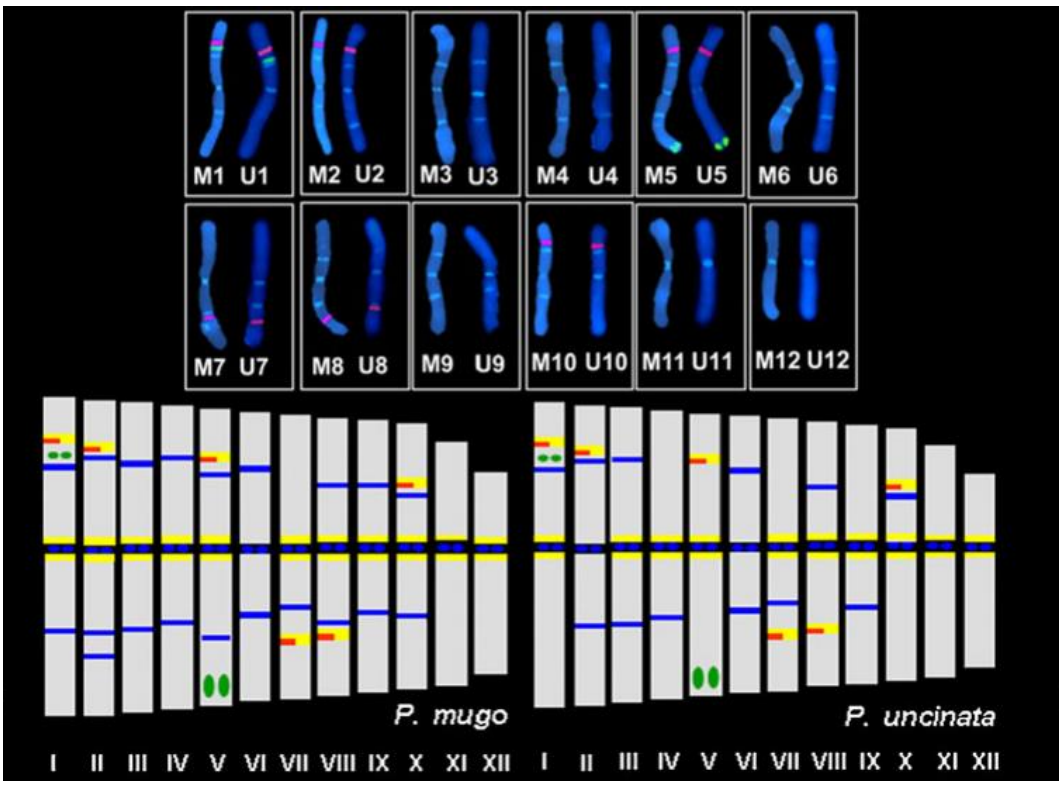
Karłowatość jest znaną formą adaptacji roślin do warunków górskich (Bomblies 2015), być może charakterystyczny pokrój kosodrzewiny jest właśnie jej przystosowaniem do porywistych wiatrów, niskich temperatur, czy długo zalegającej pokrywy śnieżnej. Dzięki zdolności ukorzeniania pędów (Jenik 1994), porasta podłoże tworząc zwarte kobierce (Ryc. 3). Sosna hakowata ma pokrój drzewiasty, ‘choinkowaty’, z jednym prostym pędem głównym sięgającym od kilkunastu do nawet 25 m wysokości i jajowato-stożkową koroną (Businsky i Kirschner 2010). Tworzy dość luźne lasy, szczególnie na większych wzniesieniach.

Szyszki żeńskie obu sosen rosną na krótkich szypułkach i rozwijają się podobnie – przez większość czasu są wyprostowane i skierowane ku górze, a wiosną drugiego roku dojrzewania charakterystycznie odginają się i zaczynają otwierać (Businsky i Kirschner 2010). Szyszki kosodrzewiny są jednak mniejsze i mają symetrię promienistą, apofizy łusek nasiennych są u niej płaskie i słabiej wykształcone, natomiast u sosny hakowatej szyszki mają jedną płaszczyznę symetrii – są zakrzywione, poza tym każda z apofiz jest znacznie bardziej uwydatniona i haczykowato wygięta (Ryc. 4). Gatunki różnią się nieco sposobem łuszczenia ich ciemnej, szarawo-brązowej do czarnej kory, która u sosny hakowatej odrywa się trochę większymi, zwijającymi się płatami (Businsky i Kirschner 2010).




**Ryc. 4.** Pokroje szyszek żeńskich u przedstawicieli kompleksu *Pinus mugo*; sosna hakowata skrajnie po lewej, sosna górska skrajnie po prawej. Dolny panel przedstawia różnice w kształcie apofizy łuski nasiennej – widok odpowiednio od strony grzbietowej (u góry) i z boku (u dołu). Rycina pochodzi z publikacji Hamernika i Musila (2007) i jest modyfikacją oryginalnych rycin Christensena (1987) i Kindel-a (1995).

W mniejszym stopniu gatunki różnicuje również anatomia igieł, w tym liczba kanałów żywicznych oraz budowa i rozmieszczenie komórek sklerenchymatycznych (Boratyńska i Boratyński 2007; Boratyńska i in. 2015; Businsky i Kirschner 2010). Hodowane w jednakowych warunkach eksperymentu szklarniowego, młodociane osobniki kosodrzewiny i sosny hakowatej wykazywały różnice w fenologii oraz tempie wzrostu (Wachowiak i in. 2018). Mimo mocno konserwatywnej organizacji zestawu chromosomów u sosen (2n=24), Bogunić i in. (2011) obserwowali rozbieżności we wzorach wybarwienia prążków heterochromatyny, świadczące o postępującym różnicowaniu taksonów również na poziomie kariotypu (Ryc. 5).




**Ryc. 5.** Zestawienie wzorców wybarwienia chromosomów metafazowych sosny górskiej (po lewej, M) i sosny hakowatej (po prawej, U), ukazujące zasadnicze podobieństwo kariotypów i niewielkie zróżnicowanie miejsc wybarwienia heterochromatyny (niebieskie prążki na schemacie). Więcej szczegółów w artykule Bogunić i in. (2011), z którego pochodzi rycina.

Natomiast dotychczasowe analizy biochemiczne i genetyczne wskazywały przede wszystkim na duże podobieństwo pomiędzy badanymi taksonami oraz wyraźny brak, w większości zastosowanych markerów, istotnych różnic. Relacje analizowane z wykorzystaniem markerów

izoenzymatycznych wykazały ich wysoką zmienność wewnątrz populacji, ale niewielką pomiędzy taksonami (Lewandowski i in. 2000). Na poziomie sekwencji DNA, żadne z licznych badań prowadzonych zarówno z wykorzystaniem rejonów jądrowych (Monteleone i in. 2006; Wachowiak i in. 2015), jak i cytoplazmatycznych (Dzialuk i in. 2017; Heuertz i in. 2010; Sokołowska i in. 2021), nie pozwoliło na wyłonienie jednoznacznych, specyficznych gatunkowo markerów. Większość z tych badań, mimo, że ujawniała pewne zróżnicowanie genetyczne istniejące pomiędzy taksonami, podkreślała fakt, że przeważająca część przebadanych alleli segreguje w ich populacjach z podobnymi częstościami.

Duże podobieństwo genetyczne badanych taksonów może być efektem relatywnie niedawnej dywergencji tych organizmów (poniżej 5 mln lat; Wachowiak i in. 2011), charakteryzujących się dodatkowo powolnym rozwojem i dojrzewaniem (ok. 10 lat u *P. mugo* i 15 -25 u *P. sylvestris*), a także długim czasem trwania pokolenia (20-25 lat), które przekładają się na niepełne rozdzielenie linii ewolucyjnych i wciąż segregującą w nich zmienność ancestralną (Ballian i in. 2016; Carlisle i Brown 1968; Mátyás i in. 2003; Wachowiak i in. 2011). Podobieństwo to może wynikać również ze zmian zasięgów tych gatunków, powodujących wtórny kontakt i hybrydyzację pomiędzy wcześniej izolowanymi populacjami. Takie fluktuacje zasięgów mogły wielokrotnie mieć miejsce na przestrzeni ostatnich kilku milionów lat, zwłaszcza w niestabilnym okresie Plejstoceńskich glacjałów i interglacjałów (Boratyńska i in. 2014). Współcześnie sympatryczny zasięg występowania kosodrzewiny i sosny hakowatej jest ograniczony przede wszystkim do obszaru Alp Zachodnich (Ryc. 2; Businsky i Kirschner 2010; Jalas i Suominen 1973; Monteleone i in. 2006), przy czym nie jest jasne, czy jest to miejsce ich wtórnego kontaktu, czy może obszar, z którego te dwie linie wzięły początek. Businsky i Kirschner (2010) scharakteryzowali występujące tam typowe mieszańce omawianych sosen górskich nadając im nazwę *Pinus* × *ascendens* Businsky. Opisywali przy tym populacje mieszańcowe, w których obecne są jednocześnie hybrydy i osobniki czystych gatunków, jak również takie, w których występują jedynie mieszańce. Znane są także miejsca w Alpach, gdzie w niewielkich odległościach współwystępują czyste populacje jednego i drugiego gatunku, ale nie widać w nich oznak introgresji (Monteleone i in. 2006). Z kolei w rejonie Apeninów, w izolowanej populacji kosodrzewiny z Abruzji, wiele osobników wykazuje cechy właściwe raczej sośnie hakowatej (Boratyńska i in. 2005;

Boratyńska i in. 2015). Dane te wskazują na potencjalnie duże znaczenie procesu hybrydyzacji w kształtowaniu zmienności genetycznej sosen górskich z kompleksu *P. mugo*.

Dotychczas jednak niewiele wiadomo na temat genetycznych podstaw przystosowania omawianych taksonów do ich specyficznych warunków siedliskowych. Nie prowadzono do tej pory studiów z zakresu genomiki populacyjnej dotyczących tej problematyki, ani nawet badań biometrycznych dotyczących *stricte* przystosowań wysokogórskich w tej grupie roślin. Jedyną wiedzę na ten temat czerpać można ze wspomnianych wcześniej cech, takich jak: pokrój roślin, właściwości ich igieł, dobrze rozwinięty i silnie rozgałęziony system korzeniowy, oraz ich reakcji na wysokie stężenia ozonu (Ballian i in. 2016; Businsky i Kirschner 2010; Diaz-de-Quijano i in. 2019).

**Cel i zakres badań**

W prezentowanej pracy wykorzystano zmienność szeregu rejonów genomowych celem analizy relacji filogenetycznych istniejących pomiędzy gatunkami sosny górskiej (*Pinus mugo* Turra) oraz sosny hakowatej (*P. uncinata* Ramond ex DC.), a także oceny zmienności genetycznej ich populacji w kontekście procesów demograficznych i selekcyjnych. Celem badań było poszukiwanie i przybliżenie odpowiedzi na szereg pytań, m.in.: Jaki jest poziom zróżnicowania genetycznego badanych taksonów? Jakie są ich relacje genetyczne względem sosny zwyczajnej? Czy istnieją markery diagnostyczne przydatne w identyfikacji molekularnej badanych gatunków? Jak wygląda struktura genetyczna ich populacji i jakie są wzajemne relacje filogeograficzne pomiędzy stanowiskami górskimi w szerokim zasięgu występowania gatunków? Czy struktura genetyczna populacji obecna w mitochondrialnym DNA pokrywa się ze wzorcem struktury genetycznej obserwowanym w genomie jądrowym? Czy w genomie jądrowym istnieją rejony, o wzorze zmienności odbiegającym od modelu ewolucji neutralnej?

W badaniach wyłoniono i przetestowano następujące hipotezy badawcze:

Hipoteza 1

Sosna górska i sosna hakowata wykazują istotne zróżnicowanie genetyczne w wyniku długotrwałej izolacji i adaptacji do odmiennych nisz ekologicznych, a zastosowane markery genetyczne grupują osobno populacje badanych gatunków.

Alternatywa

Pule genowe obu gatunków są podobne, na podstawie użytych markerów nie da się wyróżnić dwóch osobnych grup odpowiadających analizowanym taksonom.

Hipoteza 2

Rozkład zmienności genetycznej w obrębie gatunku jest niejednorodny, a utrwalone różnice w pewnych rejonach genomowych odzwierciedlają istnienie struktury i odmienne historie ich subpopulacji, wynikające ze zmian zasięgu gatunków i izolacji stanowisk w górach Europy.

Alternatywa

W genomach omawianych sosen brak jest różnic, które mógłby stanowić markery diagnostyczne w identyfikacji prób pochodzących z określonych rejonów geograficznych, co może być wynikiem ich potencjalnie niedawnej izolacji i efektywnego przepływu genów pomiędzy populacjami.

Hipoteza 3

W genomie jądrowym badanych sosen można zidentyfikować loci poddane presji selekcyjnej w wyniku adaptacji do warunków wysokogórskich, których polimorfizm odbiega od neutralnego tła genetycznego populacji i gatunków.

Alternatywa

Zmienność adaptacyjna jest wynikiem niewielkiego, addytywnego wpływu wielu genów i zastosowane metody nie pozwalają na identyfikację form allelicznych o potencjalnie silnym efekcie fenotypowym.

Hipoteza 4

Presja selekcyjna na poziomie molekularnym dotyczy nie tylko polimorfizmu form allelicznych genów (zmian strukturalnych sekwencji), ale również różnic w poziomie ekspresji genów.

Alternatywa

Poziom ekspresji genów badanych taksonów odzwierciedla ich bliskie relacje genetyczne i brak różnic na poziomie transkryptomu, które można by uznać za diagnostyczne i specyficzne gatunkowo.

**Zarys metodologii**

Do testowania przedstawionych hipotez wykorzystano szereg populacji naturalnych obu gatunków (a także taksonu referencyjnego – sosny zwyczajnej) oraz różnorodne markery genetyczne, które pozwoliły na oszacowanie między- i wewnątrzgatunkowej zmienności neutralnej (co posłużyło do badania tła genetycznego i wzajemnych relacji) oraz adaptacyjnej (służącej wnioskowaniu o potencjalnym wpływie doboru na ewolucję populacji i taksonów). U sosen markery takie pochodzić mogą z trzech genomów – jądrowego (*n*DNA), chloroplastowego (*cp*DNA) i mitochondrialnego (*mt*DNA), z których każdy posiada nieco inny zestaw cech, warunkujących ich różną użyteczność w badaniach genetycznych. Ich krótką charakterystykę zamieszczono poniżej. W niniejszej pracy wykorzystano markery *mt*DNA i *n*DNA oraz odniesiono się do zmienności markerów *cp*DNA, znanej z wcześniejszych badań.

Genomy cytoplazmatyczne, tj. chloroplastowy i mitochondrialny dziedziczone są po jednym z rodziców, u sosen pierwszy w linii ojcowskiej, drugi w linii matecznej (Wang i in. 1996). Warto zaznaczyć, że sosny są z reguły jednopienne (Giertych 1976). Obydwa genomy są haploidalne i z zasady nie podlegają rekombinacji, za to mają niższą (1/4) efektywną wielkość populacji w stosunku do diploidalnego genomu jądrowego i tym samym wykazują większą podatność na perturbacje demograficzne i działanie dryfu. Chloroplastowy DNA, przekazywany z pyłkiem na znaczne odległości (nawet do 100 km; Robledo-Arnuncio 2011), dobrze odzwierciedla potencjał krzyżowania się populacji tych wiatropylnych roślin, za to ma ograniczoną przydatność w badaniach struktury populacji i filogeografii. Znacznie bardziej

użyteczne w tym względzie są markery *mt*DNA, które rozprzestrzeniane są wraz z nasionami. Nasiona tej grupy również roznoszone są przez wiatr (Benkman 1995; Dzialuk i in. 2012), jednak z racji ich znacznie większych rozmiarów i ciężaru, migrują na dystanse znacznie mniejsze niż pyłek (Tóth i in. 2017). Zmienność tego genomu odzwierciedla nie tylko pokrewieństwo osobników wewnątrz i między populacjami, ale też pozwala na śledzenie dróg migracji nasion, konkretnych zmian zasięgów populacji, jak również ocenę potencjału dyspersyjnego gatunku – to nasiona są diasporami z których rozwiną się kolejne pokolenia. Niestety, mitogenom sosen jest bardzo słabo poznany a opisane wcześniej rejony cechuje zazwyczaj stosunkowo niskie tempo mutacji i niski poziom polimorfizmu sekwencji (Donnelly i in. 2017; Wolfe i in. 1987;), co utrudnia opracowanie odpowiednich markerów genetycznych. W obu genomach cytoplazmatycznych obecne są geny kodujące, niemniej przyjmuje się, że ewolucja zasadniczej części markerów w nich zawartych zbliżona jest do neutralnej.

Genom jądrowy tych sosen jest ogromnych rozmiarów – na haploidalny zestaw 12 chromosomów, składa się około 21-27 mld pz (oszacowania dla sosny zwyczajnej; Donnelly i in. 2017). Jest on również bardzo złożony – bogaty w obszary niekodujące i repetytywne, w większości będące pozostałościami po retrotranspozonach i procesach duplikacji genów (Neal i in. 2014; Voronova i in. 2017). W masie *n*DNA znajdują się rozproszone geny kodujące białka oraz różne rodzaje funkcjonalnego RNA. Liczba genów odpowiedzialnych za produkcję białek u roślin jest uznawana za dość konserwatywną i szacowana jest na 26 do 28,5 tys. (Cheng i in. 2016; Gonzales-Ibeas i in. 2016; Nystedt i in. 2013). Ze względu na duże rozmiary i złożoność, stosunkowo kompletne sekwencje referencyjne genomów jądrowych zostały dotychczas opisane jedynie dla 10 przedstawicieli drzew iglastych (Neale i in. 2022). Jak dotąd nie opisano sekwencji referencyjnej dla sosny zwyczajnej, ani żadnego z taksonów kompleksu *P. mugo*. Ponieważ w diploidalnych jądrach komórkowych sporofitów sosen mieszają się geny ojcowskie, migrujące z pyłkiem, i mateczne, rozsiewane z nasionami, które później podlegają szybkiej i efektywnej rekombinacji (Pyhäjärvi i in. 2019), genom jądrowy jest uznawany za mniej użyteczny w zakresie odtwarzania historii populacji. Sekwencje *n*DNA są natomiast potencjalnie cennym źródłem informacji na temat procesów selekcyjnych jakim podlegają populacje i gatunki, a współczesne wysokoprzepustowe technologie sekwencjonowania i genotypowania DNA pozwalają na pozyskiwanie dużej ilości danych molekularnych w krótkim czasie i relatywnie niewielkim kosztem.

W dalszej części rozprawy przedstawiono najważniejsze założenia i wyniki badań stanowiących efekt mojej pracy doktorskiej, które opublikowałam w kolejnych trzech artykułach naukowych.

**Publikacja pierwsza**

Pierwszym krokiem w badaniach genetycznych podstaw zmienności adaptacyjnej i relacji filogenetycznych jest rozpoznanie wariancji tła genetycznego i pokrewieństwa istniejącego pomiędzy osobnikami i populacjami gatunku w efekcie historycznych zmian ich liczebności i zasięgów. Próby takie podejmowano wcześniej dla kosodrzewiny i sosny hakowatej, wykorzystując w nich głównie markery biometryczne, ale także molekularne, m.in. rejony mikrosatelitarne DNA z genomu plastydowego i jądrowego (Dzialuk i in. 2009; Dzialuk i in. 2017; Heuertz i in. 2010). Jednakże informacje uzyskane z tych badań były niespójne ze względu na wysoce mobilny charakter i nikłą wartość dyskryminacyjną zastosowanych markerów. Stąd celem pierwszej części prowadzonych przeze mnie badań, której poświęcony jest artykuł opublikowany w *Journal of Systematics and Evolution* (Zaborowska i in. 2019 – *Population history of European mountain pines Pinus mugo and Pinus uncinata revealed by mitochondrial DNA markers*), było oszacowanie zmienności i zróżnicowania genetycznego omawianej pary taksonów górskich w zakresie ich neutralnego tła genetycznego. W pracy tej podjęto próbę weryfikacji **hipotez 1** i **2** dotyczących zróżnicowania obu taksonów oraz obecności struktury genetycznej w zasięgu ich występowania. Przebadano w tym celu kilkaset osobników pochodzących z 27 populacji obu gatunków, reprezentatywnych dla całości ich naturalnych zasięgów, i wykorzystano zmienność sekwencji mitochondrialnego DNA. Dotychczas w badaniach filogeograficznych przedstawicieli rodzaju *Pinus* funkcjonowały zasadniczo jedynie dwa markery *mt*DNA o nikłej rozdzielczości – w rejonach *nad1 intron B/C* oraz *nad7 intron 1* (Jaramillo-Correa i in. 2004; Soranzo i in. 2000). W prezentowanej pracy, bazując na wynikach Donnelly i in. (2017), opracowałam 15 nowych markerów *mt*DNA: 14 typu RFLP oraz 1 typu IN-DEL. Na ich podstawie zidentyfikowałam łącznie 31 mitotypów, wykorzystanych m.in. do przeprowadzenia analiz filogenetycznych, wariancji molekularnej, analizy struktury populacji oraz oceny izolacji genetycznej w zasięgu występowania populacji.

Przeprowadzone badania umożliwiły wstępną weryfikację postawionych hipotez. Zmienność DNA mitochondrialnego pozwoliła na rozróżnienie dwóch grup genetycznych wśród przebadanych osobników i populacji, które blisko odpowiadały wcześniejszej, fenotypowej identyfikacji taksonomicznej i ich klasyfikacji do dwóch gatunków. Wyjątek stanowiła pojedyncza populacja z Alp Kotyjskich (obszaru nakładania się zasięgów taksonów), uznawana za populację sosny hakowatej, w której wykryto wyłącznie mitotypy właściwe kosodrzewinie. W zestawie znalazły się dwa loci z niemalże utrwalonymi, odmiennymi allelami u sosny górskiej i sosny hakowatej (*Pr13* oraz *nad1 intron B/C*), które analizowane wspólnie mogą służyć za markery diagnostyczne dla linii *mt*DNA tych dwóch taksonów. Ponadto, uzyskane wyniki wykazały obecność w populacjach obu gatunków wyraźnej struktury genetycznej, co potwierdziły również dalsze analizy z wykorzystaniem markerów jądrowych (Zaborowska i in. 2021).

**Publikacja druga**

Kolejna część pracy, zaprezentowana w publikacji w *International Journal of Molecular Sciences* (Zaborowska i in. 2021 – *Candidate genes for the high-altitude adaptations of two mountain pine taxa*), poświęcona została analizom zmienności genetycznej w skali całego genomu jądrowego omawianych sosen. Badanie miało na celu ocenę przydatności skanowania tysięcy pojedynczych miejsc zmiennych (SNPs) do rozróżniania badanych sosen górskich na poziomie gatunków oraz rozpoznania istniejącej struktury genetycznej w ich populacjach (weryfikacja **hipotez 1** i **2**). Celem analiz była również identyfikacja genów i ich grup funkcyjnych, posiadających w swoich sekwencjach markery SNP o wzorcach zmienności – częstościach alleli – odbiegających od oczekiwanych na podstawie zmienności tła genetycznego (weryfikacja **hipotezy 3**). Badania oparto o dane z platformy Affymetrix – z zaprojektowanej w tym celu płytki Axiom_PineGAP z 49,829 markerami SNP (Perry i in. 2020). Przeanalizowano w nich 524 osobniki potomne, pochodzące z 4 populacji *P. mugo*, 5 populacji *P. uncinata*, jak również z 5 populacji *P. sylvestris* (sosnę zwyczajną wykorzystano jako takson referencyjny). Nasiona, z których zostały wyhodowane drzewa pochodziły z populacji naturalnych badanych gatunków, natomiast same siewki wzrastały w kontrolowanych warunkach eksperymentalnych.

Uzyskane, przefiltrowane pod kątem jakości genotypowania dane zostały wykorzystane do przeprowadzenia podstawowych analiz zmienności i zróżnicowania populacji, analizy wariancji, a także ich grupowania oraz weryfikacji zależności pomiędzy odległościami fizycznymi i genetycznymi istniejącymi między populacjami. Ponadto, przeprowadzona została identyfikacja loci wykazujących odstępstwa częstości alleli SNP od uśrednionych częstości tła genetycznego, tj. identyfikacja tzw. outlierów. Poszukiwane były markery o skrajnych częstościach, zarówno pomiędzy sosnami górskimi, jak również w porównaniu do taksonu referencyjnego (*P. sylvestris*). Dla każdego odznaczającego się markera SNP odszukano sekwencję DNA, z której pochodził, a następnie przyrównano ją do znanych sekwencji białkowych obecnych w zasobach recenzowanych baz danych i określono jej potencjalne funkcje.

Wyniki badania pokazały, że większość zmienności genetycznej jądrowego DNA segreguje u tych sosen wewnątrz populacji (73% wg. analizy wariancji), potwierdzając wcześniejsze obserwacje (Dzialuk i in. 2017; Heuertz i in. 2010). Niemniej, udało się wykazać, że zróżnicowanie markerów SNP między trzema taksonami jest wystarczające dla ich jednoznacznego rozróżnienia, a także, że w populacjach obu gatunków górskich, jak również sosny zwyczajnej, zmienność genetyczna nie jest rozlokowana losowo i pozwala na odróżnienie niektórych ze stanowisk. Badania potwierdziły bliższe pokrewieństwo dwóch gatunków górskich względem siebie w porównaniu do taksonu referencyjnego, a także ujawniły pewne wzorce pokrewieństwa poszczególnych populacji. Co ciekawe, nie wykryto natomiast ani jednego locus, w którym byłby obecny allel prywatny i jednocześnie utrwalony wewnątrz któregokolwiek gatunku. Zidentyfikowanych zostało 35 genów kandydackich (zawierających polimorfizmy odbiegające od modelu ewolucji neutralnej) pomiędzy dwoma gatunkami górskimi i sosną zwyczajną oraz 75 genów potencjalnie odpowiadających za zróżnicowanie sosny górskiej i hakowatej. Mimo, że samo zróżnicowanie częstości alleli nie jest ostatecznym dowodem działania doboru na dany gen, to stanowi silną przesłankę dla weryfikacji takiej zależności, zwłaszcza, jeśli funkcje kodowanego białka/ białek są spójne z działaniem czynników środowiskowych znanych z wywierania wpływu na dany organizm. Stąd też opisane geny kandydackie stanowią dogodny obiekt badań zmierzających do określenia efektu fenotypowego wykrytych form allelicznych (Zaborowska i in. 2021).

**Publikacja trzecia**

W trzeciej części pracy, której poświęcony został artykuł opublikowany w *Journal of Systematics and Evolution* (Zaborowska i in. 2022 – *Evolutionary targets of gene expression divergence in a complex of closely related pine species*), przeprowadzono ocenę zmienności genetycznej istniejącej pomiędzy badanymi taksonami i jej wpływu na dostosowanie do warunków górskich na poziomie ekspresji genów. Celem weryfikacji **hipotezy 4** wykorzystano dane transkryptomowe pochodzące z sekwencjonowania nowej generacji (RNA-Seq Illumina). Źródłem RNA użytego w badaniach były dwuletnie osobniki omawianych sosen górskich oraz sosny zwyczajnej, pochodzące z nasion zebranych w populacjach naturalnych i wzrastające w jednakowych warunkach szklarniowych (stanowiące podzbiór prób użytych w drugiej publikacji). Taki zabieg pozwolił na ujednolicenie warunków środowiskowych i identyfikację wzorów ekspresji genów będących wynikiem zróżnicowania genetycznego badanych taksonów (i ewentualnych wpływów czynników epigenetycznych).

Surowe odczyty sekwencji (ang. *reads*) zostały oczyszczone, a na ich podstawie złożono (ang. *assembling*) wspólny transkryptom. Na tym etapie, dla zwiększenia szans odtworzenia poprawnych, pełnej długości sekwencji, użyto dodatkowej próby referencyjnej sosny zwyczajnej, o znacznie wyższym pokryciu (ang. *coverage*) transkryptomu. Oszacowano poziomy ekspresji sekwencji i usunięto te najsłabiej reprezentowane. Pozostałe sekwencje zostały porównane pomiędzy trzema parami gatunków, a także pomiędzy sosnami górskimi wspólnie skontrastowanymi z sosną zwyczajną. Modele genów poddano analizie funkcjonalnej poprzez odszukanie w nich otwartych ramek odczytu (ORFs od ang. *open reading frames*), przyrównanie (ang. *alignment*) fragmentów kodujących do sekwencji białek znanych z uznanych baz danych, a następnie wnioskowano o ich funkcji na podstawie białek pokrewnych oraz wykrytych domen strukturalnych.

Wykonane testy pozwoliły na wyłonienie 393 modeli genów, które wykazały istotną statystycznie różnicę we względnym poziomie ekspresji w którejkolwiek z czterech porównywanych par taksonów, najwięcej pomiędzy sosną górską i sosną zwyczajną. Spośród tej liczby, 121 modeli genów zawierało ORF, a potencjalną rolę kodowanego białka można było wywnioskować w 74 przypadkach (odpowiednio: 12 w porównaniu *P. mugo* i *P. uncinata*

z *P. sylvestris*, 18 w porównaniu *P. mugo* z *P. uncinata*, 54 w porównaniu *P. mugo* z *P. sylvestris* oraz 13 w porównaniu *P. uncinata* z *P. sylvestris*). Geny różnicujące gatunki okazały się należeć do wielu grup funkcyjnych, między innymi odpowiadających za cechy drewna, odpowiedź na niekorzystne czynniki abiotyczne takie jak stres oksydacyjny, zasolenie, susza czy skrajna temperatura, a także na negatywne czynniki biotyczne. Uzyskane wyniki potwierdzają duże podobieństwo pomiędzy badanymi gatunkami na poziomie ekspresji genów, dostarczając zarazem informacji o genach kandydackich, których poziomy ekspresji mogą odzwierciedlać utrwalone ewolucyjnie różnice międzygatunkowe leżące u podstaw ich dywergencji (Zaborowska i in. 2022).

**Podsumowanie**

Badania przeprowadzone w ramach prezentowanej pracy doktorskiej pozwoliły na rozpoznanie poziomu zróżnicowania genetycznego pomiędzy sosną górską i sosną hakowatą, dwoma blisko spokrewnionymi taksonami, poddanymi podobnej presji środowiskowej. Uzyskane wyniki wskazują na istotną rozłączność tych linii ewolucyjnych – większy udział wariancji genetycznej pomiędzy taksonami niż ich populacjami, zaznaczający się zarówno w markerach mitochondrialnych, jak i zmienności pojedynczych miejsc nukleotydowych genomu jądrowego. Uzyskane wyniki dla różnych typów markerów podkreślają generalnie bliższe pokrewieństwo sosny zwyczajnej (użytej w badaniach jako grupa referencyjna) z sosną hakowatą niż z sosną górską. Taki wzorzec podobieństwa genetycznego wskazuje, że przodkowie sosny hakowatej wywodzą się od wspólnego przodka z sosną zwyczajną, a rozdział dający początek linii ewolucyjnej *P. mugo* jest relatywnie młodszy. Zastosowane markery genetyczne potwierdziły również obecność struktury genetycznej w populacjach obu omawianych taksonów. Markery *mt*DNA okazały się nieco lepiej rozróżniać populacje *P. mugo*, podczas gdy odrębność populacji *P. uncinata* była silniej zaznaczona w *n*DNA. Tylko w mitogenomie udało się zidentyfikować utrwalone warianty diagnostyczne dla badanych taksonów, których nie stwierdzono dla żadnego z ok. 6 tys. badanych polimorfizmów SNP *n*DNA. Uzyskane dane wskazują na odmienne trajektorie migracji gatunków, niż dotychczas przyjmowano, np. pochodzenie współczesnych populacji *P. mugo* w Sudetach i Karpatach z półwyspu Bałkańskiego i Alp, oraz znacznie większy historyczny zasięg *P. mugo*,

szczególnie rozciągnięty na zachód względem współczesnego rozmieszczenia gatunku, co znajduje potwierdzenie w wynikach badań pelobotanicznych (Farjon i in. 2019). Opracowany zestaw markerów *mt*DNA stanowi nowe, użyteczne narzędzie w badaniach z zakresu genetyki populacyjnej badanych sosen, w tym filogenezy kompleksu *P. mugo*. Dokonana rekonstrukcja zmian demograficznych omawianych taksonów może pomóc również lepiej zrozumieć możliwe zmiany ich zasięgów w obliczu postępujących zmian środowiskowych. Dane ukazujące niewielką zmienność i przypuszczalnie niski potencjał adaptacyjny niektórych populacji *P. uncinata*, mogą stanowić naukowe podstawy dla zarządzania zasobami genowymi i ochrony tych cennych przyrodniczo stanowisk. W tym przypadku populacje z Pirenejów, tj. centrum zasięgu występowania gatunku, wykazały niski poziom różnorodności mitotypów (zaledwie 4 w 7 populacjach) w porównaniu do silnie izolowanej populacji z Masywu Centralnego (7 mitotypów).

Zrealizowane prace badawcze pozwoliły zidentyfikować w genomie jądrowym dwóch sosen górskich szereg sekwencji kodujących, których poziomy zmienności mierzone frekwencją poszczególnych alleli oraz wzorem ekspresji genów odbiegały istotnie od opisanej dla tła genetycznego. Określenie ich funkcji biologicznych i aktywności molekularnych, dostarcza nowych danych na temat możliwych rejonów genomowych poddanych presji selekcyjnej u roślin iglastych. Należą do nich m.in. czynniki związane z transkrypcją genów, odpowiadające za metabolizm białek, związków azotu i fosforu, transport i odpowiedź na czynniki abiotyczne i biotyczne, które mogą być związane z odpowiedzią adaptacyjną gatunków do panujących w ich populacjach warunków środowiskowych. W aspekcie analiz czynników genetycznych potencjalnie odpowiedzialnych za różnice fenotypowe pomiędzy badanymi sosnami górskimi, w tym szczególnie karłowaty pokrój kosodrzewiny, udało się udokumentować obecność istotnie zróżnicowanych częstości alleli SNP w dwóch czynnikach transkrypcyjnych działających w odpowiedzi na gibereliny (znane z regulacji wzrostu roślin i mutacji skutkujących karłowatością; Ford i in. 2018; Schomburg i in. 2003), jak również wykazać odmienną ekspresję transkryptów regulujących metabolizm tych hormonów.

W badaniach wykorzystano populacje sosny zwyczajnej jako próby referencyjnej w analizach porównawczych zmienności genetycznej kosodrzewiny i sosny hakowatej, i ich odpowiedzi adaptacyjnej na warunki wysokogórskie. Istotnie zróżnicowane allele SNP

wykryto w genach kodujących białka biorące udział w fotosyntezie, regulacji fotorespiracji, utrzymaniu w komórkach równowagi między produkcją i redukcją reaktywnych form tlenu (homeostazy red-ox), a także w regulacji transkrypcji genów i późniejszej edycji mRNA. Większość z nich dobrze wpisuje się w wyniki wcześniejszych badania dotyczące przystosowań roślin do wzmożonego promieniowania słonecznego, czy obniżonego ciśnienia cząsteczkowego dwutlenku węgla. Natomiast wśród sekwencji charakteryzujących się odmienną ekspresją u taksonów górskich względem referencyjnego, znalazły się przede wszystkim geny odpowiadające za rozwój i właściwości drewna, reakcje na stres oksydacyjny, oraz biotyczne i abiotyczne stresogenne bodźce środowiskowe (gł. czynniki związane z warunkami temperaturowymi, hydrologicznymi i zasoleniem). Wyniki porównań sekwencji transkryptomowych sugerują istotną rolę regulacji ekspresji genów i modyfikacji RNA w historii rozchodzenia się linii ewolucyjnych w tej grupie taksonomicznej. Niemniej, zwiększenie próby populacyjnej oraz szersze przebadanie zmienności strukturalnej sekwencji wyłonionych genów kandydackich, jest wskazane dla odróżnienia wpływu działania doboru naturalnego od innych czynników mogących generować podobne, obserwowane różnice w poziomie ekspresji genów i/ lub rozkładzie częstości alleli, oraz dla określenia znaczenia tych dwu w ewolucji analizowanych gatunków.

Choć zaprezentowane badania tylko w niewielkim stopniu obejmowały problematykę obszaru sympatrycznego, tj. miejsc współwystępowania sosny górskiej i hakowatej, uzyskane wyniki mogą stanowić punkt odniesienia dla dalszych analiz w tym zakresie. Sugerowane w przedstawionych badaniach zjawisko introgresji i przechwycenia (ang. *genome capture*) genomu mitochondrialnego *P. mugo* przez populacje fenotypowo odpowiadające *P. uncinata*, jest zbieżne pod względem kierunku przepływu genów do zjawisk obserwowanych u innych gatunków iglastych (Petrova i in. 2018; Wang i Wang 2014). Introgresji nie wykryto natomiast w analizie zmienności markerów SNP w populacji kosodrzewiny z Abruzji, co do której istniało szereg przesłanek, głównie biometrycznych, o jej hybrydowym charakterze. Innym interesującym wątkiem dotyczącym wymiany genów między taksonami jest także powinowactwo sosen hakowatych z dwóch marginalnych populacji gatunku – zlokalizowanych w Masywie Centralnym i paśmie Sierra de Gudar – do sosny zwyczajnej. Te dwa stanowiska wydają się być starsze ewolucyjnie od populacji z Pirenejów (dzielą wiele markerów, których brak w Pirenejach, są także bogatsze pod względem różnorodności

*mt*DNA). Niemniej odpowiedź na pytanie, czy wzajemne podobieństwo genetyczne jest zaszłością historyczną, czy raczej wynika ze współczesnej admiksji z *P. sylvestris*, wymaga dodatkowych badań.

Konkludując, przedstawione powyżej badania pogłębiają wiedzę na temat historii populacji i procesów ewolucyjnych dwóch europejskich taksonów z rodzaju *Pinus* – *P. mugo* i *P. uncinata*. Dostarczają one nowych danych dotyczących filogeografii taksonów w pasmach górskich Europy, ich możliwych dróg migracji i zmian zasięgów, wpływu zjawiska hybrydyzacji na zmienność genetyczną populacji sympatrycznych, a także genetycznych podstaw ich adaptacji do specyficznych wymagań siedliska. Weryfikacja użyteczności macierzy markerów SNP u badanych gatunków stwarza możliwość pogłębionych analiz relacji filogenetycznych łączących także inne jednostki taksonomiczne w tej grupie roślin. Z kolei nowy zestaw złożonych sekwencji transkryptomowych stanowi dogodny punkt odniesienia dla dalszych analiz genetycznych prowadzonych wśród tych i innych taksonów z kompleksu *P. mugo*. Zastosowane markery genetyczne i przedstawiona metodyka badań stanowią również punkt wyjścia w innych badaniach z zakresu genetyki populacyjnej i genomiki drzew leśnych, zmierzających do efektywnego gospodarowania istniejącymi zasobami genowymi tych gatunków w obliczu obserwowanych, i odwiecznych, zmian środowiskowych.

# 2. BIBLIOGRAFIA


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

2. 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.

3. Alexandrov AH, von Wühlisch G, Vendramin GG. 2019. Technical guidelines for genetic conservation and use of mountain pine (*Pinus mugo*). European Forest Genetic Resources Programme (EUFORGEN), European Forest Institute: Bonn, Germany.

4. Anacker BL, Strauss SY. 2014. The geography and ecology of plant speciation: Range overlap and niche divergence in sister species. *Proceedings of the Royal Society B: Biological Sciences* 281: 20132980.

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

6. Arenas M, Ray N, Currat M, Excoffier L. 2012. Consequences of range contractions and range shifts on molecular diversity. *Molecular Biology and Evolution* 29: 207–218.

7. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT et al. 2000. Gene Ontology: Tool for the unification of biology. *Nature Genetics* 25: 25–29.

8. Ballian D, Ravazzi C, de Rigo D, Caudullo G. 2016. *Pinus mugo* in Europe: distribution, habitat, usage and threats. In: San-Miguel-Ayanz J, de Rigo D, Caudullo G, Houston Durrant T, Mauri A eds. European Atlas of Forest Tree Species. Publications Office of the EU: Luxembourg.

9. Bandelt HJ, Forster P, Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* 16: 37–48.

10. Basset GJC, Ravanel S, Quinlivan EP, White R, Giovannoni JJ, Rebeille F, Nichols BP, Shinozaki K, Seki M, Gregory JF et al. 2004. Folate synthesis in plants: The last step of the p-aminobenzoate branch is catalyzed by a plastidial aminodeoxychorismate lyase. *The Plant Journal* 40: 453–461.

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

12. Behringer D, Zimmermann H, Ziegenhagen B, Liepelt S. 2015. Differential gene expression reveals candidate genes for drought stress response in *Abies alba* (Pinaceae). *PLoS ONE* 10: e0124564.

13. Benito Garzon M, Sanchez de Dios R, Sainz Ollero H. 2007. Predictive modelling of tree species distributions on the Iberian Peninsula during the Last Glacial Maximum and Mid-Holocene. *Ecography* 30: 120–134.

14. Benkman CW. 1995. Wind dispersal capacity of pine seeds and the evolution of different seed dispersal modes in pines. *Oikos* 73: 221–224.

15. Bičárová S, Sitková Z, Pavlendová H, Fleischer P (Jr), Fleischer P (Sr), Bytnerowicz A. 2019. The role of environmental factors in ozone uptake of *Pinus mugo* Turra. *Atmospheric Pollution Research* 10: 283-293.

16. Billings WD. 1974. Adaptations and Origins of Alpine Plants. *Arctic and Alpine Research* 6: 129–142.

17. BioBam Bioinformatics SL [online]. 2020. OmicsBox—Bioinformatics Made Easy. Available from www.biobam.com/omicsbox [accessed 11 June 2020].

18. Blada I. 2017. The short needle stone pine (*Pinus microphylla* species nova), a new dwarf two-needle hard pine species in the Pietrosul Rodnei Massif flora. Editura Silvica: Bucharest, Romania.

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

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

21. Bomblies K. 2015. The High Life: Alpine Dwarfism in *Arabidopsis*. *Plant Physiology* 168: 767.

22. Boratyńska K, Bobowicz MA. 2001. *Pinus uncinata* Ramond taxonomy based on needle characters. *Plant Systematics and Evolution* 227: 183–194.

23. Boratyńska K, Boratyński 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.

24. Boratyńska K, Dzialuk A, Lewandowski A, Marcysiak K, Jasińska AK, Sobierajska K, Tomaszewski D, Burczyk J, Boratyński A. 2014. Geographic distribution of quantitative traits variation and genetic variability in natural populations of *Pinus mugo* in Central Europe. *Dendrobiology* 72: 65–84.

25. Boratyńska K, Jasińska AK, Boratyński A. 2015. Taxonomic and geographic differentiation of *Pinus mugo* complex on the needle characteristics. *Systematics and Biodiversity* 13: 901–915.

26. Boratyńska K, Marcysiak K, Boratyński A. 2005. *Pinus mugo* (Pinaceae) in the Abruzzi mountains: high morphological variation in isolated populations. *Botanical Journal of the Linnean Society* 147: 309–316.

27. Boratyńska K, Muchewicz E, Drojma M. 2004. *Pinus mugo* Turra geographic differentiation based on needle characters. *Dendrobiology* 51: 9–17.

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

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

30. Businský R, Kirschner J. 2010. *Pinus mugo* and *P. uncinata* as parents of hybrids: A taxonomic and nomenclatural survey. *Phyton* 50: 27–57.

31. Camarero JJ, Gutierrez E, Fortin MJ, Ribbens E. 2005. Spatial patterns of tree recruitment in a relict population of *Pinus uncinata*: forest expansion through stratified diffusion. *Journal of Biogeography* 32: 1979–1992.

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

33. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021. eggNOG-mapper v2: functional annotation, orthology assignments, and prediction at the metagenomic scale. *Molecular Biology and Evolution* 38: 5825-5829.

34. Carbon S, Douglass E, Dunn N, Good B, Harris NL, Lewis SE, Mungall CJ, Basu S, Chisholm RL, Dodson RJ et al. 2019. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Research* 47: D330–D338.

35. Carcaillet C, Fauvart N, Roiron P, Terral JF, Ali AA. 2009. A new, isolated and endangered relict population of dwarf pine (*Pinus mugo* Turra) in the northwestern Alps. *Comptes Rendus Biologies* 332: 456–463.

36. Carlise A, Brown AHF. 1968. *Pinus sylvestris* L. *Journal of Ecology* 56: 269-307.

37. Charrier O, Dupont P, Pornon A, Escaravage N. 2014. Microsatellite marker analysis reveals the complex phylogeographic history of *Rhododendron ferrugineum* (Ericaceae) in the pyrenees. *PLoS One* 9: e92976.

38. Cheddadi R, Birks HJB, Tarroso P, Liepelt S, Gomory D, Dullinger S, Meier ES, Hülber K, Maiorano L, Laborde H. 2014. Revisiting tree migration rates: *Abies alba* (Mill.), a case study. *Vegetation History and Archaeobotany* 23: 113–122.

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

40. 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.

41. 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.

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

43. Christensen KI. 1987b. Taxonomic revision of the *Pinus mugo* complex and *P. rhaetica* (*P. mugo* x *sylvestris*). *Nordic Journal of Botany* 7: 338–408.

44. Chun HJ, Lim LH, Cheong MS, Baek D, Park MS, Cho HM, Lee SH, Jin BJ, No DH, Cha YJ et al. MC. 2021. *Arabidopsis* CCoAOMT1 plays a role in drought stress response via ROS- and ABA-dependent manners. *Plants* 10: 831.

45. Coomes DA, Jenkin KL, Cole LES. 2007. Scaling of tree vascular transport systems along gradients of nutrient supply and altitude. *Biology Letters* 3: 86–89.

46. Critchfield WB, Little EL. 1966. Geographic distribution of the pines of the world. Department of Agriculture, Forest Service: Washington, USA.

47. Cudlin P, Klopcic M, Tognetti R, Malis F, Alados CL, Bebi P, Grunewald K, Zhiyanski M, Andonowski V, La Porta N et al. 2017. Drivers of treeline shift in different European mountains. *Climate Research* 73: 135–150.

48. Cumming JR. 1996. Phosphate-limitation physiology in ectomycorrhizal pitch pine (*Pinus rigida*) seedlings. *Tree Physiology* 16: 977–983.

49. Currat M, Ruedi M, Petit RJ, Excoffier L. 2008. The hidden side of invasions: massive introgression by local genes. *Evolution* 62: 1908–1920.

50. Dainou K, Blanc-Jolivet C, Degen B, Kimani P, Ndiade-Bourobou D, Donkpegan AS, Tosso F, Kaymak E, Bourland N, Doucet JL et al. 2016. Revealing hidden species diversity in closely related species using nuclear SNPs, SSRs and DNA sequences – A case study in the tree genus *Milicia*. *BMC Evolutionary Biology* 16: 259.

51. 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.

52. De Abreu-Neto JB, Turchetto-Zolet AC, de Oliveira LFV, Zanettini MHB, Margis-Pinheiro M. 2013. Heavy metal-associated isoprenylated plant protein (HIPP): Characterization of a family of proteins exclusive to plants. *The FEBS Journal* 280: 1604–1616.

53. Diaz-de-Quijano M, Kefauver S, Ogaya R, Vollenweider P, Ribas Á, Peñuelas 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.

54. Donnelly K, Cottrell J, Ennos RA, Vendramin GG, A'Hara S, King S, Perry A, Wachowiak W, Cavers S. 2017. Reconstructing the plant mitochondrial genome for marker discovery: A case study using *Pinus*. *Molecular Ecology Resources* 17: 943–954.

55. Dostalek J, Frantíkb T, Pospiskova M, Krizovac M. 2018. Population genetic structure and delineation of conservation units in European larch (*Larix decidua* Mill.) across its native range. *Flora* 246-247: 26–32.

56. Dupanloup I, Schneider S, Excoffier L. 2002. A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology* 11: 2571–2581.

57. 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.

58. Dzialuk A, Boratyńska K, Romo A, Boratyński A. 2017. Taxonomic and geographic variation of the *Pinus mugo* complex on chloroplast microsatellite markers. *Systematics and Biodiversity* 15: 464–479.

59. Dzialuk A, Boratyński A, Boratyńska K, Burczyk J. 2012. Geographic patterns of genetic diversity of *Pinus mugo* (Pinaceae) in Central European mountains. *Dendrobiology* 68: 31–41.

60. Dzialuk A, Muchewicz E, Boratyński A, Montserrat JM, Boratyńska K, Burczyk J. 2009. Genetic variation of *Pinus uncinata* (Pinaceae) in the Pyrenees determined with cpSSR markers. *Plant Systematics and Evolution* 277: 197–205.

61. Earl DA, Vonholdt BM. 2012. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4: 359–361.

62. Ecker S, Pancaldi V, Valencia A, Beck S, Paul DS. 2018. Epigenetic and transcriptional variability shape phenotypic plasticity. *Bioessays* 40: 1700148.

63. European Forest Genetic Resources Programme (EUFORGEN) [online]. 2009. Distribution map of Scots pine (*Pinus sylvestris*). European Forest Genetic Resources Programme. Available from www.euforgen.org [accessed 22 Steptember 2009].

64. Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology* 14: 2611–2620.

65. Eveno E, Collada C, Guevara MA, Leger V, Soto A, Diaz L, Leger P, Gonzalez-Martinez SC, Cervera MT, Plomion C et al. 2008. Contrasting patterns of selection at *Pinus pinaster* Ait. drought stress candidate genes as revealed by genetic differentiation analyses. *Molecular Biology and Evolution* 25: 417–437.

66. Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10: 564–567.

67. Falush D, Stephens M, Pritchard JK. 2007. Inference of population structure using multilocus genotype data: Dominant markers and null alleles. *Molecular Ecology Notes* 7: 574–578.

68. Fariello MI, Boitard S, Naya H, SanCristobal M, Servin B. 2013. Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. *Genetics* 193: 929–941.

69. Farjon A. 2001. World checklist and bibliography of conifers, 2nd edn. The Royal Botanic Gardens: Kew, UK.

70. Farjon A. 2018. The Kew review: Conifers of the world. *Kew Bulletin* 73: 8.

71. Farjon A. Horne DJ, Parfitt SA, Buckland PI, Lewis MD. 2020. Early Pleistocene conifer macrofossils from Happisburgh, Norfolk, UK, and their environmental implications for early hominin occupation. *Quaternary Science Reviews* 232: 106115.

72. Fauvart N, Ali AA, Terral JF, Roiron P, Blarquez O, Carcaillet C. 2012. Holocene upper tree-limits of *Pinus* section *sylvestris* in the Western Alps as evidenced from travertine archives. *Reviews of Paleobotany and Palynology* 169: 96–102.

73. Feldeverd E, Porter BW, Yuen CYL, Iwai K, Carrillo R, Smith T, Barela C, Wong K, Wang Pe, Kang B-H et al. 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.

74. Foll M, Gaggiotti O. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics* 180: 977–993.

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

76. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J et al. 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology* 5: R80.

77. Gerlitz TGM, Gerlitz A. 1997. Phosphate uptake and polyphosphate metabolism of mycorrhizal and nonmycorrhizal roots of pine and of *Suillus bovinus* at varying external pH measured by in vivo P-31-NMR. *Mycorrhiza* 7: 101–106.

78. Gernandt DS, Geada Lopez G, Ortiz Garcia S, Liston A. 2005. Phylogeny and classification of *Pinus*. *TAXON* 54: 29-42.

79. Gibert JM, Mouchel-Vielh E, De Castro S, Peronnet F. 2016. Phenotypic plasticity through transcriptional regulation of the evolutionary hotspot gene *tan* in *Drosophila melanogaster*. *PLoS Genetics* 12: e1006218.

80. Giertych M. 1976. Sexual expression in pines in relation to their ontogenetic development. Acta Horticulturae 56: 99-104.

81. Gonzalez-Ibeas D, Martinez-Garcia PJ, Famula RA, Delfino-Mix A, Stevens KA, Loopstra CA, Langley CH, Neale DB, Wegrzyn JL. 2016. Assessing the gene content of the megagenome: sugar pine (*Pinus lambertiana*). *G3: Genes, Genomes, Genetics* 6: 3787-3802.

82. González-Martínez 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.

83. Gonzalo-Turpin H, Hazard L. 2009. Local adaptation occurs along altitudinal gradient despite the existence of gene flow in the alpine plant species *Festuca eskia*. *Journal of Ecology* 97: 742–751.

84. Gorelova V, Ambach L, Rebeille F, Stove C, Van Der Straeten D. 2017. Folates in plants: Research advances and progress in crop biofortification. *Frontiers in Chemistry* 5: 21.

85. Grotkopp E, Rejmanek M, Sanderson MJ, Rost TL. 2004. Evolution of genome size in pines (*Pinus*) and its life-history correlates: Supertree analyses. *Evolution* 58: 1705–1729.

86. Gugerli F, Ruegg M, Vendramin GG. 2009. Gradual decline in genetic diversity in Swiss stone pine populations (*Pinus cembra*) across Switzerland suggests postglacial re-colonization into the Alps from a common eastern glacial refugium. *Botanica Helvetica* 119: 13–22.

87. Haas B [online]. 2018. TransDecoder. Available from www.github.com/transdecoder/transdecoder/wiki [accessed 22 October 2018].

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

89. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M et al. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* 8: 1494-512.

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

91. Hanson AD, Gregory JF. 2002. Synthesis and turnover of folates in plants. *Current Opinion in Plant Biology* 5: 244–249.

92. 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.

93. Heer K, Behringer D, Piermattei A, Bassler C, Brandl R, Fady B, Jehl H, Liepelt S, Lorch S, Piotti A et al. 2018. Linking dendroecology and association genetics in natural populations: Stress responses archived in tree rings associate with SNP genotypes in silver fir (*Abies alba* Mill.). *Molecular Ecology* 27: 1428–1438.

94. Hernandez L, Camarero JJ, Gil-Peregrin E, San Sanchez MA, Canellas I, Montes F. 2019. Biotic factors and increasing aridity shape the altitudinal shifts of marginal Pyrenean silver fir populations in Europe. *Forest Ecology and Management* 432: 558–567.

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

96. Heuertz M, Teufel J, González-Martínez SC, Soto A, Fady B, Alía 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.

97. Hicks MA, Barber AE, Babbitt PC. 2013. The nucleophilic attack six-bladed-propeller (N6P) superfamily. In: Orengo C, Bateman A eds. Protein families: Relating protein sequence, structure, and function. John Wiley & Sons, Inc.: Chichester, UK; pp. 125–158.

98. Hicks MA, Barber AE, Giddings LA, Caldwell J, O'Connor SE, Babbitt PC. 2011. The evolution of function in strictosidine synthase-like proteins. *Proteins* 79: 3082–3098.

99. Hijimans R, Williams E, Vennes C [online]. 2019. Geosphere. Available from www.cran.r-project.org/package=geosphere [accessed 12 October 2020].

100. Hirst J, Bright NA, Rous B, Robinson MS. 1999. Characterization of a fourth adaptor-related protein complex. *Molecular Biology of the Cell* 10: 2787–2802.

101. 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.

102. Huang W, Yang YJ, Hu H, Zhang SB. 2016. Seasonal variations in photosystem I compared with photosystem II of three alpine evergreen broad-leaf tree species. *Journal of Photochemistry and Photobiology B: Biology* 165: 71–79.

103. Hubisz MJ, Falush D, Stephens M, Pritchard JK. 2009. Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources* 9: 1322–1332.

104. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, MendeDR , Letunic I, Rattei T, Jensen LJ et al. 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.

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

106. Jaramillo-Correa JP, Beaulieu J, Bousquet J. 2004. Variation in mitochondrial DNA reveals multiple distant glacial refugia in black spruce (*Picea mariana*), a transcontinental north American conifer. *Molecular Ecology* 13: 2735–2747.

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

108. Jeník J. 1994. Clonal growth in woody plants: A review. *Folia Geobotanica et Phytotaxonomical* 29: 291-306.

109. Jombart T, Ahmed I. 2011. adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics* 27: 3070–3071.

110. Jombart T, Pontier D, Dufour AB. 2009. Genetic markers in the playground of multivariate analysis. *Heredity* 102: 330–341.

111. Jombart T. 2008. adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* 24: 1403–1405.

112. Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, Swofford R, Pirun M, Zody MC, White S et al. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484: 55-61.

113. Junttila O. 1991. Gibberellins and the regulation of shoot elongation in woody plants. In: Takahashi N, Phinney BO, MacMillan J eds. Gibberellins. Springer: New York, USA; pp. 199–210.

114. Kates M, Pugh EL, Ferrante G. 1984. Regulation of membrane fluidity by lipid desaturases. In: Kates M, Manson LA eds. Biomembranes vol 12: Membrane fluidity. Plenum Press: New York, USA; pp. 379-395.

115. Keles Y, Everest A. 2008. Relation to altitude adaptation and antioxidant defence system in five shrubs and trees species from middle Taurus Mountains. *International Journal of Natural and Engineering Sciences* 2: 45–49.

116. Keller F, Korner C. 2003. The role of photoperiodism in alpine plant development. *Arctic, Antarctic and Alpine Research* 35: 361–368.

117. Kew and Missouri Botanical Gardens [online]. 2013. The Plant List: *Pinus sylvestris*. Version 1.1. Available from www.theplantlist.org [accessed 10 March 2021].

118. Kormuťák A, Galgoci M, Manka P, Koubova M, Jopcik M, Sukenikova D, Bolecek P, Gomory D. 2017. Field-based artificial crossings indicate partial compatibility of reciprocal crosses between *Pinus sylvestris* and *Pinus mugo* and unexpected chloroplast DNA inheritance. *Tree Genetics & Genomes* 13: 68.

119. Kormuťák A, Galgoci M, Sukeniková D, Bolecek P, Libantova J, Gomory D. 2018. Maternal inheritance of chloroplast DNA in *Pinus mugo* Turra: A case study of *Pinus mugo* x *Pinus sylvestris* crossing. *Plant Systematics and Evolution* 304: 71–76.

120. Kormuťák A, Ostrolucká M, Vooková B, Preťová A, Fečková M. 2005. Artificial hybridization of *Pinus sylvestris* L. and *Pinus mugo* Turra. *Acta Biologica Cracoviensia s. Botanica* 47: 129-134.

121. 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.

122. Kujala ST, Savolainen O. 2012. Sequence variation patterns along a latitudinal cline in Scots pine (*Pinus sylvestris*): Signs of clinal adaptation? *Tree Genetics & Genomes* 8: 1451–1467.

123. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* 33: 1870–1874.

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

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

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

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

128. Latalowa M, Tobolski K, Nalepka D. 2004. *Pinus* L. subgenus *Pinus* (subgen. *Diploxylon* (Koehne) Pilger) – Pine. In: Ralska-Jasiewiczowa M ed. Late Glacial and Holocene history of vegetation in Poland based on isopollen maps. W. Szafer Institute of Botany Polish Academy of Sciences: Krakow, Poland; pp. 165–177.

129. Lee SR, Gaskin JF, Kim YD. 2019. Molecular diagnosis for a *Tamarix* species from two reclaimed lands along the Yellow Sea in Korea inferred from genome wide SNP markers. *Journal of Systematics and Evolution* 57: 247–255.

130. Leigh JW, Bryant D. 2015. POPART: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution* 6: 1110–1116.

131. Lewandowski A, Boratyński A, Mejnartowicz L. 2000. Allozyme investigations on the genetic differentiation between closely related pines - *Pinus sylvestris*, *P. mugo*, *P. uncinata*, and *P. uliginosa* (Pinaceae). *Plant Systematics and Evolution* 221: 15–24.

132. 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.

133. 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.

134. Librado P, Rozas J. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452.

135. Liepelt S, Bialozyt R, Ziegenhagen B. 2002. Wind-dispersed pollen mediates postglacial gene flow among refugia. *Proceedings of the National Academy of Sciences USA* 99: 14590–14594.

136. Liepelt S, Cheddadi R, de Beaulieu JL, Fady B, Gomory D, Hussendorfer E, Konnert M, Litt T, Longauer R, Terhürne-Berson R et al. 2009. Postglacial range expansion and its genetic imprints in *Abies alba* (Mill.) — A synthesis from palaeobotanic and genetic data. *Review of Palaeobotany and Palynology* 153: 139–149.

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

138. Lischer HEL, Excoffier L. 2012. PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* 28: 298–299.

139. Little CHA, MacDonald JE. 2003. Effects of exogenous gibberellin and auxin on shoot elongation and vegetative bud development in seedlings of *Pinus sylvestris* and *Picea glauca*. *Tree Physiology* 23: 73–83.

140. Liu L, Hao ZZ, Liu YY, Wei XX, Cun YZ, Wang XQ. 2014. Phylogeography of *Pinus armandii* and its relatives: heterogeneous contributions of geography and climate changes to the genetic differentiation and diversification of Chinese white pines. *PLoS One* 9: e85920.

141. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15: 550.

142. Luo YH, Dong XW, Yu TY, Shi X, Li ZY, Yang WC, Widmer A, Karrenberg S. 2015. A Single nucleotide deletion in *gibberellin20-oxidase1* causes alpine dwarfism in *Arabidopsis*. *Plant Physiology* 168: 930–937.

143. Lutz C, Engel L. 2007. Changes in chloroplast ultrastructure in some high-alpine plants: Adaptation to metabolic demands and climate? *Protoplasma* 231: 183–192.

144. Ma L, Sun XD, Kong XX, Galvan JV, Li X, Yang SH, Yang YQ, Yang YP, Hu XY. 2015. Physiological, biochemical and proteomics analysis reveals the adaptation strategies of the alpine plant *Potentilla saundersiana* at altitude gradient of the Northwestern Tibetan Plateau. *Journal of Proteomics* 112: 63–82.

145. 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.

146. 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.

147. Manni M, Berkeley MR, Seppey M, Simão 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.

148. Mantel N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Research* 27: 209–220.

149. Marcysiak K, Boratyński A. 2007. Contribution to the taxonomy of *Pinus uncinata* (Pinaceae) based on cone characters. *Plant Systematics and Evolution* 264: 57–73.

150. Martinez I, Gonzalez-Taboada F, Wiegand T, Camarero JJ, Gutierrez E. 2012. Dispersal limitation and spatial scale affect model based projections of *Pinus uncinata* response to climate change in the Pyrenees. *Global Change Biology* 18: 1714–1724.

151. Mátyás C, Ackzell L, Samuel CJA. 2004. EUFORGEN Technical Guidelines for genetic conservation and use for Scots pine (*Pinus sylvestris*). International Plant Genetic Resources Institute: Rome, Italy.

152. 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. Springer: Heidelberg, Germany.

153. Mayr E. 1942. Systematics and the origin of species. Columbia University Press: New York, USA.

154. Mitić ZS, Jovanović SČ, Zlatković BK, Nikolić BM, Stojanović GS, Marin PD. 2017. Needle terpenes as chemotaxonomic markers in *Pinus*: subsections *Pinus* and *Pinaster*. *Chemistry & Biodiversity* 14: e1600453.

155. Mitić ZS, Nikolić JS, Zlatković BK, Milanovici SJ, Jovanović SČ, Nikolić BM, Stojanović GS, Marin PD. 2018. Epicuticular waxes provide insights into phytochemical differentiation of natural populations of *Pinus mugo* Turra sensu stricto. *Chemistry & Biodiversity* 15: e1800378.

156. 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.

157. Mosca E, Eckert AJ, Di Pierro EA, Rocchini D, La Porta N, Belletti P, Neale DB. 2012. The geographical and environmental determinants of genetic diversity for four alpine conifers of the European Alps. *Molecular Ecology* 21: 5530–5545.

158. 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.

159. 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.

160. Nakamura Y, Koizumi R, Shui GH, Shimojima M, Wenk MR, Ito T, Ohta H. 2009. *Arabidopsis* lipins mediate eukaryotic pathway of lipid metabolism and cope critically with phosphate starvation. *Proceedings of the National Academy of Sciences USA* 106: 20978–20983.

161. Nakamura Y. 2013. Phosphate starvation and membrane lipid remodeling in seed plants. *Progress in Lipid Research* 52: 43–50.

162. Nakatsu F, Ohno H. 2003. Adaptor protein complexes as the key regulators of protein sorting in the post-Golgi network. *Cell Structure and Function* 28: 419–429.

163. Nathan R, Muller-Landau HC. 2000. Spatial patterns of seed dispersal, their determinants and consequences for recruitment. *Trends in Ecology and Evolution* 15: 278–285.

164. Naydenov K, Senneville S, Beaulieu J, Tremblay F, Bousquet J. 2007. Glacial vicariance in Eurasia: mitochondrial DNA evidence from Scots pine for a complex heritage involving genetically distinct refugia at mid-northern latitudes and in Asia Minor. *BMC Evolutionary Biology* 7: 233.

165. Neale DB, Martínez-García 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.

166. Neale DB, Wegrzyn JL, Stevens KA, Zimin AV, Puiu D, Crepeau MW, Cardeno C, Koriabine M, Holtz-Morris AE, Liechty JD et al. 2014. Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome Biology* 15: R59.

167. Neale DB, Zimin AV, Zaman S, Scott AD, Shrestha B, Workman RE, Puiu D, Allen BJ, Moore ZJ, Sekhwal MK et al. 2022. Assembled and annotated 26.5 Gbp coast redwood genome: A resource for estimating evolutionary adaptive potential and investigating hexaploid origin. *G3: Genes, Genomes, Genetics* 12: jkab380.

168. Nicotra AB, Chong C, Bragg JG, Ong CR, Aitken NC, Chuah A, Lepschi B, Borevitz JO. 2016. Population and phylogenomic decomposition via genotyping-by-sequencing in Australian Pelargonium. *Molecular Ecology* 25: 2000–2014.

169. Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin YC, Scofield DG, Vezzi F, Delhomme N, Giacomello S, Alexeyenko A et al. 2013 The Norway spruce genome sequence and conifer genome evolution. *Nature* 497: 579–584.

170. Obidowicz A. 1996. A Late Glacial-Holocene history of the formation of vegetation belts in the Tatra Mts. *Acta Paleobotanica* 36: 159–206.

171. Ohsawa T, Ide Y. 2008. Global patterns of genetic variation in plant species along vertical and horizontal gradients on mountains. *Global Ecology and Biogeography* 17: 152–163.

172. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D et al. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research* 44: D733-745.

173. Oleksyn J, Modrzyński J, Tjoelker MG, Żytkowiak R, Reich PB, Karolewski P. 1998. Growth and physiology of *Picea abies* populations from elevational transects: Common garden evidence for altitudinal ecotypes and cold adaptation. *Functional Ecology* 12: 573–590.

174. Oncel I, Yurdakulol E, Keles Y, Kurt L, Yildiz A. 2004. Role of antioxidant defense system and biochemical adaptation on stress tolerance of high mountain and steppe plants. *Acta Oecologica* 26: 211–218.

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

176. 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.

177. 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.

178. Peakall R, Smouse PE. 2006. GenAlEx 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6: 288–295.

179. Peakall R, Smouse PE. 2012. GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research - an update. *Bioinformatics* 28: 2537–2539.

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

181. Perry A, Wachowiak W, Beaton J, Iason G, Cottrell J, Cavers S. 2022. Identifying and testing marker-trait associations for growth and phenology in three pine species: Implications for genomic prediction. *Evolutionary Applications* 15: 330-348.

182. 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.

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

184. Petrova EA, Zhuk EA, Popov AG, Bondar AA, Belokon MM, Goroshkevich SN, Vasilyeva GV. 2018. Asymmetric introgression between *Pinus sibirica* and *Pinus pumila* in the Aldan plateau (Eastern Siberia). *Silvae Genetica* 67: 66–71.

185. Phinney BO. 1985. Gibberellin $A_1$ dwarfism and shoot elongation in higher plants. *Biologia Plantarum* 27: 172–179.

186. Pinosio S, González-Martínez SC, Bagnoli F, Cattonaro F, Grivet D, Marroni F, Lorenzo Z, Pausas JG, Verdú 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.

187. Plomion C, Bastien C, Bogeat-Triboulot MB, Bouffier L, Déjardin A, Duplessis S, Fady B, Heuertz M, Le Gac AL, Le Provost G et al. 2016. Forest tree genomics: 10 achievements from the past 10 years and future prospects. *Annals of Forest Science* 73: 77-103.

188. Polle A, Baumbusch LO, Oschinski C, Eiblmeier M, Kuhlenkamp V, Vollrath B, Scholz F, Rennenberg H. 1999. Growth and protection against oxidative stress in young clones and mature spruce trees (*Picea abies* L.) at high altitudes. *Oecologia* 121: 149–156.

189. Polle A, Rennenberg H. 1992. Field studies on Norway spruce trees at high-altitudes: II. Defense systems against oxidative stress in needles. *New Phytologist* 121: 635–642.

190. Pons O, Petit RJ. 1996. Measuring and testing genetic differentiation with ordered versus unordered alleles. *Genetics* 144: 1237–1245.

191. Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.

192. Proost S, Van Bel M, Vaneechoutte D, Van de Peer Y, Inzé D, Mueller-Roeber B, Vandepoele K. 2015. PLAZA 3.0: An access point for plant comparative genomics. *Nucleic Acids Research* 43: D974–D981.

193. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ et al. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* 81: 559–575.

194. Pyhäjärvi T, Kujala ST, Savolainen, O. 2019. 275 years of forestry meets genomics in *Pinus sylvestris*. *Evolutionary Applications* 13: 11– 30.

195. Pyhäjärvi T, Salmela M J, Savolainen O. 2008. Colonization routes of *Pinus sylvestris* inferred from distribution of mitochondrial DNA variation. *Tree Genetics & Genomes* 4: 247-254.

196. R Core Team [online]. 2019. R: A language and environment for statistical computing. Available from www.r-project.org [accessed 22 July 2019].

197. Ramasamy RK, Ramasamy S, Bindroo BB, Naik VG. 2014. STRUCTURE PLOT: A program for drawing elegant STRUCTURE bar plots in user friendly interface. *Springerplus* 3: 431.

198. Ramond L. 1805. In: De Candolle AP & Lamarck J-B eds. Flore Française. Troisième Édition 3: 726.

199. 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.

200. Ripley B, Venables B, Bates DM, Hornik K, Gebhardt A, Firth D [online]. 2020. MASS. Available from www.cran.r-project.org/package=MASS [accessed 12 October 2020].

201. 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.

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

203. Robledo-Arnuncio JJ. 2011. Wind pollination over mesoscale distances: An investigation with Scots pine. *New Phytologist* 190: 222–233.

204. RStudio Team [online]. 2018. RStudio: Integrated development for R. Available from www.rstudio.com [accessed 22 July 2019].

205. Sandoz H. 1987. Recerches taxonomiques, biogéographiques et phytoécologiques sur les principaux conifères subalpins des Alpes: Mélèze d'Europe, Pin cembro, Pin à crochets et Pin mugho. Ph.D. dissertation at Université Paul Cézanne: Marseille, France.

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

207. Schomburg FM, Bizzell CM, Lee DJ, Zeevaart JAD, Amasino RM. 2003. Overexpression of a novel class of gibberellin 2-oxidases decreases gibberellin levels and creates dwarf plants. *Plant Cell* 15: 151–163.

208. Semerikov VL, Putintseva YA, Oreshkova NV, Semerikova SA, Krutovsky KV. 2015. Development of new mitochondrial DNA markers in Scots pine (*Pinus sylvestris* L.) for population and phylogeographic studies. *Russian Journal of Genetics* 51: 1199–1203.

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

210. Sinclair WT, Morman JD, Ennos RA. 1998. Multiple origins for Scots pine (*Pinus sylvestris* L.) in Scotland: evidence from mitochondrial DNA variation. *Heredity* 80: 233–240.

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

212. Smith CCR, Tittes S, Mendieta JP, Collier-Zans E, Rowe HC, Rieseberg LH, Kane NC. 2018. Genetics of alternative splicing evolution during sunflower domestication. *Proceedings of the National Academy of Sciences USA* 115: 6768-6773.

213. Smith WK, Johnson DM, Reinhardt K. 2008. Alpine forest. In: Jørgensen SE, Fath BD eds. Encyclopedia of Ecology. Academic Press: Oxford, UK; pp. 144–153.

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

215. Soranzo N, Alia R, Provan J, Powell W. 2000. Patterns of variation at a mitochondrial sequence-tagged-site locus provides new insights into the postglacial history of European *Pinus sylvestris* populations. *Molecular Ecology* 9: 1205–1211.

216. Sork VL. 2016. Gene flow and natural selection shape spatial patterns of genes in tree populations: implications for evolutionary processes and applications. Evolutionary Applications 9: 291-310.

217. Soto A, Robledo-Arnuncio JJ, Gonzalez-Martinez SC, Smouse PE, Alia R. 2010. Climatic niche and neutral genetic diversity of the six Iberian pine species: A retrospective and prospective view. *Molecular Ecology* 19: 1396–1409.

218. Stocklin J, Kuss P, Pluess AR. 2009. Genetic diversity, phenotypic variation and local adaptation in the alpine landscape: Case studies with alpine plant species. *Botanica Helvetica* 119: 125–133.

219. Streb P, Shang W, Feierabend J, Bligny R. 1998. Divergent strategies of photoprotection in high-mountain plants. *Planta* 207: 313–324.

220. 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.

221. Supek F, Bošnjak M, Škunca N, Šmuc T. 2011. REVIGO summarizes and visualizes long lists of Gene Ontology terms. *PLoS ONE* 6: e21800.

222. Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF. 1998. Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* 7: 453–464.

223. Takahashi K, Yoshida S. 2009. How the scrub height of dwarf pine *Pinus pumila* decreases at the treeline. *Ecological Research* 24: 847–854.

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

225. Turra A. 1764. Dei Vegetabili di Monte Baldo. *Giornale d'Italia, spettante alla Scienza naturale, e principalmente all' Agricoltura, alle Arti, ed al Commercio* 1: 152.

226. Vaattovaara A, Brandt B, Rajaraman S, Safronov O, Veidenberg A, Luklová M, Kangasjärvi J, Löytynoja A, Hothorn M, Salojärvi J et al. 2019. Mechanistic insights into the evolution of DUF26-containing proteins in land plants. *Communications Biology* 2: 56.

227. Voronova A, Belevich V, Korica A, Rungis D. 2017. Retrotransposon distribution and copy number variation in gymnosperm genomes. *Tree Genetics & Genomes* 13: 88.

228. Wachowiak W, Balk PA, Savolainen O. 2009. Search for nucleotide diversity patterns of local adaptation in dehydrins and other cold-related candidate genes in Scots pine (*Pinus sylvestris* L.). *Tree Genetics & Genomes* 5: 117–132.

229. Wachowiak W, Boratyńska 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.

230. Wachowiak W, Leśniewicz K, Odrzykoski IJ, Augustyniak H, Prus-Głowacki 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.

231. Wachowiak W, Palmé 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.

232. 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.

233. Wachowiak W, Perry A, Zaborowska J, González-Martínez SC Cavers S. 2022. Admixture and selection patterns across the European distribution of Scots pine, *Pinus sylvestris* (Pinaceae). *Botanical Journal of the Linnean Society*: boac016.

234. Wachowiak W, Prus-Głowacki W. 2008. Hybridisation processes in sympatric populations of pines *Pinus sylvestris* L., *P. mugo* Turra and *P. uliginosa* Neumann. *Plant Systematics and Evolution* 271: 29–40.

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

236. Wachowiak W, Zaborowska J, Łabiszak B, Perry A, Zucca GM, González-Martínez SC, Cavers S. 2018. Molecular signatures of divergence and selection in closely related pine taxa. *Tree Genetetics & Genomes* 14: 83.

237. Wachowiak W, Żukowska WB, Wójkiewicz B, Cavers S, Litkowiec M. 2016. Hybridization in contact zone between temperate European pine species. *Tree Genetics & Genomes* 12: 48.

238. Wagner S, Litt T, Sanchez-Goni MF, Petit RJ. 2015. History of *Larix decidua* Mill. (European larch) since 130 ka. *Quaternary Science Reviews* 124: 224–247.

239. Walther GR, Beissner S, Burga CA. 2005. Trends in the upward shift of alpine plants. *Journal of Vegetation Science* 16: 541–548.

240. Wang BS, Wang XR. 2014. Mitochondrial DNA capture and divergence in *Pinus* provide new insights into the evolution of the genus. *Molecular Phylogenetics and Evolution* 80: 20–30.

241. Wang H, Prentice IC, Davis TW, Keenan TF, Wright IJ, Peng CH. 2017. Photosynthetic responses to altitude: An explanation based on optimality principles. *New Phytologist* 213: 976–982.

242. Wang X-R, Szmidt AE, Lu M-Z. 1996. Genetic evidence for the presence of cytoplasmic DNA in pollen and megagametophytes and maternal inheritance of mitochondrial DNA in *Pinus*. *Forest Genetics* 3: 37-44.

243. Weigand H, Leese F. 2018. Detecting signatures of positive selection in non-model species using genomic data. *Zoological Journal of the Linnean Society of London* 184: 528–583.

244. Wieser G, Tausz M. 2007. Trees at their upper limit: Treelife limitation at the alpine timberline. Springer GmbH: Heidelberg, Germany.

245. Willi Y, Van Buskrik J, Hoffmann AA. 2006. Limits to the adaptive potential of small populations. *Annual Review of Ecology, Evolution and Systematics* 37: 433–458.

246. Willis JK, van Andel TH. 2004. Trees or no trees? The environments of central and eastern Europe during the Last Glaciation. *Quaternary Science Reviews* 23: 2369–2387.

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

248. Wolfe KH, Li WH, Sharp PM. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences USA* 84: 9054–9058.

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

250. Wrzaczek M, Brosché M, Salojärvi J, Kangasjärvi S, Idänheimo N, Mersmann S, Robatzek S, Karpiński S, Karpińska B, Kangasjärvi 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.

251. Xu N, Gao XQ, Zhao XY, Zhu DZ, Zhou LZ, Zhang XS. 2011. *Arabidopsis* AtVPS15 is essential for pollen development and germination through modulating phosphatidylinositol 3-phosphate formation. *Plant Molecular Biology* 77: 251–260.

252. 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.

253. Zaborowska J, Łabiszak 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.

254. Zaborowska J, Łabiszak 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.

255. 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 *Eucalyptus* species. *BMC Plant Biology* 22: 1.

256. 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.

257. Zhang ZL, Ogawa M, Fleet CM, Zentella R, Hu JH, Heo JO, Lim J, Kamiya Y, Yamaguchi S, Sun TP. 2011. SCARECROWLIKE 3 promotes gibberellin signaling by antagonizing master growth repressor DELLA in *Arabidopsis. Proceedings of the National Academy of Sciences USA* 108: 2160–2165.

258. Zhou XM, Zhao P, Wang W, Zou J, Cheng TH, Peng XB, Sun MX. 2015. A comprehensive, genome-wide analysis of autophagyrelated genes identified in tobacco suggests a central role of autophagy in plant response to various environmental cues. *DNA Research* 22: 245–257.

259. Zhou YF, Zhang LR, Liu JQ, Wu GL, Savolainen O. 2014. Climatic adaptation and ecological *divergence between two closely related pine species in Southeast China. Molecular* Ecology 23: 3504–3522.

260. Zimin AV, Stevens KA, Crepeau MW, Puiu D, Wegrzyn JL, Yorke JA, Langley CH, Neale DB, Salzberg SL. 2017. An improved assembly of the loblolly pine mega-genome using long-read single-molecule sequencing. *Gigascience* 6: 1–4.

261. Żukowska WB, Boratyńska K, Wachowiak W. 2017. Comparison of range-wide chloroplast microsatellite and needle trait variation patterns in *Pinus mugo* Turra (dwarf mountain pine). *iForest* 10: 250–258.

262. Żukowska WB, Wachowiak W. 2017. Nuclear microsatellite markers reveal the low genetic structure of *Pinus mugo* Turra (dwarf mountain pine) populations in Europe. *Plant Systematics and Evolution* 303: 641–651.

## 3. ROZPRAWA DOKTORSKA

### oryginalne publikacje wchodzące w skład dysertacji

Lista publikacji wchodzących w skład rozprawy:

1. Zaborowska J, Łabiszak 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

2. Zaborowska J, Łabiszak 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

3. Zaborowska J, Perry A, Cavers S, Wachowiak W. 2022. **Evolutionary targets of gene expression divergence in a complex of closely related pine species**. *Journal of Systematics and Evolution* [in press, DOI: https://doi.org/10.1111/jse.12896].

## 3.1. PUBLIKACJA PIERWSZA

Julia Zaborowska, Bartosz Łabiszak & Witold Wachowiak

# Population history of European mountain pines *Pinus mugo* and *Pinus uncinata* revealed by mitochondrial DNA markers

*Journal of Systematics and Evolution* 58: 474-486

https://doi.org/10.1111/jse.12520

2019

JSE Journal of Systematics and Evolution

doi: 10.1111/jse.12520

Research Article

# Population history of European mountain pines *Pinus mugo* and *Pinus uncinata* revealed by mitochondrial DNA markers

Julia Zaborowska[1*], Bartosz Łabiszak[1], and Witold Wachowiak[1,2]

[1]Institute of Environmental Biology, Adam Mickiewicz University in Poznan, Uniwersytetu Poznanskiego 6, 61-614 Poznan, Poland
[2]Institute of Dendrology, Polish Academy of Sciences, Parkowa 5, 62-035 Kornik, Poland
*Author for correspondence. E-mail: julia.zaborowska@amu.edu.pl
Received 15 March 2019; Accepted 22 May 2019; Article first published online 6 June 2019

**Abstract** The dwarf mountain pine (*Pinus mugo*) and the Pyrenean pine (*P. uncinata*) constitute a pair of closely related coniferous taxa of poorly resolved evolutionary history and affinity, which inhabit numerous stands scattered over subalpine environments of European mountain ranges. The aim of the study was to investigate their phylogeography and mutual relationships, shedding new light on their taxonomy and the past of the alpine flora. Previous evolutionary reconstructions of the mountain pines relied mainly on bi-parentally or paternally inherited markers that quickly homogenize between populations, showing rather shallow and recent differentiation of gene pools. Therefore, to contrast these pictures, we analyzed diversity and differentiation within a large set of new mitochondrial loci, inherited in maternal line and distributed by seeds at short geographical distances. Samples of the taxa were taken from 27 natural populations representing their range-wide distributions—17 populations of *P. mugo* and 10 of *P. uncinata*. All markers appeared polymorphic, providing a total of 31 multilocus haplotypes. Two of the loci proved to be species-diagnostic and nearly fixed between analyzed samples. Distribution of mitotypes indicate that allopatric populations of the taxa constitute separate mitochondrial haplogroups, and the two mountain pines have independent evolutionary history. However, introgression of *P. mugo* mitotypes by *P. uncinata* specimens revealed in the species contact zone in Western Alps shows that their speciation is not fully completed.

**Key words:** introgression, mitochondrial DNA, phylogeography, *Pinus mugo*, *Pinus uncinata*, subalpine taxa.

## 1 Introduction

Trees influence ecosystems through multiple biotic and abiotic interactions. However, the patterns and processes shaping changes in their distribution ranges, and those that influence their diversity and divergence at heritable phenotypic traits are still poorly understood. Establishment and maintenance of local adaptations is particularly interesting in mountain trees due to their features and habitat specificities. Many tree species, including pines, are characterized by large effective population sizes, outcrossing, and considerable dispersion potential of pollen and seeds, which contribute to the rapid homogenization of gene pools between populations (Petit & Hampe, 2006). At the same time, they maintain high fertility and strong selection between recruited individuals, facilitating adaptation to local conditions. Moreover, in mountain habitats, populations are subjected to rapid changes in environmental factors in the elevation gradient, and their ranges usually have an insular character. These features affect gene exchange and promote diversity between stands, but limit variation within them (Ohsawa & Ide, 2008). In search for the signatures of selection at molecular level, the first step is to learn about the variation of neutral genetic background that results from species population history.

The European mountain chains constitute habitats for a variety of plant taxa including coniferous dwarf mountain pine (*Pinus mugo* Turra) and its close relative—mountain or Pyrenean pine (*P. uncinata* Ramond). The two differ considerably in phenotype, however, they are not always considered as separate species and are often classified into a broader group recognized as the *P. mugo* complex (Christensen, 1987; Hamernik & Musil, 2007; Businsky & Kirschner, 2010). The complex split from *P. sylvestris* at the end of Tertiary and begun to diverge about 5 million years BP (Wachowiak et al., 2011). Since then, insufficient lineage sorting delayed by long life-span of pines and relatively small number of generations that passed, together with possible hybridization events between populations coming into secondary contacts during Pleistocene glacial and interglacial periods, had blurred their later quaternary history (Boratynska et al., 2014).

In the present study, we focused on *P. mugo* and *P. uncinata* as two well-recognized entities of the *P. mugo* complex, both restricted to subalpine habitats and facing similar environmental pressures. The dwarf mountain pine is

a shrubby, polycormic conifer up to 3 m high (Monteleone et al., 2006; Hamernik & Musil, 2007). Species distribution ranges from Western Alps, through Sudetes and Carpathians in Central Europe, to the Balkans and Apennines on the south. These shrubs inhabit harsh high-altitude stands above upper forest limits (Alexandrov et al., 2019). The Pyrenean pine is erect, 4 to 25 m tall, usually monocormic. It inhabits Western parts of the *P. mugo* complex range, mainly subalpine areas (1400–2700 m a.s.l.) in Pyrenees and Western Alps, with smaller endemic populations within Iberian Peninsula, Massif Central, Ligurian Apennines, Jura and Vosges (Jalas & Suominen, 1973; Boratynska et al., 2015). It grows on rocks and debris, where it forms more or less dense forests and sets the upper tree line.

The debate regarding taxonomic position of *P. mugo* and *P. uncinata* is vivid (Hamernik & Musil, 2007; Businsky & Kirschner, 2010). The two pines share common ancestry that manifests in similar karyotype structure of $2n = 24$ chromosomes (Grotkopp et al., 2004) and high genetic similarity of molecular and biochemical markers (Lewandowski et al., 2000; Monteleone et al., 2006; Wachowiak et al., 2013). They show incomplete reproductive isolation, can be artificially crossed and are able to hybridize in contact zones (Monteleone et al., 2006; Wachowiak & Prus-Glowacki, 2008). However, they differ significantly in a number of characteristics. Their geographic distributions are generally disjunctive, they express other habits and show differentiation at phenology and growth traits under standardized conditions (Wachowiak et al., 2018). They differ by a set of more discreet morphological (e.g., Sandoz, 1987 after Boratynska & Bobowicz, 2001; Monteleone et al., 2006; Marcysiak & Boratynski, 2007; Boratynska et al., 2015) and karyotype (Bogunic et al., & Medjedovic, 2011) characteristics, but also in allelic frequency variation at diverse types of biochemical and molecular markers (Lewandowski et al., 2000; Monteleone et al., 2006; Wachowiak et al., 2015; Dzialuk et al., 2017).

The main goal of this study was to obtain insights into the phylogeographic structure and history of the range-wide sampled populations of the two species. Such reconstruction of the present structure and the demographic past of populations is needed to bring valuable information for genetic studies on their mutual relationships and inner variation, including information about neutral background of these populations, which is essential for studies of molecular basis of their local adaptations. Research on this particular pine example can also help to better understand the impact of past and future environmental changes on other species currently restricted to higher mountain habits. For these reasons, we looked at the variation in mitochondrial DNA (mtDNA) that is characterized by lower effective population sizes, making it more prone to differentiation by drift and better suited to depict demographical changes as compared to nuclear markers. Furthermore, mtDNA regions proved to be effective in population structure investigations of pines as their mitogenomes are inherited in maternal line and transmitted by seeds on shorter distances compared to pollen mediated markers (Liepelt et al., 2002; Robledo-Arnuncio, 2011; Liu et al., 2014; Wang & Wang, 2014). So far, the attempts for development of high-resolution markers in *mt*DNA that would be suitable for investigations in this group of pines were limited by the lack of access to larger fragments of the genome and its low base mutation rate (Wolfe et al., 1987; Sinclair et al., 1998; Soranzo et al., 2000; Semerikov et al., 2015).

Here we took advantage of a new set of mitochondrial polymorphic regions described in these pines (Donnelly et al., 2017), to scan neutral genetic variation of their scattered populations. We were interested in answering question if the taxa share gene pool of mtDNA markers. We wanted to identify putative regions of restricted gene flow between stands, and follow traces of horizontal and vertical shifts during the climatic oscillations of Pleistocene and the later warming in Holocene. Populations of mountain species can often be isolated even within large massifs and the seed dispersal capabilities are the crucial for new settlement establishment or repopulation (Nathan & Muller-Landau, 2000). Therefore, in discussing results, we also paid attention to a few more remote populations of *P. mugo* and *P. uncinata*, isolated by at least 100 km from other species stands, with practically no probability of current seed transfer and little inflow of pollen (Robledo-Arnuncio, 2011; Alexandrov et al., 2019). Although the region of sympatry was not of first importance for us, we included one stand from the Western Alps for each species to recognize the identity of their matrilineal backgrounds.

## 2 Material and Methods

### 2.1 Sample collection and DNA extraction

Four hundred forty-nine samples from 17 natural populations of *Pinus mugo* and 10 populations of *P. uncinata* (8–22 and 12–22 individuals per population, respectively) were examined, covering most of the species ranges (Table 1 and Fig. 1). Sampled individuals were separated at a distance of at least 30 m. Genomic DNA was isolated from needles with DNeasy Plant Mini Kit (Qiagen) and its quality was evaluated using BioPhotometer (Eppendorf).

### 2.2 Genotyping methods

A set of 17 mtDNA markers was used including two insertion-deletion (IN-DEL) loci *nad1 intron B/C* (Soranzo et al., 2000) and *nad7 intron 1* (Jaramillo-Correa et al., 2004; Naydenov et al., 2007), and 15 newly developed based on polymorphism described by Donnelly et al. (2017): one IN-DEL locus (*Pr34*) and 14 single nucleotide polymorphisms (SNPs; Tables S1 and S2). New variants were genotyped by PCR-RFLP technique, except region *Pr34* sequenced with Sanger's method. Standardized 15 µL polymerase chain reaction (PCR) reaction mix contained 1× PCR buffer, 1× bovine serum albumin, 1.5 µmol/L $MgCl_2$, 40 µmol/L dNTP mix, 0.2 µmol/L of both forward and reverse primers (Genomed, Poland), 0.15 U Taq DNA polymerase (Novazym, Poland) and about 15 ng of DNA template. Reaction conditions were unified with 3 minutes of initial denaturation at 94 °C, 35 cycles with 30-second denaturation at 94 °C, 30-second annealing at 57 °C and 1 minute 20-second elongation at 72 °C, with the final extension for 5 minutes at 72 °C. Regions *nad1* and *nad7* were amplified according to methods described in Soranzo et al. (2000) and Jaramillo-Correa et al. (2004), respectively.

**Table 1** Locations of 27 studied populations of *Pinus mugo* and *P. uncinata* from European mountains with estimates of nucleotide and haplotype diversity at 17 mitochondrial DNA markers

| Taxon | Acr. | Location (lat./long./alt.) | *N* | *S* | $S_p$ | *d* | $H_n$ | $H_p$ | $H_d$ (SD) | $_uH$ |
|---|---|---|---|---|---|---|---|---|---|---|
| *Pinus mugo* | M1 | Sudetes, Slaskie Kamienie & Czarny Kociol Jagniatkowski (50°46'35"/15°36'08"/1300) | 10 | 1 | 0 | 0.200 | 2 | 1 | 0.200 (0.154) | 0.222 |
| | M2 | Tatra Mountains, Piec Stawow Polskich (49°13'09"/20°03'05"/1700) | 11 | 11 | 0 | 3.382 | 4 | 1 | 0.673 (0.123) | 0.740 |
| | M3 | Lower Tatra Mts., Lysa above Jasna (49°00'43"/19°39'35"/1500) | 12 | 9 | 0 | 3.833 | 3 | 1 | 0.712 (0.069) | 0.777 |
| | M4 | Gorgany Mountains, Osmoloda (48°40'33"/23°55'19"/830) | 6 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| | M5 | Muntii Rodnei, Pasul Prislop (47°34'03"/24°48'00"/1720) | 13 | 1 | 0 | 0.154 | 2 | 1 | 0.154 (0.126) | 0.167 |
| | M6 | Muntii Bucegi, Busteni (45°25'55"/25°27'06"/2070) | 13 | 5 | 1 | 2.564 | 4 | 2 | 0.769 (0.072) | 0.833 |
| | M7 | Rila Mountains, Belica (42°04'01"/23°30'00"/2100) | 12 | 2 | 0 | 1.076 | 3 | 0 | 0.621 (0.087) | 0.678 |
| | M8 | Pirin Mountains, Vikhren (41°46'07"/23°25'22"/2000) | 17 | 2 | 0 | 0.618 | 2 | 0 | 0.309 (0.122) | 0.328 |
| | M9 | Durmitor Mountains, Zhabjak (43°09'33"/19°05'27"/2100) | 18 | 2 | 0 | 0.634 | 3 | 0 | 0.582 (0.061) | 0.616 |
| | M10 | Bjelasnica Mountain near Sarajevo (43°45'00"/18°13'08"/2120) | 11 | 2 | 0 | 0.873 | 4 | 0 | 0.600 (0.154) | 0.660 |
| | M11 | Kamnik–Savinja Alps, Kamniska Bistrica (46°05'10"/14°37'51"/1600) | 10 | 9 | 0 | 3.378 | 3 | 1 | 0.622 (0.138) | 0.691 |
| | M12 | Karnishe Alps, Passo di Pramollo (46°32'45"/13°15'35"/1530) | 12 | 2 | 0 | 0.333 | 3 | 1 | 0.318 (0.164) | 0.347 |
| | M13 | Bavarian Alps, Kreuzspitze (47°31'30'/10°55'12"/1870) | 19 | 2 | 0 | 0.737 | 3 | 0 | 0.632 (0.073) | 0.667 |
| | M14 | Karwendel Gebirge, Scharnitz (47°22'42"/11°17'45"/1400) | 10 | 1 | 0 | 0.356 | 2 | 0 | 0.356 (0.159) | 0.395 |
| | M15 | Dolomites, Brenta Group, Lago di Tovel (46°15'39"/10°56'46"/1200) | 12 | 5 | 0 | 1.833 | 6 | 3 | 0.879 (0.060) | 0.959 |
| | M16 | Maritime Alps, Coll de Tende (44°08'00"/07°22'30"/2000) | 11 | 3 | 0 | 1.055 | 4 | 0 | 0.745 (0.098) | 0.820 |
| | M17 | Abruzzi Mountains, La Maiella (41°46'20"/13°58'30"/2200) | 15 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| *Pinus mugo* overall | | | 212 | 15 | 11 | 2.618 | 23 | 22 | 0.863 (0.018) | 0.867 |
| *Pinus uncinata* | U1 | Cottian Alps, Passo del Monginevro, Claviere (44°55'58"/06°44'10"/1800) | 11 | 3 | 0 | 1.382 | 3 | 1 | 0.691 (0.086) | 0.760 |
| | U2 | Massif Central, Col de la Croix de Morand (45°36'00"/02°50'59"/1400) | 21 | 6 | 0 | 1.448 | 7 | 2 | 0.790 (0.060) | 0.830 |
| | U3 | East Pyrenees, Col de Jau (42°39'19"/02°15'22"/1520) | 12 | 1 | 0 | 0.167 | 2 | 0 | 0.167 (0.134) | 0.182 |
| | U4 | East Pyrenees, Vall de Nuria (42°20'45"/02°06'15"/2200) | 12 | 1 | 0 | 0.409 | 2 | 0 | 0.409 (0.133) | 0.446 |
| | U5 | East Pyrenees, Vall de Ransol (42°35'02"/01°38'21"/2025) | 17 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| | U6 | East Pyrenees, San Miguel de Engolasters (42°31'28"/01°34'12"/2000) | 18 | 2 | 0 | 0.405 | 3 | 1 | 0.386 (0.128) | 0.408 |
| | U7 | Central Pyrenees, Port de la Bonaiqua (42°39'48"/00°57'44"/2100) | 12 | 1 | 0 | 0.409 | 2 | 0 | 0.409 (0.133) | 0.446 |
| | U8 | Central Pyrenees, Benasque (42°37'58"/00°39'44"/2000) | 12 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| | U9 | Western Pyrenees, La Trapa near Jaca (42°41'19"/00°32'12"/1720) | 22 | 1 | 0 | 0.368 | 2 | 0 | 0.368 (0.100) | 0.385 |
| | U10 | Sierra de Gudar, Valdelinares (40°28'49"/00°41'51'/2000) | 21 | 1 | 0 | 0.429 | 2 | 0 | 0.429 (0.089) | 0.450 |
| *Pinus uncinata* overall[†] | | | 147 | 6 | 2 | 0.838 | 8 | 7 | 0.549 (0.041) | 0.553 |
| Total | | | 370 | 17 | na | 2.961 | 31 | na | 0.879 (0.010) | 0.881 |

Acr., population acronym; alt., altitude [m a.s.l.]; *d*, average number of nucleotide differences between two sequences; $H_d$, haplotype diversity; $H_n$, number of haplotypes; $H_p$, number of private haplotypes; $_uH$, unbiased gene diversity; lat., latitude; long., longitude; *N*, number of individuals genotyped; na, not applied; *S*, number of segregating sites; SD, standard deviation; $S_p$, number of private polymorphisms; Tax., taxon name.

[†]Population U1 was excluded from calculations. The samples were provided by Krystyna Boratynska from the Institute of Dendrology, Polish Academy of Sciences.

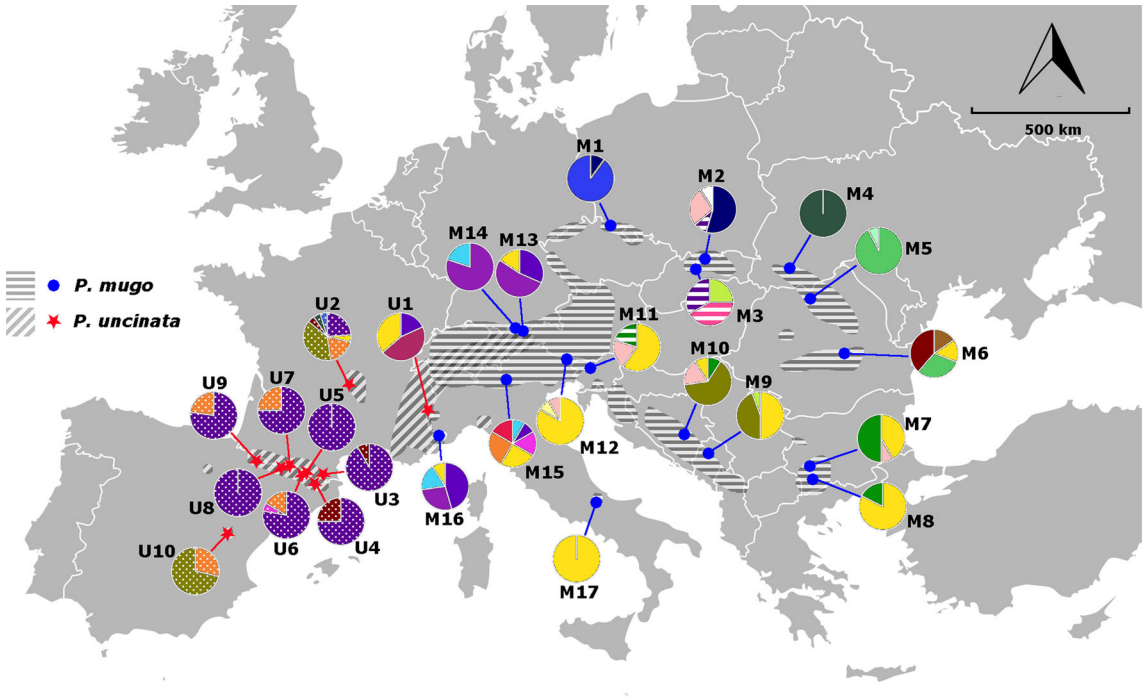



**Fig. 1.** Distribution of 27 sampled populations of *Pinus mugo* (blue dots) and *P. uncinata* (red stars) in Europe with mitochondrial DNA haplotype frequencies. Acronyms as in Table 1, circle sizes are standardized and do not account for population sampling, shading corresponds to the colour-code used on Fig. 2. Geographic ranges of the species after Jalas & Suominen (1973, modified), map from Wikimedia Commons.

PCR-RFLP genotyped markers were digested in 10 µL reaction volume, containing 2× buffer, 2.5 U of relevant restriction enzyme (Table S2) and 3 µL of PCR product. Samples were incubated for at least 10 hours, following manufacturer conditions for respective enzymes (Thermo Fisher Scientific) and electrophoretically separated on 1.5–2% agarose gel, stained with GelRed in UV. For sequencing demands, *Pr34* amplification products were purified by Exonuclease I (New England Biolabs) and TSAP Thermosensitive Alkaline Phosphatase (Promega). About 20 ng of purified template was used in 10 µL sequencing reaction with Big Dye Terminator DNA Sequencing Kit (Applied Biosystems) conducted by Genomed. The sequences were aligned and edited using CodonCode Aligner (CodonCode Corporation).

### 2.3 Diversity and differentiation measures

Both types of polymorphisms, SNPs and IN-DELs, were coded as single-mutation events, and distance estimates were measured based on raw numbers of differences. For tests based on haplotype information, we constructed individual mitotypes by concatenation of alleles identified within each marker. Connections among obtained haplotypes were analyzed by median joining network constructed in PopART 1.7 (Bandelt et al., 1999; Leigh & Bryant, 2015).

For each population, species, and the whole probe, we calculated the number of segregating sites ($S$), number of private variants ($S_p$), and mean within group distance ($d$) in DnaSP 5 (Librado & Rozas, 2009) and MEGA 7 (Kumar et al., 2016). Similarly, we counted the number of haplotypes ($N_h$) and private haplotypes ($N_p$) present in populations and species. To better assess the scale of genetic variation segregating in the taxa, haplotype frequencies and population gene diversities ($H_d$) were estimated in DnaSP.

Based on the mean between population distances ($d_{xy}$) calculated in MEGA, three analyses were performed to find best pattern of genetic relationships among samples. First, the unweighted pair group method with arithmetic mean (UPGMA) clustering of populations was conducted in MEGA, to see if there is any tendency of populations to group by species identities or by geographical locations. Second, the principal coordinate analysis (PCoA) implemented in GenAlEx 6.503 software (Peakall & Smouse, 2006; Peakall & Smouse, 2012) was used separately for populations within the two species and on the whole data set level, to see if the inter-population affinity changes when they are analyzed in a broader context. Finally, to verify the hypothesis of isolation by distance, Mantel tests of the correlation between genetic and geographic distances among populations were performed (GenAlEx, with 999 permutations). These tests were also examined on the whole data set and for each taxon.

To properly reconstruct genetic structure present in the analyzed group of populations, we verified their mutual differentiation and checked the distribution of genetic variation among different levels of population hierarchy. Estimates of $F_{st}$ and $G_{st}$ were calculated in DnaSP and the differentiation of populations was investigated by the hierarchical analysis of molecular variance (AMOVA) using Arlequin 3.5.22 (Excoffier & Lischer, 2010). In the initial test populations were grouped by morphological species identity, but we also tested the spatial variant of the analysis in

SAMOVA 2.0 (Dupanloup et al., 2002). It defines *K* maximally genetically differentiated groups of populations, simultaneously maintaining the highest possible geographical and genetic homogeneity of populations within each group. We looked for the number of such clusters running analyses with *K* values from two to 20. Additionally, to test how haplotype diversity is distributed among populations, two estimates of population differentiation: $G_{st}$—based on haplotype frequencies alone, and $N_{st}$—accounting also for their similarity, were calculated and compared following the method of Pons & Petit (1996) with 1000 permutations performed in PermutCpSSR 2.0. Significantly higher value of $N_{st}$ over $G_{st}$ indicates non-random occurrence of more genetically similar variants within populations, suggesting existence of barriers for migration or faster production of new variants than their exchange between stands.

## 3 Results

### 3.1 Polymorphism and diversity levels

In the present study, we applied a set of newly developed markers to investigate variation in mtDNA within the natural ranges of two closely related mountain conifers—dwarf and Pyrenean pines. From the total number of 449 samples genotyped in all 17 markers, 212 biometrically assigned specimens of *Pinus mugo* and 158 of *P. uncinata* were genotyped successfully, with at least six individuals per population. The samples with missing data in more than 10% positions were excluded from the analysis, while remaining singular missing data points were replaced by variants from most similar haplotypes and/or those most frequent in the same population, avoiding creation of *de novo* combinations. All markers proved to be variable and parsimony informative (Table S2), resulting in the total number of 31 multilocus haplotypes (Table S3). In reference to previous studies, we investigated variation in formerly used markers *nad1* and *nad7*. Both variants of *intron B/C* described by Soranzo et al. (2000) in locus *nad1* were found in our probe: short variant *a* appeared to be fixed in *P. mugo* populations, while variant *b* dominated in *P. uncinata*. At *intron 1* of *nad7*, variants *A* and *B* (according to Naydenov et al., 2007) were found in *P. mugo*, while the second species possessed only the *A* allele. We found only one nucleotide variant private for a single population (*Pr31* allele *T* in M6), but at the haplotype level, nearly half of the alleles occurred just in one location. When species were examined for private polymorphisms, most appeared to segregate in *P. mugo*, while *P. uncinata* had only two such variants. These (allele *b* in *nad1* and allele *A* in *Pr13*) were fixed in all *P. uncinata* populations except U2 in Massif Central (with one individual possessing the alternative alleles) and U1 from Western Alps. This particular population U1, defined at biometry and in plastid markers as *P. uncinata* (Boratynska et al., 2015; Dzialuk et al., 2017), but possessing exclusively mtDNA variants specific to *P. mugo*, was excluded from all analytical comparisons between species, however, left for overall and regional genetic variation estimates. The median joining network revealed existence of three arbitrary defined mitotype groups (Fig. 2). Group GrI being central, GrII separated from it by two mutations in *nad1* and *Pr13*, and GrIII being more distant with three discriminating alleles in *nad7*, *Pr15*, and *Pr20*, possessing also one other private SNP.

The number of segregating sites varied substantially among populations (Table 1). Four populations (M4, M17,

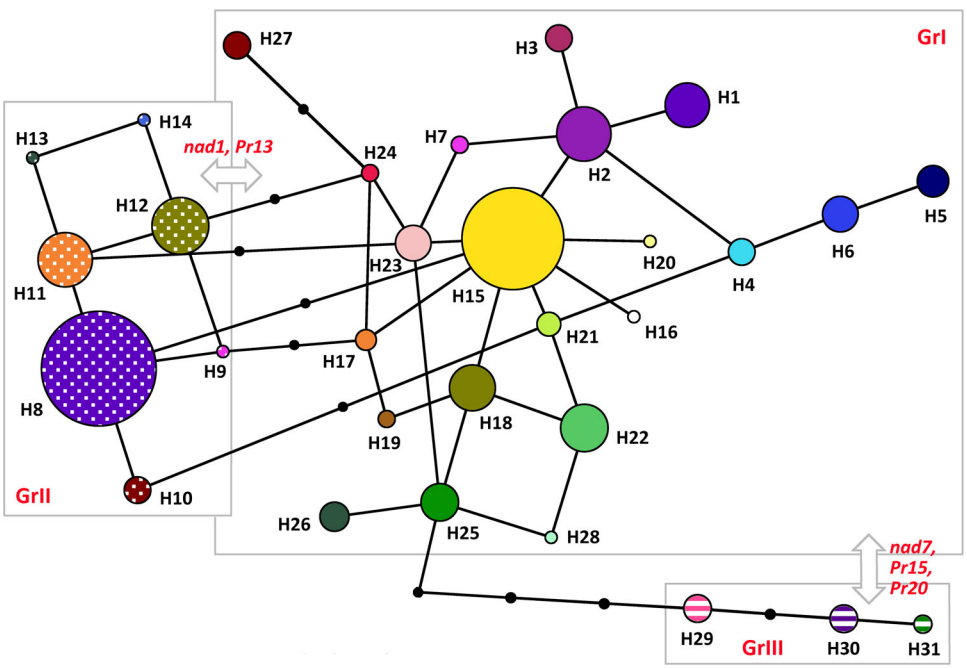



**Fig. 2.** Median joining network of *Pinus mugo* and *P. uncinata* mitotypes. Circle sizes represent relative frequencies of haplotypes in general population, lines between two circles represent singular mutation events, and black dots substitute absent haplotypes. Three haplogroups (GrI-GrIII) and discriminating them markers are presented.

U5, and U8) had just single haplotype, in contrast, 11 polymorphisms were identified in population M3. The level of polymorphisms differed between species: *P. mugo* had 15 polymorphic sites, while in *P. uncinata* five SNPs and one INDEL were variable (Tables 1 and Table S2). Mean number of differences between two samples reached $d = 2.961$ in the total probe, 0.838 in the Pyrenean pine and 2.618 in the second taxon, ranging between zero and 3.833 within populations. The highest numbers of haplotypes were found in locations M15 and U2 (six and seven mitotypes, respectively; Table 1 and Fig. 1). *Pinus mugo* possessed, in total, 23 mitotypes, almost three times as much as found in *P. uncinata*. The most common haplotype H8 (94 individuals) occurred in *P. uncinata* populations from Pyrenees and Massive Central. The second one (H15 with 74 occurrences) was much more widely distributed and present in the majority of *P. mugo* stands. It should be noted, that this result may be biased due to some inequality of sample sizes taken from populations of the two species. Gene diversity over all samples was high ($H_d = 0.879 \pm 0.01$ SD), and *P. mugo* showed a higher level of haplotype variation ($H_d = 0.863 \pm 0.018$) as compared to *P. uncinata* ($H_d = 0.549 \pm 0.041$).

### 3.2 Population grouping and structure

Mean nucleotide differences between populations ($d_{xy}$) ranged from zero (U5 vs. U8 possessing the same haplotype) up to 8.817 among M1 and M3. It averaged at 1.996 changes between any two populations and 3.876 between populations of the two taxa. The UPGMA tree construction and PCoA indicated similar population grouping, with the existence of two bigger clades and a few independent populations (Figs. 3, 4). Dwarf pine populations, except those from Sudetes and Western Carpathians, formed one group, incorporating also stand U1. Within this cluster, another subunit including majority of the populations from Alpine region and stands in Abruzzi and Balkans was clearly distinguishable on the UPGMA tree. Results of the two analyses differed mainly in positioning of the most distinct populations: M1, M2, and M3. Both methods clustered *P. uncinata* populations together, except U1 placed among *P. mugo* samples, and both showed that populations U2 and U10 differ from other stands being relatively close together. PCoA analysis performed on the whole data closely resembled the UPGMA groupings, and this pattern did not change considerably when taxa were tested separately (compare Fig. 4 with Figs. S1 and S2).

A significant correlation between genetic and geographic distances among populations was observed in the Mantel test for the probe as a whole ($r^2 = 0.2662$, $P \le 0.001$) and in *P. uncinata* ($r^2 = 0.7071$, $P \le 0.001$), but not in *P. mugo* ($r^2 = 0.0136$, $P = 0.195$). In turn, comparison of population differentiation estimates identified non-random distribution of mitotypes ($N_{st} > G_{st}$, $P = 0.01$), with similar variants grouping together within populations. However, on the species level such dependency was found only in *P. mugo*.

Pairwise differentiation ($F_{st}$) between populations was high and significant in the majority of comparisons, but not within *P. uncinata* (Tables 2, S4A, S4B). AMOVA with no hierarchy imposed revealed that most of the variation segregate among populations rather than within them ($F_{st} = 0.70$, $P < 0.001$). When populations were hierarchically grouped according to their species origin (with exclusion of U1), among taxa differentiation appeared to be much higher (51.67%, $P < 0.001$) than within-species differentiation between populations (26.61%), providing evidence for genetic distinction of mitochondrial gene pools of the two taxa. SAMOVA analysis with and without constraint for the geographic composition of the groups did not detect a single unambiguous optimal $K$ value, because the among group variation and the $F_{ct}$ statistics gradually increased with $K$ (Figs. S3A and S3C). However, when $K = 2$ was tested, both analysis variants separated the populations according to their phenotype species identity, except U1 stand ($F_{ct} = 0.52$; $P < 0.001$). In case of $K = 6$, distinguished by the lowest value of $\Delta F_{sc}$ and decreased level of variation among populations within groups (Figs. S3B and S3D), most of the variance was found among six groups: 62.61% ($P < 0.001$). The remaining 12.32% and 25.07% of variation was placed among populations within groups and inside them, respectively. For $K = 6$ samples of *P. uncinata* split into two entities: populations from Pyrenees (U3 – U9) vs. stands from Massive Central and Sierra de Gudar together (U2 and U10). *Pinus mugo* was separated into two individual populations (M1 in Sudetes and M3 in Tatra Mountains) and two bigger clusters: one in North Western Alps (including M13, M14, M16, and U1), and the second consisting of all the remaining populations: from Apennines (M17), south-eastern Alps (M11, M12, and M15), Dinaric Alps (M9 and M10) and Carpathians (M2, M4, M5, M6, M7 and M8; Fig. S4).

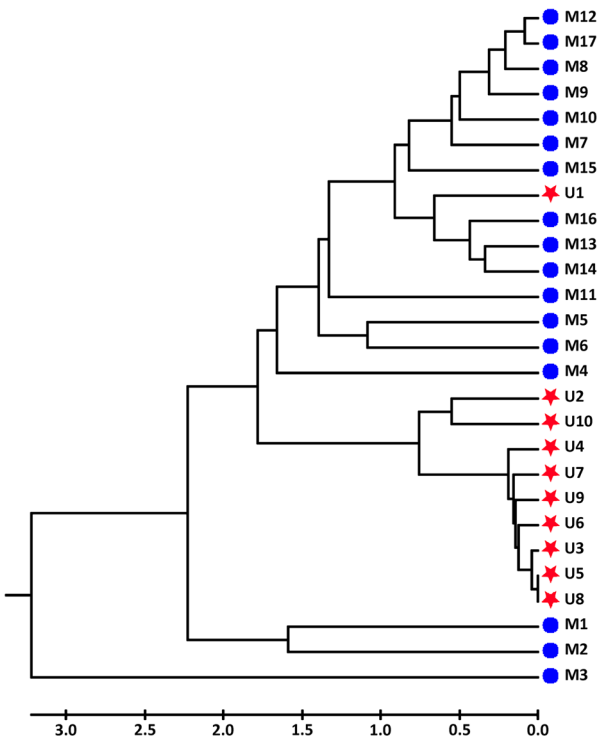



**Fig. 3.** Unweighted pair group method with arithmetic mean (UPGMA) tree depicting mutual distances in mitochondrial DNA markers between *Pinus mugo* (blue dots) and *P. uncinata* (red stars) populations in Europe.

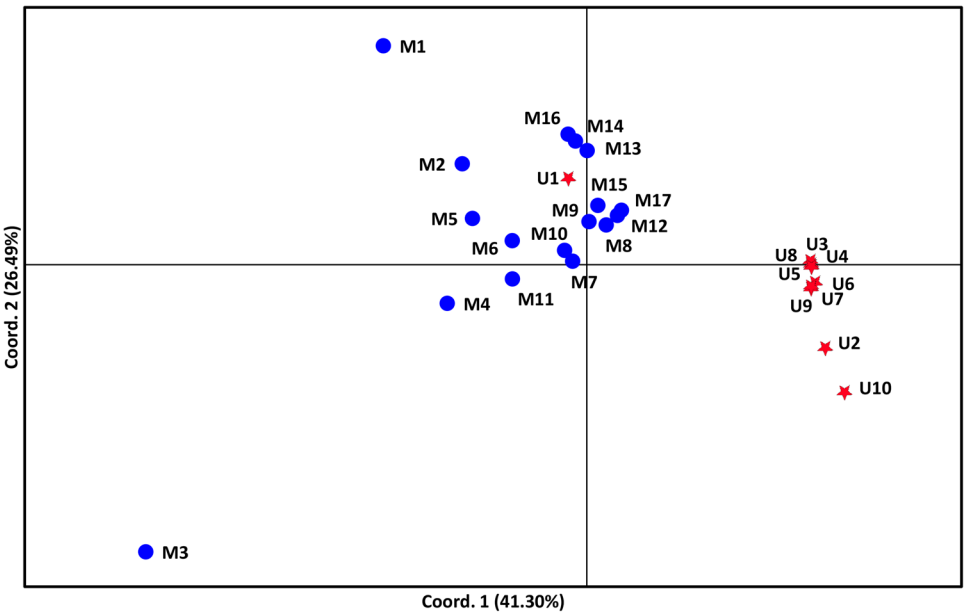



**Fig. 4.** Principal component analysis (PCoA) of population relationships based on mitochondrial DNA variation, showing dissonance between two European mountain pine species and grouping of *Pinus uncinata* Alpine accession U1 among *P. mugo* populations.

**Table 2** Population differentiation estimates within and between *Pinus mugo* and *P. uncinata* pine species; number of compared groups is specified in brackets

| | $F_{st}$ | $G_{st}$ |
|---|---|---|
| *P. mugo* (17) | 0.576 | 0.431 |
| *P. uncinata* (9)† | 0.482 | 0.348 |
| Between species (2)† | 0.561 | 0.161 |

$F_{st}$, population differentiation based on sequence data; $G_{st}$, population differentiation based on haplotype data.
†Population U1 was excluded from calculations.

## 4 Discussion

### 4.1 Distribution of mitochondrial DNA diversity among the mountain pines populations

The performed analyses allowed us to broadly assess variability and divergence of 27 studied populations of the two European mountain pines. Overall, in contrast to plastid microsatellite markers (Dzialuk et al., 2009; Heuertz et al., 2010; Dzialuk et al., 2017), most of the diversity in mitogenomes segregates between populations. Furthermore, we found that similar mtDNA variants co-occur in close proximities as shown in the patterns of allocation of molecular variance and its spatial distribution, together with significant results of the Mantel test and comparisons of $N_{st}$ and $G_{st}$ indices. The research suggests existence of phylogeographic structure in examined populations and confirms the utility of mtDNA markers in population history studies as compared to markers transmitted by mobile pollen grains or biparentally. The data provide basis for phylogeographic retracement of recolonization routes and identification of potential refugia sites of these species. This should be noted, however, that we lack reliable dating of the putative demographic changes due to the shortage of macrofossil data and uncertainties in palynological assignments (Obidowicz, 1996; Latalowa et al., 2004; Willis & van Andel, 2004; Benito Garzon et al., 2007; Fauvart et al., 2012). Our study also significantly advance earlier search for diagnostic biometric and genetic characters discriminating the species (e.g., Lewandowski et al., 2000; Boratynska & Bobowicz, 2001; Monteleone et al., 2006; Heuertz et al., 2010; Dzialuk et al., 2017). Divergence can be slow in case of long-lived organisms like pines, characterized by long generation times, outcrossing reproductive mode, and large effective population sizes (Petit & Hampe, 2006). Nevertheless, despite relatively low number of generations since the species divergence (Wachowiak et al., 2011), some fixation of genetic differences can be expected, especially in non-recombinant organelle genomes.

### 4.2 Genetic differentiation in *Pinus mugo*

Dwarf mountain pine exhibited generally high level of mtDNA diversity, although, its populations differed greatly in this regard. We found the most common and widely distributed mitotype H15 (Figs. 1, 2) to occupy the central position within the haplotype network and to have the highest number of direct connections. Therefore, we propose that this particular combination of alleles could represent ancestral state of dwarf pine, as the haplotypes with such interior positions are regarded older than external ones (Liu et al., 2014). Its distribution pattern and highest frequencies reached in the mountain systems located around the Adriatic Sea points these wide area as primary location and potential source region for other species' settlements.

Although weak isolation by distance between *P. mugo* populations was found, the comparisons of differentiation indices indicated that similar mitotypes tend to group together. Existence of small spatial clusters within species

range was revealed by SAMOVA, PCoA, and UPGMA analysis. Grouping of the easternmost accessions and the recognition of two major clusters present within the Alpine range, follow in general, the patterns of differentiation described at nuclear (Mosca et al., 2012) and plastid genomes (Dzialuk et al., 2017). The result suggests that the Alpine massif could have been colonized from two sides, and itself could form a barrier for more effective gene flow. We see one of the possible source locations in a proximity of the population M15 in Dolomites, where considerable contribution of private haplotypes and the highest inner variation among all sampled *P. mugo* stands were found. Interestingly, closely located population (about 50 km south) was described as the least diverged in cpSSR and nSSR loci and differentiated from other *P. mugo* populations due to possible results of a bottleneck effect after contraction (Zukowska & Wachowiak, 2017; Zukowska et al., 2017). The second source for colonization of Alps was likely localized in Maritime Alps, on west, where we observed considerable mitochondrial diversity, and high number of private cpDNA haplotypes was found formerly (Heuertz et al., 2010). The discovered pattern differs from those revealed so far for other coniferous species inhabiting these mountain range. Silver fir (*Abies alba* Mill.) colonized Alps most probably form single refugium in central part of the northern Apennines, and after crossing this range spread to the west, to the north and eastwards (Liepelt et al., 2009; Cheddadi et al., 2014), similarly Swiss stone pine (*P. cembra* L.) migrated to the current stands in Swiss Alps from one direction, likely located at the (south-) eastern periphery of the Alps (Gugerli et al., 2009), while European larch (*Larix decidua* Mill.) persisted in a few, presumably 3–5, different refugia situated within the southern Alpine foothills (Wagner et al., 2015; Dostalek et al., 2018).

The close affinity of populations from both sides of Adriatic Sea uncovered here contrasts with previous findings in which close similarity of Abbruzian population to the dwarf pines from Maritime Alps was described based on needle characteristics and cpSSRs, along with marked distinction of these two from other species populations (Boratynska et al., 2015; Dzialuk et al., 2017; Zukowska et al., 2017). The authors considered their results as potential signs of different origin of dwarf pines in these two locations or the influence of interspecific gene flow. However, as the remote stand in Abruzzi appeared fixed for single mitotype H15, but retained considerable amounts of phenotypic (Boratynska et al., 2005, Boratynska et al., 2015) and genetic (Zukowska & Wachowiak, 2017; Zukowska et al., 2017) variation, we argue, that the revealed pattern might represent a case of the founder effect and reflect long isolation from other seed sources. Simultaneously, the population must had been supplied by effective gene flow of pollen from *P. mugo* stands in Alps (Dzialuk et al., 2017), or even from *P. uncinata* as they reveal similarity in morphology of needles and cones and some genomic markers (Boratynska et al., 2015, 2005; Zukowska et al., 2017). The colonizers of this population in Apennines may had appeared there during one of the more severe glacial periods, when water level of Adriatic Sea was significantly lowered and some accessible dry land on shallow shelves was formed (Taberlet et al., 1998). This remote settlement might also be a result of long distance seed dispersal mediated by animals. However, none of these two seems a sufficiently comprehensive explanation for *P. mugo,* as this high-mountain species could shifted to lower altitudes (Benito Garzon et al., 2007), but unlikely as far down, and animal dispersion of seeds is considered non-significant in this group of pines (Benkman, 1995; Dzialuk et al., 2012).

Our results also shed light on the colonization routes of *P. mugo* in Central Europe. Most likely dwarf pines settled Sudetes and Carpathians expanding eastward from Alps and northward from the Balkans, and formed a broad secondary contact zone. The first source is supported in our data by the dominance in Sudetes and presence in Tatra Mts. of mitotypes H5 and H6, both closely related to H4 common in North Western Alps (Figs. 1, 2). It was previously suspected (Zukowska & Wachowiak, 2017) that *P. mugo* ancestors in Sudetes could had originally derived from the Alpine range, and experienced episode of isolation afterward, for they are clearly diverged at neutral nuclear markers. It is interesting in comparison with *Larix decidua* recolonization history, although a migration flow of unrecognized direction between north-eastern Alpine and Carpathian ranges of this species was described based on nDNA (Dostalek et al., 2018), there is a clear distinction between the two mountain massifs in the mitogenomes and other set of nuclear markers (Wagner et al., 2015).

On the other hand, the inflow of migrants from Balkans into the Carpathian region is indicated by shared SAMOVA group for $K = 6$ (Fig. S4) and the putative location of ancient species population on the south, as indicated by distribution of the central haplotype H15. The presented picture contrasts with earlier suggestions (Boratynska et al., 2015; Dzialuk et al., 2017; Mitic et al., 2018) that discussed the opposite direction of the gene flow—from Carpathians and Alps to the Balkan Mountains. However, this does not necessarily lead to disagreement, for the variation in mitochondrial markers may represent deeper evolutionary history compared to the other markers, and additionally, this direction of migration corroborates with findings regarding history of other coniferous high-altitude European species—the silver fir (Liepelt et al., 2009; Cheddadi et al., 2014).

Our findings support possible existence of a much wider distribution of *P. mugo* in the past (Dzialuk et al., 2012; Boratynska et al., 2015), followed by a period of separation, when the populations in Sudetes, Western Carpathians, Eastern Carpathians, and Balkans had undergone independent differentiation from each other and from the Alpine stands. Interestingly, high morphological dissimilarity of the populations set in East and South Carpathians underlined in recent studies (Boratynska et al., 2015, 2005), found no support in the mitochondrial genome diversity. Nevertheless, this region seems worth deeper investigation, especially in the light of new shrubby pine species (*P. microphylla* Blada) reported lately from the Pietrosul Rodnei Massif, different from *P. mugo* in number of biometric traits (Blada, 2017).

The above findings could, however, only partly explain the far and discrepant positions of northern dwarf pine accessions, i.e., populations M1, M2, and M3, as shown on the PCoA plots, UPGMA tree, and by their varying allocation by SAMOVA. Distinct character of these populations comes also from the presence of haplotypes from haplogroup GrIII,

composed of polymorphism combinations somewhat resembling five Finnish *P. sylvestris* haplotypes identified by Donnelly et al. (2017). Therefore, it is possible that part of this remarkable distinctness of northern *P. mugo* populations results from genetic contribution from Scots pines, as these species can still hybridize in zones of sympatry and this particular direction of gene flow—pollination of Scots pine cones by dwarf pine pollen—is more often noted (Wachowiak & Prus-Glowacki, 2008; Heuertz et al., 2010; Wachowiak et al., 2016; Kormutak et al., 2017; Kormutak et al., 2018).

### 4.3 Phylogeography of *Pinus uncinata*

Pyrenean pine showed moderate amounts of genetic diversity and weak support for phylogeographic structure. Generally, populations of this mountain pine formed one clade, except the population U1 from Cottian Alps. In all of the tests, the same subdivision was recognized with two populations from Massive Central and Sierra de Gudar grouped together based on their congruent differentiation from the main range in Pyrenees. High genetic similarity was observed also between two other marginal populations of the species from the two sides of Pyrenean chain (Dzialuk et al., 2009; Dzialuk et al., 2017). Such distribution of variants suggest existence of much wider range of mountain pine in past, supporting some of the earlier suggestions (Monteleone et al., 2006; Benito Garzon et al., 2007; Soto, Robledo-Arnuncio et al., 2010), rather than homoplasy in the slowly evolving mitochondrial genome (Wolfe et al., 1987). Interestingly, the relict stand U10 located inside Iberian Peninsula, most geographically isolated form others, appeared not to have the lowest genetic variation of mtDNA, what could be expected based on its small size and isolation (Willi et al., 2006). It was the other remote population—in Massif Central, which showed the highest variability, despite its marginal position. This stand, however, could be currently influenced by gene flow from Scots pine populations growing in a proximity, as already pointed for nearby populations (Heuertz et al., 2010), or historically by *P. mugo* (see later discussion). Nevertheless, more intriguing is the fact of generally low diversity found in the Pyrenean massif, where two of the examined populations appeared fixed for the same haplotype, and the others proved not much more polymorphic. Reveled pattern of *P. uncinata* genetic variation contrasts with findings from other species of similar subalpine distribution in European mountains—*Rhododendron ferrugineum*. Based on a set of nuclear microsatellite markers, Charrier, Dupont, Pornon and Escaravage (2014) described the existence of two separate genetic clusters in Alpine vs. Pyrenean populations of this evergreen shrub species. Furthermore, authors underlined high diversity and differentiation of the latter range, where five subgroups were noted.

We hypothesize that previously wide Pyrenean pine distribution became disjunctive in past, and that the separation covered one or more interglacial periods (Dzialuk et al., 2009), however, predating Holocene warming (Heuertz et al., 2010). Then, the Pyrenean populations underwent severe shrinkage followed by a quick reconstruction from a small number of survivors or an inflow from north. This scenario seems most likely as the populations are characterized by decreased variation within this relatively large and coherent group of stands, and because they share more mitochondrial and plastid (Dzialuk et al., 2017) haplotypes with the marginal northern stand U2. The populations within Iberian Peninsula, i.e., in Sierra de Gudar and Sierra Cebollera, remained isolated for much longer, and although maintained some variation in mitochondria, they lost most diversity in plastids (Dzialuk et al., 2009; Heuertz et al., 2010).

### 4.4 Interspecific gene exchange

Our study provides evidence for interspecific gene exchange between the two investigated mountain pines. We revealed fixation of the population U1 from the Alpine edge of *P. uncinata* range with mitotypes from the haplogroup Grl. Furthermore, the presence of similar individual of *P. uncinata* phenotype but *P. mugo*-like haplotype was observed in the second easternmost of the Pyrenean pine populations. Simultaneously, both of this stands, and other in the area, showed genetic similarity of paternally inherited markers to the western populations in Pyrenees (Heuertz et al., 2010; Dzialuk et al., 2017). We found no instances of opposite incongruence in cytoplasmic genetic markers. Such geographical distribution and only one way direction of inconsistency provide evidence of hybridization event between the two taxa, and another example of mitochondrial genome capture in pines (Wang & Wang, 2014). It also suggest that secondary contact and introgression occurred between already diverged species ancestors, taking place during one of the colder periods, when taxa could possibly met around the Rhône valley, moving down the slopes of their mountain ranges. In such case, at least the range of dwarf pine would require bigger, more westerly extended populations, because this species acted as the donor of less mobile seeds (Currat et al., 2008). Although the isolated populations of *P. mugo* found recently in Mont Cenis (Carcaillet et al., 2009) do not exceed to the mentioned valley, they indicate that this species could reach further west than generally accepted.

The disclosed unidirectional character of the hybridization between the two European conifers, and often lack of phenotypically intermediate individuals despite close proximity of both species within populations (Monteleone et al., 2006), while swarms of morphologically intermediate pines are common in other areas of sympatry (Christensen, 1987; Monteleone et al., 2006), is interesting also in another context. Recent study by Petrova et al. (2018) conducted on similar system of two closely related pines—*P. sibirica* and *P. pumila*, coming into secondary contact on Asian plains, revealed somehow analogous pattern. In both cases, only one species serves as mitochondrial genome donor, while their phenotypes, if are not intermediate (in some populations), remain the other taxon (in other places). In both systems, and in additional one referred to by Petrova et al. (2018), it is the prostrate species whose genome is captured and the upright whose genes are expressed.

For the best resolution of the ancestry and evolution of different ambiguous populations of the two mountain pines studied here (similarly in the case of mentioned Asian pines), tests of dense nuclear markers supported by genotyping of both cytoplasmic genomes conducted on individuals within the contact zone and in allopatric reference populations would be need. Specifically, the analysis could allow testing our hypothesis that primary

population of dwarf mountain pine specimens had become pollinated by invading *P. uncinata* pollen (Currat et al., 2008), and the observed pattern is not an example of incomplete lineage sorting.

### 4.5 Conservation implications

The two investigated taxa inhabit subalpine areas in European mountain massifs and chains, where their ranges remain patchy, composed of numerous scattered, often limited in size populations. The specific character of species distributions makes these pines susceptible to negative effects of isolation—consequences of genetic drift and diversity loss (Arenas et al., 2012), but they also must face the contemporary upward shifts of mountain vegetation and competition (Walther et al., 2005; Martinez et al., 2012; Hernandez et al., 2019). As both taxa play important ecological roles as shelters and food suppliers for alpine fauna, and they protect mountain habitats against soil erosion, torrents, and avalanches (Christensen, 1987; Monteleone et al., 2006; Alexandrov et al., 2019), we wanted to verify some of the previous suggestions regarding their conservation priorities (Camarero et al., 2005; Heuertz et al., 2010; Martinez et al., 2012). Our data show, that *P. mugo* maintained generally higher genetic polymorphism and higher intra-population diversity than *P. uncinata*. In regard to the dwarf pine preservation needs, as first we would focus on the secluded Apennine population that appeared devoid of mtDNA variation. Second most important stand, fixed with one private mitotype and showing considerable morphological distinction from other populations (Boratynska et al., 2004), is the Ukrainian M4. In *P. uncinata*, we identified shortage of polymorphism in the seven Pyrenean populations with four mitotypes in total. Modest diversity within this massif rises the need for better supervision of Pyrenean pines, as they may suffer weakened adaptive potential (Willi et al., 2006). The promise source for species reinforcements or reintroductions, if cautiously managed, might be seen in the most diverged population U2. Hybridizing populations, as U1 in North Western Alps and other with unrecognized taxonomy (Monteleone et al., 2006), where interspecific gene flow may bring both negative and positive effects, seems especially interesting for investigation of the adaptive variation of the species contact zones.

## 5 Conclusions

The applied new set of mtDNA markers proved useful for phylogeographic and evolutionary investigations among populations of the dwarf mountain pine and the Pyrenean mountain pine, improving also efficiency of previous research in this matter. The results reinforce hypothesis about their highly independent evolution. We found evidence for significant differentiation of mitochondrial genomes of the two conifers, as they exhibit separate mitotype gene pools in vast majority of their populations. In our study, opposite alleles at *Pr13* and *nad1* loci were found in the investigated taxa, that could be used as additional species diagnostic characters. Data acquired provide also support for wider distribution of both species in the past. Demographic fluctuations caused by constant glacial and interglacial environmental changes during Pleistocene translated into up and down shifts of these trees on mountain slopes. Those range shifts created areas of species secondary contacts that were most likely located around the area of current south-east France, where mitotypes of both species are present. However, the true identity of the putative *Pinus uncinata* and morphologically transient populations situated therein and in the Alpine range have to be verified with use of additional nuclear markers.

## Acknowledgements

The authors acknowledge with thanks the financial support from the Polish National Science Centre (Grant No. UMO-2015/19/B/NZ9/00024). The authors also thank Krystyna Boratyńska from the Institute of Dendrology, Polish Academy of Sciences, for providing plant material. None of the authors have any conflict of interest to declare.

## References

Alexandrov AH, von Wuhlisch G, Vendramin GG. 2019. *EUFORGEN Technical guidelines for genetic conservation and use of mountain pine (Pinus mugo)*. Bonn: European Forest Genetic Resources Programme (EUFORGEN), European Forest Institute.

Arenas M, Ray N, Currat M, Excoffier L. 2012. Consequences of range contractions and range shifts on molecular diversity. *Molecular Biology and Evolution* 29: 207–218.

Bandelt HJ, Forster P, Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* 16: 37–48.

Benito Garzon M, Sanchez de Dios R, Sainz Ollero H. 2007. Predictive modelling of tree species distributions on the Iberian Peninsula during the Last Glacial Maximum and Mid-Holocene. *Ecography* 30: 120–134.

Benkman CW. 1995. Wind dispersal capacity of pine seeds and the evolution of different seed dispersal modes in pines. *Oikos* 73: 221–224.

Blada I. 2017. *The short needle stone pine (*Pinus microphylla *species nova), a new dwarf two-needle hard pine species in the Pietrosul Rodnei Massif flora*. Voluntari: Institutul National de Cercetare Dezvoltare in Silvicultura 'Marin Dracea'.

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.

Boratynska K, Bobowicz MA. 2001. *Pinus uncinata* Ramond taxonomy based on needle characters. *Plant Systematics and Evolution* 227: 183–194.

Boratynska K, Dzialuk A, Lewandowski A, Marcysiak K, Jasinska AK, Sobierajska K, Tomaszewski D, Burczyk J, Boratynski A. 2014. Geographic distribution of quantitative traits variation and genetic variability in natural populations of *Pinus mugo* in Central Europe. *Dendrobiology* 72: 65–84.

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.

Boratynska K, Marcysiak K, Boratynski A. 2005. *Pinus mugo* (Pinaceae) in the Abruzzi mountains: high morphological variation in isolated populations. *Botanical Journal of the Linnean Society* 147: 309–316.

Boratynska K, Muchewicz E, Drojma M. 2004. *Pinus mugo* Turra geographic differentiation based on needle characters. *Dendrobiology* 51: 9–17.

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

Camarero JJ, Gutierrez E, Fortin MJ, Ribbens E. 2005. Spatial patterns of tree recruitment in a relict population of *Pinus uncinata*: forest expansion trough stratified diffusion. *Journal of Biogeography* 32: 1979–1992.

Carcaillet C, Fauvart N, Roiron P, Terral JF, Ali AA. 2009. A new, isolated and endangered relict population of dwarf pine (*Pinus mugo* Turra) in the northwestern Alps. *Comptes Rendus Biologies* 332: 456–463.

Charrier O, Dupont P, Pornon A, Escaravage N. 2014. Microsatellite marker analysis reveals the complex phylogeographic history of *Rhododendron ferrugineum* (Ericaceae) in the pyrenees. *PLoS One* 9: e92976.

Cheddadi R, Birks HJB, Tarroso P, Liepelt S, Gomory D, Dullinger S, Meier ES, Hülber K, Maiorano L, Laborde H. 2014. Revisiting tree-migration rates: *Abies alba* (Mill.), a case study. *Vegetation History and Archaeobotany* 23: 113–122.

Christensen K. 1987. Taxonomic revision of the *Pinus mugo* complex and *P. rhaetica* (*P. mugo* x *sylvestris*). *Nordic Journal of Botany* 7: 338–408.

Currat M, Ruedi M, Petit RJ, Excoffier L. 2008. The hidden side of invasions: massive introgression by local genes. *Evolution* 62: 1908–1920.

Donnelly K, Cottrell J, Ennos RA, Vendramin GG, A'Hara S, King S, Perry A, Wachowiak W, Cavers S. 2017. Reconstructing the plant mitochondrial genome for marker discovery: a case study using *Pinus*. *Molecular Ecology Resources* 17: 943–954.

Dostalek J, Frantíkb T, Pospiskova M, Krizovac M. 2018. Population genetic structure and delineation of conservation units in European larch (*Larix decidua* Mill.) across its native range. *Flora* 246-247: 26–32.

Dupanloup I, Schneider S, Excoffier L. 2002. A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology* 11: 2571–2581.

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.

Dzialuk A, Boratynski A, Boratynska K, Burczyk J. 2012. Geographic patterns of genetic diversity of *Pinus mugo* (Pinaceae) in Central European mountains. *Dendrobiology* 68: 31–41.

Dzialuk A, Muchewicz E, Boratynski A, Montserrat JM, Boratynska K, Burczyk J. 2009. Genetic variation of *Pinus uncinata* (Pinaceae) in the Pyrenees determined with cpSSR markers. *Plant Systematics and Evolution* 277: 197–205.

Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10: 564–567.

Fauvart N, Ali AA, Terral JF, Roiron P, Blarquez O, Carcaillet C. 2012. Holocene upper tree-limits of *Pinus* section *sylvestris* in the Western Alps as evidened from travertine archives. *Reviews of Paleobotany and Palynology* 169: 96–102.

Grotkopp E, Rejmanek M, Sanderson MJ, Rost TL. 2004. Evolution of genome size in pines (*Pinus*) and its life-history correlates: Supertree analyses. *Evolution* 58: 1705–1729.

Gugerli F, Ruegg M, Vendramin GG. 2009. Gradual decline in genetic diversity in Swiss stone pine populations (*Pinus cembra*) across Switzerland suggests postglacial re-colonization into the Alps from a common eastern glacial refugium. *Botanica Helvetica* 119: 13–22.

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

Hernandez L, Camarero JJ, Gil-Peregrin E, San Sanchez MA, Canellas I, Montes F. 2019. Biotic factors and increasing aridity shape the altitudinal shifts of marginal Pyrenean silver fir populations in Europe. *Forest Ecology and Management* 432: 558–567.

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.

Jalas J, Suominen J eds. 1973. *Atlas Florae Europeae: Distribution of vascular plants in Europe*. Helsinki: Cambridge University Press. 2.

Jaramillo-Correa JP, Beaulieu J, Bousquet J. 2004. Variation in mitochondrial DNA reveals multiple distant glacial refugia in black spruce (*Picea mariana*), a transcontinental north american conifer. *Molecular Ecology* 13: 2735–2747.

Kormutak A, Galgoci M, Manka P, Koubova M, Jopcik M, Sukenikova D, Bolecek P, Gomory D. 2017. Field-based artificial crossings indicate partial compatibility of reciprocal crosses between *Pinus sylvestris* and *Pinus mugo* and unexpected chloroplast DNA inheritance. *Tree Genetics & Genomes* 13: 68.

Kormutak A, Galgoci M, Sukenikova D, Bolecek P, Libantova J, Gomory D. 2018. Maternal inheritance of chloroplast DNA in *Pinus mugo* Turra: a case study of *Pinus mugo* x *Pinus sylvestris* crossing. *Plant Systematics and Evolution* 304: 71–76.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* 33: 1870–1874.

Latalowa M, Tobolski K, Nalepka D. 2004. *Pinus* L. subgenus *Pinus* (subgen. *Diploxylon* (Koehne) Pilger) – Pine. In: Ralska-Jasiewiczowa M ed. *Late Glacial and Holocene history of vegetation in Poland based on isopollen maps*. Krakow: W. Szafer Institute of Botany Polish Academy of Sciences. 165–177.

Leigh JW, Bryant D. 2015. POPART: full-feature software for haplotype network construction. *Methods in Ecology and Evolution* 6: 1110–1116.

Lewandowski A, Boratynski A, Mejnartowicz L. 2000. Allozyme investigations on the genetic differentiation between closely related pines - *Pinus sylvestris*, *P. mugo*, *P. uncinata*, and *P. uliginosa* (Pinaceae). *Plant Systematics and Evolution* 221: 15–24.

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452.

Liepelt S, Bialozyt R, Ziegenhagen B. 2002. Wind-dispersed pollen mediates postglacial gene flow among refugia. *Proceedings of the National Academy of Sciences USA* 99: 14590–14594.

Liepelt S, Cheddadi R, de Beaulieu JL, Fady B, Gomory D, Hussendorfer E, Konnert M, Litt T, Longauer R, Terhürne-Berson R, Ziegenhagen B, Ziegenhagen B. 2009. Postglacial range expansion and its genetic imprints in *Abies alba* (Mill.) — A synthesis from palaeobotanic and genetic data. *Review of Palaeobotany and Palynology* 153: 139–149.

Liu L, Hao ZZ, Liu YY, Wei XX, Cun YZ, Wang XQ. 2014. Phylogeography of *Pinus armandii* and its relatives: heterogeneous contributions of geography and climate changes to the genetic differentiation and diversification of Chinese white pines. *PLoS One* 9: e85920.

Marcysiak K, Boratynski A. 2007. Contribution to the taxonomy of *Pinus uncinata* (Pinaceae) based on cone characters. *Plant Systematics and Evolution* 264: 57–73.

Martinez I, Gonzalez-Taboada F, Wiegand T, Camarero JJ, Gutierrez E. 2012. Dispersal limitation and spatial scale affect model based projections of *Pinus uncinata* response to climate change in the Pyrenees. *Global Change Biology* 18: 1714–1724.

Mitic ZS, Nikolic JS, Zlatkovic BK, Milanovici SJ, Jovanovic SC, Nikolic BM, Stojanovic GS, Marin PD. 2018. Epicuticular waxes provide insights into phytochemical differentiation of natural populations of *Pinus mugo* Turra sensu stricto. *Chemistry & Biodiversity* 15: e1800378.

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.

Mosca E, Eckert AJ, Di Pierro EA, Rocchini D, La Porta N, Belletti P, Neale DB. 2012. The geographical and environmental determinants of genetic diversity for four alpine conifers of the European Alps. *Molecular Ecology* 21: 5530–5545.

Nathan R, Muller-Landau HC. 2000. Spatial patterns of seed dispersal, their determinants and consequences for reqruitment. *Trends in Ecology and Evolution* 15: 278–285.

Naydenov K, Senneville S, Beaulieu J, Tremblay F, Bousquet J. 2007. Glacial vicariance in Eurasia: mitochondrial DNA evidence from Scots pine for a complex heritage involving genetically distinct refugia at mid-northern latitudes and in Asia Minor. *BMC Evolutionary Biology* 7: 233.

Obidowicz A. 1996. A Late Glacial-Holocene history of the formation of vegetation belts in the Tatra Mts. *Acta Paleobotanica* 36: 159–206.

Ohsawa T, Ide Y. 2008. Global patterns of genetic variation in plant species along vertical and horizontal gradients on mountains. *Global Ecology and Biogeography* 17: 152–163.

Peakall R, Smouse PE. 2006. GenAlEx 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6: 288–295.

Peakall R, Smouse PE. 2012. GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research - an update. *Bioinformatics* 28: 2537–2539.

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

Petrova EA, Zhuk EA, Popov AG, Bondar AA, Belokon MM, Goroshkevich SN, Vasilyeva GV. 2018. Asymmetric introgression between *Pinus sibirica* and *Pinus pumila* in the Aldan plateau (Eastern Siberia). *Silvae Genetica* 67: 66–71.

Pons O, Petit RJ. 1996. Measuring and testing genetic differentiation with ordered versus unordered alleles. *Genetics* 144: 1237–1245.

Robledo-Arnuncio JJ. 2011. Wind pollination over mesoscale distances: an investigation with Scots pine. *New Phytologist* 190: 222–233.

Sandoz H. 1987. *Recerches taxonomiques, biogéographiques et phytoécologiques sur les principaux conifères subalpins des Alpes: Mélèze d'Europe, Pin cembro, Pin à crochets et Pin mugho*. Ph.D. Dissertation. Marseille: Université Paul Cézanne.

Semerikov VL, Putintseva YA, Oreshkova NV, Semerikova SA, Krutovsky KV. 2015. Development of new mitochondrial DNA markers in Scots pine (*Pinus sylvestris* L.) for population and phylogeographic studies. *Russian Journal of Genetics* 51: 1199–1203.

Sinclair WT, Morman JD, Ennos RA. 1998. Multiple origins for Scots pine (*Pinus sylvestris* L.) in Scotland: evidence from mitochondrial DNA variation. *Heredity* 80: 233–240.

Soranzo N, Alia R, Provan J, Powell W. 2000. Patterns of variation at a mitochondrial sequence-tagged-site locus provides new insights into the postglacial history of European *Pinus sylvestris* populations. *Molecular Ecology* 9: 1205–1211.

Soto A, Robledo-Arnuncio JJ, Gonzalez-Martinez SC, Smouse PE, Alia R. 2010. Climatic niche and neutral genetic diversity of the six Iberian pine species: a retrospective and prospective view. *Molecular Ecology* 19: 1396–1409.

Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF. 1998. Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* 7: 453–464.

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, 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.

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, Prus-Glowacki W. 2008. Hybridisation processes in sympatric populations of pines *Pinus sylvestris* L., *P. mugo* Turra and *P. uliginosa* Neumann. *Plant Systematics and Evolution* 271: 29–40.

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

Wachowiak W, Zukowska WB, Wojkiewicz B, Cavers S, Litkowiec M. 2016. Hybridization in contact zone between temperate European pine species. *Tree Genetics & Genomes* 12: 48.

Wagner S, Litt T, Sanchez-Goni MF, Petit RJ. 2015. History of *Larix decidua* Mill. (European larch) since 130 ka. *Quaternary Science Reviews* 124: 224–247.

Walther GR, Beissner S, Burga CA. 2005. Trends in the upward shift of alpine plants. *Journal of Vegetation Science* 16: 541–548.

Wang BS, Wang XR. 2014. Mitochondrial DNA capture and divergence in *Pinus* provide new insights into the evolution of the genus. *Molecular Phylogenetics and Evolution* 80: 20–30.

Willi Y, Van Buskrik J, Hoffmann AA. 2006. Limits to the adaptive potential of small populations. *Annual Review of Ecology, Evolution and Systematics* 37: 433–458.

Willis JK, van Andel TH. 2004. Trees or no trees? The environments of central and eastern Europe during the Last Glaciation. *Quaternary Science Reviews* 23: 2369–2387.

Wolfe KH, Li WH, Sharp PM. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences USA* 84: 9054–9058.

Zukowska WB, Boratynska K, Wachowiak W. 2017. Comparison of range-wide chloroplast microsatellite and needle trait variation patterns in *Pinus mugo* Turra (dwarf mountain pine). *iForest-Biogeosciences and Forestry* 10: 250–258.

Zukowska WB, Wachowiak W. 2017. Nuclear microsatellite markers reveal the low genetic structure of *Pinus mugo* Turra (dwarf mountain pine) populations in Europe. *Plant Systematics and Evolution* 303: 641–651.

## Supplementary Material

The following supplementary material is available online for this article at http://onlinelibrary.wiley.com/doi/1.1111/jse.12520/ suppinfo:

**Table S1.** Forward and reverse primers targeting 4 mitochondrial SNP loci analyzed here and by Donnelly et al. (2017), but not specified therein. Amplification conditions are provided in the main text.

**Table S2.** Specification of *mt*DNA markers and frequency of identified variants within studied populations of two European mountain pines.

**Table S3.** Mitotypes identified in 27 populations of dwarf mountain pine and Pyrenean pine, their relative frequencies in total sample and both species.

**Table S4.** Intraspecific differentiation of the *mt*DNA markers among (A) *Pinus mugo* and (B) *P. uncinata* (U1 was excluded) populations from European mountain ranges; only statistically significant ($p < 0.05$) result are shown.

**Fig. S1.** Principal Component Analysis (PCoA) of 17 *P. mugo* populations in Europe showing clear mtDNA markers distinction of the northern populations in Sudetes (M1) and Western Carpathians (M2 and M3) from the remaining species stands.

**Fig. S2.** Principal Component Analysis (PCoA) of mtDNA variation in 9 *P. uncinata* populations (U1 was not considered). Note high percentage of differentiation assigned to the first coordinate compared with the second, which separates the Pyrenean stands from two satellite populations in Iberian Peninsula (U10) and Massive Central (U2).

**Fig. S3.** Distribution of genetic variation and differentiation indices in spatial AMOVA clustering of 27 mountain pine populations based on their mitotype variation, whit K values tested within range of two to 20. Variation distribution indices and change in their values between following *K* values are present on (A) and (B) plots, respectively; cluster differentiation indices and their changes are shown on (C) and (D) graphs.

**Fig. S4.** Spatial distribution of genetically homogenous clusters identified by SAMOVA analysis for *K*- value equal to 6 based on mtDNA variation in populations of two European mountain pine taxa; distinct colors indicate different genetic groups.

**Table S1.** Forward and reverse primers targeting 4 mitochondrial SNP loci analyzed here and by Donnelly et al. (2017), but not specified therein. Amplification conditions are provided in the main text.

| Locus | Forward (5' -> 3') | Reverse (3' -> 5') | Reference contig |
|---|---|---|---|
| ***Pr20*** | GTTCCTACGATCCAGCCAGG | ACCATGGATTCTTCGGACGG | Ptaeda_M_Contig28_A_F |
| ***Pr30*** | ACTTACATTGACCGGCGGAT | CACACATCTAGGGCACAGGG | Ptaeda_M_Contig399_B_F |
| ***Pr32*** | ACCCTCCTTCAACTGATGCG | CCTCAACCAACCGTCAGTCA | Ptaeda_M_Contig431_A_F |
| ***Pr34*** | GAACCCCCTCTTGCCTTGAT | TTCGTGACGGTCCAATTCCA | Ptaeda_M_Contig431_C_F |

**Table S2.** Specification of *mt*DNA markers and frequency of identified variants within studied populations of two European mountain pines.

| Locus | Locus & reference† | Genotyping method, enzyme | Variant position‡ | Product size [bp] | Common / rare variant§ | No. bands | No. missing variants¶ | Freq. in *P. mugo* Turra⊖ | Freq. in *P. uncinata* Ramond⊖,¥ |
|---|---|---|---|---|---|---|---|---|---|
| ***Pr5*** | locus 3, [1] | PCR-RFLP, DraI | 279 | 570 | *G* / *T* | 1 / 2 | 6 | 205 / 7 | 147 / 0 |
| ***Pr7*** | locus 5, [1] | PCR-RFLP, DraI | 409 | 514 | *C* / *A* | 1 / 2 | 4 | 156 / 56 | 147 / 0 |
| ***Pr13*** | locus 11, [1] | PCR-RFLP, MseI | 246 | 369 | *C* / *A* | 2 / 1 | 22 | 212 / 0 | 1 / 146 |
| ***Pr14*** | locus 12, [1] | PCR-RFLP, MseI | 382 | 548 | *G* / *T* | 1 / 2 | 3 | 165 / 47 | 101 / 46 |
| ***Pr15*** | locus 13, [1] | PCR-RFLP, BpiI | 155 | 358 | *G* / *T* | 2 / 1 | 2 | 200 / 12 | 147 / 0 |
| ***Pr19*** | locus 17, [1] | PCR-RFLP, DraI | 413 | 593 | *G* / *T* | 1 / 2 | 0 | 163 / 49 | 142 / 5 |
| ***Pr20*** | this study, [Table S2] | PCR-RFLP, MvaI | 173 | 382 | *C* / *A* | 3 / 2 | 5 | 200 / 12 | 147 / 0 |
| ***Pr21*** | locus 18, [1] | PCR-RFLP, HincII | 421 | 522 | *G* / *T* | 2 / 1 | 0 | 204 / 8 | 147 / 0 |
| ***Pr24*** | locus 19, [1] | PCR-RFLP, BshNI | 272 | 434 | *T* / *G* | 1 / 2 | 1 | 200 / 12 | 147 / 0 |
| ***Pr25*** | locus 20, [1] | PCR-RFLP, BtsCI | 123 | 398 | *G* / *T* | 1 / 2 | 2 | 150 / 62 | 147 / 0 |
| ***Pr29*** | locus 22, [1] | PCR-RFLP, BtsCI | 226 | 558 | *G* / *T* | 1 / 2 | 3 | 200 / 12 | 145 / 2 |
| ***Pr30*** | this study, [Table S2] | PCR-RFLP, Alw26I | 244 | 401 | *G* / *T* | 2 / 1 | 1 | 200 / 12 | 122 / 25 |
| ***Pr31*** | locus 23, [1] | PCR-RFLP, VspI | 274¤ | 558 | *G* / *T* | 1 / 2 | 0 | 207 / 5 | 147 / 0 |
| ***Pr32*** | this study, [Table S2] | PCR-RFLP, PfeI | 177 | 507 | *G* / *T* | 2 / 1 | 13 | 197 / 15 | 147 / 0 |
| ***Pr34*** | this study, [Table S2] | Sanger sequencing | 176-185 | 398 | *--A--AAGA* / *TCATCAATC* | na | 15 | 195 / 17 | 147 / 0 |
| ***nad1*** | *intron B/C*, [2] | PCR-RFLP, DraII | 84-115 | 217 / 248 | del (*a* in [2]) / in (*b* in [2]) | 2 / 3 | 2 | 212 / 0 | 1 / 146 |
| ***nad7*** | *intron 1*, [3, 4, 5] | PCR-RFLP, DraII | 622-254, 674-679 | 1175 / 1170 / 1143 | in in (*A* in [4]) / in del (*B* in [4]) / del in (*C* in [4]) | 3 (long) / 2 / 3 (short) | 7 | 200 / 12 / 0 | 147 / 0 / 0 |

† [1] Donnelly et al. (2017), [2] Soranzo et al. (2000; primers H/I), [3] Jaramillo-Correa et al. (2004), [4] Naydenov et al. (2007), [5] Pyhajarvi et al. (2008); ‡ 5' –> 3' (primers excluded); § del – deletion, in-insertion; ¶ these positions were complemented in haplotypes; ⊖ after haplotype complementation; ¥ population U1 excluded; ¤ note that this is another SNP position within the locus than the variant *293* (*C* / *A*) analyzed by Donnelly et al. (2017).

**Table S3.** Mitotypes identified in 27 populations of dwarf mountain pine and Pyrenean pine, their relative frequencies in total sample and both species.

| Mitotype | Locus | | | | | | | | | | | | | | | | | Total probe freq. [%] | *P. mugo* Turra freq. [%] | *P. uncinata* Ramond[¶] freq. [%] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ***Pr5*** | ***Pr7*** | ***Pr13*** | ***Pr14*** | ***Pr15*** | ***Pr19*** | ***Pr20*** | ***Pr21*** | ***Pr24*** | ***Pr25*** | ***Pr29*** | ***Pr30*** | ***Pr31*** | ***Pr32*** | ***Pr34***[†] | ***nad1***[‡] | ***nad7***[§] | | | |
| ***H1*** | *G* | *A* | *C* | *G* | *G* | *G* | *C* | *G* | *G* | *G* | *G* | *G* | *G* | *G* | 0 | *a* | *A* | 3.8 | 5.7 | 0.0 |
| ***H2*** | *G* | *A* | *C* | *G* | *G* | *G* | *C* | *G* | *T* | *G* | *G* | *G* | *G* | *G* | 0 | *a* | *A* | 5.7 | 9.9 | 0.0 |
| ***H3*** | *G* | *A* | *C* | *G* | *G* | *G* | *C* | *G* | *T* | *G* | *T* | *G* | *G* | *G* | 0 | *a* | *A* | 1.4 | 0.0 | 0.0 |
| ***H4*** | *G* | *A* | *C* | *G* | *G* | *T* | *C* | *G* | *T* | *G* | *G* | *G* | *G* | *G* | 0 | *a* | *A* | 1.4 | 2.4 | 0.0 |
| ***H5*** | *G* | *A* | *C* | *G* | *G* | *T* | *C* | *G* | *T* | *G* | *G* | *G* | *G* | *G* | 1 | *a* | *A* | 1.9 | 3.3 | 0.0 |
| ***H6*** | *G* | *A* | *C* | *G* | *G* | *T* | *C* | *G* | *T* | *G* | *G* | *G* | *G* | *T* | 1 | *a* | *A* | 2.4 | 4.2 | 0.0 |
| ***H7*** | *G* | *A* | *C* | *T* | *G* | *G* | *C* | *G* | *T* | *G* | *G* | *G* | *G* | *G* | 0 | *a* | *A* | 0.5 | 0.9 | 0.0 |
| ***H8*** | *G* | *C* | *A* | *G* | *G* | *G* | *C* | *G* | *T* | *G* | *G* | *G* | *G* | *G* | 0 | *b* | *A* | 25.4 | 0.0 | 63.9 |
| ***H9*** | *G* | *C* | *A* | *G* | *G* | *G* | *C* | *G* | *T* | *G* | *G* | *T* | *G* | *G* | 0 | *b* | *A* | 0.3 | 0.0 | 0.7 |
| ***H10*** | *G* | *C* | *A* | *G* | *G* | *T* | *C* | *G* | *T* | *G* | *G* | *G* | *G* | *G* | 0 | *b* | *A* | 1.4 | 0.0 | 3.4 |
| ***H11*** | *G* | *C* | *A* | *T* | *G* | *G* | *C* | *G* | *T* | *G* | *G* | *G* | *G* | *G* | 0 | *b* | *A* | 5.7 | 0.0 | 14.3 |
| ***H12*** | *G* | *C* | *A* | *T* | *G* | *G* | *C* | *G* | *T* | *G* | *G* | *T* | *G* | *G* | 0 | *b* | *A* | 6.2 | 0.0 | 15.6 |
| ***H13*** | *G* | *C* | *A* | *T* | *G* | *G* | *C* | *G* | *T* | *G* | *T* | *G* | *G* | *G* | 0 | *b* | *A* | 0.3 | 0.0 | 0.7 |
| ***H14*** | *G* | *C* | *A* | *T* | *G* | *G* | *C* | *G* | *T* | *G* | *T* | *T* | *G* | *G* | 0 | *b* | *A* | 0.3 | 0.0 | 0.7 |
| ***H15*** | *G* | *C* | *C* | *G* | *G* | *G* | *C* | *G* | *T* | *G* | *G* | *G* | *G* | *G* | 0 | *a* | *A* | 20.0 | 32.5 | 0.7 |
| ***H16*** | *G* | *C* | *C* | *G* | *G* | *G* | *C* | *G* | *T* | *G* | *G* | *G* | *G* | *G* | 1 | *a* | *A* | 0.3 | 0.5 | 0.0 |
| ***H17*** | *G* | *C* | *C* | *G* | *G* | *G* | *C* | *G* | *T* | *G* | *G* | *T* | *G* | *G* | 0 | *a* | *A* | 0.8 | 1.4 | 0.0 |
| ***H18*** | *G* | *C* | *C* | *G* | *G* | *G* | *C* | *G* | *T* | *T* | *G* | *G* | *G* | *G* | 0 | *a* | *A* | 4.1 | 7.1 | 0.0 |
| ***H19*** | *G* | *C* | *C* | *G* | *G* | *G* | *C* | *G* | *T* | *T* | *G* | *T* | *G* | *G* | 0 | *a* | *A* | 0.5 | 0.9 | 0.0 |
| ***H20*** | *G* | *C* | *C* | *G* | *G* | *G* | *C* | *T* | *T* | *G* | *G* | *G* | *G* | *G* | 0 | *a* | *A* | 0.3 | 0.5 | 0.0 |
| ***H21*** | *G* | *C* | *C* | *G* | *G* | *T* | *C* | *G* | *T* | *G* | *G* | *G* | *G* | *G* | 0 | *a* | *A* | 1.1 | 1.9 | 0.0 |
| ***H22*** | *G* | *C* | *C* | *G* | *G* | *T* | *C* | *G* | *T* | *T* | *G* | *G* | *G* | *G* | 0 | *a* | *A* | 4.3 | 7.5 | 0.0 |
| ***H23*** | *G* | *C* | *C* | *T* | *G* | *G* | *C* | *G* | *T* | *G* | *G* | *G* | *G* | *G* | 0 | *a* | *A* | 2.4 | 4.2 | 0.0 |
| ***H24*** | *G* | *C* | *C* | *T* | *G* | *G* | *C* | *G* | *T* | *G* | *G* | *T* | *G* | *G* | 0 | *a* | *A* | 0.5 | 0.9 | 0.0 |
| ***H25*** | *G* | *C* | *C* | *T* | *G* | *G* | *C* | *G* | *T* | *T* | *G* | *G* | *G* | *G* | 0 | *a* | *A* | 2.7 | 4.7 | 0.0 |
| ***H26*** | *G* | *C* | *C* | *T* | *G* | *G* | *C* | *G* | *T* | *T* | *G* | *G* | *G* | *T* | 0 | *a* | *A* | 1.6 | 2.8 | 0.0 |
| ***H27*** | *G* | *C* | *C* | *T* | *G* | *T* | *C* | *G* | *T* | *G* | *G* | *T* | *T* | *G* | 0 | *a* | *A* | 1.4 | 2.4 | 0.0 |
| ***H28*** | *G* | *C* | *C* | *T* | *G* | *T* | *C* | *G* | *T* | *T* | *G* | *G* | *G* | *G* | 0 | *a* | *A* | 0.3 | 0.5 | 0.0 |
| ***H29*** | *G* | *C* | *C* | *T* | *T* | *G* | *A* | *G* | *T* | *T* | *T* | *G* | *G* | *G* | 0 | *a* | *B* | 1.4 | 2.4 | 0.0 |
| ***H30*** | ***T*** | *C* | *C* | *T* | *T* | *G* | *A* | *T* | *T* | *T* | *T* | *G* | *G* | *G* | 0 | *a* | *B* | 1.4 | 2.4 | 0.0 |
| ***H31*** | ***T*** | *C* | *C* | *T* | *T* | *T* | *A* | *T* | *T* | *T* | *T* | *G* | *G* | *G* | 0 | *a* | *B* | 0.5 | 0.9 | 0.0 |

† '0' stands for '*--A—AAGA*' sequence allele, and '1' for '*TCATCAATC*' variant; ‡ allele acronyms as in Soranzo et al. (2000); § allele acronyms according to Naydenov et al. (2007; see Table S2); ¶ population U1 excluded from calculation on species level.

**Table S4.** Intraspecific differentiation of the *mt*DNA markers among (a) *Pinus mugo* and (b) *P. uncinata* (U1 was excluded) populations from European mountain ranges; only statistically significant ($p < 0.05$) result are shown.

**(a)**

| | M1 | M2 | M3 | M4 | M5 | M6 | M7 | M8 | M9 | M10 | M11 | M12 | M13 | M14 | M15 | M16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **M2** | 0.422 | - | | | | | | | | | | | | | | |
| **M3** | 0.755 | 0.446 | - | | | | | | | | | | | | | |
| **M4** | 0.975 | 0.570 | 0.558 | - | | | | | | | | | | | | |
| **M5** | 0.956 | 0.516 | 0.640 | 0.963 | - | | | | | | | | | | | |
| **M6** | 0.697 | 0.273 | 0.492 | 0.567 | 0.374 | - | | | | | | | | | | |
| **M7** | 0.864 | 0.331 | 0.497 | 0.643 | 0.710 | 0.304 | - | | | | | | | | | |
| **M8** | 0.891 | 0.365 | 0.613 | 0.830 | 0.798 | 0.394 | 0.209 | - | | | | | | | | |
| **M9** | 0.889 | 0.406 | 0.616 | 0.822 | 0.730 | 0.385 | 0.264 | 0.101 | - | | | | | | | |
| **M10** | 0.887 | 0.374 | 0.513 | 0.724 | 0.691 | 0.321 | ns | 0.266 | ns | - | | | | | | |
| **M11** | ns | 0.147 | 0.270 | 0.478 | 0.492 | 0.201 | ns | ns | 0.184 | 0.155 | - | | | | | |
| **M12** | 0.933 | 0.353 | 0.614 | 0.925 | 0.892 | 0.425 | 0.389 | ns | 0.257 | 0.469 | ns | - | | | | |
| **M13** | 0.838 | 0.363 | 0.694 | 0.867 | 0.847 | 0.576 | 0.608 | 0.550 | 0.586 | 0.633 | 0.419 | 0.573 | - | | | |
| **M14** | 0.897 | 0.274 | 0.659 | 0.946 | 0.916 | 0.540 | 0.675 | 0.670 | 0.686 | 0.716 | 0.402 | 0.749 | 0.162 | - | | |
| **M15** | 0.749 | 0.209 | 0.537 | 0.672 | 0.689 | 0.238 | 0.252 | 0.198 | 0.322 | 0.341 | 0.124 | 0.204 | 0.313 | 0.355 | - | |
| **M16** | 0.802 | 0.280 | 0.638 | 0.850 | 0.825 | 0.507 | 0.595 | 0.577 | 0.601 | 0.621 | 0.370 | 0.601 | ns | ns | 0.303 | - |
| **M17** | 0.980 | 0.434 | 0.677 | 1.000 | 0.966 | 0.502 | 0.540 | ns | 0.340 | 0.612 | 0.219 | ns | 0.654 | 0.881 | 0.302 | 0.701 |

**(b)**

| | U2 | U3 | U4 | U5 | U6 | U7 | U8 | U9 |
|---|---|---|---|---|---|---|---|---|
| **U3** | 0.368 | - | | | | | | |
| **U4** | 0.362 | ns | - | | | | | |
| **U5** | 0.428 | ns | 0.233 | - | | | | |
| **U6** | 0.268 | ns | ns | ns | - | | | |
| **U7** | 0.219 | ns | ns | ns | ns | - | | |
| **U8** | 0.384 | ns | ns | ns | ns | ns | - | |
| **U9** | 0.282 | 0.116 | 0.190 | ns | ns | ns | ns | - |
| **U10** | 0.149 | 0.817 | 0.786 | 0.863 | 0.726 | 0.713 | 0.844 | 0.733 |

**Figure S1.** Principal Component Analysis (PCoA) of 17 *P. mugo* populations in Europe showing clear *mt*DNA markers distinction of the northern populations in Sudetes (M1) and Western Carpathians (M2 and M3) from the remaining species stands.

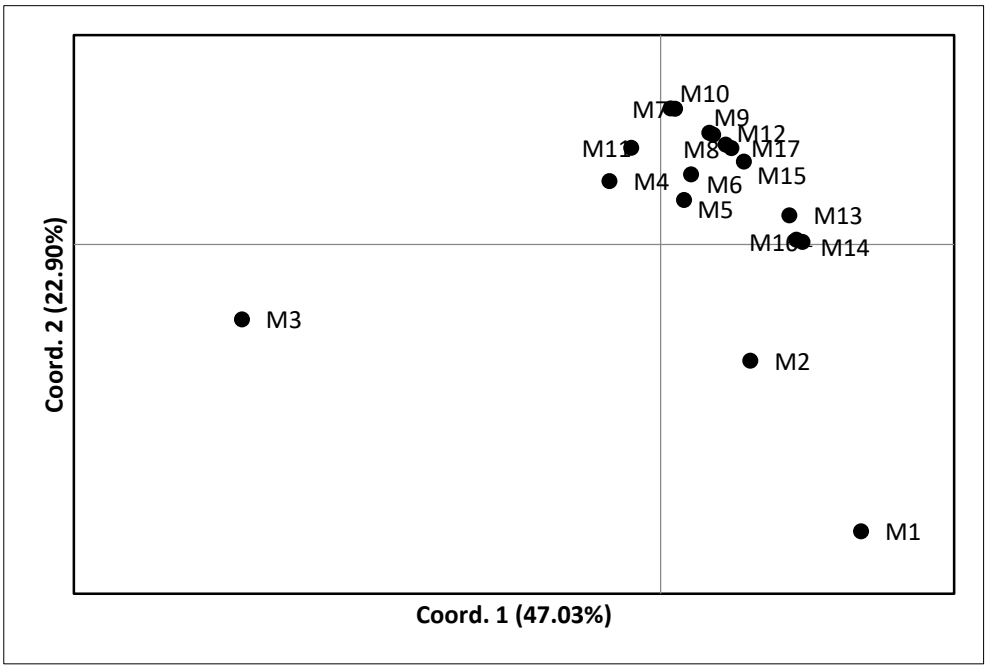



**Figure S2.** Principal Component Analysis (PCoA) of *mt*DNA variation in 9 *P. uncinata* populations (U1 was not considered). Note high percentage of differentiation assigned to the first coordinate compared with the second, which separates the Pyrenean stands from two satellite populations in Iberian Peninsula (U10) and Massive Central (U2).

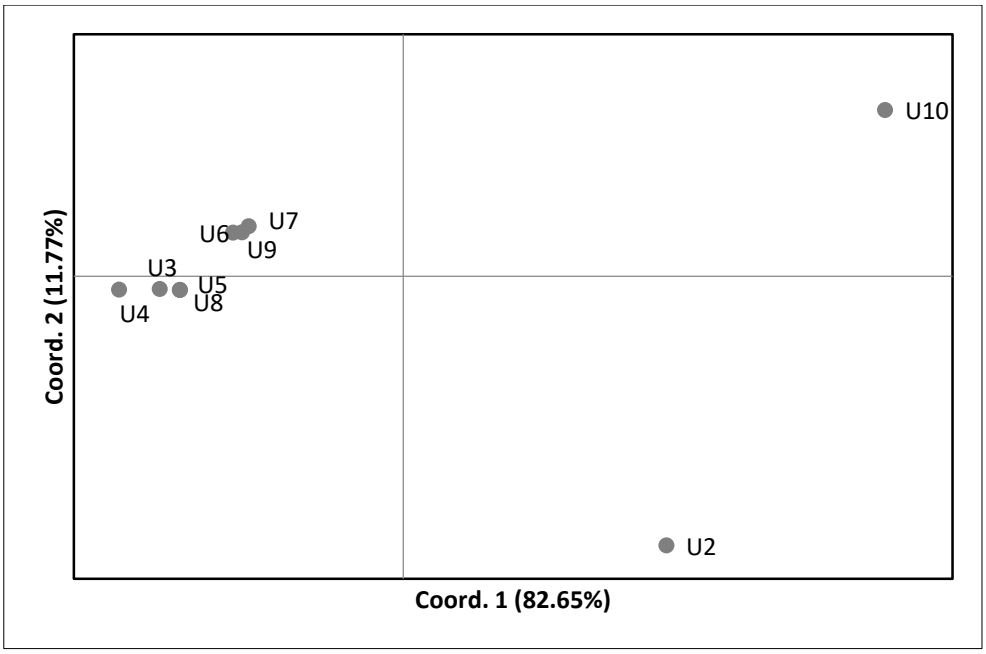
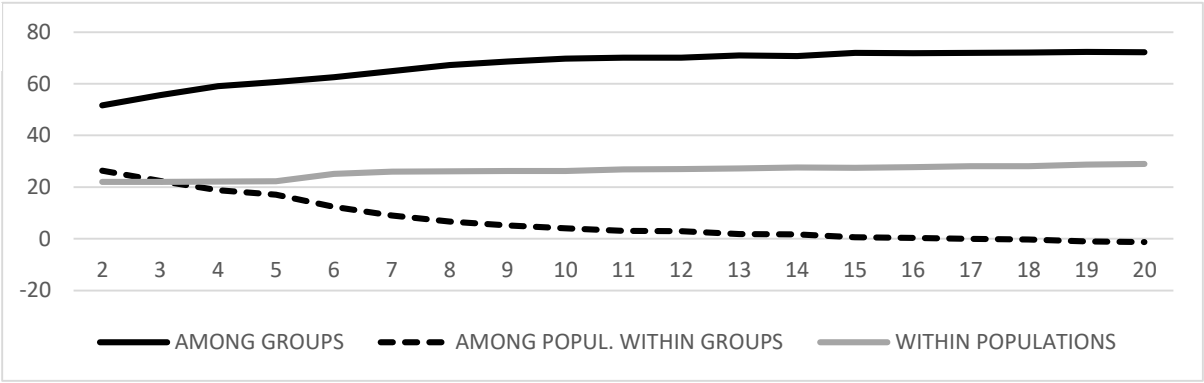
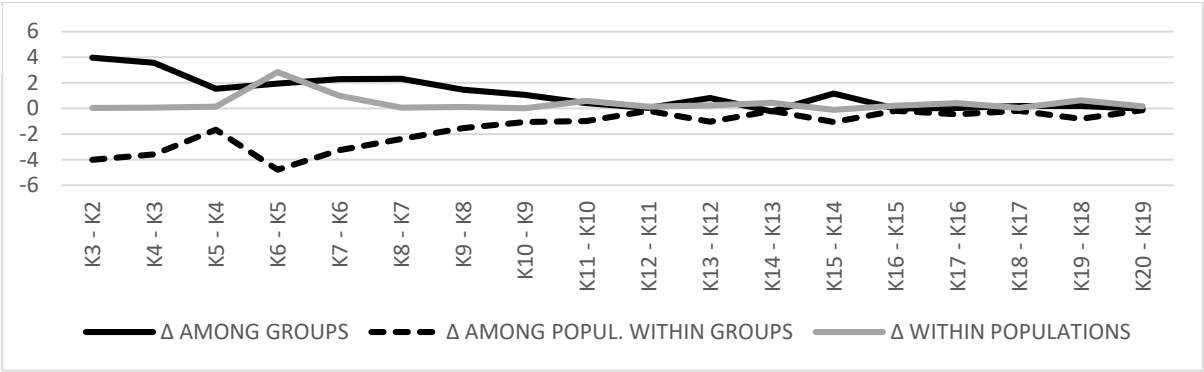


**Figure S3.** Distribution of genetic variation and differentiation indices in spatial AMOVA clustering of 27 mountain pine populations based on their mitotype variation, whit K values tested within range of two to 20. Variation distribution indices and change in their values between following K values are present on (a) and (b) plots, respectively; cluster differentiation indices and their changes are shown on (c) and (d) graphs.

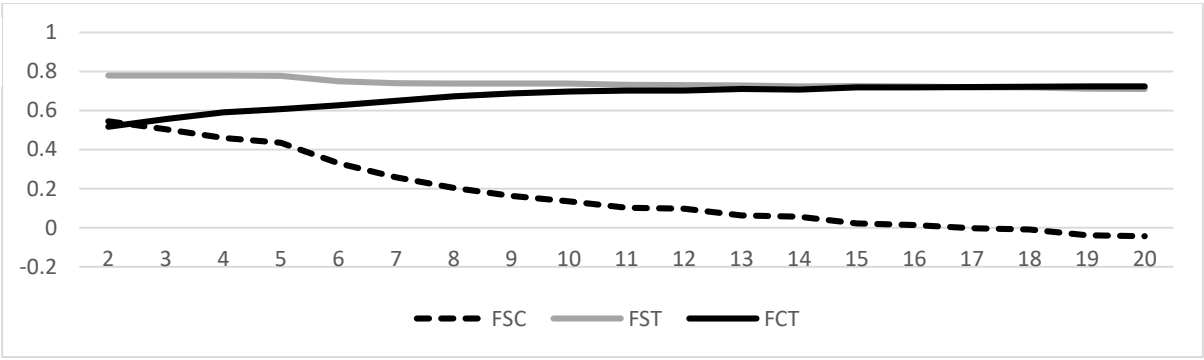
**(a)**




**(b)**




**(c)**




**(d)**

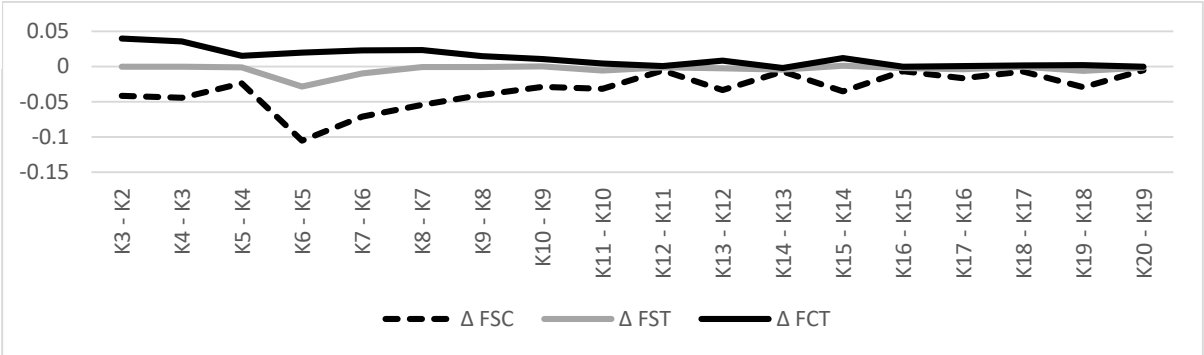
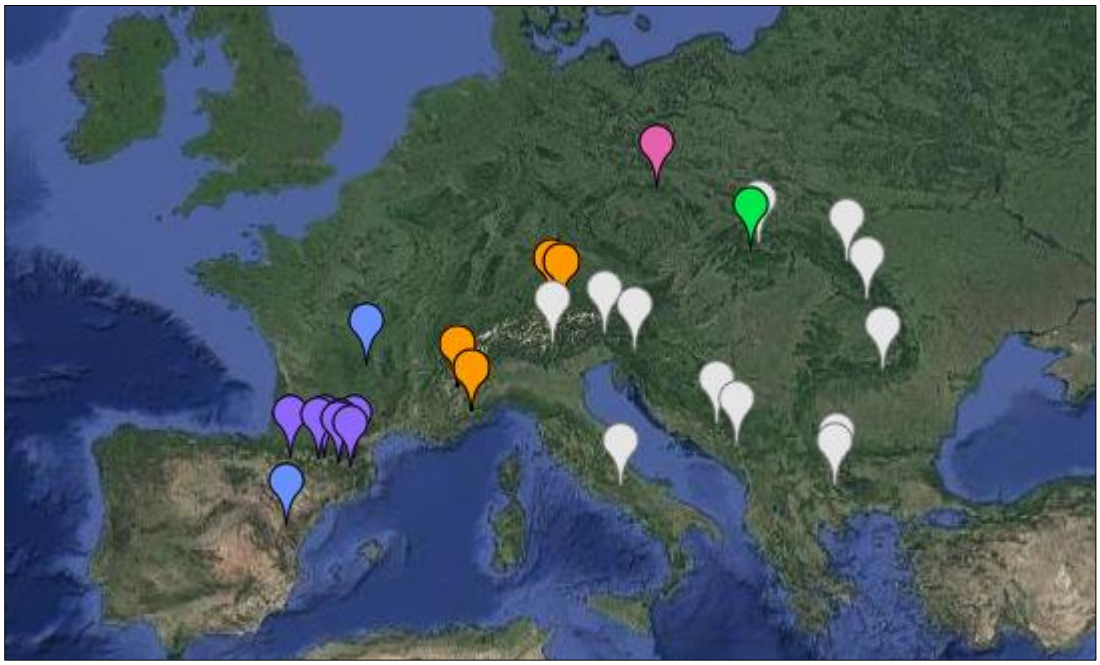


**Figure S4.** Spatial distribution of genetically homogenous clusters identified by SAMOVA analysis for K-value equal to 6 based on *mt*DNA variation in populations of two European mountain pine taxa; distinct colors indicate different genetic groups.



## 3.2. PUBLIKACJA DRUGA

Julia Zaborowska, Bartosz Łabiszak, Annika Perry, Stephen Cavers & Witold Wachowiak

# Candidate genes for the high-altitude adaptations of two mountain pine taxa

International Journal of Molecular Sciences 22: 3477

https://doi.org/10.3390/ijms22073477

2021

International Journal of
*Molecular Sciences*

MDPI

*Article*

# Candidate Genes for the High-Altitude Adaptations of Two Mountain Pine Taxa

**Julia Zaborowska [1,*], Bartosz Łabiszak [1], Annika Perry [2], Stephen Cavers [2] and Witold Wachowiak [1,3]**

[1] Institute of Environmental Biology, Adam Mickiewicz University in Poznań, Uniwersytetu Poznańskiego 6, 61-614 Poznań, Poland; bartosz.labiszak@amu.edu.pl (B.Ł.); witold.wachowiak@amu.edu.pl (W.W.)
[2] Centre for Ecology and Hydrology-Edinburgh Site, Bush Estate, Penicuik, Midlothian EH26 0QB, UK; annt@ceh.ac.uk (A.P.); scav@ceh.ac.uk (S.C.)
[3] Institute of Dendrology, Polish Academy of Sciences, Parkowa 5, 62-035 Kórnik, Poland
* Correspondence: julia.zaborowska@amu.edu.pl

**Abstract:** Mountain plants, challenged by vegetation time contractions and dynamic changes in environmental conditions, developed adaptations that help them to balance their growth, reproduction, survival, and regeneration. However, knowledge regarding the genetic basis of species adaptation to higher altitudes remain scarce for most plant species. Here, we attempted to identify such corresponding genomic regions of high evolutionary importance in two closely related European pines, *Pinus mugo* and *P. uncinata*, contrasting them with a reference lowland relative—*P. sylvestris*. We genotyped 438 samples at thousands of single nucleotide polymorphism (SNP) markers, tested their genetic differentiation and population structure followed by outlier detection and gene ontology annotations. Markers clearly differentiated the species and uncovered patterns of population structure in two of them. In *P. uncinata* three Pyrenean sites were grouped together, while two outlying populations constituted a separate cluster. In *P. sylvestris*, Spanish population appeared distinct from the remaining four European sites. Between mountain pines and the reference species, 35 candidate genes for altitude-dependent selection were identified, including such encoding proteins responsible for photosynthesis, photorespiration and cell redox homeostasis, regulation of transcription, and mRNA processing. In comparison between two mountain pines, 75 outlier SNPs were found in proteins involved mainly in the gene expression and metabolism.

**Keywords:** candidate genes; high-altitude adaptations; mountain pines; outlier loci; *Pinus*; SNP genotyping array



**Citation:** Zaborowska, J.; Łabiszak, B.; Perry, A.; Cavers, S.; Wachowiak, W. Candidate Genes for the High-Altitude Adaptations of Two Mountain Pine Taxa. *Int. J. Mol. Sci.* **2021**, *22*, 3477. https://doi.org/10.3390/ijms22073477

Academic Editor: Matthias Fladung

Received: 5 February 2021
Accepted: 24 March 2021
Published: 27 March 2021



## 1. Introduction

The sessile lifestyle of plants forces them to develop adaptations, enabling them to quickly, reversibly, and often enduringly cope with environmental changes or colonize new niches to avoid competition. Because photoperiod and temperature are important environmental factors for plants, long-lived trees must make trade-offs between survival, regeneration, development, and reproduction, fitting annual growth cycles in response to seasonal length variations [1]. For alpine species, in addition to usually strong contraction of the vegetation time, the most serious stressors include reduced atmospheric pressure, photooxidative stress, lowered temperatures, and demanding substrates [2–4]. In addition, plants are frequently exposed to strong winds, torrential rains, and avalanches, particularly affecting trees above the forest line. Mountain trees not only need to balance growth and dormancy periods through phenological adaptations but must also evolve morphological, anatomical, and physiological modifications to adjust their metabolisms or withstand mechanical damage [5–8]. Therefore, we were interested in exploring the genetic basis of the adaptations of such trees to high altitudes. Many tree species, including the pines on which the survey was conducted, maintain high fertility and enforce strong selection between recruited individuals, which facilitates adaptation to new environments. At the

same time, these trees are characterized by large effective population sizes, considerable dispersion potential of pollen and seeds and outcrossing, all of which contribute to rapid homogenization of gene pools within their populations, thereby hindering the consolidation of beneficial novelties [1]. In mountain habitats, the populations are usually insular and are subjected to rapid changes in environmental factors along the elevation gradient. These features affect gene exchange and promote the diversification of stands but tend to limit variation within them [9–11]. Biometric and quantitative genetic studies have helped to elucidate patterns of differentiation; however, knowledge remains scarce regarding the genetic basis of plants adaptation to higher altitudes and patterns of differentiation at genomic regions of high adaptive importance.

The genus *Pinus* in the Pinaceae family of conifers has a widespread distribution, mostly in the Northern Hemisphere, with nearly 100 species being described [12], including taxa associated with mountain habitats. In Europe, two well-defined high-altitude sister taxa from the *Pinus mugo* complex are known: the dwarf mountain pine (*Pinus mugo* (Turra)) and the Pyrenean mountain pine (*P. uncinata* (Ramond)). The dwarf mountain pine is a shrubby species that creates dense carpets on the ground, occasionally reaching 3 m in height. This species inhabits stands above tree lines, up to about 2700 m [13,14]. The Pyrenean mountain pine is a typical erect tree up to 25 m in height, which itself establishes the upper limits of forest and trees between altitudes of 1400 and 2700 m. Both species grow on rocks and debris, and their ecological niches appear to be comparable; however, to the best of our knowledge, this similarity was never specifically tested. The taxa meet in a wide area of the Western Alps, where both populations of exclusively one species and mixed stands, often with ongoing hybridization, are observed [15–17], except that their current ranges are essentially disjointed. The major populations of dwarf pine occupy subalpine regions of the Alps, Sudetes, and Carpathians and southern mountain chains in Romania and the Balkan Peninsula, as well as several smaller, remote populations, as in Abruzzo in Italy [13,18]. The Pyrenean pine, aside from stands in the Western Alps and the core populations in the Pyrenees, inhabits higher parts of the Massif Central, Ligurian Apennines, Jura and Vosges, as well as several more isolated areas spread over the Iberian Peninsula [19,20].

Although these two pines share a common history, have similar genomic backgrounds, and occupy comparable environments that demand the same specific adaptations, they are differentiated by a number of phenotypic and genetic characteristics. Biometric studies showed that in addition to their different growth forms, these two species differ in various cone and needle characteristics [15,18,21]. Under a common garden experiment, they exhibited notable differences in the pace of growth of young trees and in important phenological properties related to the timing of bud set and bud burst [22]. The species possess the same numbers of chromosomes, but ongoing divergence has been found in their karyotypes [23]. More detailed molecular studies, although underlining common history and genetic background, also indicated some significant changes that appeared in the genomes of these pines [24–26], such as fixed variants at mitochondrial DNA [17].

As the species overlap to a certain extent across altitude, temperature, and rainfall ranges, we might expect patterns of adaptive phenotypic variation to be similar for both. Generally, variation between species may be a consequence of adaptive changes, including selection for species-specific or high-frequency divergent alleles accumulated over the time of independent evolution, but this variation may partially result from mutation load, recombination, and demography. To date, attempts to identify potential drivers of the adaptation of these trees to mountain habits or to investigate their mutual dissimilarity have been notably limited and have focused on a small number of genetic regions. This small number of studies reflects the overall challenge in pine research related to their longevity and long generation time, which make experimental investigations impractical. Molecular studies focused on identifying potential loci under selection are also demanding, considering the large and complex genomes of pines, that hamper the acquisition of a reliable whole genome reference assembly (over 20 gigabases, rich in retrotransposons and

repetitive sequences [27–29]). Although some effort has been made to explore the patterns of neutral genetic variation between populations at plastid, nuclear, and mitochondrial markers [16,17,30], considerable research remains to be performed. To date, studies investigating genetic variation at the coding regions of these taxa have included sets of 12 and 79 candidate genes, respectively [25,31].

In the present study, we attempted to broaden our knowledge regarding the patterns of polymorphisms and divergence in two subalpine taxa to identify subjects of selection potentially responsible for their adaptations and speciation. We analyzed these two species because their evolutionary history has not been fully elucidated, and because they are both potentially threatened by ongoing habitat loss [14,20,32]. We intentionally focused on interspecies comparisons, contrasting not only the two mountain pines with each other, but also with respect to the Scots pine (*P. sylvestris* L.). The Scots pine is a species with vast but mostly lowland distribution and is the closest living relative of both mountain pines from which they diverged approximately 5 million years ago [33,34]. We conducted genome-wide analysis using a genotyping array recently developed for these pines [35]. This new Axiom_PineGAP chip (Affymetrix, Thermo Fisher Scientific, Waltham, MA, USA) was composed of 49,829 single nucleotide polymorphism (SNP) markers derived from the reference transcriptome [26], some resequenced genes and other markers known from these and other pine genomes, which appears to be especially useful for comparative evolutionary investigations. To elucidate the drivers of adaptation to higher altitudes and identify regions responsible for the further differentiation of these species, most likely caused by adjustments to local conditions, we first verified application of the SNP array to distinguish the species and subsequently evaluated levels of polymorphism and divergence along with the population structure of analyzed stands. Finally, using several approaches, we investigated outlier SNPs that strongly differ between investigated species, assessed the functions of the source genes and compared the results with previous findings concerning these taxa and other high-altitude or similarly phenotypically differentiated taxa.

## 2. Results

### 2.1. Genetic Diversity

After quality filtering of SNPs and individuals following our criteria of genomic origin and potential linkage of markers, frequency of alleles (MAF), percentage of individuals properly genotyped per marker, and percentage of SNPs genotyped per individual, the data set consisted of 6003 polymorphisms and 438 samples: 79 dwarf pines, 182 Pyrenean pines, and 177 Scots pines, derived from 14 populations spread across Europe (Figure 1, Supplementary Table S1).

The populations varied considerably in terms of the percentage of polymorphic loci. This index ranged from 61.1% to 91.3%, exhibiting a mean over the 14 populations that was equal to 80.4%, but higher averages within species reached above 90% (Table 1). The majority of the variants were shared, and none of the variants was fixed in a species while being absent from others. The taxa differed considerably in the numbers of unshared alleles (Supplementary Figure S1), which in most cases were species- and not population-specific (Table 1). The measures of mean within group distances (*d*), which in the case of Scots pine's populations did not exceed 1000, ranged between 1193.4 (M14) and 1437.1 (U18) in the stands of other two species. The observed heterozygosity (*Ho*) was generally higher than the unbiased expected heterozygosity (*uHe*), resulting in multiple negative values of fixation indices (*F*) in populations. Two heterozygosity estimates did not differ notably in terms of mean values between species; however, in all cases, except for *uHe* between mountain pines, the differences were highly statistically significant ($p < 0.001$). In populations, *Ho* exhibited the lowest levels in Scots pine (from 0.160), while populations of the two mountain pines exhibited higher values (up to 0.338; Table 1). An analogous pattern of low variability in Scots pine and larger variability in the remaining taxa was observed in the case of the expected heterozygosity (*uHe*); however, in the latter two species, the statistics decreased in relation to *Ho*, while the Scots pine's average increased (Table 1). The

fixation index (*F*) calculated on the whole dataset appeared to be positive (0.081), which was also observed in Scots pine (0.067) and its two populations; nevertheless, the majority of investigated stands exhibited greater diversity than was expected.

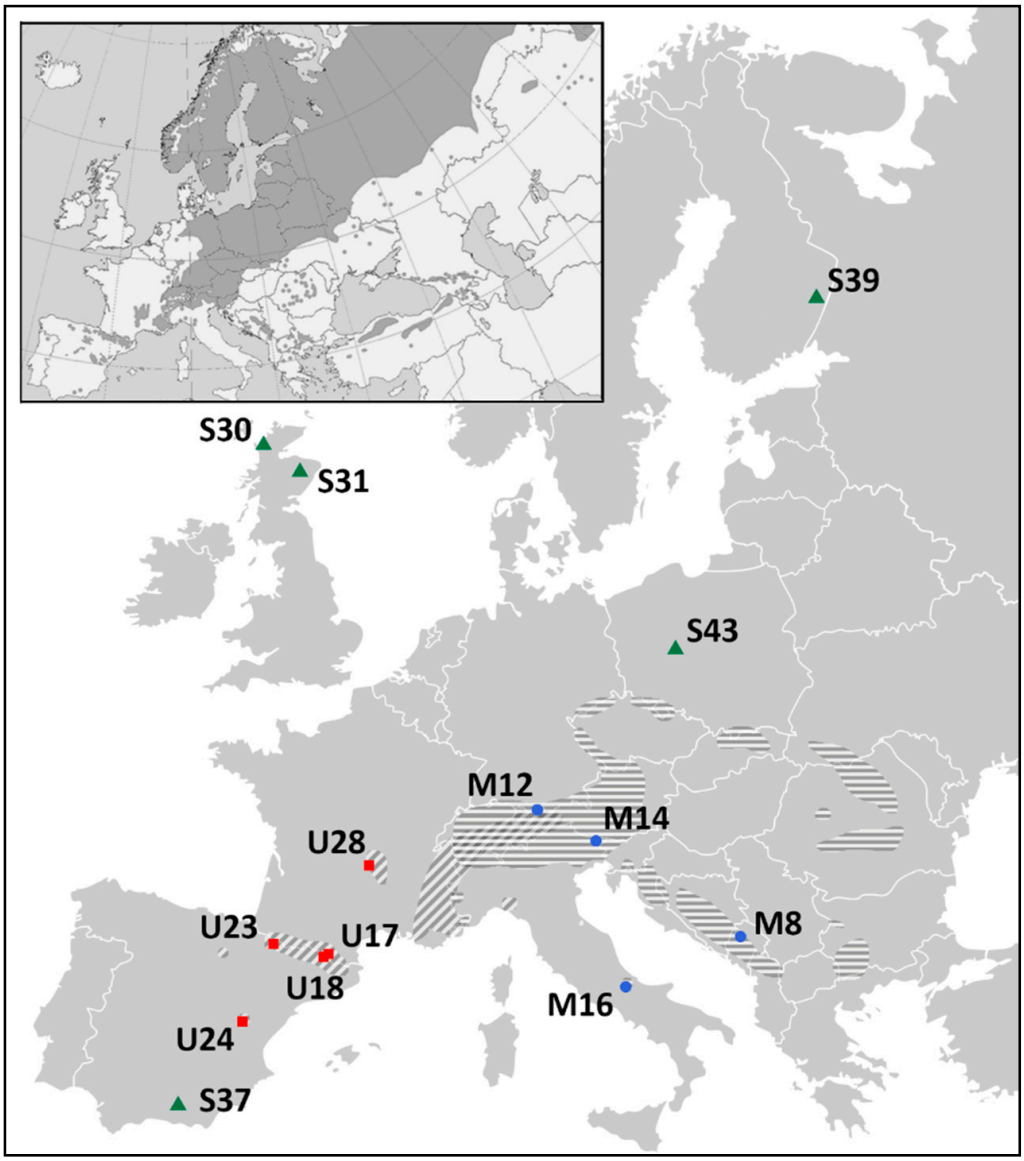



**Figure 1.** Localization of 14 studied populations with indication of species distributions. Dwarf pine—blue dots and horizontal shading, Pyrenean pine—red squares and diagonal shading, Scots pine—green triangles and dark shading on attached miniature presenting species' European range. The mountain pines' distribution map was created by the authors based on the Empty Political Map of Europe iso3166-1, downloaded from Wikimedia Commons (https://commons.wikimedia.org, accessed 4 October 2018; [36]), and the information on species' ranges taken from [19]. The Scots pines' distribution map has been obtained by courtesy of the EUFORGEN (EUFORGEN 2009, www.euforgen.org, accessed 22 September 2009; [37]), and adapted.

### *2.2. Differentiation and Grouping of Populations*

To validate the robustness of our marker set in terms of discrimination power among analyzed species and populations, and to utilize it for the reconstruction of relationships among them, we assessed our data using a series of distance, differentiation, and clustering methods.

**Table 1.** Genetic variation estimates of 14 populations and three pine species investigated.

| Group | N | %P | $S_p$ | d | $H_o$ | $uH_e$ | F |
|---|---|---|---|---|---|---|---|
| M8 | 7 | 75.4 | 1 | 1262.7 | 0.338 | 0.278 | −0.272 |
| M12 | 22 | 90.2 | 2 | 1325.5 | 0.313 | 0.274 | −0.123 |
| M14 | 20 | 82.5 | 1 | 1193.4 | 0.292 | 0.253 | −0.141 |
| M16 | 30 | 91.3 | 4 | 1282.4 | 0.285 | 0.261 | −0.060 |
| U17 | 23 | 85.2 | 0 | 1368.4 | 0.287 | 0.259 | −0.105 |
| U18 | 42 | 87.5 | 0 | 1437.1 | 0.277 | 0.256 | −0.075 |
| U23 | 47 | 84.9 | 0 | 1393.2 | 0.245 | 0.247 | −0.006 |
| U24 | 33 | 85.9 | 1 | 1264.1 | 0.298 | 0.257 | −0.137 |
| U28 | 37 | 88.8 | 1 | 1236.1 | 0.316 | 0.264 | −0.158 |
| S30 | 22 | 61.1 | 0 | 906.8 | 0.182 | 0.168 | −0.085 |
| S31 | 37 | 70.6 | 2 | 946.0 | 0.192 | 0.177 | −0.072 |
| S37 | 44 | 62.5 | 1 | 867.4 | 0.160 | 0.156 | −0.035 |
| S39 | 32 | 80.3 | 1 | 994.8 | 0.182 | 0.180 | 0.037 |
| S43 | 42 | 79.6 | 0 | 994.5 | 0.176 | 0.177 | 0.058 |
| *P. mugo* | 79 | 97.5 | 75 | 1308.7 | 0.299 | 0.272 | −0.065 |
| *P. uncinata* | 182 | 96.8 | 12 | 1426.8 | 0.280 | 0.270 | −0.030 |
| *P. sylvestris* | 177 | 92.8 | 25 | 1003.5 | 0.177 | 0.182 | 0.067 |
| Total | 438 | 100 | na | 1459.6 | 0.239 | 0.274 | 0.081 |

N—number of samples successfully genotyped in a group; *%P*—proportion of polymorphic sites; *Sp*—number of private polymorphisms; *d*—mean pairwise distance within group; *Ho*—observed heterozygosity; *uHe*—unbiased expected heterozygosity; *F*—fixation index; na—not applied. Population acronyms as in the Supplementary Table S1.

First, distances based on numbers of differences within and among groups were calculated and compared. The overall mean distance (d = 1459.6) was higher than that observed between individuals of Scots (d = 1003.5) or dwarf mountain pine (d = 1308.7) and only slightly larger than observed among individuals of Pyrenean pine (d = 1426.8); however, it was markedly lower than the distances between taxa (Table 2). The largest gap among species ($d_{xy}$ = 1761.9) was observed between Scots and dwarf pines; the second largest gap, which was considerably smaller, was between Scots and Pyrenean pines with $d_{xy}$ = 1552.3, while the two mountain species were separated by 1523.5 nucleotides of differences on average. A similar pattern was also reflected in the results obtained from respective comparisons of population pairs (Supplementary Table S2a,b), with exceptions only being observed in cases of populations U28 and U24, which showed a closer relation with Scots pines' populations than to populations of dwarf pine and the majority of populations from their own taxon. Based on these distances, a UPGMA tree was constructed (Supplementary Figure S2a), which depicted precise separation of Scots pine and dwarf pine with a less clear position of the third species. Populations of Pyrenean pine appeared to be split into two clusters, with one, consisting of populations U17, U18, and U23 (all from the Pyrenees) being set at the base of the dwarf pine branch, and the other—populations U24 and U28—being set at the base of the Scots pine part. On the dendrogram constructed without Scots pine, dwarf, and Pyrenean pines occupied separate branches (Supplementary Figure S2b).

**Table 2.** Genetic distance and differentiation between the three investigated pine species.

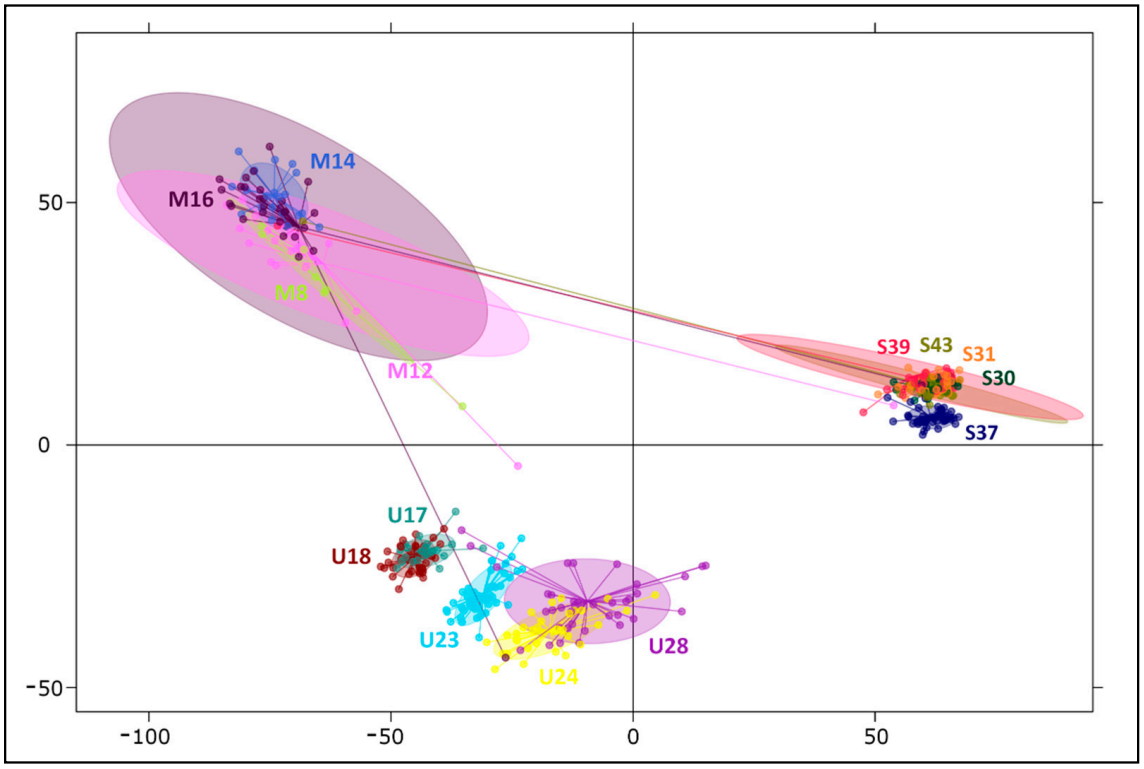
| Comparison | $d_{xy}$ | *Fst* (a/b) | Outlier SNPs (2**/2*/3*) |
|---|---|---|---|
| *P. mugo* vs. *P. uncinata* | 1523.5 | 0.1061/0.1510 | 12/75/9 |
| *P. mugo* vs. *P. sylvestris* | 1761.9 | 0.3612/0.3937 | 0/1/0 |
| *P. uncinata* vs. *P. sylvestris* | 1552.3 | 0.2162/0.2629 | 12/29/6 |
| mountain pines vs. *P. sylvestris* | 1615.7 | 0.2324/0.2969 | 7/35/2 |

$d_{xy}$—mean distance between populations from two species; *Fst*—differentiation index (calculation method: a—species considered as one group/b—hierarchical structure with populations nested in species), all results were significant at $p$ value $\leq 0.001$; numbers of outlier SNPs identified at: 2**—statistical significance level of $p/q$ value $\leq 0.01$ and minimum two methods simultaneously; 2*—$p/q$ value $\leq 0.05$ and minimum two methods in concert; 3*—minimum three methods at $p/q$ value $\leq 0.05$.

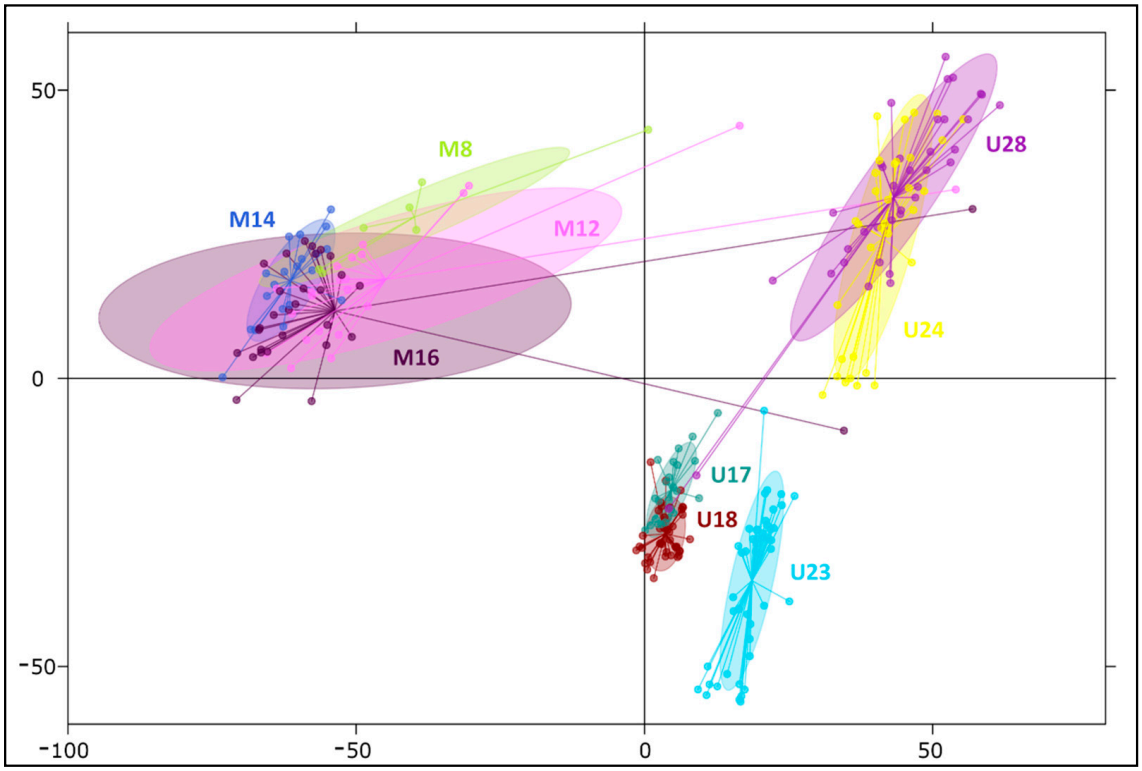
Analysis of molecular variance (AMOVA) with populations hierarchically grouped within their respective species demonstrated that the majority of variance present in our data segregated inside populations (73.28%), and the variation between them was considerably lower (5.43%) than that among taxa (21.28%); all these results were statistically significant ($p < 0.001$). In the AMOVA with no imposed structure, 22.51% of the variation differentiated taxa and 77.49% segregated within them. The overall *Fst* fixation index values in these two versions of the test were equal to 0.2672 and 0.2251, respectively. More specific calculations of differentiation indices (Table 2, Supplementary Table S2), supported the pattern demonstrated by distance analysis. The most differentiated of all interspecific comparisons was the dwarf and Scots pine pair (*Fst* = 0.3612), and taxa from the *P. mugo* complex reached a considerably lower level of *Fst* = 0.1061 (or 0.1510 when population structure was included). Comparisons between individual populations, where *Fst* values ranged between 0.015 (M8 vs. M12) and 0.451 (M14 vs. S37), generally confirmed previous findings. One exception was found in the case of Pyrenean pine population U28, which exhibited a closer relationship with most Scots pine's stands than with the Alpine dwarf pine provenance M14. Moreover, most of the populations in this species showed genetic similarity to dwarf pine population M8, often exhibiting higher similarity than to other populations in their own taxon. All the *Fst* values between taxa and among populations were significant ($p < 0.001$).

Among clustering analyses, two multivariate methods were implemented. To determine how our SNPs discriminate populations of studied taxa, principal coordinate analysis (PCoA) was conducted, and principal component analysis (PCA) was subsequently employed to characterize variation within populations and to elucidate individual genotypes. In comparison between all 14 investigated populations, the three first axes of PCoA explained 58.8%, 12.6%, and 5.7% of the original variation, respectively, while in PCA, considering variation among individuals, these values decreased to 12.3%, 3.5%, and 2.0% (Figure 2a and Supplementary Figure S3a), confirming that a greater part of the diversity segregates within populations. These results confirm the distinction of three gene pools correlating with the studied species. Generally, horizontal axes separated dwarf pine from Scots pine, and Pyrenean pine was distant from them along the second axis, remaining somewhat closer to the other mountain taxon. The PCA plot (Figure 2a) exhibited greater variability within species and populations and facilitated the identification of some outlier samples. Five individuals grouped with populations of other species, indicating their putative hybrid origin, or less likely, mixture of the samples. In multivariate inquiries focused on two mountain pines, PCA axes accounted for 6.6%, 3.4%, and 2% of variation (Figure 2b), and in PCoA, they explained 44.5%, 16.0%, and 11.6% of diversity (Supplementary Figure S3b), in both cases distinguishing first between taxa, then between Pyrenean pine stands (U17, U18, and U23 vs. U24 and U28). The second axis also demonstrated only slight distinction of the dwarf pine stand M8 in the Dinaric Alps from the rest of its populations; however, 95% confidence intervals of the dwarf pine populations remained overlapping.

To verify the hypothesis of isolation by distance (IBD), i.e., to formally determine whether genetic differences are larger between populations when geographic gaps among them increase, we performed Mantel tests of correlation between genetic and geographic distances. This analysis was needed to ensure that STRUCTURE analyses would not be affected by bias caused by data with allele frequencies varying gradually, as the model underlying this clustering method is not well suited for such scenarios. These analyses were computed for the whole set of 14 populations, as well as in two-species comparisons and within taxa. Except for intraspecies tests in all other cases, these correlations were positive and significant (Supplementary Table S3 and Figure S6a,b). We consider the three negative outcomes—weak and insignificant correlations of within-species inspections; and the fact that the positive outputs might have been biased by specific geographic distributions of studied populations and generally unequal ranges of the taxa to be sufficient to not expect any disturbances in further tests.



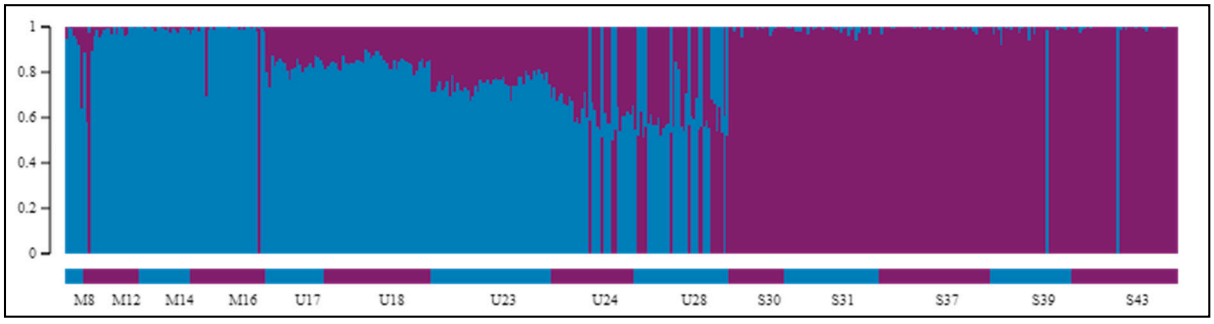

(**a**)



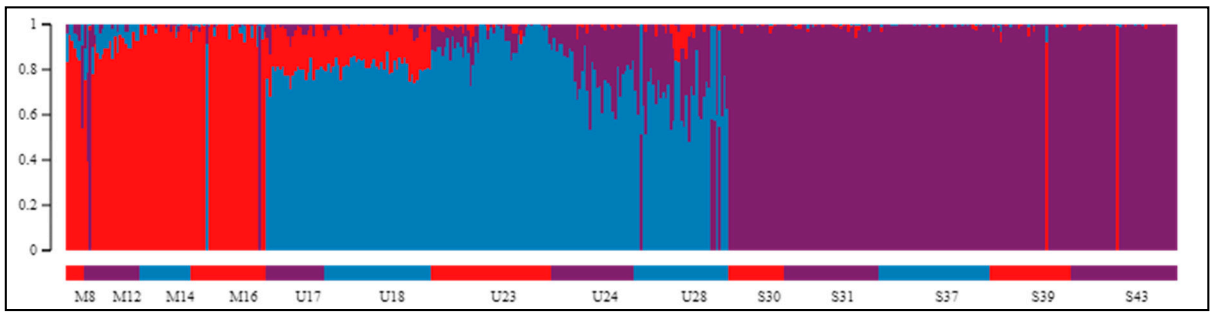

(**b**)

**Figure 2.** Principal component analysis (PCA) revealing relationships between and within the studied pine species. (**a**) Division of the three investigated taxa, first and second coordinates (horizontal and vertical axes, respectively) correspond to 12.3% and 3.5% of variation; (**b**) separation of two mountain species with 6.6% of variation explained by the first and 3.4% by the second component. Ellipses indicate 95% confidence intervals.

The Bayesian clustering method implemented in STRUCTURE, which was performed on our SNP data, helped to elucidate the genetic structuring of individuals. Upon two employed methods (Evanno and L(K) method) for indication of best genetically explicit grouping from the examined range of *K* values between 1 and 10, division into two groups (Figure 3a) was indicated to be optimal (Supplementary Figure S4a,b). According to STRUCTURE HARVESTER, the step between $K = 1$ and $K = 2$ was characterized by the largest change in likelihood distribution and the highest Delta*K* score. When we excluded $K = 1$ from the analysis, because this point dimmed the picture with its markedly low anticipated and obtained probability of data, the program suggested division into five clusters as the best (Figure 3b). Splitting the samples into two groups ($K = 2$, Figure 3a) validated the distinction between populations of the *P. mugo* complex and Scots pine. The division was clear in the case of the dwarf pines that had very little contribution from the second cluster, whilst among specimens of Pyrenean pine, mixing of both clusters was manifested. In these individuals, dwarf pine participation dominated based on *Q* values constituting, on average, 71% (SD = 13.5%) of their genomes. In the case of $K = 3$ (Supplementary Figure S4c), each species established its own cluster, and only in Pyrenean pine was potential gene flow from the other groups found. In populations from the Eastern Pyrenees (U17 and U18), the influence of dwarf pine was primarily observed; at the same time, populations from more remote stands (U24 and U28) displayed a considerable contribution of Scots pine in their genomes. Generally, outcomes from tests with higher *K* values (Supplementary Figure S4d–g) were characterized by the existence of a stable, coherent group of dwarf pine samples and increasing differentiation within Scots pine while still maintaining the consistency of individual populations, especially in Pyrenean pine, where three provenances showed greater variability and inner differentiation. In the exact outcome of clustering samples into five groups (Figure 3b), which was second-best based on the Delta*K* criterion, Pyrenean pine was divided into two parties, and Spanish population S37 was separated from the rest of the Scots pine's samples.

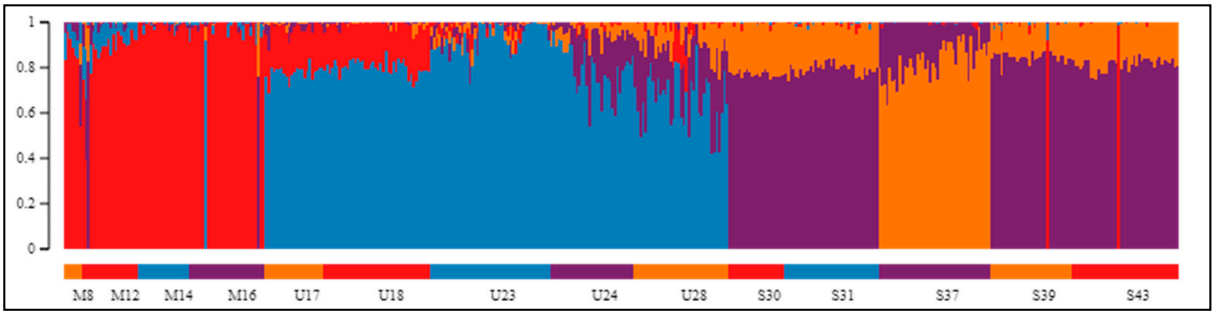



(**a**)

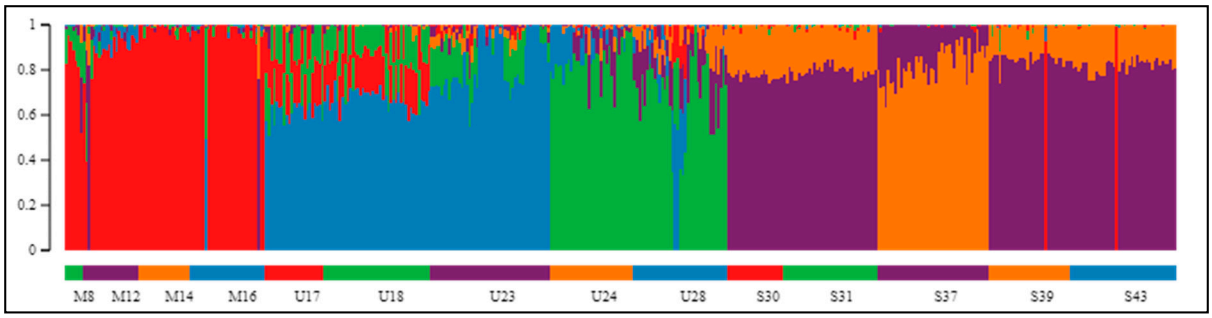



(**b**)

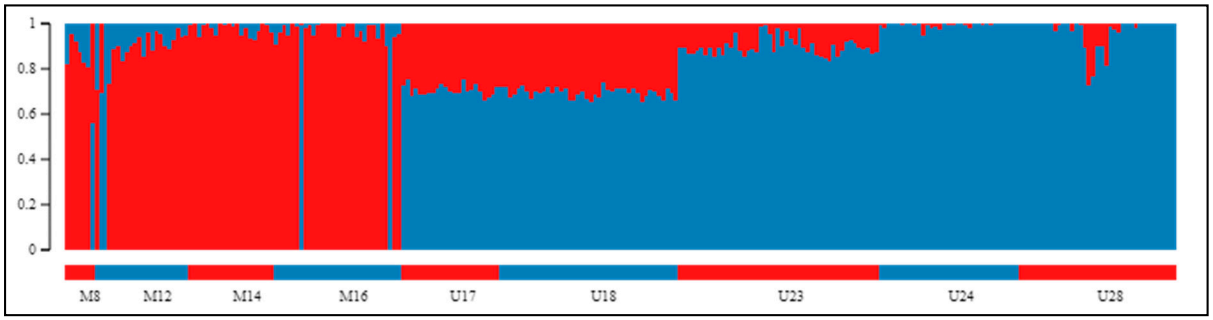
**Figure 3.** *Cont.*



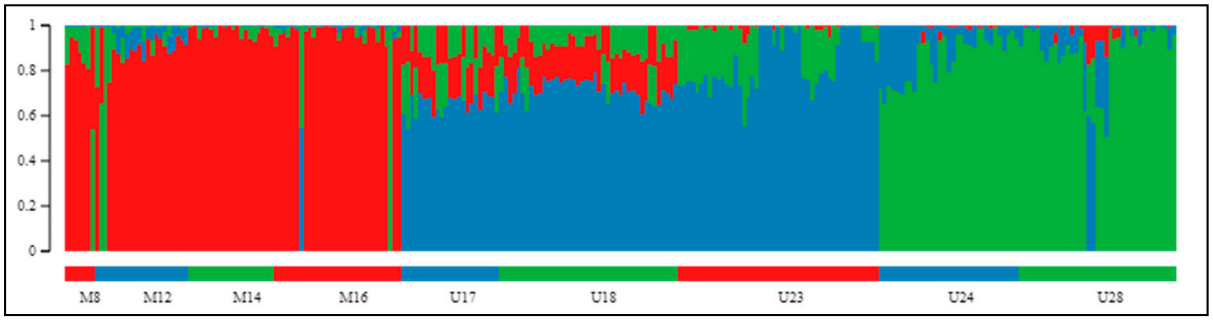

(**c**)




(**d**)




(**e**)




(**f**)

**Figure 3.** Clustering of 14 inspected pine populations, as proposed by the STRUCTURE algorithm. Separation of populations of three species into (**a**) two, (**b**) three, (**c**) four, or (**d**) five distinct groups. $K = 2$ and $K = 5$ were pointed out as the best supported clustering, depending on whether the result for $K = 1$ was under consideration. (**e**) After exclusion of Scots pine samples, $K = 2$ was indicated as the most likely assemble, closely corresponding with the two mountain taxa. (**f**) The second best supported structure of the tested *P. mugo* complex populations consisted of three groups ($K = 3$): dwarf pine specimens, Pyrenean pine representatives from Pyrenees and its accessions from isolated populations in Sierra de Gudar and Massif Central together. Scale on the left and vertical bars represent the proportion of each genome being composed of variants specific to particular clusters. Horizontal color bars at the bottom of each chart label distinct populations.

The results of analyses focused on two mountain taxa yielded very similar patterns. First, clear discrimination between dwarf and Pyrenean pines was obtained for $K = 2$ (Figure 3c; Supplementary Figure S5a,b). The second most likely structure indicated by STRUCTURE HARVESTER was composed of three clusters, supporting strong subdivision within Pyrenean pine (Figure 3d). Models assuming the existence of more groups confirmed the substructure in Pyrenean pine populations, and only once, for $K = 7$, was differentiation in dwarf pine detected (Supplementary Figure S5c–h). Interestingly, at each grouping variant, the same few individuals were consistently observed to manifest affiliation to a cluster corresponding with a species different from the one expected, and these were exactly the same samples that already appeared inconsistent on PCA plots.

### 2.3. Outlier SNPs and Their Functional Annotation

From the pairwise comparisons between taxa, the most differentiated markers based on their frequencies and high *Fst* values were obtained. Outlier SNPs were detected with three methods—BayeScan, FLK, and FDist. We focused on the markers that were indicated by at least two methods at a threshold level of significance equal to 0.05 (depending on the analysis used, it was either $p$ or $q$ value). We later performed a more in-depth examination of the loci that were found simultaneously in three analyses or were significant at more stringent level ($p/q$ values = 0.01). In total, in all the comparisons conducted between species, 120 outlier SNPs were confirmed by at least two methods at the significance level of 0.05 (Supplementary Tables S4 and S5).

When two high-altitude taxa were together compared with the Scots pine populations, 35 detected SNPs corresponded in at least two tests at a significance level of 0.05. These markers closely correlated with those found to differentiate between Pyrenean and Scots pines (29 outliers), while only one such marker was detected between the last species and dwarf pine (Supplementary Table S4). Between the two mountain species 75 highly divergent SNPs were identified (Figure 4). However, only 9 SNPs were indicated by all three methods at the same level, whereas at a more stringent threshold of 0.01 at least two methods agreed on 9 markers; eight were common to these two sets and were considered to be the most reliable outliers (Table 3 and Supplementary Table S4). No fixed SNP allele was observed in any of three taxa that would be absent from the others (Supplementary Table S5).

Each of the 120 highly divergent outliers was subjected to Gene Ontology annotation. Sequences from the reference transcriptome (contigs) on which the SNPs were called and derived from were subjected to OmixBox, where they underwent phases of BLAST search, GO mapping, gene annotation, InterProScan annotation and final merging of the results from last two steps. From the initial input 5 contigs yielded no alignment in the xblast search (additional screening of the expressed sequence tags (EST) database with NCBI blastn to identify related nucleotide sequences that could substitute the original queries proved equally inefficient). Furthermore, 15 sequences failed at the BLAST quality filtering, mapping or annotation step, leaving 100 loci that successfully passed the whole annotation process, described by 239 different GO terms in total (Supplementary Table S4). The determined gene ontologies represented each of the three general GO domains—biological processes (BP), cellular components (CC), and molecular functions (MF), and numbers of terms assigned to sequences varied strongly reaching maximum of 14 ontologies in one case.

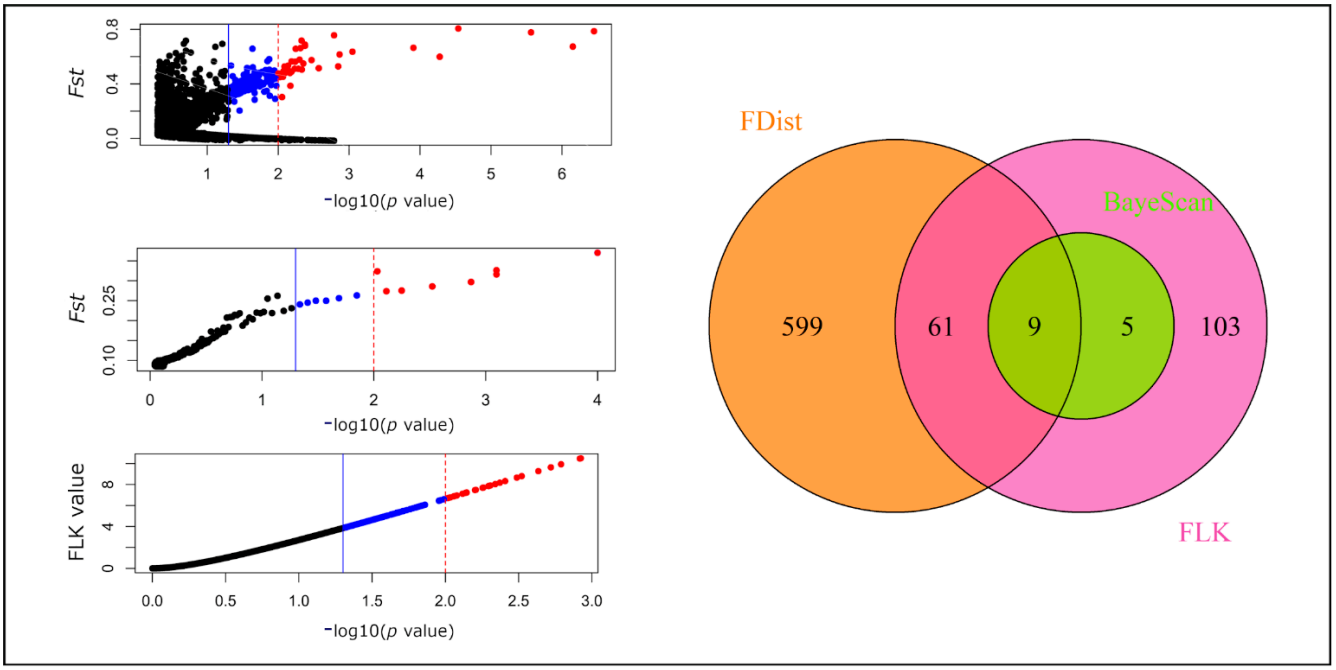



**Figure 4.** Outlier SNPs detected by three methods in comparison between populations of dwarf and Pyrenean pine. Three panels on the left represent outlier markers (significant at $p/q$ value $\leq 0.05$—blue and red points, the first values in brackets, or $p/q$ value $\leq 0.01$—red points only, the second values in brackets) revealed by FDist (207/43; **top**), BayeScan (14/8; **middle**) and FLK (178/28; **bottom**) methods. The right panel with a Venn diagram shows how the results from the less stringent level of significance in the three analyses overlap. From the output of the FDist method, only SNPs with *Fst* above 0.02 were considered significant and are colored.

**Table 3.** Description of sequences containing the eight most credible outlier SNPs differentiating the two studied mountain pine species.

| Sequence Name | Sequence Description and Definitions of Related GO Terms (Order: BP/MF/CC) |
|---|---|
| comp20176_c0_seq1 | AP-4 complex subunit mu: protein targeting; Golgi to lysosome transport/*nd*/cytosol; cytoplasmic vesicle; trans-Golgi network; AP-4 adaptor complex; clathrin adaptor complex |
| comp39941_c0_seq1 | glucan endo-1,3-beta-glucosidase 8-like: carbohydrate metabolic process/hydrolase activity, hydrolyzing O-glycosyl compounds/anchored component of plasma membrane |
| comp41821_c0_seq1 | heavy metal-associated isoprenylated plant protein 36-like: metal ion transport/metal ion binding/nd |
| comp44835_c0_seq1 | protein strictosidine synthase-like 3 biosynthetic process/strictosidine synthase activity/nd |
| comp50552_c0_seq1 | phosphatidate phosphatase PAH2-like: hydrolase activity/cellular metabolic process/nd |
| comp52994_c0_seq1 | unknown: transcription coregulator activity/regulation of transcription, DNA-templated/nd |
| comp53610_c0_seq1 | aminodeoxychorismate synthase, chloroplastic isoform X1: biosynthetic process; carboxylic acid metabolic process/nd/nd |
| comp54487_c0_seq1 | serine/threonine-protein kinase VPS15-like isoform X1: protein kinase activity; protein binding; endosome; autophagy/vacuolar transport; phosphorylation/phosphatidylinositol 3-kinase complex, class III; intracellular transport |

BP—Biological Process; CC—Cellular Component; MF—Molecular Function; nd—no data for particular GO domain.

Three comparisons among Scots and mountain pines (MU vs. S, M vs. S, and U vs. S) yielded only one marker that was supported by them all, and it was the singular marker that diverged between dwarf and Scots pines. This particular SNP was observed in a transcript (comp28590_c0_seq1) with very good xblast and annotation quality. The sequence was described as photosystem I P700 chlorophyll A apoprotein A2 (phyA), an integral component of the thylakoid membrane, which participates in electron transfer during photosynthesis. Except for that particular sequence, both alpine taxa were differentiated from Scots pine in 34 other loci, of which 29 were correctly annotated. Half of these markers (17) were also found at variance in Pyrenean pine and its lowland relative when they were analyzed independently along with 12 additional outlier loci not found in any other comparison (Supplementary Table S4).

From genes with SNPs differentiating two focal pines, 241 terms were elucidated. More of the ontologies belonged to MF (45 different terms found for 50 out of 64 efficiently annotated transcripts) and BP domains (42 terms for 48 transcripts) than were affiliated with CC (23 terms found for 38 sequences). Molecular functions were most commonly represented by the binding of organic cyclic compounds, heterocyclic compounds, or ions and by transferase, transcription regulator, and catalytic activities (Supplementary Figure S7). Among biological processes, metabolic processes (organic substances, cellular, primary, or nitrogen compound metabolic processes) and biosynthesis were most heavily represented. Gene products of the sequences with outlier SNPs identified were located mainly in organelles, the membrane, often constituting its intrinsic components, and cytoplasm. In addition to examining the most frequent ontologies, we also inspected the GO terms identified for the eight most significant outliers (Table 3), which were supported by all three methods applied for detection and, in most of the cases, were significant at $p/q$ value of 0.01 too.

## 3. Discussion

### 3.1. Discrimination of the Species

In this study, we analyzed genetic relationships between a pair of closely related coniferous taxa—dwarf mountain pine and Pyrenean mountain pine. Studies of the adaptive variation of these species are especially important due to environmental changes and ongoing habitat loss in mountainous areas. Although these taxa face rather similar environmental pressures and share a common genetic background [25,31], they differ significantly in a number of characteristics and are widely recognized to be two separate taxa [38]. For evolutionary assessments we used a novel SNP array and the reference Scots pine to, first, determine the discrimination power of this new tool in the studied system, and to elucidate the genetic differentiation and divergence of these species. We further looked for candidate loci that could be responsible for ecological adaptation and speciation of these pines.

The number of finally used markers (6003 SNPs) appeared low compared to the number of those primary arranged on the array (almost 50 thousand). This drop was a result of a combination of factors, first of all of a medium-low conversion rate on the test set of samples (about 42%, better described in [35]). The second drop, from the 20,795 markers that passed the initial test, to the final six thousand, resulted from the worse quality of *P. mugo* genotyping in this study, probably an outcome of imbalanced mixing of species samples during array development. To keep as many individuals of this focal species as possible, we made a trade off on the number of markers, excluding those that performed the worst in *P. mugo*.

Patterns of genome-wide polymorphism at a set of nuclear SNPs in the analyzed pines showed that most variation (over 73% in AMOVA analysis) was distributed within populations. This result is in line with patterns previously found in the case of pollen transmitted chloroplast markers [16,30,39] and biparentally inherited nuclear loci [25,26,31], but is different from those observed in less mobile, seed-mediated mitochondrial genomes [17]. The levels of variation were comparable between species, and there were relatively low

numbers of unshared polymorphisms detected in species and individual populations. The genetic diversity was sufficient, however, to delineate three investigated taxa and many of their populations, supporting the genetic differentiation between specimens of the Scots pine and two mountain pine species. Moreover, although some particular interspecies pairwise population comparisons deviated from the general pattern, populations within the *P. mugo* complex exhibited more common variants and their allele frequencies were more evenly distributed, as observed in distance and differentiation indices, than between any of these samples and the five Scots pine populations.

For the majority of statistical analyses, a closer relationship between Pyrenean pine and Scots pine was observed, rather than between the latter and dwarf pine, and the pattern was mostly affected by two Pyrenean pine populations. A close genetic relation between these taxa, which could have preserved more of the ancestral polymorphisms, was also observed in earlier comparative analyses of transcriptomes, phenology, and certain phenotypic traits (growth form and rate, needle characters [18,22,26]). Interestingly, notably more outliers were found in this instance in comparisons between Pyrenean and Scots pines than between the reference species and dwarf pine—1 and 35, respectively.

Our study provides a new example [40–42] of a successful application of a genome-wide scan for delineation of phylogenetic relationships among taxonomically challenging plant groups which, in this study, were from the *P. mugo* complex [28,38]. The two sister taxa appeared separated in multivariate analysis plots, having disjointed 95% confidence intervals in PCA tests. Definite distinction into two genetically distinct groups was also determined by Bayesian clustering of dwarf and Pyrenean pines. The species showed moderate but significant differentiation by AMOVA (*Fst* = 0.151); moreover, a larger proportion of variation was found among species (9.14%) than among populations (5.95%). Similar to previous resequencing studies of nuclear loci [25,31], no fixed differences between these pines have been demonstrated.

### 3.2. Inner Genetic Variation of Three Pine Species Populations

Our data showed only minor differentiation between dwarf pine populations. This differentiation was quite visible in constant homogeneity of dwarf pine samples on STRUCTURE charts but was also marked by the lowest values of pairwise *Fst* estimates—on average, half the size of that observed in the two other species. Only the Abruzzian site M16 emerged as an individual cluster in Bayesian clustering under $K = 7$. Earlier studies showed that trees from that population exhibit many phenotypic [18,21] and genetic [43,44] characteristics of Pyrenean pines. The present investigation on nuclear markers, however, cannot conclusively confirm the introgression hypothesis, as the population as a whole did not differ from others of its species, neither being more distant nor showing exceptional pairwise *Fst* estimates. This population also did not show signs of closer affinity to Pyrenean pine such that more than that one specimen from this stand was grouped with samples of the sister species in PCoA analysis (Figure 2) and on charts obtained from STRUCTURE (Figure 3 and Supplementary Figure S5). Therefore, the characteristics of this population most likely result from its isolation in remote location [14,45], rather than inflow from Pyrenean pine. Apart from that finding, in multivariate analyses focused on two mountain species another stand; namely, M8 from the Dinaric Alps, displayed some distinction. Nevertheless, on both variants of UPGMA trees all dwarf pine sites constituted one coherent group, which appeared to have formed in a relatively short and distant period of diversification.

More differentiated were populations of Pyrenean pine. Each of the populations displayed some symptoms of inner variation or stronger differentiation from others on STRUCTURE plots, and the highest mean within and between group distances were observed in this species and its populations. Two sets of populations, one including Pyrenean sites and the other composed of stands in Massif Central and Sierra de Gudar, were distinguished by the Bayesian method and multivariate tests. Further differentiation of these populations was observed in the Pyrenean cluster, with the northwestern population U23

being distinct from two Andorran sites, which was in keeping with prior observations of a significantly slower growth rate of pines in this region [22].

Two remote Pyrenean pine stands from France and southern Spain (U24 and U28, respectively) appeared particularly similar to each other, as they grouped together on PCA and PCoA plots, the phylogenetic dendrograms, and equally clearly in STRUCTURE analysis. These stands also exhibited the greatest diversities of individual genotypes, well expressed in measures of heterozygosity and on PCA plots. Moreover, the two populations were characterized by considerably higher similarity to Scots pine, and on the UPGMA tree with all three taxa included, they were even located at the branch specific to Scots pine. Distinctness of the two stands and their striking genetic similarity, especially in the context of significant mutual geographical distance, was observed earlier based on mitochondrial and chloroplast markers [17,30,39] and some biometric traits [18], suggesting a long evolutionary relationship and/or past introgression, especially since plausible mixing of Pyrenean and Scots pines within stands in Massif Central has already been considered [17,24].

Although Scots pine was not the focal species in our study, some observations regarding its genetic variation may be noted based on the conducted analyses. Distinct patterns of differentiation of populations within species were obtained using our two approaches. Estimates based on numbers of nucleotide differences ($d_{xy}$) showed little variation among Scots pine populations compared to the other taxa, whilst conclusions based on allele frequency differentiations (*Fst*) indicated that this species is the least consistent of all. Although Scots pine appeared very consistent not only on the UPGMA tree but also on the plots obtained with multivariate analyses, STRUCTURE showed divergence of the Iberian Peninsula population (S37) that also exhibited the highest *Fst* in pairwise tests in relation to provenances of its own and of other species, which can be interpreted as a signal of its stronger spatial isolation from other species' sites [46].

### 3.3. Candidates for Drivers of High-Altitude Adaptations in the Studied Subalpine Pines

Loci, from which the 35 outlier SNPs differentiating two mountain species from the reference taxon were derived may serve as potential targets of selection that have shaped today's diversity of these pines. Transcripts for which gene ontology annotation was possible primarily represent proteins taking part in photosynthesis, cell redox homeostasis, photorespiration or oxidant detoxification, regulation of transcription and mRNA processing. Some of these genes encoded diverse protein-related agents (with proteolytic, kinase or methylotransferase activity), molecules participating in transport or in cortical microtubule organization. The set of identified genes corresponds to previous studies in alpine conifers and other plants, in which the existence of various adaptive strategies for reducing photodamage (a serious stressor under extended periods of high irradiance, reduced partial pressure of $CO_2$, and low temperature, which are often present in high-altitude areas) has been demonstrated [47–53].

Transcripts containing outlier SNPs identified between dwarf and Pyrenean pines were commonly involved in metabolic and biosynthetic processes. The largest group, representing approximately one-third, was related to gene expression, and the others were responsible for the metabolism of proteins, nitrogen compounds or phosphorus, transport, and responses to stimuli. Complementarily, the prevailing molecular functions of genes were related to nucleic acid binding and transcription regulator activities. Other frequent MFs were associated with metal ion binding, protein binding and diverse catalytic activities. Additionally, one of the most significant outliers observed in this study (see comp52994_c0_seq1 in Table 3) and a homeobox domain-containing protein discovered as strongly diverged in previous interspecies studies on these pines (marker Pr1_10 in [31]) are also both involved in the regulation of transcription. This finding indicates that an important role is played by the differentiated expression of genes in the evolutionary history of the studied organisms and may partly explain the overall low differentiation of these species at the sequence level [31]. This finding may also indicate an important

role played by the phenotypic plasticity mediated by transcriptional regulation in diverse responses of these pines to changes in environmental conditions [54,55]. Clearly, more detailed comparative transcriptomic studies are warranted, for which the foundation has already been established [26].

Along with the mentioned transcript encoding homeobox domain containing protein, seven other sequences containing SNPs that were highly diverged between two mountain pines, supported by three methods and significant at the 1% level, were annotated as follows: AP-4 complex subunit mu, glucan endo-1,3-beta-glucosidase 8-like, heavy metal-associated isoprenylated plant protein 36-like, strictosidine synthase-like 3 protein, phosphatidate phosphatase PAH2-like, aminodeoxychorismate synthase (chloroplastic isoform X1) and serine/threonine-protein kinase VPS15-like (isoform X1). The first molecule is an element of the adaptor protein (AP) complex and most likely functions as a vesicle coat component involved in targeting proteins from the trans-Golgi network (TGN) to the endosomal-lysosomal system [56,57], thereby participating in the regulation of intracellular transport. The second protein, recognized as glucan eno-1,3-beta-glucosidase 8, is an enzyme with hydrolase activity bound to the extracellular side of the lipid bilayer of the cell membrane. This protein catalyzes hydrolysis of beta-D-glucose units from the nonreducing ends of 1,3-beta-D-glucans, thereby releasing glucose. This protein was previously found to exhibit significant covariance with altitude and/or be related to drought stress responses in plants [58], including conifers *Pinus pinaster* [59] and *Abies alba* [60,61]. This protein was also indicated to be a climate-related candidate gene subjected to differentiating selection in two closely related Asian species of high- and low-altitude pines, namely, *Pinus hwangshanensis* and *P. massoniana* [62]. Although both species investigated in this study are restricted to mountain conditions, divergence observed between them and not between them and the reference Scots pine populations may indicate the importance of the water conditions in the speciation history of Pyrenean and dwarf pines. Another candidate was found among members of a large family of heavy metal-associated isoprenylated plant proteins (HIPPs). These metal ion-binding and transporting molecules (metallochaperones) play important roles in the development of vascular plants and their responses to environmental changes [63]. The molecules generally act in two ways, being involved in either heavy metal homeostasis and mechanisms of detoxification (in particular, tolerance to cadmium) or in the regulation of transcriptional responses to cold and drought. Strictosidine synthase-like 3 protein, as its name suggests, participates in biosynthesis of strictosidine, a precursor molecule in the monoterpenoid indole alkaloid biosynthesis pathway; however, according to Hicks et al. [64], this protein may catalyze hydrolytic reactions typical of other members of functionally diverse N6P superfamily [65], to which strictosidine synthase-like proteins belong; therefore, its role remains unclear. Another outlier-containing transcript identified in this study in comparisons of two mountain pines was described as potentially encoding phosphatidate phosphatase PAH2. This enzyme regulates cellular metabolic processes through catalysis of the dephosphorylation of phosphatidate, thereby yielding diacylglycerol. PAH2 represses the biosynthesis of phospholipids at the endoplasmic reticulum, and is involved in galactolipid synthesis, which is required for membrane remodeling [66]. This mechanism is considered an essential adaptation to cope with phosphate starvation [67], and although the chemical composition of the substrates on which the two studied plants grow was not specifically tested previously, it would be worthwhile to investigate whether the soil conditions or mycorrhizal symbionts do play important roles as ecophysiological factors determining the speciation of these pines [68–71]. Aminodeoxychorismate synthase (chloroplastic isoform X1) participates in carboxylic acid metabolic processes in plastids, where it catalyzes the biosynthesis of 4-amino-4-deoxychorismate and L-glutamate from chorismate and glutamine, part of a pathway for the biosynthesis of para-aminobenzoic acid, a precursor for folate production [72]. Folates constitute an essential family of cofactors, and are involved in virtually every aspect of plant physiology [73,74]. Disturbances in folate metabolism usually strongly inhibit growth of a plant; therefore, their altered function could be responsible for the dwarfism of the dwarf pine. The last of the eight most

significant outliers was found in the transcript for a protein similar to serine/threonine-protein kinase VPS15. This enzyme is located in endosomes and is required for transport from cytoplasm to vacuole and for autophagy, while autophagy has been shown to perform numerous functions, both in normal plant growth and development and in the plant response to environmental stresses [75,76].

Considering that these species exhibit substantial distinctions in growth forms, some anatomical and morphological traits and phenology, certain genes may be responsible for the observed diversity. Good candidates could potentially be among factors regulating the development and receptivity of individuals to periodic life cycle events, such as different kinds of phytohormones. Among the uncovered outliers, only three annotated sequences matched the prediction directly; one was annotated as the ethylene-responsive transcription factor ERF112, and the other two were transcription factors acting in response to gibberellins (two isoforms of scarecrow-like protein 3 (scl3)). Although the role of the first protein is less clear, the involvement of gibberellin signaling pathways is notable. Gibberellins (GA) constitute a group of plant hormones important for the regulation of various developmental processes, such as flowering, germination, senescence of fruits and leaves, dormancy or stem elongation; therefore, genes affected by GA appear to be likely candidates for controllers behind the observed shifts in phenology and phenotypes of pines. Mutants in GA biosynthesis and downstream responses have been observed in a variety of taxa, often causing dwarfism or semi-dwarfism of plants, including alpine plants [77,78]. The role played by scl3 in enhancing dwarf phenotypes of GA signaling pathway mutants was already shown in *Arabidopsis* [79]. Alpine dwarfism is a widely observed phenomenon of plant height decrease along altitudinal gradients, often regarded as an evolutionary adaptation to extreme environmental conditions, enabling plants to take advantage of higher ambient temperatures near the ground, reduce evaporation, counteract mechanical damage from wind or snow cover, and invest more resources in reproduction [78]. It was formerly reported for other shrubby alpine pine species—*Pinus pumilla* in Asia, that their mean stem height was approximately four-fold lower at the upper distribution limit (49 cm) than at the lower limit (187 cm), and the ratio of stem height to length was also lower at the upper distribution boundary [80]. Moreover, it was demonstrated that the less developed stature of scrubs at the upper distribution limit was due to shorter stem age, more creeping stems and lower shoot elongation rates. This latter finding fits well with the known role of gibberellins in regulating the elongation of shoots [81,82], among others, in pines [83], suggesting that the scl3 protein may be involved in shrubby and prostrate development of dwarf pine, which is not observed in Pyrenean mountain pines. However, it is interesting that the same two outliers were not found in comparisons between dwarf and the reference Scots pine. The above candidate genes represent good targets for further nucleotide polymorphism and gene expression studies that may be equally important in the development of adaptive divergence among these species.

In the future, information regarding the patterns of expression of genes important in speciation history, including those demonstrated here, and of epigenetic marks could help to elucidate the evolutionary trajectories of these plants. As these differences cannot be fully explained by nucleotide variation alone, these factors are usually important drivers of phenotypic and phenological variation in long-lived, sessile organisms. A better ecological description of habitats occupied by mountain species is needed because more detailed eco-physiological findings may help to elucidate the mechanistic connection between the identified diverged sequences and the selection pressures acting on the trees. Moreover, correlation studies conducted on a larger set of populations to check if the environmental factors and genetic diversity of investigated pines vary dependently would be required, with special focus on the revealed outlier sequences. Another, well-suited test for altitude-related adaptations could incorporate association studies performed on repeated pairs of genetically homogenous populations, coming from lower versus higher altitude locations.

### 3.4. Conclusions

The nuclear loci investigated in this study showed largely homogeneous patterns of variation between the taxa; nevertheless, using a novel set of SNPs, we managed to delineate the three species. Differentiation within species appeared moderate; however, some of the patterns of population structure appeared consistent between analyses, especially the differentiation found in the Pyrenean pine population. The outlier loci possibly targeted by selection and variable between two mountain species and the reference, represented two main groups of biological functions—were involved either in oxidation-reduction processes or regulation of transcription, including transcripts engaged in the gibberellin signaling pathway, known of its important function in plant development and of its connection with dwarf phenotypes, which are important adaptations observed in high-mountain species.

Considering the longevity of pines, their predominantly outcrossing mating system with considerable pollen transport abilities and effective population sizes, our results indicate that the SNP array, providing genotyping data directly comparable between different experiments, may serve as a new diagnostic tool enhancing the discrimination power of previously limited biometric and genetic features of the species that may be useful in a variety of applications in population genetics.

## 4. Materials and Methods

### 4.1. Materials

Specimens of three closely related European pine species, i.e., dwarf mountain pine (*Pinus mugo* Turra), Pyrenean pine (*P. uncinata* Ramond), and Scots pine (*P. sylvestris* L.) were investigated. Initially, cones from five mother trees, distanced from each other by a minimum of 50 m, were collected from 14 natural populations of the taxa—four or five stands per species. The formal biometric identification of the mother trees was conducted by Prof. Krystyna Boratyńska from the Institute of Dendrology, Polish Academy of Sciences, who generously shared the material (in case of this collection, no voucher specimen has been deposited in the Institutes' herbarium). Seedlings obtained from the open-pollinated seeds were grown under unified conditions of a common garden experiment at a glasshouse facility of the Centre for Ecology and Hydrology, Edinburgh, UK (for more details about the samples, see Wachowiak et al. [22]). From the total number of seedlings, 524 individuals were selected (141 dwarf, 201 Pyrenean and 182 Scots pines), and 22 to 47 samples per population were collected (Figure 1, Supplementary Table S1).

### 4.2. SNP Array Genotyping

DNA was extracted from needles using a DNeasy 96 Plant Kit (Qiagen, Hilden, Germany). Needles were dried on silica gel prior to extraction and DNA was quantified using a Qubit spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA, USA) to ensure a minimum standardized concentration of 35 ng/µL. We genotyped 49,829 SNPs that were used for the development of the Axiom_PineGAP SNP array (Thermo Fisher Scientific Inc., Waltham, MA, USA). Details about the array design and initial quality assessments performed on the called SNPs are described in Perry et al. [35]. Briefly, the array comprises polymorphisms from transcriptome sequencing [26] and candidate genes resequenced in previous population genetic studies of pine species [28,84–86]. Genotyping was performed at Edinburgh Genomics following DNA amplification, fragmentation, chip hybridization, single-base extension through DNA ligation, and signal amplification performed according to the Affymetrix Axiom Assay protocol. Genotyping was performed in 384-well format on a GeneTitan (Thermo Fisher Scientific Inc., Waltham, MA, USA) according to the manufacturer's procedure. Genotype calls were obtained using Axiom Analysis Suite v.4.0.3.3 software as recommended by the manufacturer (Thermo Fisher Scientific Inc., Waltham, MA, USA).

From the initial set of 49,829 SNP markers designed on the array and 20,795 loci that passed basic quality tests (details available in [35]), the number of markers was further reduced by removal of (1) redundant SNPs developed from the same original

sequence, (2) SNPs putatively originating from one of the organelle genomes, (3) SNPs that were monomorphic or did not pass the threshold of 1% minor allele frequency (MAF), and (4) loci with more than 15% missing data. In detail, the first step was taking one, randomly chosen, marker per contig to ensure independence between polymorphisms. Although pines are generally known for their rapid LD decay, given the limited information regarding the physical connection of the SNPs, we preferred to choose conservative approach eliminating possible effects of linkage on the analyses, even at a cost of weaker association signal. Additionally, individuals with genotypes missing at more than 15% of SNPs were also rejected. Different cut-off levels were tested (between 5% and 20%), and the chosen levels (15% per loci and 15% per individual) were considered to be the most reasonable trade-off between the quality of genotyping data and the number of excluded markers and samples. Data processing was performed using PLINK v.1.07 (http://pngu.mgh.harvard.edu/purcell/plink/, accessed on 29 August 2019; [87]), PGDSpider v.2.1.1.5 (http://www.cmpg.unibe.ch/software/PGDSpider/, accessed on 30 June 2017; [88]), and R v.3.6.1 (https://cran.r-project.org/bin/windows/base/old/3.6.1/, accessed on 22 July 2019; [89,90]).

### 4.3. Analysis of Genetic Diversity, Differentiation, and Population Grouping

We calculated several statistics of genetic diversity of the SNP dataset to assess the overall level of variation within the whole sample and on the species and population levels. Within each analyzed combination of samples, we checked (1) percentage of polymorphic sites (*%P*), (2) number of private polymorphisms (*Sp*), (3) mean within group distance (*d*), (4) observed heterozygosity (*Ho*), (5) unbiased expected heterozygosity (*uHe*) taking into account sample sizes, and (6) fixation index (*F*). Calculations of *%P*, *Sp*, *Ho*, *uHe* and *F* were conducted in GenAlEx v.6.503 (https://biology-assets.anu.edu.au/GenAlEx/Welcome.html, accessed 4 December 2016; [91]), and of *d* in MEGA v.7.0.21 (https://www.megasoftware.net/, accessed 15 March 2017; [92]).

To reconstruct the genetic structure of the analyzed group of populations, and thereby to assess the robustness of our SNPs set in terms of discrimination power for studied populations and species and to delineate more subtle genetic relationships among them, we employed a number of clustering methods. In these tests, we checked the distribution of genetic variation among different levels of the population hierarchy, i.e., among (1) the three species (herein MUS), (2) two mountain pines vs. Scots pine (MU vs. S), (3) two focal species (M vs. U), (4) dwarf vs. Scots pine (M vs. S), (5) Pyrenean vs. Scots pine (U vs. S), and (6) populations within each taxon separately afterwards (M, U, S). To investigate whether populations tend to group by species identities or other patterns, the unweighted pair group method with arithmetic mean (UPGMA) phylogenetic tree was calculated based on the mean between population distances ($d_{xy}$) in MEGA, both for three species and for the two mountain pines only. Differentiation of populations was investigated by the hierarchical analysis of molecular variance (AMOVA) using Arlequin v.3.5.2.2 (http://cmpg.unibe.ch/software/arlequin35/, accessed on 15 March 2017 with updates; [93], and *Fst* statistics between pairs of populations or pairs of species were estimated in this package. We also verified these relationships by conducting two variants of multivariate analysis [94] in the adegenet package for R v.2.1.3 (http://adegenet.r-forge.r-project.org/, accessed on 13 October 2020; [95,96]). First, principal component analysis (PCA) was employed, using Edward's Euclidean distances between samples. Next, the comparisons were repeated in principal coordinate analysis (PCoA) with analogous information on the population level, which additionally showed the distribution of genetic diversity within stands. Allele frequencies were scaled and centered prior multivariate calculations, and missing data were replaced by the mean frequencies. Additionally, to verify the hypothesis of isolation by distance (IBD) and determine if nearby populations have greater genetic similarity than those geographically separated, we employed Mantel tests [97] to determine the correlation between genetic and geographic distances among populations. The analysis was performed in

adegenet, on Edward's distances with 1000 permutations, and packages MASS v.7.3-53 (https://CRAN.R-project.org/package=MASS, accessed on 12 October 2020; [98]) and geosphere v.1.5-10 (https://cran.r-project.org/web/packages/geosphere/index.html, accessed on 12 October 2020; [99]) were used to manipulate geographical data. Finally, grouping of the samples was investigated by STRUCTURE analysis with the individual-based Bayesian clustering method implemented in [100–102]. In these trials, we tested all 14 populations with $K$ values from 1 to 10. The admixture model with correlated allele frequencies was chosen, and no prior regarding population origin was used. For each $K$, four runs were performed, each staring with 10,000 burn-ins followed by 10,000 iterations. To determine the best $K$ value and the grouping that optimally fit our data, two attempts were applied: the L($K$) method [100] and the Evanno method based on an ad hoc measure of the rate of change in the log probability of data between subsequent $K$ values—Delta$K$ [103]. Both were run with STRUCTURE HARVESTER v.0.6.94 (http://taylor0.biology.ucla.edu/structureHarvester/, accessed 26 June 2020; [104]), while the results of the analyses were plotted using STRUCTURE PLOT v.2.0 (http://omicsspeaks.com/strplot2/, accessed 11 July 2020; [105]). An analogous model-based Bayesian test was performed for representatives of the *Pinus mugo* complex only to more specifically investigate the discrimination power of the array within this aggregate. In this case, $K$ values between 1 and 9 were inspected, with each employing four runs with the same parameters as above.

### 4.4. Outlier SNP Detection and Functional Annotations

To identify putative selection targets differentiating inspected taxa, we searched for outlier SNPs among them, comparing allele frequencies in four setups: (1) MU vs. S, (2) M vs. S, (3) U vs. S, and (4) M vs. U. For the analysis, we used three methodological approaches, all based on *Fst* statistics, namely, FDist, BayeScan, and FLK (for comparison of the methods please see [106]). Tests implementing the FDist method were run in Arlequin; for comparisons, we set a hierarchical island model with populations nested in species. Parameters were set to 100,000 coalescent simulations, with 100 dems simulated per group and 10 groups; pairwise differences were used as the distance method for AMOVA calculations. The SNP data were considered multilocus with an unknown gametic phase, and the allowed missing level per site was changed to 0.1. For BayeScan calculations run within BayeScan v.2.1 (http://cmpg.unibe.ch/software/BayeScan/download.html, accessed 2 March 2020; [107]), and in the case of analysis implementing the FLK method in hapFLK v.1.4 (https://forge-dga.jouy.inra.fr/projects/hapflk, accessed 9 March 2020; [108]), we employed default settings. In the tests, we further focused only on the markers that differentiated our groups, concentrating on potential directional selection targets. We searched for outliers looking at two levels of significance, where $p$ or $q$ values, depending on the test, were set to the thresholds of 0.05 or 0.01. We were most interested in these SNPs that were detected by more than one method at the level of $p/q$ value 0.05 and/or were confirmed on the more stringent level.

Respective contigs containing outlier SNPs revealed in those comparisons were imported into OmicsBox v.1.2 (https://www.biobam.com/omicsbox/, accessed on 11 June 2020; [109]) and analyzed for Gene Ontology (GO) terms [110,111]. Although the transcripts were already annotated in 2013 at the phase of the original RNA-Seq assembly [26], we decided to repeat the analysis for this small subset of sequences, as the GO vocabularies are continuously updated. We followed the standard procedure of (1) BLAST—finding sequences similar to the queries; in our case, blastx-fast was used, (2) GO mapping—retrieving GO terms associated with the hits obtained by BLAST, (3) gene annotation—assigning selected terms from the obtained GO pool to the query sequences, and (4) InterProScan annotation—improving the annotation results by collecting domain/motif information for analyzed query sequences. Default settings were applied, except that we narrowed BLAST searches to Viridiplantae (taxon ID: 33090). As successfully annotated, we regarded only these sequences that were blasted with high confidence (besides default

filtering, we applied an additional cut-off for mean BLAST similarity above 50%), and passed two following steps, regardless of the InterPro search support. However, if InterProScan annotation was effective, the outcomes were merged with those obtained through basic gene annotation.

**Supplementary Materials:** The following are available online at https://www.mdpi.com/article/10.3390/ijms22073477/s1. Table S1: List and description of 14 source populations from which offspring genotypes were investigated in the study; Figure S1: Shared and unique SNP polymorphisms segregating in populations of dwarf, Pyrenean and Scots pines; Table S2: Distances and differentiation between investigated pine populations measured by (a) $d_{xy}$—mean number of differences between groups, (b) *Fst*—inbreeding coefficient within subpopulations; Figure S2: Evolutionary relationships of populations inferred using the UPGMA method calculated on the mean between group distances ($d_{xy}$) for (a) three studied taxa, (b) two mountain pines only; Figure S3: Principal coordinate analysis (PCoA) revealing relationships between 14 studied populations. (a) Discrimination among three pine species, (b) distinction within the *P. mugo* complex; Figure S4: Results of clustering analysis performed with STRUCTURE for 14 studied populations. (a) Delta*K* estimates and (b) mean values of natural logarithm of likelihood, as indicated by STRUCTURE HARVESTER for *K* groups between 1 and 10. Detailed results of analysis for specific *K* values: (c) *K* = 6, (d) *K* = 7, (e) *K* = 8, (f) *K* = 9, (g) *K* = 10; Figure S5: Results of clustering analysis performed with STRUCTURE for 9 populations of *P. mugo* complex. (A) Delta*K* estimates and (b) mean values of the natural logarithm of likelihood, as indicated by STRUCTURE HARVESTER for *K* groups between 1 and 9. Detailed results of analysis for specific *K* values: (c) *K* = 4, (d) *K* = 5, (e) *K* = 6, (f) *K* = 7, (g) *K* = 8 and (h) *K* = 9; Table S3: Results of Mantel tests verifying the isolation by distance (IBD) hypothesis performed on Edwards' genetic and log standardized geographic distances for different combinations of studied pine populations; Figure S6: Correlation plots of the geographic and genetic distances verifying the isolation by distance (IBD) hypothesis in three pine species. (a) Mantel test including comparisons within and between populations of all species, (b) results of three independent tests performed within taxa; Table S4: Gene ontology (GO) annotation of sequences with outlier polymorphisms detected; Table S5: Frequencies of outlier SNP alleles; Figure S7: Distribution of the most common third-level gene ontologies annotated for 64 sequences in which outlier SNPs were detected between dwarf and Pyrenean pine, shown by domains.

**Author Contributions:** J.Z. and W.W. designed the study; W.W. supervised the work; A.P. and S.C. provided genotyping results from Axiom array; J.Z. performed genotyping data quality control and analyses; B.Ł. helped with data analysis in R and STRUCTURE; J.Z. wrote the original manuscript; W.W. and B.Ł. critically reviewed the article. All authors have read and agreed to the published version of the manuscript.

**Funding:** Analyses of the obtained genotypes, interpretation of data, writing of the article and APC were funded by the Polish National Science Centre (UMO-2017/27/B/NZ9/00159). The data acquisition was supported by the Natural Environment Research Council (NE/K012177/1) and the PROTREE project (BB/L012243/1), which funded generation of the SNP array and genotyping.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The dataset supporting the results of this publication is available from the corresponding author upon request.

**Acknowledgments:** The authors would like to express their thanks to Krystyna Boratyńska from the Institute of Dendrology, Polish Academy of Sciences, for providing part of the plant material used in this work.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References


1. Petit, R.J.; Hampe, A. Some evolutionary consequences of being a tree. *Annu. Rev. Ecol. Evol. Syst.* **2006**, *37*, 187–214. [CrossRef]
2. Billings, W.D. Adaptations and Origins of Alpine Plants. *Arct. Alp. Res.* **1974**, *6*, 129–142. [CrossRef]
3. Wieser, G.; Tausz, M. *Trees at Their Upper Limit: Treelife Limitation at the Alpine Timberline*; Springer GmbH: Heidelberg, Germany, 2007.
4. Smith, W.K.; Johnson, D.M.; Reinhardt, K. Alpine Forest. In *Encyclopedia of Ecology*; Jørgensen, S.E., Fath, B.D., Eds.; Academic Press: Oxford, UK, 2008; pp. 144–153.
5. Oleksyn, J.; Modrzynski, J.; Tjoelker, M.G.; Zytkowiak, R.; Reich, P.B.; Karolewski, P. Growth and physiology of *Picea abies* populations from elevational transects: Common garden evidence for altitudinal ecotypes and cold adaptation. *Funct. Ecol.* **1998**, *12*, 573–590. [CrossRef]
6. Keller, F.; Korner, C. The role of photoperiodism in alpine plant development. *Arct. Antarct. Alp. Res.* **2003**, *35*, 361–368. [CrossRef]
7. Coomes, D.A.; Jenkins, K.L.; Cole, L.E.S. Scaling of tree vascular transport systems along gradients of nutrient supply and altitude. *Biol. Lett.* **2007**, *3*, 86–89. [CrossRef]
8. Lutz, C.; Engel, L. Changes in chloroplast ultrastructure in some high-alpine plants: Adaptation to metabolic demands and climate? *Protoplasma* **2007**, *231*, 183–192. [CrossRef] [PubMed]
9. Ohsawa, T.; Ide, Y. Global patterns of genetic variation in plant species along vertical and horizontal gradients on mountains. *Glob. Ecol. Biogeogr.* **2008**, *17*, 152–163. [CrossRef]
10. Gonzalo-Turpin, H.; Hazard, L. Local adaptation occurs along altitudinal gradient despite the existence of gene flow in the alpine plant species *Festuca eskia*. *J. Ecol.* **2009**, *97*, 742–751. [CrossRef]
11. Stocklin, J.; Kuss, P.; Pluess, A.R. Genetic diversity, phenotypic variation and local adaptation in the alpine landscape: Case studies with alpine plant species. *Bot. Helv.* **2009**, *119*, 125–133. [CrossRef]
12. Farjon, A. The Kew Review: Conifers of the World. *Kew Bull.* **2018**, *73*, 8. [CrossRef]
13. Critchfield, W.B.; Little, E.L. *Geographic Distribution of the Pines of the World*; Department of Agriculture, Forest Service: Washington, DC, USA, 1966.
14. Alexandrov, A.H.; von Wühlisch, G.; Vendramin, G.G. *EUFORGEN Technical Guidelines for Genetic Conservation and Use of Mountain Pine (Pinus mugo)*; European Forest Genetic Resources Programme (EUFORGEN), European Forest Institute: Bonn, Germany, 2019.
15. Monteleone, I.; Ferrazzini, D.; Belletti, P. Effectiveness of neutral RAPD markers to detect genetic divergence between the subspecies *uncinata* and *mugo* of *Pinus mugo* Turra. *Silva Fenn.* **2006**, *40*, 391–406. [CrossRef]
16. Heuertz, M.; Teufel, J.; Gonzalez-Martinez, S.C.; Soto, A.; Fady, B.; Alia, R.; Vendramin, G.G. Geography determines genetic relationships between species of mountain pine (*Pinus mugo* complex) in western Europe. *J. Biogeogr.* **2010**, *37*, 541–556. [CrossRef]
17. Zaborowska, J.; Labiszak, B.; Wachowiak, W. Population history of European mountain pines *Pinus mugo* and *Pinus uncinata* revealed by mitochondrial DNA markers. *J. Syst. Evol.* **2020**, *58*, 474–486. [CrossRef]
18. Boratynska, K.; Jasinska, A.K.; Boratynski, A. Taxonomic and geographic differentiation of *Pinus mugo* complex on the needle characteristics. *Syst. Biodivers.* **2015**, *13*, 901–915. [CrossRef]
19. Jalas, J.; Suominen, J. *Atlas Florae Europaeae: Gymnospermae (Pinaceae to Ephedraceae)*; Committee for Mapping the Flora of Europe & Suomen Biologian Seura Vanamo: Helsinki, Finland, 1972.
20. Camarero, J.J.; Gutierrez, E.; Fortin, M.J.; Ribbens, E. Spatial patterns of tree recruitment in a relict population of *Pinus uncinata*: Forest expansion through stratified diffusion. *J. Biogeogr.* **2005**, *32*, 1979–1992. [CrossRef]
21. Boratynska, K.; Marcysiak, K.; Boratynski, A. *Pinus mugo* (Pinaceae) in the Abruzzi Mountains: High morphological variation in isolated populations. *Bot. J. Linn. Soc.* **2005**, *147*, 309–316. [CrossRef]
22. Wachowiak, W.; Perry, A.; Donnelly, K.; Cavers, S. Early phenology and growth trait variation in closely related European pine species. *Ecol. Evol.* **2018**, *8*, 655–666. [CrossRef] [PubMed]
23. Bogunic, F.; Siljak-Yakovlev, S.; Muratovic, E.; Pustahija, F.; Medjedovic, S. Molecular cytogenetics and flow cytometry reveal conserved genome organization in *Pinus mugo* and *P. uncinata*. *Ann. For. Sci.* **2011**, *68*, 179–187. [CrossRef]
24. Lewandowski, A.; Boratynski, A.; Mejnartowicz, L. Allozyme investigations on the genetic differentiation between closely related pines—*Pinus sylvestris*, *P. mugo*, *P. uncinata*, and *P. uliginosa* (Pinaceae). *Plant Syst. Evol.* **2000**, *221*, 15–24. [CrossRef]
25. Wachowiak, W.; Boratynska, K.; Cavers, S. Geographical patterns of nucleotide diversity and population differentiation in three closely related European pine species in the *Pinus mugo* complex. *Bot. J. Linn. Soc.* **2013**, *172*, 225–238. [CrossRef]
26. Wachowiak, W.; Trivedi, U.; Perry, A.; Cavers, S. Comparative transcriptomics of a complex of four European pine species. *BMC Genom.* **2015**, *16*, 234. [CrossRef]
27. Nystedt, B.; Street, N.R.; Wetterbom, A.; Zuccolo, A.; Lin, Y.C.; Scofield, D.G.; Vezzi, F.; Delhomme, N.; Giacomello, S.; Alexeyenko, A.; et al. The Norway spruce genome sequence and conifer genome evolution. *Nature* **2013**, *497*, 579–584. [CrossRef]
28. Neale, D.B.; Wegrzyn, J.L.; Stevens, K.A.; Zimin, A.V.; Puiu, D.; Crepeau, M.W.; Cardeno, C.; Koriabine, M.; Holtz-Morris, A.E.; Liechty, J.D.; et al. Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome Biol.* **2014**, *15*, R59. [CrossRef]
29. Zimin, A.V.; Stevens, K.A.; Crepeau, M.W.; Puiu, D.; Wegrzyn, J.L.; Yorke, J.A.; Langley, C.H.; Neale, D.B.; Salzberg, S.L. An improved assembly of the loblolly pine mega-genome using long-read single-molecule sequencing. *Gigascience* **2017**, *6*, 1–4. [CrossRef]

30. Dzialuk, A.; Boratynska, K.; Romo, A.; Boratynski, A. Taxonomic and geographic variation of the *Pinus mugo* complex on chloroplast microsatellite markers. *Syst. Biodivers.* **2017**, *15*, 464–479. [CrossRef]
31. Wachowiak, W.; Zaborowska, J.; Labiszak, B.; Perry, A.; Zucca, G.M.; Gonzalez-Martinez, S.C.; Cavers, S. Molecular signatures of divergence and selection in closely related pine taxa. *Tree Genet. Genomes* **2018**, *14*, 83. [CrossRef]
32. Martinez, I.; Gonzalez-Taboada, F.; Wiegand, T.; Camarero, J.J.; Gutierrez, E. Dispersal limitation and spatial scale affect model based projections of *Pinus uncinata* response to climate change in the Pyrenees. *Glob. Chang. Biol.* **2012**, *18*, 1714–1724. [CrossRef]
33. Grotkopp, E.; Rejmanek, M.; Sanderson, M.J.; Rost, T.L. Evolution of genome size in pines (*Pinus*) and its life-history correlates: Supertree analyses. *Evolution* **2004**, *58*, 1705–1729. [CrossRef] [PubMed]
34. Wachowiak, W.; Palme, A.E.; Savolainen, O. Speciation history of three closely related pines *Pinus mugo* (T.), *P. uliginosa* (N.) and *P. sylvestris* (L.). *Mol. Ecol.* **2011**, *20*, 1729–1743. [CrossRef] [PubMed]
35. Perry, A.; Wachowiak, W.; Downing, A.; Talbot, R.; Cavers, S. Development of a SNP array for population genomic studies in four European pine species. *Mol. Ecol. Resour.* **2020**, *20*, 1697–1705. [CrossRef]
36. Wikimedia Commons, the Free Media Repository. Available online: https://commons.wikimedia.org (accessed on 22 September 2009).
37. Distribution map of Scots pine (Pinus sylvestris) EUFORGEN 2009, European Forest Genetic Resources Programme. Available online: http://euforgen.org (accessed on 4 October 2018).
38. Hamernik, J.; Musil, I. *Pinus mugo* complex—Its structuring and general overview of the used nomenclature. *J. For. Sci.* **2007**, *53*, 253–266. [CrossRef]
39. Dzialuk, A.; Muchewicz, E.; Boratynski, A.; Montserrat, J.M.; Boratynska, K.; Burczyk, J. Genetic variation of *Pinus uncinata* (Pinaceae) in the Pyrenees determined with cpSSR markers. *Plant Syst. Evol.* **2009**, *277*, 197–205. [CrossRef]
40. Dainou, K.; Blanc-Jolivet, C.; Degen, B.; Kimani, P.; Ndiade-Bourobou, D.; Donkpegan, A.S.; Tosso, F.; Kaymak, E.; Bourland, N.; Doucet, J.L.; et al. Revealing hidden species diversity in closely related species using nuclear SNPs, SSRs and DNA sequences—A case study in the tree genus *Milicia*. *BMC Evol. Biol.* **2016**, *16*, 259. [CrossRef]
41. Nicotra, A.B.; Chong, C.; Bragg, J.G.; Ong, C.R.; Aitken, N.C.; Chuah, A.; Lepschi, B.; Borevitz, J.O. Population and phylogenomic decomposition via genotyping-by-sequencing in Australian *Pelargonium*. *Mol. Ecol.* **2016**, *25*, 2000–2014. [CrossRef]
42. Lee, S.R.; Gaskin, J.F.; Kim, Y.D. Molecular diagnosis for a *Tamarix* species from two reclaimed lands along the Yellow Sea in Korea inferred from genome wide SNP markers. *J. Syst. Evol.* **2019**, *57*, 247–255. [CrossRef]
43. Zukowska, W.B.; Boratynska, K.; Wachowiak, W. Comparison of range-wide chloroplast microsatellite and needle trait variation patterns in *Pinus mugo* Turra (dwarf mountain pine). *iForest* **2017**, *10*, 250–258. [CrossRef]
44. Zukowska, W.B.; Wachowiak, W. Nuclear microsatellite markers reveal the low genetic structure of *Pinus mugo* Turra (dwarf mountain pine) populations in Europe. *Plant Syst. Evol.* **2017**, *303*, 641–651. [CrossRef]
45. Robledo-Arnuncio, J.J. Wind pollination over mesoscale distances: An investigation with Scots pine. *New Phytol.* **2011**, *190*, 222–233. [CrossRef]
46. Soto, A.; Robledo-Arnuncio, J.J.; Gonzalez-Martinez, S.C.; Smouse, P.E.; Alia, R. Climatic niche and neutral genetic diversity of the six Iberian pine species: A retrospective and prospective view. *Mol. Ecol.* **2010**, *19*, 1396–1409. [CrossRef]
47. Polle, A.; Rennenberg, H. Field Studies on Norway Spruce Trees at High-Altitudes. 2. Defense Systems against Oxidative Stress in Needles. *New Phytol.* **1992**, *121*, 635–642. [CrossRef]
48. Streb, P.; Shang, W.; Feierabend, J.; Bligny, R. Divergent strategies of photoprotection in high-mountain plants. *Planta* **1998**, *207*, 313–324. [CrossRef]
49. Polle, A.; Baumbusch, L.O.; Oschinski, C.; Eiblmeier, M.; Kuhlenkamp, V.; Vollrath, B.; Scholz, F.; Rennenberg, H. Growth and protection against oxidative stress in young clones and mature spruce trees (*Picea abies* L.) at high altitudes. *Oecologia* **1999**, *121*, 149–156. [CrossRef]
50. Oncel, I.; Yurdakulol, E.; Keles, Y.; Kurt, L.; Yildiz, A. Role of antioxidant defense system and biochemical adaptation on stress tolerance of high mountain and steppe plants. *Acta Oecol.* **2004**, *26*, 211–218. [CrossRef]
51. Keles, Y.; Everest, A. Relation to altitude adaptation and antioxidant defence system in five shrubs and trees species from middle Taurus Mountains. *Int. J. Nat. Eng. Sci.* **2008**, *2*, 45–49.
52. Huang, W.; Yang, Y.J.; Hu, H.; Zhang, S.B. Seasonal variations in photosystem I compared with photosystem II of three alpine evergreen broad-leaf tree species. *J. Photochem. Photobiol. B* **2016**, *165*, 71–79. [CrossRef]
53. Wang, H.; Prentice, I.C.; Davis, T.W.; Keenan, T.F.; Wright, I.J.; Peng, C.H. Photosynthetic responses to altitude: An explanation based on optimality principles. *New Phytol.* **2017**, *213*, 976–982. [CrossRef] [PubMed]
54. Gibert, J.M.; Mouchel-Vielh, E.; De Castro, S.; Peronnet, F. Phenotypic Plasticity through Transcriptional Regulation of the Evolutionary Hotspot Gene tan in *Drosophila melanogaster*. *PLoS Genet.* **2016**, *12*, e1006218. [CrossRef] [PubMed]
55. Ecker, S.; Pancaldi, V.; Valencia, A.; Beck, S.; Paul, D.S. Epigenetic and Transcriptional Variability Shape Phenotypic Plasticity. *Bioessays* **2018**, *40*, 1700148. [CrossRef]
56. Hirst, J.; Bright, N.A.; Rous, B.; Robinson, M.S. Characterization of a fourth adaptor-related protein complex. *Mol. Biol. Cell* **1999**, *10*, 2787–2802. [CrossRef] [PubMed]
57. Nakatsu, F.; Ohno, H. Adaptor protein complexes as the key regulators of protein sorting in the post-Golgi network. *Cell Struct. Funct.* **2003**, *28*, 419–429. [CrossRef] [PubMed]

58. Ma, L.; Sun, X.D.; Kong, X.X.; Galvan, J.V.; Li, X.; Yang, S.H.; Yang, Y.Q.; Yang, Y.P.; Hu, X.Y. Physiological, biochemical and proteomics analysis reveals the adaptation strategies of the alpine plant *Potentilla saundersiana* at altitude gradient of the Northwestern Tibetan Plateau. *J. Proteom.* **2015**, *112*, 63–82. [CrossRef]
59. Eveno, E.; Collada, C.; Guevara, M.A.; Leger, V.; Soto, A.; Diaz, L.; Leger, P.; Gonzalez-Martinez, S.C.; Cervera, M.T.; Plomion, C.; et al. Contrasting patterns of selection at *Pinus pinaster* Ait. drought stress candidate genes as revealed by genetic differentiation analyses. *Mol. Biol. Evol.* **2008**, *25*, 417–437. [CrossRef]
60. Behringer, D.; Zimmermann, H.; Ziegenhagen, B.; Liepelt, S. Differential Gene Expression Reveals Candidate Genes for Drought Stress Response in *Abies alba* (Pinaceae). *PLoS ONE* **2015**, *10*, e0124564. [CrossRef]
61. Heer, K.; Behringer, D.; Piermattei, A.; Bassler, C.; Brandl, R.; Fady, B.; Jehl, H.; Liepelt, S.; Lorch, S.; Piotti, A.; et al. Linking dendroecology and association genetics in natural populations: Stress responses archived in tree rings associate with SNP genotypes in silver fir (*Abies alba* Mill.). *Mol. Ecol.* **2018**, *27*, 1428–1438. [CrossRef]
62. Zhou, Y.F.; Zhang, L.R.; Liu, J.Q.; Wu, G.L.; Savolainen, O. Climatic adaptation and ecological divergence between two closely related pine species in Southeast China. *Mol. Ecol.* **2014**, *23*, 3504–3522. [CrossRef]
63. De Abreu-Neto, J.B.; Turchetto-Zolet, A.C.; de Oliveira, L.F.V.; Zanettini, M.H.B.; Margis-Pinheiro, M. Heavy metal-associated isoprenylated plant protein (HIPP): Characterization of a family of proteins exclusive to plants. *FEBS J.* **2013**, *280*, 1604–1616. [CrossRef] [PubMed]
64. Hicks, M.A.; Barber, A.E.; Giddings, L.A.; Caldwell, J.; O'Connor, S.E.; Babbitt, P.C. The evolution of function in strictosidine synthase-like proteins. *Proteins* **2011**, *79*, 3082–3098. [CrossRef] [PubMed]
65. Hicks, M.A.; Barber, A.E.; Babbitt, P.C. The Nucleophilic Attack Six-Bladed β-Propeller (N6P) Superfamily. In *Protein Families: Relating Protein Sequence, Structure, and Function*; Orengo, C., Bateman, A., Eds.; John Wiley & Sons, Inc.: Chichester, UK, 2013; pp. 125–158.
66. Nakamura, Y.; Koizumi, R.; Shui, G.H.; Shimojima, M.; Wenk, M.R.; Ito, T.; Ohta, H. *Arabidopsis* lipins mediate eukaryotic pathway of lipid metabolism and cope critically with phosphate starvation. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 20978–20983. [CrossRef]
67. Nakamura, Y. Phosphate starvation and membrane lipid remodeling in seed plants. *Prog. Lipid Res.* **2013**, *52*, 43–50. [CrossRef]
68. Cumming, J.R. Phosphate-limitation physiology in ectomycorrhizal pitch pine (*Pinus rigida*) seedlings. *Tree Physiol.* **1996**, *16*, 977–983. [CrossRef] [PubMed]
69. Gerlitz, T.G.M.; Gerlitz, A. Phosphate uptake and polyphosphate metabolism of mycorrhizal and nonmycorrhizal roots of pine and of *Suillus bovinus* at varying external pH measured by in vivo P-31-NMR. *Mycorrhiza* **1997**, *7*, 101–106. [CrossRef]
70. Anacker, B.L.; Strauss, S.Y. The geography and ecology of plant speciation: Range overlap and niche divergence in sister species. *Proc. R. Soc. B Biol. Sci.* **2014**, *281*, 20132980. [CrossRef] [PubMed]
71. Cudlin, P.; Klopcic, M.; Tognetti, R.; Malis, F.; Alados, C.L.; Bebi, P.; Grunewald, K.; Zhiyanski, M.; Andonowski, V.; La Porta, N.; et al. Drivers of treeline shift in different European mountains. *Clim. Res.* **2017**, *73*, 135–150. [CrossRef]
72. Basset, G.J.C.; Ravanel, S.; Quinlivan, E.P.; White, R.; Giovannoni, J.J.; Rebeille, F.; Nichols, B.P.; Shinozaki, K.; Seki, M.; Gregory, J.F.; et al. Folate synthesis in plants: The last step of the p-aminobenzoate branch is catalyzed by a plastidial aminodeoxychorismate lyase. *Plant J.* **2004**, *40*, 453–461. [CrossRef]
73. Hanson, A.D.; Gregory, J.F. Synthesis and turnover of folates in plants. *Curr. Opin. Plant Biol.* **2002**, *5*, 244–249. [CrossRef]
74. Gorelova, V.; Ambach, L.; Rebeille, F.; Stove, C.; Van Der Straeten, D. Folates in Plants: Research Advances and Progress in Crop Biofortification. *Front. Chem.* **2017**, *5*, 21. [CrossRef]
75. Xu, N.; Gao, X.Q.; Zhao, X.Y.; Zhu, D.Z.; Zhou, L.Z.; Zhang, X.S. *Arabidopsis* AtVPS15 is essential for pollen development and germination through modulating phosphatidylinositol 3-phosphate formation. *Plant Mol. Biol.* **2011**, *77*, 251–260. [CrossRef]
76. Zhou, X.M.; Zhao, P.; Wang, W.; Zou, J.; Cheng, T.H.; Peng, X.B.; Sun, M.X. A comprehensive, genome-wide analysis of autophagy-related genes identified in tobacco suggests a central role of autophagy in plant response to various environmental cues. *DNA Res.* **2015**, *22*, 245–257. [CrossRef]
77. Schomburg, F.M.; Bizzell, C.M.; Lee, D.J.; Zeevaart, J.A.D.; Amasino, R.M. Overexpression of a novel class of gibberellin 2-oxidases decreases gibberellin levels and creates dwarf plants. *Plant Cell* **2003**, *15*, 151–163. [CrossRef]
78. Luo, Y.H.; Dong, X.W.; Yu, T.Y.; Shi, X.; Li, Z.Y.; Yang, W.C.; Widmer, A.; Karrenberg, S. A Single Nucleotide Deletion in Gibberellin20-oxidase1 Causes Alpine Dwarfism in *Arabidopsis*. *Plant Physiol.* **2015**, *168*, 930–937. [CrossRef] [PubMed]
79. Zhang, Z.L.; Ogawa, M.; Fleet, C.M.; Zentella, R.; Hu, J.H.; Heo, J.O.; Lim, J.; Kamiya, Y.; Yamaguchi, S.; Sun, T.P. SCARECROW-LIKE 3 promotes gibberellin signaling by antagonizing master growth repressor DELLA in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 2160–2165. [CrossRef]
80. Takahashi, K.; Yoshida, S. How the scrub height of dwarf pine *Pinus pumila* decreases at the treeline. *Ecol. Res.* **2009**, *24*, 847–854. [CrossRef]
81. Phinney, B.O. Gibberellin A1 dwarfism and shoot elongation in higher plants. *Biol. Plant.* **1985**, *27*, 172–179. [CrossRef]
82. Junttila, O. Gibberellins and the Regulation of Shoot Elongation in Woody Plants. In *Gibberellins*; Springer: New York, NY, USA, 1991; pp. 199–210.
83. Little, C.H.A.; MacDonald, J.E. Effects of exogenous gibberellin and auxin on shoot elongation and vegetative bud development in seedlings of *Pinus sylvestris* and *Picea glauca*. *Tree Physiol.* **2003**, *23*, 73–83. [CrossRef]
84. Wachowiak, W.; Balk, P.A.; Savolainen, O. Search for nucleotide diversity patterns of local adaptation in dehydrins and other cold-related candidate genes in Scots pine (*Pinus sylvestris* L.). *Tree Genet. Genomes* **2009**, *5*, 117–132. [CrossRef]

85. Kujala, S.T.; Savolainen, O. Sequence variation patterns along a latitudinal cline in Scots pine (*Pinus sylvestris*): Signs of clinal adaptation? *Tree Genet. Genomes* **2012**, *8*, 1451–1467. [CrossRef]
86. Mosca, E.; Eckert, A.J.; Di Pierro, E.A.; Rocchini, D.; La Porta, N.; Belletti, P.; Neale, D.B. The geographical and environmental determinants of genetic diversity for four alpine conifers of the European Alps. *Mol. Ecol.* **2012**, *21*, 5530–5545. [CrossRef]
87. Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.R.; Bender, D.; Maller, J.; Sklar, P.; de Bakker, P.I.W.; Daly, M.J.; et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **2007**, *81*, 559–575. Available online: http://pngu.mgh.harvard.edu/purcell/plink/ (accessed on 29 August 2019). [CrossRef]
88. Lischer, H.E.L.; Excoffier, L. PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* **2012**, *28*, 298–299. Available online: http://www.cmpg.unibe.ch/software/PGDSpider/ (accessed on 30 June 2017). [CrossRef]
89. RStudio Team. RStudio: Integrated Development for R. 2018. Available online: http://www.rstudio.com/ (accessed on 22 July 2019).
90. R Core Team. R: A Language and Environment for Statistical Computing. 2019. Available online: http://www.R-project.org/ (accessed on 22 July 2019).
91. Peakall, R.; Smouse, P.E. GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* **2012**, *28*, 2537–2539. Available online: https://biology-assets.anu.edu.au/GenAlEx/Welcome.html (accessed on 4 December 2016). [CrossRef]
92. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. Available online: https://www.megasoftware.net/ (accessed on 15 March 2017). [CrossRef]
93. Excoffier, L.; Lischer, H.E.L. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **2010**, *10*, 564–567. Available online: http://cmpg.unibe.ch/software/arlequin35/ (accessed on 15 March 2017). [CrossRef] [PubMed]
94. Jombart, T.; Pontier, D.; Dufour, A.B. Genetic markers in the playground of multivariate analysis. *Heredity* **2009**, *102*, 330–341. [CrossRef] [PubMed]
95. Jombart, T. adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* **2008**, *24*, 1403–1405. [CrossRef] [PubMed]
96. Jombart, T.; Ahmed, I. adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics* **2011**, *27*, 3070–3071. Available online: http://adegenet.r-forge.r-project.org/ (accessed on 13 October 2020). [CrossRef] [PubMed]
97. Mantel, N. The detection of disease clustering and a generalized regression approach. *Cancer Res.* **1967**, *27*, 209–220.
98. Hijimans, R.; Williams, E.; Vennes, C. Geosphere. 2019. Available online: https://cran.r-project.org/package=geosphere (accessed on 12 October 2020).
99. Ripley, B.; Venables, B.; Bates, D.M.; Hornik, K.; Gebhardt, A.; Firth, D. MASS. 2020. Available online: https://CRAN.R-project.org/package=MASS (accessed on 12 October 2020).
100. Pritchard, J.K.; Stephens, M.; Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **2000**, *155*, 945–959. [PubMed]
101. Falush, D.; Stephens, M.; Pritchard, J.K. Inference of population structure using multilocus genotype data: Dominant markers and null alleles. *Mol. Ecol. Notes* **2007**, *7*, 574–578. [CrossRef] [PubMed]
102. Hubisz, M.J.; Falush, D.; Stephens, M.; Pritchard, J.K. Inferring weak population structure with the assistance of sample group information. *Mol. Ecol. Resour.* **2009**, *9*, 1322–1332. [CrossRef]
103. Evanno, G.; Regnaut, S.; Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* **2005**, *14*, 2611–2620. [CrossRef]
104. Earl, D.A.; Vonholdt, B.M. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* **2012**, *4*, 359–361. Available online: http://taylor0.biology.ucla.edu/structureHarvester/ (accessed on 26 June 2020). [CrossRef]
105. Ramasamy, R.K.; Ramasamy, S.; Bindroo, B.B.; Naik, V.G. STRUCTURE PLOT: A program for drawing elegant STRUCTURE bar plots in user friendly interface. *Springerplus* **2014**, *3*, 431. Available online: http://omicsspeaks.com/strplot2/ (accessed on 11 July 2020). [CrossRef] [PubMed]
106. Weigand, H.; Leese, F. Detecting signatures of positive selection in non-model species using genomic data. *Zool. J. Linn. Soc. Lond.* **2018**, *184*, 528–583. [CrossRef]
107. Foll, M.; Gaggiotti, O. A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. *Genetics* **2008**, *180*, 977–993. [CrossRef]
108. Fariello, M.I.; Boitard, S.; Naya, H.; SanCristobal, M.; Servin, B. Detecting Signatures of Selection Through Haplotype Differentiation Among Hierarchically Structured Populations. *Genetics* **2013**, *193*, 929–941. [CrossRef] [PubMed]
109. BioBam Bioinformatics SL. OmicsBox—Bioinformatics Made Easy. Available online: https://www.biobam.com/omicsbox (accessed on 11 June 2020).
110. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene Ontology: Tool for the unification of biology. *Nat. Genet.* **2000**, *25*, 25–29. [CrossRef] [PubMed]
111. Carbon, S.; Douglass, E.; Dunn, N.; Good, B.; Harris, N.L.; Lewis, S.E.; Mungall, C.J.; Basu, S.; Chisholm, R.L.; Dodson, R.J.; et al. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **2019**, *47*, D330–D338. [CrossRef]

**Table S1.** List and descriptions of the 14 source populations from which offspring genotypes were investigated in the study.

| Tax. | Acr. | Geographic localization (country code) | Latitude | Longitude | Alt. | N |
|---|---|---|---|---|---|---|
| *P. mugo* | M8 | Dinaric Alps, Durmitor Mts., Meded (ME) | 43°09'33" N | 19°05'27" E | 2100 | 23 |
| | M12 | Northern Limestone Alps, Karwendel Mts., Scharnitz (AT) | 47°22'42" N | 11°17'45" E | 1400 | 42 |
| | M14 | Carnic Alps, Nassfeld Pass, Pontebba (IT) | 46°32'45" N | 13°15'35" E | 1530 | 37 |
| | M16 | Central Apennines, Majella massif, Barrea (IT) | 41°46'20" N | 13°58'30" E | 2200 | 39 |
| *P. uncinata* | U17 | Eastern Pyrenees, Vall de Ransol (AD) | 42°35'02" N | 01°38'21" E | 2025 | 26 |
| | U18 | Eastern Pyrenees, Engolasters (AD) | 42°31'28" N | 01°34'12" E | 2000 | 42 |
| | U23 | Western Pyrenees, La Trapa near Jaca (ES) | 42°41'19" N | 00°32'12" W | 1720 | 47 |
| | U24 | Sierra de Gudar, Valdelinares (ES) | 40°28'49" N | 00°41'51" W | 2000 | 40 |
| | U28 | Massif Central, Col de la Croix-Morand (FR) | 45°36'00" N | 02°50'59" E | 1400 | 46 |
| *P. sylvestris* | S30 | Scottish Highlands, Wester Ross, Shieldaig (GB) | 57°30'35" N | 05°38'24" W | 81 | 22 |
| | S31 | Eastern Scottland, Glen Tanar (GB) | 57°02'60" N | 02°51'36" W | 160 | 38 |
| | S37 | Sierra Nevada, Trevenque (ES) | 37°05'47" N | 03°32'51" W | 1170 | 45 |
| | S39 | Finnish Lakeland, Punkaharju (FI) | 61°45'33" N | 29°23'21" E | 80 | 34 |
| | S43 | Polish Plains, Jarocin (PL) | 51°58'20" N | 17°28'40" E | 120 | 43 |

Tax. – taxon name; Acr. – acronym for the population; Alt. - altitude given in meters above sea level; N – number of sampled specimens.

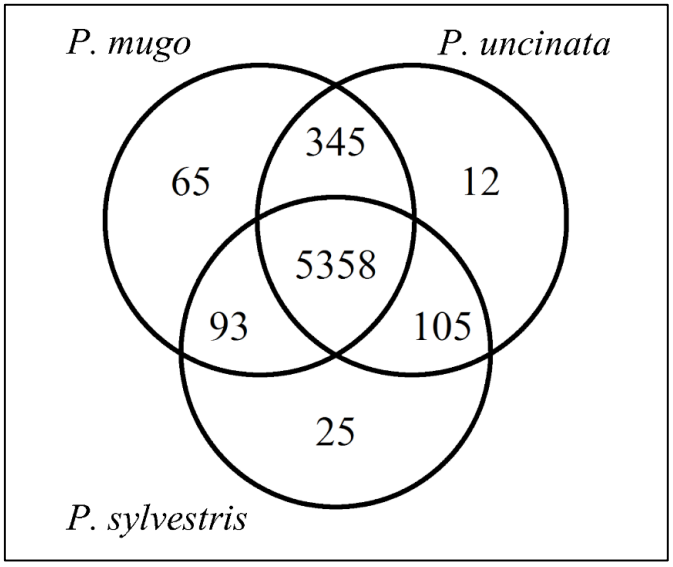



**Figure S1.** Shared and unique SNP polymorphisms segregating in populations of dwarf, Pyrenean and Scots pines.

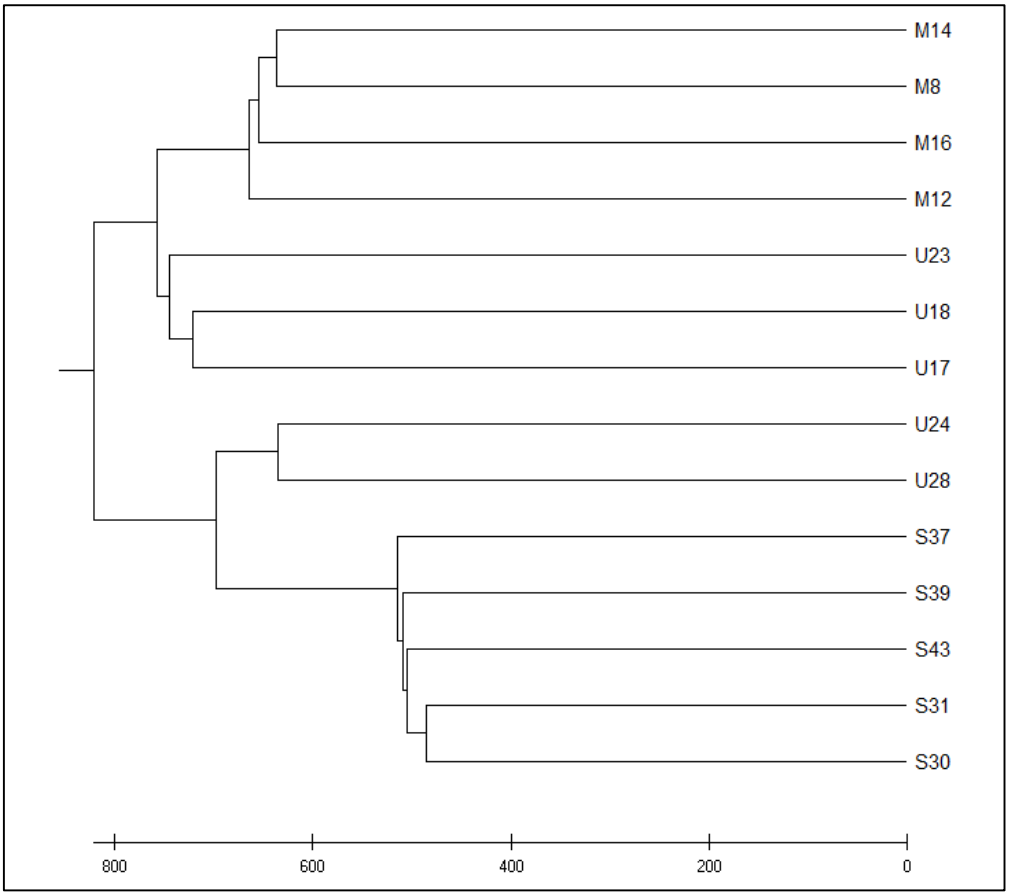
**Table S2.** Distances and differentiation between investigated pine populations measured by (**a**) $d_{xy}$ - mean number of differences between groups, (**b**) *Fst* - inbreeding coefficient within subpopulations.

(**a**)

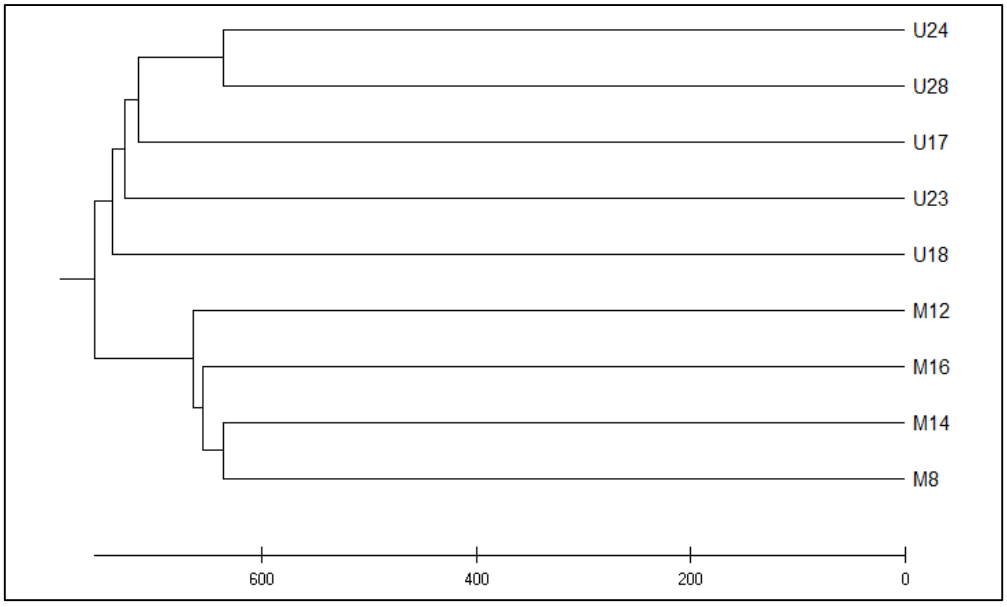
| | **M8** | **M12** | **M14** | **M16** | **U17** | **U18** | **U23** | **U24** | **U28** | **S30** | **S31** | **S37** | **S39** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **M12** | 1313.3 | | | | | | | | | | | | |
| **M14** | 1272.2 | 1311.5 | | | | | | | | | | | |
| **M16** | 1307.6 | 1357.8 | 1312.1 | | | | | | | | | | |
| **U17** | 1454.1 | 1481.2 | 1484.0 | 1488.3 | | | | | | | | | |
| **U18** | 1495.4 | 1522.7 | 1524.7 | 1528.5 | 1443.5 | | | | | | | | |
| **U23** | 1525.1 | 1554.7 | 1565.6 | 1562.5 | 1465.8 | 1512.1 | | | | | | | |
| **U24** | 1474.9 | 1518.3 | 1541.5 | 1536.5 | 1440.7 | 1488.6 | 1459.1 | | | | | | |
| **U28** | 1442.2 | 1491.1 | 1519.7 | 1514.1 | 1418.4 | 1468.1 | 1438.0 | 1270.9 | | | | | |
| **S30** | 1681.9 | 1730.0 | 1797.8 | 1775.1 | 1635.3 | 1697.4 | 1610.4 | 1436.1 | 1343.7 | | | | |
| **S31** | 1671.9 | 1722.1 | 1788.4 | 1763.0 | 1628.5 | 1694.8 | 1607.6 | 1432.8 | 1338.3 | 972.1 | | | |
| **S37** | 1697.8 | 1745.7 | 1817.5 | 1790.5 | 1637.3 | 1695.9 | 1605.8 | 1436.3 | 1342.6 | 1006.4 | 1015.4 | | |
| **S39** | 1672.8 | 1723.6 | 1788.7 | 1765.8 | 1642.0 | 1702.7 | 1624.2 | 1451.0 | 1358.6 | 1017.1 | 1013.6 | 1055.1 | |
| **S43** | 1687.1 | 1737.7 | 1803.3 | 1779.9 | 1649.1 | 1710.5 | 1626.3 | 1453.2 | 1360.1 | 1009.8 | 1010.4 | 1041.6 | 1023.1 |

(**b**)

| | **M8** | **M12** | **M14** | **M16** | **U17** | **U18** | **U23** | **U24** | **U28** | **S30** | **S31** | **S37** | **S39** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **M12** | 0.015 | | | | | | | | | | | | |
| **M14** | 0.039 | 0.042 | | | | | | | | | | | |
| **M16** | 0.029 | 0.042 | 0.058 | | | | | | | | | | |
| **U17** | 0.096 | 0.093 | 0.140 | 0.114 | | | | | | | | | |
| **U18** | 0.093 | 0.093 | 0.136 | 0.113 | 0.028 | | | | | | | | |
| **U23** | 0.127 | 0.127 | 0.173 | 0.147 | 0.060 | 0.065 | | | | | | | |
| **U24** | 0.149 | 0.153 | 0.208 | 0.177 | 0.089 | 0.093 | 0.090 | | | | | | |
| **U28** | 0.141 | 0.147 | 0.205 | 0.174 | 0.085 | 0.089 | 0.085 | 0.017 | | | | | |
| **S30** | 0.382 | 0.358 | 0.422 | 0.381 | 0.305 | 0.295 | 0.270 | 0.240 | 0.196 | | | | |
| **S31** | 0.370 | 0.354 | 0.415 | 0.377 | 0.302 | 0.295 | 0.269 | 0.233 | 0.186 | 0.048 | | | |
| **S37** | 0.413 | 0.391 | 0.451 | 0.412 | 0.335 | 0.322 | 0.294 | 0.263 | 0.218 | 0.121 | 0.109 | | |
| **S39** | 0.348 | 0.335 | 0.396 | 0.359 | 0.287 | 0.280 | 0.258 | 0.222 | 0.178 | 0.066 | 0.044 | 0.119 | |
| **S43** | 0.355 | 0.344 | 0.403 | 0.368 | 0.294 | 0.288 | 0.263 | 0.226 | 0.180 | 0.058 | 0.041 | 0.107 | 0.029 |

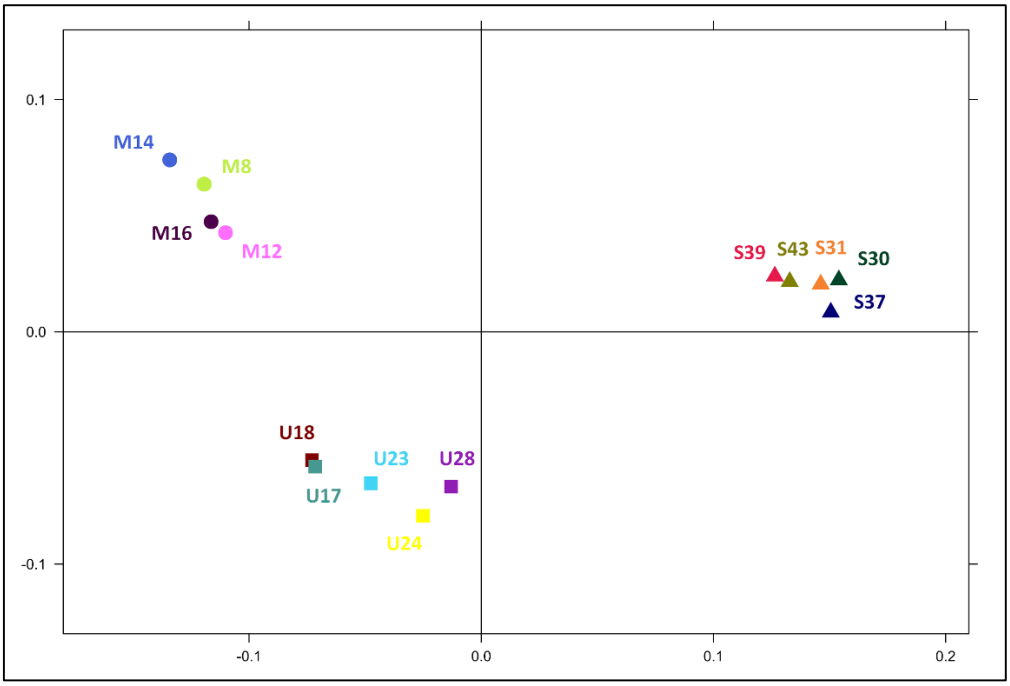



(**a**)

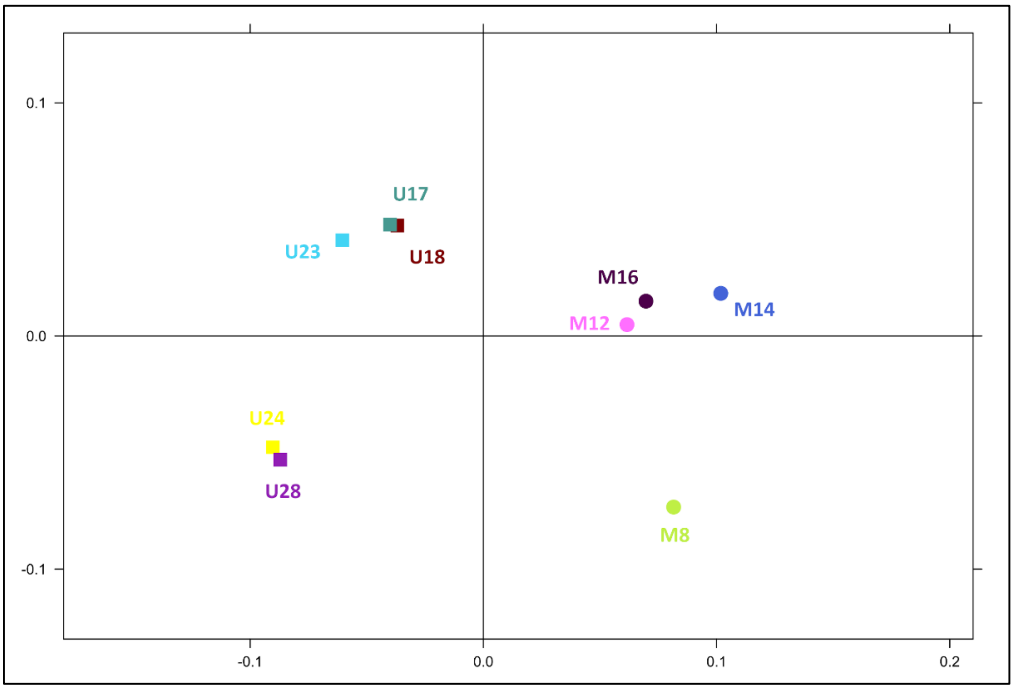



(**b**)

**Figure S2.** Evolutionary relationships of populations inferred using the UPGMA method calculated on the mean between group distances ($d_{xy}$) for (**a**) three studied taxa, (**b**) two mountain pines only.

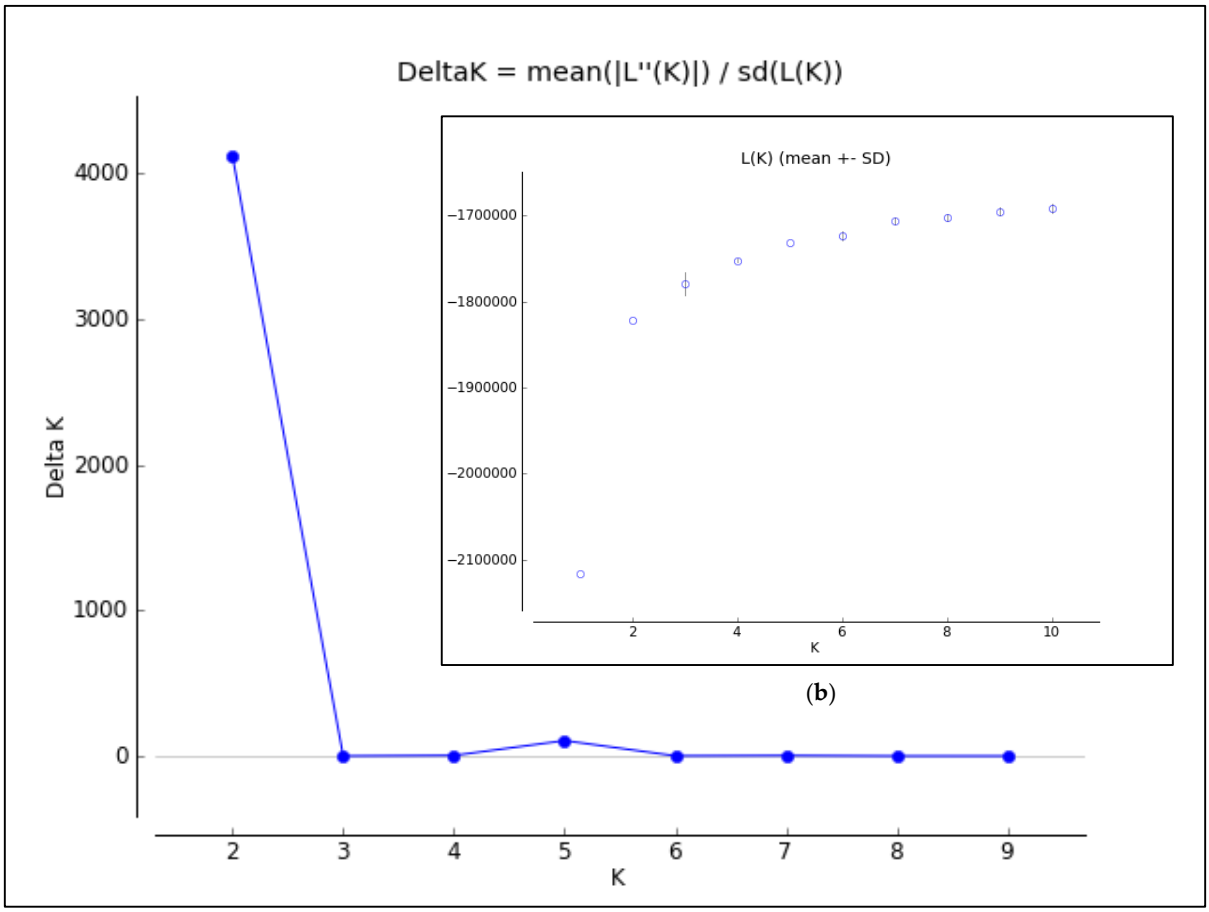



(**a**)



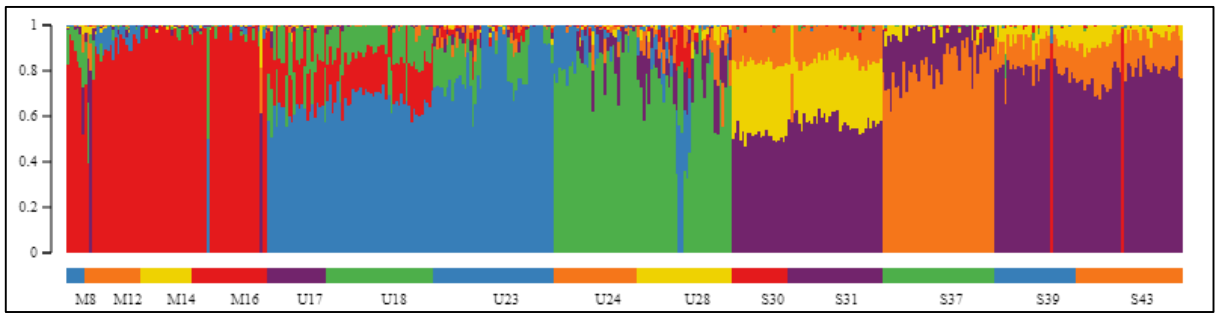

(**b**)

**Figure S3.** Principal coordinate analysis (PCoA) revealing the relationships between the 14 studied populations. (**a**) Discrimination among three pine species, horizontal axis corresponds to 58.8% of variation and vertical axis represents further 12.6%; (**b**) distinction within the *P. mugo* complex with 44.5% of variation distributed along horizontal and 16.0% along vertical axis.

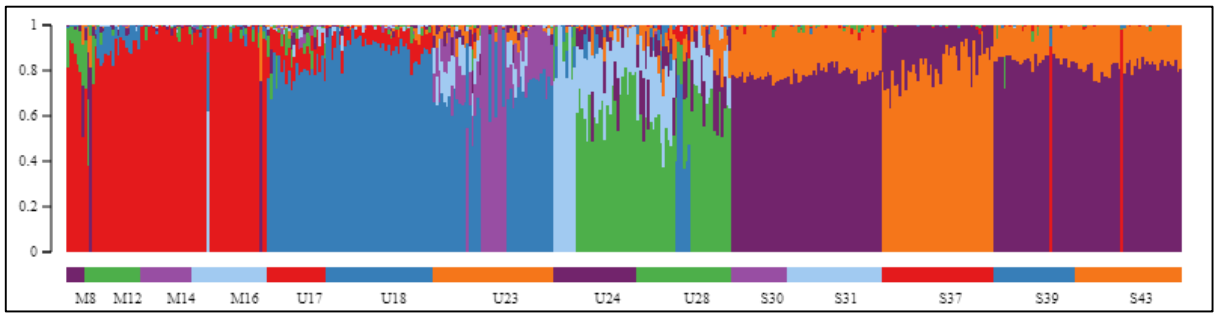

DeltaK = mean(|L''(K)|) / sd(L(K))
Delta K
4000
3000
2000
1000
0
2
3
4
5
6
7
8
9
K
L(K) (mean +- SD)
−1700000
−1800000
−1900000
−2000000
−2100000
2
4
6
8
10
K

(**b**)

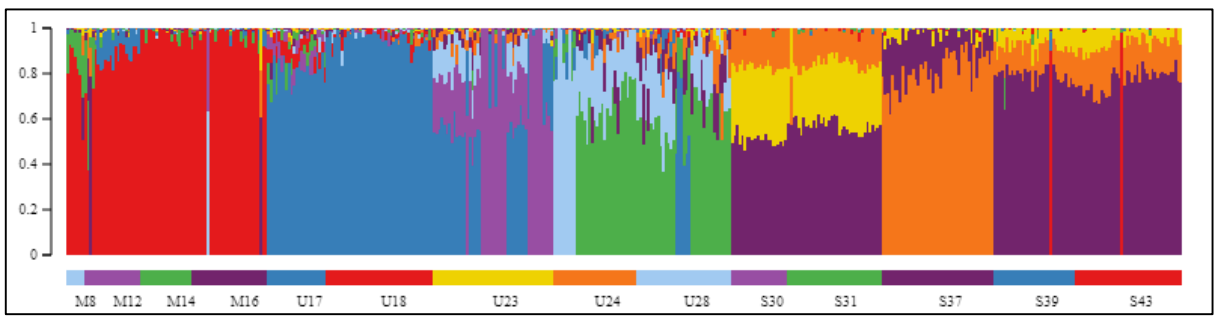
(**a**)


1
0.8
0.6
0.4
0.2
0
M8
M12
M14
M16
U17
U18
U23
U24
U28
S30
S31
S37
S39
S43

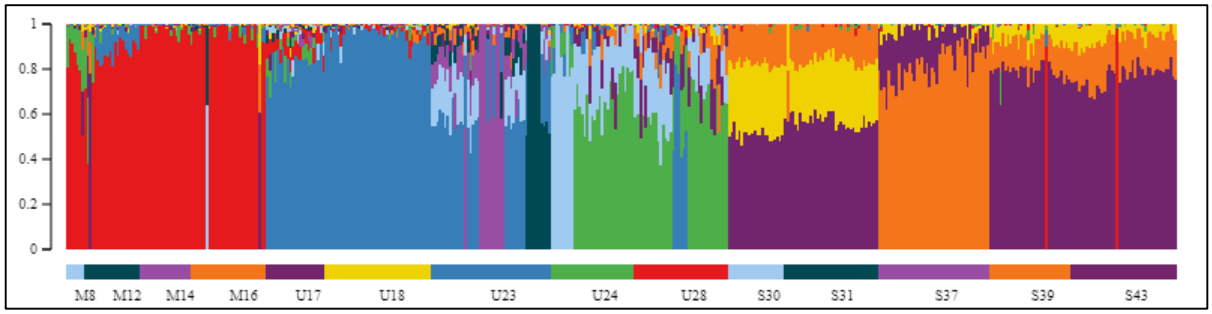
(**c**)


1
0.8
0.6
0.4
0.2
0
M8
M12
M14
M16
U17
U18
U23
U24
U28
S30
S31
S37
S39
S43

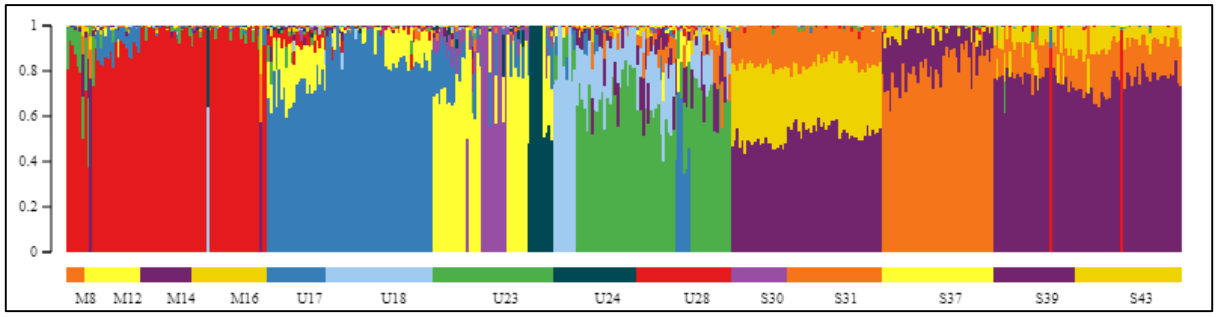
(**d**)


1
0.8
0.6
0.4
0.2
0
M8
M12
M14
M16
U17
U18
U23
U24
U28
S30
S31
S37
S39
S43

(**e**)

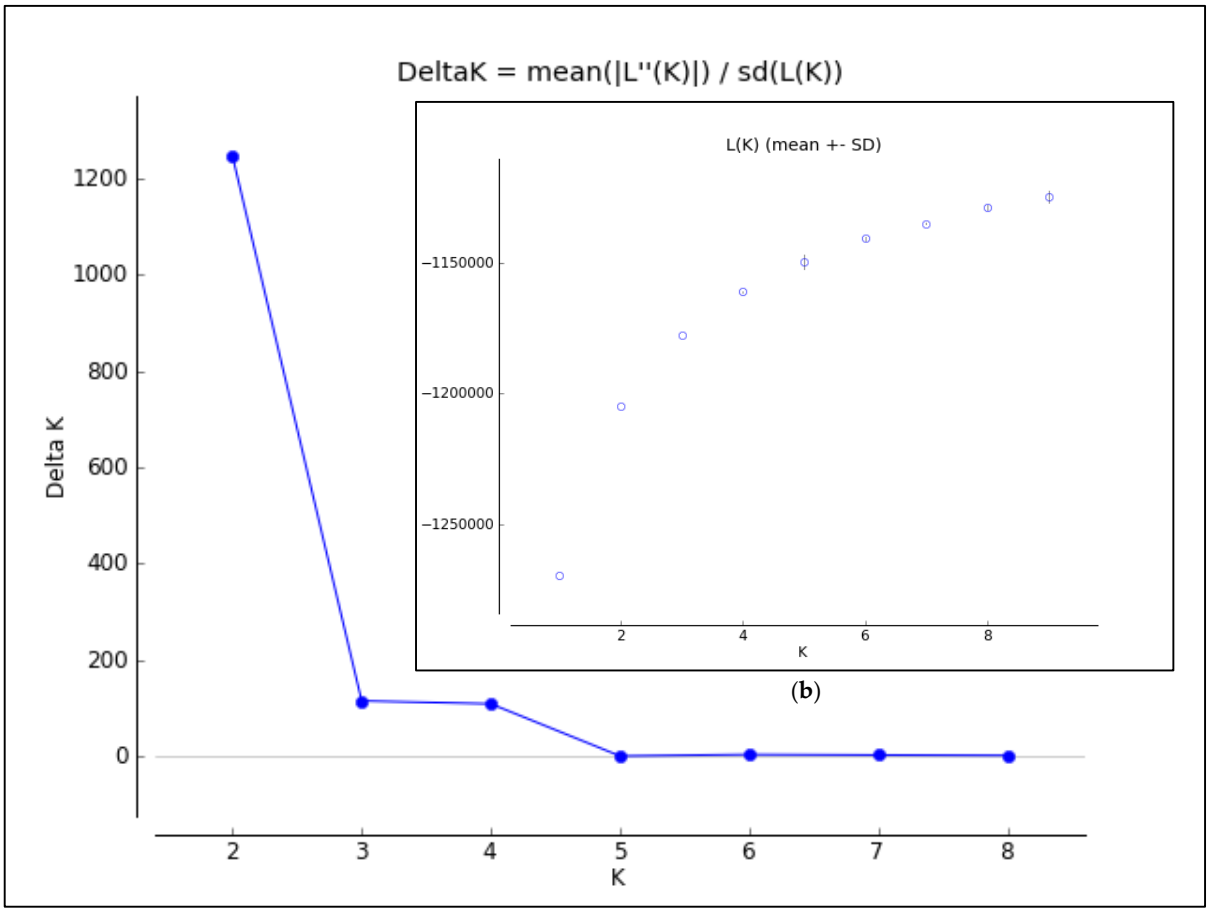



(**f**)



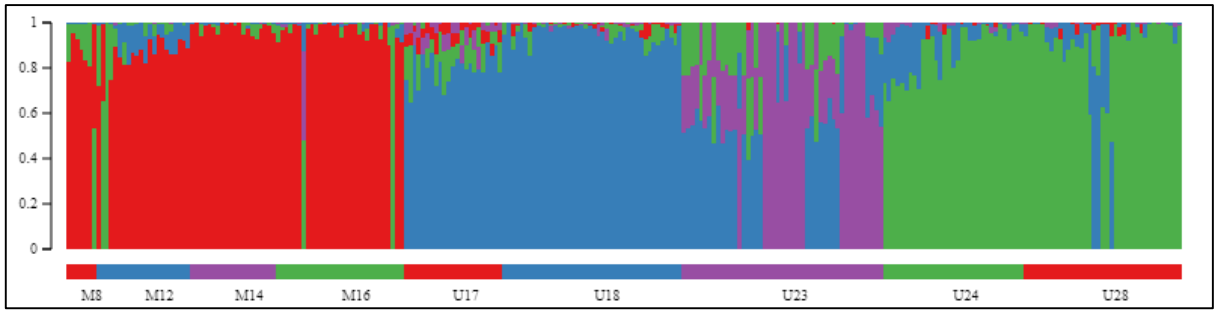

(**g**)

**Figure S4**. Results of clustering analysis performed with STRUCTURE for the 14 studied populations. (**a**) Delta*K* estimates and (**b**) mean values of natural logarithm of likelihood, as indicated by STRUCTURE HARVESTER for *K* groups between 1 and 10. Detailed results of analysis for specific *K* values: (**c**) *K* = 6, (**d**) *K* = 7, (**e**) *K* = 8, (**f**) *K* = 9, (**g**) *K* = 10. Scale on the left and vertical bars represent proportion of each genome being composed by variants specific for particular genomes. Horizontal colour bars at the bottom of each chart label individuals from distinct populations.

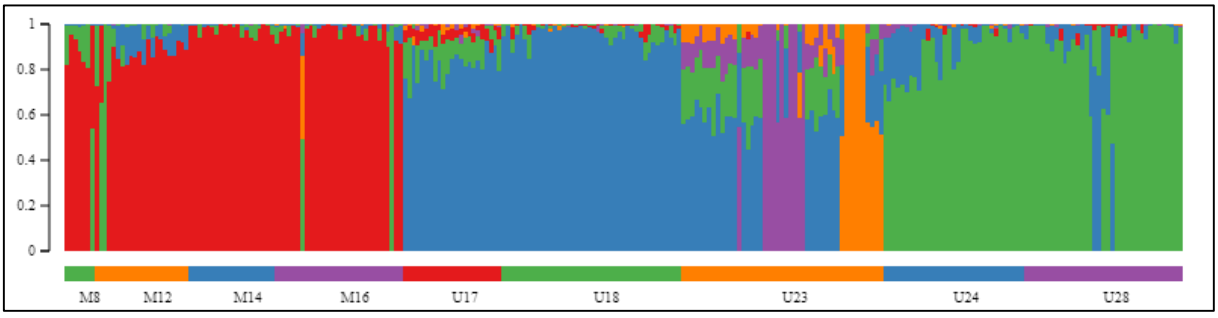

DeltaK = mean(|L''(K)|) / sd(L(K))
Delta K
K
L(K) (mean +- SD)
−1150000
−1200000
−1250000
K

(**b**)

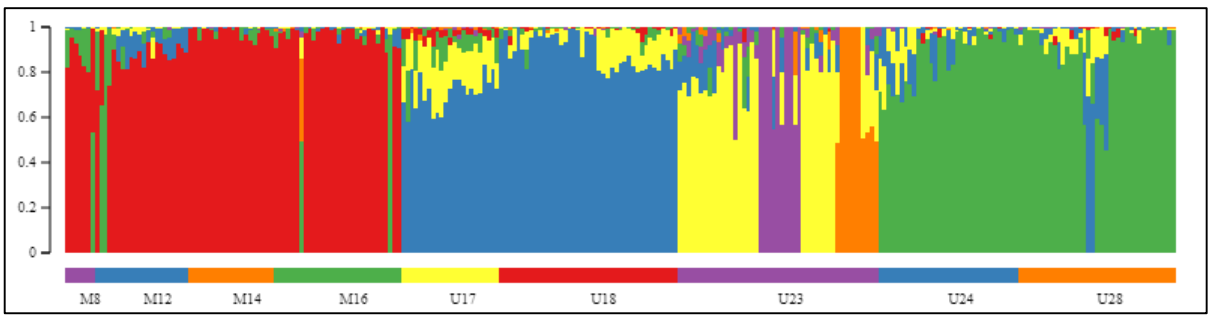
(**a**)


M8
M12
M14
M16
U17
U18
U23
U24
U28

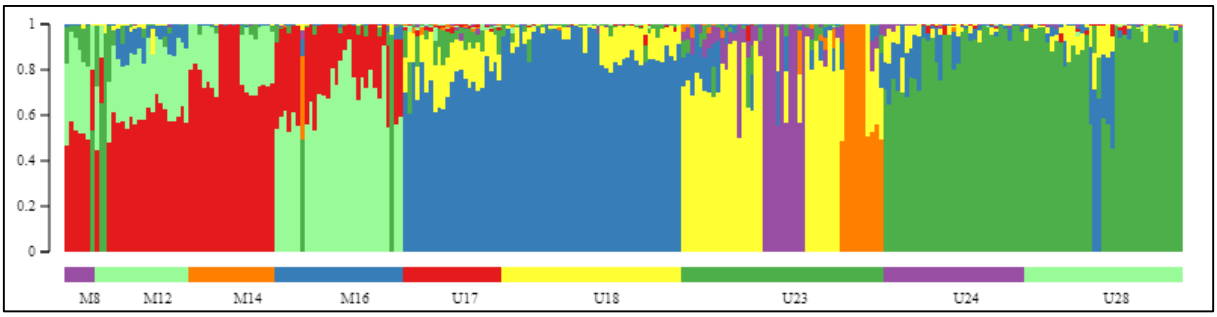
(**c**)


M8
M12
M14
M16
U17
U18
U23
U24
U28

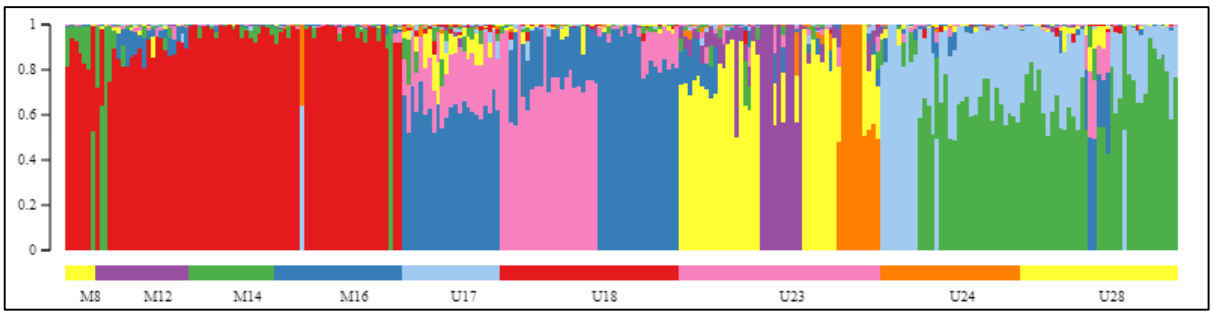
(**d**)


M8
M12
M14
M16
U17
U18
U23
U24
U28

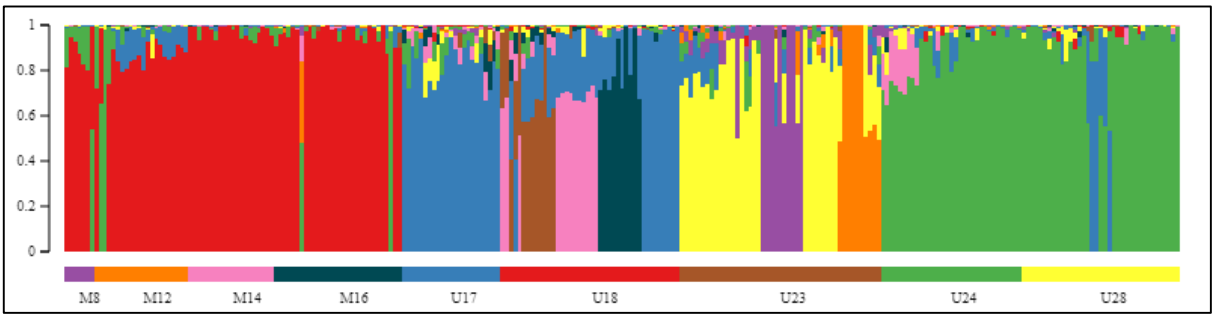
(**e**)




(**f**)




(**g**)



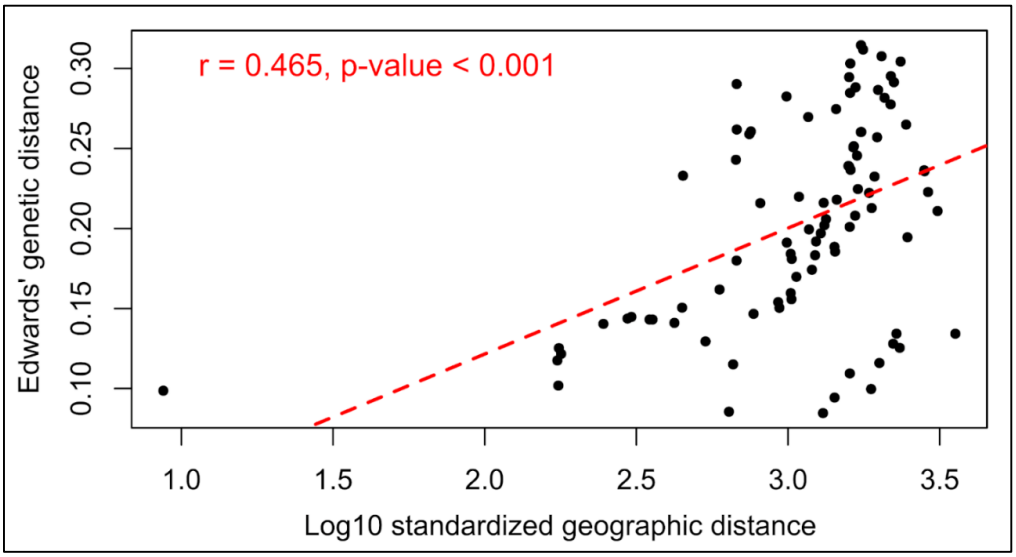

(**h**)

**Figure S5.** Results of clustering analysis performed with STRUCTURE for 9 populations of *Pinus mugo* complex. (**a**) Delta*K* estimates and (**b**) mean values of the natural logarithm of likelihood, as indicated by STRUCTURE HARVESTER for *K* groups between 1 and 9. Detailed results of analysis for specific *K* values: (**c**) *K* = 4, (**d**) *K* = 5, (**e**) *K* = 6, (**f**) *K* = 7, (**g**) *K* = 8 and (**h**) *K* = 9. Scale on the left and vertical bars represent proportion of each genome being composed by variants specific for particular genomes. Horizontal colour bars at the bottom of each chart label individuals from distinct populations.

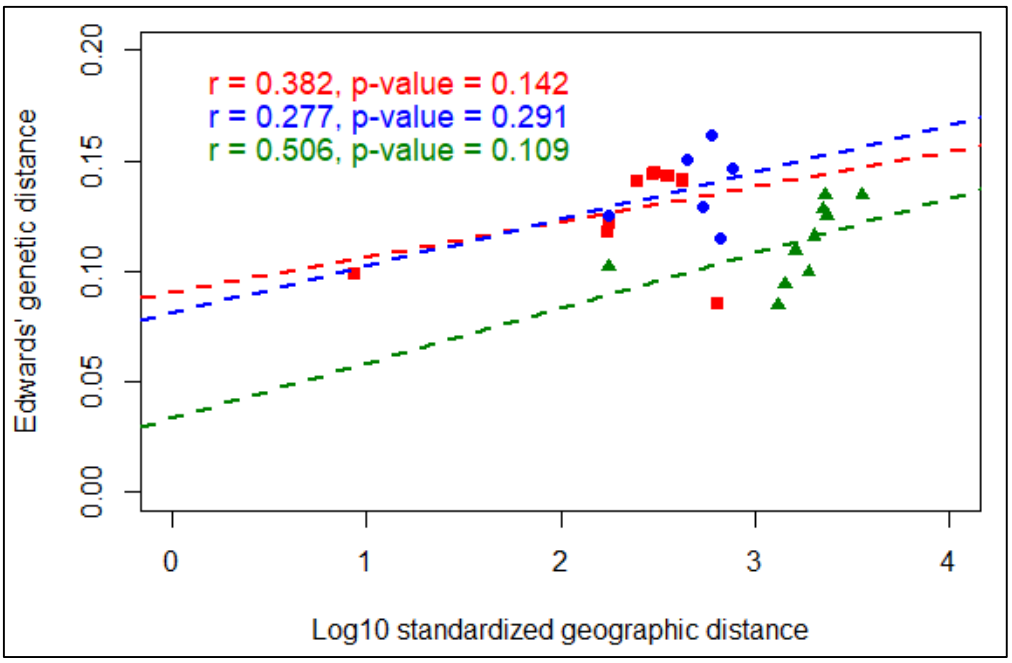
**Table S3.** Results of Mantel tests verifying the isolation by distance (IBD) hypothesis performed on Edwards' genetic and log standardized geographic distances for different combinations of studied pine populations.

| Populations analysed | N | r | *p* value |
|---|---|---|---|
| Three species | 14 | 0.465 | 0.001 |
| *P. mugo* & *P. uncinata* | 9 | 0.705 | 0.005 |
| *P. mugo* & *P. sylvestris* | 9 | 0.335 | 0.021 |
| *P. uncinata* & *P. sylvestris* | 10 | 0.398 | 0.014 |
| within *P. mugo* | 4 | 0.277 | 0.291 |
| within *P. uncinata* | 5 | 0.382 | 0.142 |
| within *P. sylvestris* | 5 | 0.506 | 0.109 |

N – number of populations tested, r - Pearson correlation coefficient.




(**a**)




(**b**)

**Figure S6.** Correlation plots of the geographic and genetic distances verifying the isolation by distance (IBD) hypothesis in three pine species. (**a**) Mantel test including comparisons within and between populations of all species, (**b**) results of three independent tests performed within taxa: blue dots – dwarf pine, red squares – Pyrenean pine, green triangles – Scots pine. Regression lines and results of the tests on bottom panel are indicated with respective colours.

**Table S4.** Gene ontology (GO) annotation of sequences with outlier polymorphisms detected. The green background distinguishes the sequence differentiating Scots pine from two mountain pines while markers differentiating the latter are fulfilled with blue; eight sequences with the most reliable outlier SNPs supported by all three methods and significant at *p*/*q* value < 0.01 are given in bold.

| **Sequence name** | **Species comparison** | | | | **Gene Ontology (GO) annotation** | |
|---|---|---|---|---|---|---|
| **(outlier SNP position)** | **M *vs* U** | **MU *vs* S** | **M *vs* S** | **U *vs* S** | **Domains** | **IDs** |
| comp17013_c0_seq1 (1608) | X* | | | | BP; CC | GO:0005774;GO:0016021;GO:0055085 |
| comp18757_c0_seq1 (2058) | | X³* | | X³ | BP; MF | GO:0008760;GO:0019277 |
| comp18889_c0_seq1 (111) | X | | | | BP; MF | GO:0016567;GO:0016874;GO:0061630 |
| comp18988_c0_seq1 (2220) | | X* | | X* | BP; CC; MF | GO:0000287;GO:0004497;GO:0005524;GO:0009535;GO:0009853;GO:0015986; GO:0016984;GO:0019253;GO:0045261;GO:0046933;GO:0055114 |
| comp19056_c0_seq1 (1447) | X | | | | BP; MF | GO:0003712;GO:0006355 |
| comp19359_c0_seq1 (402) | | | | X | BP; CC; MF | GO:0005506;GO:0009055;GO:0009507;GO:0016021;GO:0022900;GO:0043448 |
| comp19450_c0_seq1 (1154) | | X | | X | BP; MF | GO:0004252;GO:0006508 |
| comp19472_c0_seq1 (739) | X | | | | CC | GO:0016021 |
| comp19700_c0_seq1 (1179) | X | | | | BP; CC; MF | GO:0003677;GO:0003700;GO:0005634;GO:0006355 |
| comp20014_c0_seq1 (2275) | X | | | | BP; CC; MF | GO:0004675;GO:0005515;GO:0005524;GO:0005886;GO:0005975;GO:0006468 |
| comp20164_c0_seq1 (232) | X | | | | BP; MF | GO:0004857;GO:0043086 |
| **comp20176_c0_seq1 (240)** | **X³*** | | | | **BP; CC** | **GO:0005802;GO:0005829;GO:0006605;GO:0030124;GO:0030131;GO:0031410; GO:0090160** |
| comp20234_c0_seq1 (1126) | X | | | | BP; CC | GO:0005739;GO:0016554;GO:0080156 |
| comp24802_c0_seq1 (1869) | X | | | | BP; MF | GO:0005985;GO:0016157;GO:0046524 |
| comp25462_c0_seq1 (100) | X | | | | BP; CC; MF | GO:0001671;GO:0005829;GO:0006457;GO:0032781;GO:0051087;GO:0051879 |
| comp261721_c0_seq1 (255) | X | | | | CC | GO:0016020 |
| comp27292_c0_seq1 (1521) | | X | | | BP; CC; MF | GO:0002184;GO:0005829;GO:0016149;GO:0018444;GO:1990825 |
| comp28590_c0_seq1 (1341) | | X* | X | X* | BP; CC; MF | GO:0000287;GO:0008137;GO:0009055;GO:0009522;GO:0009535;GO:0015979; GO:0016021;GO:0016168;GO:0018298;GO:0022900;GO:0048038;GO:0051539 |
| comp29118_c0_seq1 (688) | X | | | | BP; CC; MF | GO:0006629;GO:0016021;GO:0016491;GO:0055114 |
| comp29127_c0_seq1 (671) | | X | | | BP; CC; MF | GO:0000209;GO:0005515;GO:0005524;GO:0005783;GO:0016021;GO:0016874; GO:0030433;GO:0042631;GO:0061631;GO:1902457 |
| comp30050_c0_seq1 (910) | | X | | X | BP; MF | GO:0003700;GO:0006352;GO:0006355 |
| comp31542_c0_seq1 (466) | X | | | | BP; CC; MF | GO:0003700;GO:0005634;GO:0006355;GO:0009739;GO:0043565 |
| comp31770_c0_seq1 (1348) | X | | | | BP; MF | GO:0003700;GO:0006355 |
| comp33894_c0_seq1 (1936) | | | | X | BP; CC; MF | GO:0004672;GO:0005515;GO:0005524;GO:0006468;GO:0016020;GO:0016491 |
| comp34171_c0_seq1 (187) | X | | | | BP; CC; MF | GO:0004674;GO:0005524;GO:0005634;GO:0006357;GO:0006468 |
| comp35189_c0_seq1 (1282) | X | | | | BP; CC; MF | GO:0003333;GO:0005886;GO:0015171;GO:0016021 |
| comp35641_c0_seq1 (1756) | X | | | | MF | GO:0046983 |

| | | | | | |
|---|---|---|---|---|---|
| comp35702_c0_seq1 (1216) | | X | | BP; MF | GO:0004190;GO:0006508;GO:0006629;GO:0030163 |
| comp358563_c0_seq1 (135) | | X | | BP; CC | GO:0010005;GO:0043622 |
| comp36656_c0_seq1 (1022) | X | | | CC | GO:0009507 |
| comp36690_c1_seq1 (361) | X | | | CC; MF | GO:0005794;GO:0008168;GO:0016020 |
| comp36863_c0_seq1 (1165) | | X | X* | BP; CC; MF | GO:0005515;GO:0005643;GO:0005654;GO:0005829;GO:0006607;GO:0008139; GO:0042025;GO:0061608 |
| comp37364_c0_seq1 (613) | X | | | CC | GO:0016021 |
| comp37406_c0_seq1 (254) | | X | | BP; CC | GO:0043622;GO:0110165 |
| comp37651_c0_seq1 (1062) | | X* | X* | BP; MF | GO:0016206;GO:0019438;GO:0032259;GO:0046983;GO:0102084;GO:0102938 |
| comp38584_c0_seq1 (1237) | | X³* | X | BP; CC; MF | GO:0004674;GO:0005524;GO:0005634;GO:0005737;GO:0005886;GO:0006468 |
| comp38629_c0_seq1 (349) | X | | | BP; CC; MF | GO:0005506;GO:0016021;GO:0016705;GO:0020037;GO:0055114 |
| comp38943_c0_seq1 (866) | | X³* | X³* | BP; CC; MF | GO:0005634;GO:0043565;GO:0045893;GO:0046982 |
| **comp39941_c0_seq1 (1816)** | **X³*** | | | **BP; CC; MF** | **GO:0004553;GO:0005975;GO:0046658** |
| comp40569_c0_seq1 (764) | | | X | BP; CC; MF | GO:0004656;GO:0005506;GO:0005783;GO:0016021;GO:0018401;GO:0031418; GO:0055114 |
| comp41803_c0_seq1 (1452) | X | | | BP; CC; MF | GO:0005737;GO:0006511;GO:0016021;GO:0043231;GO:0061630 |
| **comp41821_c0_seq1 (413)** | **X³*** | | | **BP; MF** | **GO:0030001;GO:0046872** |
| comp423545_c0_seq1 (183) | X | | | BP; MF | GO:0004190;GO:0006508 |
| comp42606_c0_seq1 (1061) | | X³* | X³* | BP; MF | GO:0009055;GO:0015035;GO:0022900;GO:0045454 |
| comp42726_c0_seq1 (307) | | X³* | X³* | BP; MF | GO:0004252;GO:0006508 |
| comp42805_c1_seq1 (2699) | X | | | BP; MF | GO:0003700;GO:0006355;GO:0043565 |
| comp42950_c0_seq1 (425) | X* | | | BP; MF | GO:0000155;GO:0000160;GO:0006355;GO:0009584;GO:0009585;GO:0009881; GO:0017006;GO:0018298;GO:0023014;GO:0042803 |
| comp43820_c0_seq1 (103) | | X* | X* | BP; MF | GO:0006482;GO:0016740;GO:0051723 |
| comp44465_c0_seq1 (1110) | X | | | BP; MF | GO:0016702;GO:0046872;GO:0055114 |
| **comp44835_c0_seq1 (339)** | **X³*** | | | **BP; MF** | **GO:0009058;GO:0016844** |
| comp44981_c0_seq1 (281) | X | | | CC | GO:0016021 |
| comp45024_c0_seq1 (839) | | X | | MF | GO:0003677 |
| comp45463_c0_seq1 (2507) | | X | | BP; MF | GO:0008143;GO:0043488;GO:1900364 |
| comp45510_c0_seq1 (218) | X | | | MF | GO:0003824 |
| comp45924_c0_seq1 (434) | | | X | BP; MF | GO:0003712;GO:0009910;GO:1903506 |
| comp46344_c0_seq1 (1108) | | X | | BP; CC | GO:0016020;GO:0071782;GO:0071786 |
| comp47007_c0_seq1 (1300) | X | | | BP | GO:0001522 |
| comp47328_c0_seq1 (518) | X | | | BP; CC; MF | GO:0006857;GO:0015333;GO:0016021;GO:0042937;GO:0055085 |
| comp47467_c0_seq1 (1161) | | X | | BP; CC; MF | GO:0005525;GO:0005737;GO:0006886;GO:0016192 |
| comp47663_c0_seq1 (1253) | | | X | CC; MF | GO:0005737;GO:0008168;GO:0016020 |

| | | | | | |
|---|---|---|---|---|---|
| comp48051_c0_seq1 (220) | X | | | CC | GO:0016020 |
| comp48330_c0_seq1 (140) | | X[3]* | X[3]* | BP; MF | GO:0004096;GO:0005515;GO:0020037;GO:0055114;GO:0098869 |
| comp48723_c0_seq2 (83) | X | | | BP; CC; MF | GO:0003735;GO:0006364;GO:0006412;GO:0022627;GO:0032040;GO:0042274 |
| comp48942_c0_seq1 (1943) | X | | | BP; MF | GO:0003700;GO:0006355;GO:0043565 |
| comp49679_c0_seq1 (4572) | X | | | BP; MF | GO:0004674;GO:0005524;GO:0018105;GO:0018107 |
| comp49745_c0_seq1 (703) | X | | | BP | GO:0006355 |
| comp49772_c0_seq1 (1003) | X[3] | | | BP; MF | GO:0006807;GO:0016151 |
| comp50296_c0_seq1 (1883) | X | | | BP; CC; MF | GO:0001193;GO:0003676;GO:0003899;GO:0005665;GO:0005730;GO:0006283; GO:0006367;GO:0006379;GO:0008270;GO:0016021 |
| comp50473_c0_seq1 (669) | X | | | BP; CC; MF | GO:0003700;GO:0005634;GO:0030154;GO:0043565 |
| **comp50552_c0_seq1 (1098)** | **X[3]*** | | | **BP; MF** | **GO:0016787;GO:0044237** |
| comp50851_c0_seq1 (996) | | X | | MF | GO:0000166;GO:0047710 |
| comp50905_c0_seq4 (942) | | | X | BP; MF | GO:0006479;GO:0008276 |
| comp51215_c0_seq1 (1350) | X | | | BP; MF | GO:0006810;GO:0022857 |
| comp51336_c0_seq1 (1958) | X | | | BP; CC; MF | GO:0004497;GO:0005506;GO:0016021;GO:0016705;GO:0020037;GO:0055114 |
| comp51669_c0_seq1 (313) | X | | | BP; CC; MF | GO:0003723;GO:0005654;GO:0043484;GO:1990904 |
| comp51783_c0_seq1 (508) | X | | | BP; CC; MF | GO:0005829;GO:0009854;GO:0016618;GO:0030267;GO:0047995;GO:0051287; GO:0055114;GO:0102742 |
| comp52585_c0_seq1 (127) | X | | | BP; CC; MF | GO:0004160;GO:0005507;GO:0009082;GO:0009553;GO:0009555;GO:0009570; GO:0009651;GO:0048364 |
| comp52700_c0_seq1 (1559) | X | | | CC; MF | GO:0003677;GO:0005634;GO:0042025;GO:0046872 |
| **comp52994_c0_seq1 (245)** | **X[3]*** | | | **BP; MF** | **GO:0003712;GO:0006355** |
| comp53206_c1_seq1 (1087) | X | | | CC; MF | GO:0003924;GO:0005525;GO:0005739 |
| comp53528_c0_seq1 (1313) | X | | | MF | GO:0003676 |
| comp53591_c0_seq1 (320) | X | | | CC; MF | GO:0003677;GO:0005515;GO:0005634;GO:0042025 |
| **comp53610_c0_seq1 (276)** | **X[3]*** | | | **BP** | **GO:0009058;GO:0019752** |
| comp53749_c0_seq7 (2413) | | X* | X* | BP; MF | GO:0003676;GO:0006397;GO:0046872 |
| comp53870_c0_seq1 (109) | X | | | BP; CC; MF | GO:0004222;GO:0006508;GO:0006518;GO:0009507;GO:0016021;GO:0046872 |
| comp54118_c0_seq1 (2585) | X | | | MF | GO:0005515 |
| **comp54487_c0_seq1 (3338)** | **X[3]*** | | | **BP; CC; MF** | **GO:0004672;GO:0005515;GO:0005768;GO:0006914;GO:0007034;GO:0016310; GO:0035032;GO:0046907** |
| comp55083_c0_seq1 (1074) | X | | | MF | GO:0003676 |
| comp56793_c0_seq1 (670) | | | X | BP; CC; MF | GO:0006605;GO:0009306;GO:0009535;GO:0015450;GO:0016021;GO:0071806 |
| comp57453_c0_seq1 (2049) | | X | X | BP | GO:0007275;GO:0030154 |
| comp58612_c0_seq1 (681) | X* | | | BP; CC | GO:0005789;GO:0005886;GO:0016021;GO:0061817;GO:0090158 |
| comp59746_c0_seq1 (1253) | | X | | BP; CC; MF | GO:0005829;GO:0006508;GO:0016920 |

| | | | | | |
|---|---|---|---|---|---|
| comp59967_c0_seq1 (1910) | | X³ | X | BP; MF | GO:0003700;GO:0006355 |
| comp60629_c0_seq1 (1021) | X | | | BP; CC; MF | GO:0015095;GO:0015693;GO:0016021 |
| comp61109_c0_seq1 (1237) | | X | X* | MF | GO:0005515 |
| comp63308_c0_seq1 (1097) | X* | | | BP; CC | GO:0000398;GO:0000932;GO:0005688;GO:0046540;GO:0071011;GO:0071013; GO:1990726 |
| comp64225_c0_seq1 (1194) | X | | | MF | GO:0003700 |
| comp71632_c0_seq1 (1208) | X | | | BP; CC; MF | GO:0000014;GO:0000110;GO:0000413;GO:0000712;GO:0000720;GO:0000724; GO:0003684;GO:0003697;GO:0003755;GO:0006296;GO:0006979;GO:0009314; GO:0016021;GO:1901255 |
| comp72132_c0_seq1 (584) | X | | | BP; CC; MF | GO:0000166;GO:0003700;GO:0005634;GO:0006355;GO:0009536;GO:0009739; GO:0016874;GO:0043565 |
| comp74829_c0_seq1 (165) | | X | | BP; CC; MF | GO:0005768;GO:0005802;GO:0008757;GO:0016021;GO:0032259;GO:0052546 |
| comp1178_c0_seq1 (146) | | X | X | *undetermined* | |
| comp166107_c0_seq1 (81) | X | | | *undetermined* | |
| comp19558_c0_seq1 (688) | | X* | | *undetermined* | |
| comp21559_c0_seq1 (165) | X | | | *undetermined* | |
| comp227564_c0_seq1 (556) | | X | | *undetermined* | |
| comp29071_c0_seq1 (277) | X | | | *undetermined* | |
| comp37643_c0_seq1 (1161) | X | | | *undetermined* | |
| comp39687_c0_seq1 (706) | | X | | *undetermined* | |
| comp40831_c0_seq1 (622) | | | X | *undetermined* | |
| comp43127_c0_seq1 (312) | | | X | *undetermined* | |
| comp45714_c0_seq1 (1064) | | | X | *undetermined* | |
| comp48574_c0_seq1 (1639) | X | | | *undetermined* | |
| comp49514_c0_seq1 (63) | | X | | *undetermined* | |
| comp49840_c0_seq1 (2135) | | X | X³* | *undetermined* | |
| comp51626_c0_seq1 (1245) | X | | | *undetermined* | |
| comp52601_c0_seq1 (205) | X | | | *undetermined* | |
| comp58030_c0_seq1 (550) | X | | | *undetermined* | |
| comp66672_c0_seq1 (661) | X | | | *undetermined* | |
| 0_4394_01_1 (522) | X | | | *undetermined* | |
| CL12234_1_1 (626) | X | | | *undetermined* | |

X - SNP statistically significant at *p/q* value of 0.05 in at least two tests; X³ - significant at *p/q* value of 0.05 in all three tests; X* - significant at *p/q* value of 0.01 in at least two tests; M *vs* U – dwarf pine *vs* Pyrenean pine; MU *vs* S – mountain taxa *vs* Scots pine; M *vs* S – dwarf *vs* Scots pine; U *vs* S – Pyrenean *vs* Scots pine.

**Table S5.** Frequency of alleles in the loci with outlier SNPs identified. The green background distinguishes the sequence differentiating Scots pine from two mountain pines while markers differentiating the latter are fulfilled with blue; eight sequences with the most reliable outlier SNPs supported by all three methods and significant at *p*/*q* value < 0.01 are given in bold.

| Sequence name (SNP) | Allele pair | *P. mugo* | *P. uncinata* | *P. mugo* complex | *P. sylvestris* |
|---|---|---|---|---|---|
| comp17013_c0_seq1 (1608) | G / A | 0.02 / 0.98 | 0.22 / 0.78 | 0.15 / 0.85 | 0.33 / 0.67 |
| comp18757_c0_seq1 (2058) | G / A | 0.83 / 0.17 | 0.9 / 0.1 | 0.88 / 0.12 | 0.99 / 0.01 |
| comp18889_c0_seq1 (111) | C / T | 0.08 / 0.92 | 0.15 / 0.85 | 0.13 / 0.87 | 0.13 / 0.87 |
| comp18988_c0_seq1 (2220) | C / A | 0.5 / 0.5 | 0.75 / 0.25 | 0.67 / 0.33 | 0.72 / 0.28 |
| comp19056_c0_seq1 (1447) | G / T | 0.12 / 0.88 | 0.35 / 0.65 | 0.28 / 0.72 | 0 / 1 |
| comp19359_c0_seq1 (402) | A / G | 0.9 / 0.1 | 0.89 / 0.11 | 0.89 / 0.11 | 1 / 0 |
| comp19450_c0_seq1 (1154) | G / A | 0.87 / 0.13 | 0.95 / 0.05 | 0.93 / 0.07 | 1 / 0 |
| comp19472_c0_seq1 (739) | A / T | 0.28 / 0.72 | 0.18 / 0.82 | 0.21 / 0.79 | 0.01 / 0.99 |
| comp19700_c0_seq1 (1179) | C / A | 0.89 / 0.11 | 0.94 / 0.06 | 0.93 / 0.07 | 0.99 / 0.01 |
| comp20014_c0_seq1 (2275) | C / T | 0.56 / 0.44 | 0.36 / 0.64 | 0.42 / 0.58 | 0.43 / 0.57 |
| comp20164_c0_seq1 (232) | C / A | 0.79 / 0.21 | 0.98 / 0.02 | 0.93 / 0.07 | 1 / 0 |
| **comp20176_c0_seq1 (240)** | **T / C** | **0.05 / 0.95** | **0.31 / 0.69** | **0.23 / 0.77** | **0.6 / 0.4** |
| comp20234_c0_seq1 (1126) | G / A | 0.01 / 0.99 | 0.03 / 0.97 | 0.03 / 0.97 | 0.11 / 0.89 |
| comp24802_c0_seq1 (1869) | C / G | 0.17 / 0.83 | 0.04 / 0.96 | 0.08 / 0.92 | 0.04 / 0.96 |
| comp25462_c0_seq1 (100) | T / G | 0.82 / 0.18 | 0.85 / 0.15 | 0.84 / 0.16 | 0.9 / 0.1 |
| comp261721_c0_seq1 (255) | A / G | 0.01 / 0.99 | 0.02 / 0.98 | 0.02 / 0.98 | 0.03 / 0.97 |
| comp27292_c0_seq1 (1521) | C / T | 0.17 / 0.83 | 0.49 / 0.51 | 0.4 / 0.6 | 0.42 / 0.58 |
| comp28590_c0_seq1 (1341) | C / A | 0.79 / 0.21 | 0.87 / 0.13 | 0.85 / 0.15 | 0.78 / 0.22 |
| comp29118_c0_seq1 (688) | C / G | 0.05 / 0.95 | 0.03 / 0.97 | 0.03 / 0.97 | 0.01 / 0.99 |
| comp29127_c0_seq1 (671) | A / G | 0.98 / 0.02 | 0.95 / 0.05 | 0.96 / 0.04 | 1 / 0 |
| comp30050_c0_seq1 (910) | T / C | 0.67 / 0.33 | 0.58 / 0.42 | 0.61 / 0.39 | 0.83 / 0.17 |
| comp31542_c0_seq1 (466) | G / C | 0.9 / 0.1 | 0.94 / 0.06 | 0.93 / 0.07 | 0.75 / 0.25 |
| comp31770_c0_seq1 (1348) | G / C | 0.07 / 0.93 | 0.06 / 0.94 | 0.06 / 0.94 | 0.08 / 0.92 |
| comp33894_c0_seq1 (1936) | A / G | 0.93 / 0.07 | 0.8 / 0.2 | 0.83 / 0.17 | 0.5 / 0.5 |
| comp34171_c0_seq1 (187) | C / T | 0.32 / 0.68 | 0.36 / 0.64 | 0.34 / 0.66 | 0.2 / 0.8 |
| comp35189_c0_seq1 (1282) | G / C | 0.09 / 0.91 | 0.02 / 0.98 | 0.04 / 0.96 | 0.02 / 0.98 |
| comp35641_c0_seq1 (1756) | G / T | 0.93 / 0.07 | 0.96 / 0.04 | 0.95 / 0.05 | 0.93 / 0.07 |
| comp35702_c0_seq1 (1216) | C / A | 0.39 / 0.61 | 0.44 / 0.56 | 0.42 / 0.58 | 0.03 / 0.97 |
| comp358563_c0_seq1 (135) | A / G | 0.66 / 0.34 | 0.91 / 0.09 | 0.84 / 0.16 | 0.97 / 0.03 |
| comp36656_c0_seq1 (1022) | G / A | 0.86 / 0.14 | 0.96 / 0.04 | 0.93 / 0.07 | 1 / 0 |
| comp36690_c1_seq1 (361) | G / A | 0.92 / 0.08 | 0.7 / 0.3 | 0.77 / 0.23 | 0.73 / 0.27 |
| comp36863_c0_seq1 (1165) | G / A | 0.53 / 0.47 | 0.32 / 0.68 | 0.38 / 0.62 | 0.11 / 0.89 |

| | | | | | |
|---|---|---|---|---|---|
| comp37364_c0_seq1 (613) | G / T | 0.31 / 0.69 | 0.29 / 0.71 | 0.29 / 0.71 | 0.01 / 0.99 |
| comp37406_c0_seq1 (254) | G / A | 0.16 / 0.84 | 0.58 / 0.42 | 0.45 / 0.55 | 0.96 / 0.04 |
| comp37651_c0_seq1 (1062) | T / C | 0.9 / 0.1 | 1 / 0 | 0.97 / 0.03 | 1 / 0 |
| comp38584_c0_seq1 (1237) | T / C | 0.27 / 0.73 | 0.55 / 0.45 | 0.47 / 0.53 | 0.79 / 0.21 |
| comp38629_c0_seq1 (349) | C / G | 0.89 / 0.11 | 0.81 / 0.19 | 0.83 / 0.17 | 0.99 / 0.01 |
| comp38943_c0_seq1 (866) | A / G | 0.12 / 0.88 | 0.07 / 0.93 | 0.08 / 0.92 | 0.12 / 0.88 |
| **comp39941_c0_seq1 (1816)** | **G / A** | **0.17 / 0.83** | **0.39 / 0.61** | **0.32 / 0.68** | **0.44 / 0.56** |
| comp40569_c0_seq1 (764) | T / C | 0.94 / 0.06 | 0.6 / 0.4 | 0.7 / 0.3 | 0.96 / 0.04 |
| comp41803_c0_seq1 (1452) | C / T | 0.18 / 0.82 | 0.08 / 0.92 | 0.11 / 0.89 | 0 / 1 |
| **comp41821_c0_seq1 (413)** | **C / T** | **0.77 / 0.23** | **0.94 / 0.06** | **0.89 / 0.11** | **1 / 0** |
| comp423545_c0_seq1 (183) | C / T | 0.28 / 0.72 | 0.1 / 0.9 | 0.15 / 0.85 | 0 / 1 |
| comp42606_c0_seq1 (1061) | G / T | 0.27 / 0.73 | 0.23 / 0.77 | 0.24 / 0.76 | 0.01 / 0.99 |
| comp42726_c0_seq1 (307) | G / A | 0.64 / 0.36 | 0.28 / 0.72 | 0.38 / 0.62 | 0.05 / 0.95 |
| comp42805_c1_seq1 (2699) | A / G | 0.34 / 0.66 | 0.54 / 0.46 | 0.49 / 0.51 | 0.98 / 0.02 |
| comp42950_c0_seq1 (425) | T / C | 0.97 / 0.03 | 0.98 / 0.02 | 0.98 / 0.02 | 0.88 / 0.12 |
| comp43820_c0_seq1 (103) | C / G | 0.75 / 0.25 | 0.91 / 0.09 | 0.87 / 0.13 | 1 / 0 |
| comp44465_c0_seq1 (1110) | A / G | 0.83 / 0.17 | 0.74 / 0.26 | 0.76 / 0.24 | 0.99 / 0.01 |
| **comp44835_c0_seq1 (339)** | **C / G** | **0.91 / 0.09** | **0.54 / 0.46** | **0.65 / 0.35** | **0.01 / 0.99** |
| comp44981_c0_seq1 (281) | C / T | 0.7 / 0.3 | 0.43 / 0.57 | 0.51 / 0.49 | 0.58 / 0.42 |
| comp45024_c0_seq1 (839) | T / C | 0.02 / 0.98 | 0 / 1 | 0.01 / 0.99 | 0.02 / 0.98 |
| comp45463_c0_seq1 (2507) | G / A | 0.92 / 0.08 | 0.82 / 0.18 | 0.85 / 0.15 | 0.77 / 0.23 |
| comp45510_c0_seq1 (218) | G / C | 0.84 / 0.16 | 0.46 / 0.54 | 0.58 / 0.42 | 0.24 / 0.76 |
| comp45924_c0_seq1 (434) | T / C | 0.31 / 0.69 | 0.41 / 0.59 | 0.38 / 0.62 | 0.01 / 0.99 |
| comp46344_c0_seq1 (1108) | C / T | 0.04 / 0.96 | 0.03 / 0.97 | 0.03 / 0.97 | 0.2 / 0.8 |
| comp47007_c0_seq1 (1300) | G / T | 0.49 / 0.51 | 0.22 / 0.78 | 0.3 / 0.7 | 0.52 / 0.48 |
| comp47328_c0_seq1 (518) | G / T | 0.38 / 0.63 | 0.58 / 0.42 | 0.52 / 0.48 | 0.6 / 0.4 |
| comp47467_c0_seq1 (1161) | A / G | 0.38 / 0.63 | 0.61 / 0.39 | 0.54 / 0.46 | 0.6 / 0.4 |
| comp47663_c0_seq1 (1253) | G / C | 0.37 / 0.63 | 0.39 / 0.61 | 0.39 / 0.61 | 0.01 / 0.99 |
| comp48051_c0_seq1 (220) | C / T | 0.86 / 0.14 | 0.7 / 0.3 | 0.75 / 0.25 | 0.59 / 0.41 |
| comp48330_c0_seq1 (140) | T / C | 0.44 / 0.56 | 0.38 / 0.62 | 0.4 / 0.6 | 0.03 / 0.97 |
| comp48723_c0_seq2 (83) | G / T | 0.13 / 0.87 | 0.22 / 0.78 | 0.19 / 0.81 | 0.99 / 0.01 |
| comp48942_c0_seq1 (1943) | T / C | 0.94 / 0.06 | 0.87 / 0.13 | 0.89 / 0.11 | 0.81 / 0.19 |
| comp49679_c0_seq1 (4572) | A / G | 0.94 / 0.06 | 0.87 / 0.13 | 0.89 / 0.11 | 0.98 / 0.02 |
| comp49745_c0_seq1 (703) | G / C | 0.19 / 0.81 | 0.2 / 0.8 | 0.19 / 0.81 | 0.01 / 0.99 |
| comp49772_c0_seq1 (1003) | A / G | 0.72 / 0.28 | 0.7 / 0.3 | 0.71 / 0.29 | 0.41 / 0.59 |
| comp50296_c0_seq1 (1883) | T / C | 0.78 / 0.22 | 0.68 / 0.32 | 0.71 / 0.29 | 0.76 / 0.24 |

| | | | | | |
|---|---|---|---|---|---|
| comp50473_c0_seq1 (669) | A / C | 0.06 / 0.94 | 0.07 / 0.93 | 0.07 / 0.93 | 0.03 / 0.97 |
| **comp50552_c0_seq1 (1098)** | **A / G** | **0.68 / 0.32** | **0.57 / 0.43** | **0.6 / 0.4** | **0.4 / 0.6** |
| comp50851_c0_seq1 (996) | A / G | 0.11 / 0.89 | 0.5 / 0.5 | 0.39 / 0.61 | 0.98 / 0.02 |
| comp50905_c0_seq4 (942) | A / G | 0.05 / 0.95 | 0.14 / 0.86 | 0.12 / 0.88 | 0.02 / 0.98 |
| comp51215_c0_seq1 (1350) | C / G | 0.72 / 0.28 | 0.83 / 0.17 | 0.8 / 0.2 | 0.99 / 0.01 |
| comp51336_c0_seq1 (1958) | T / G | 0.16 / 0.84 | 0.23 / 0.77 | 0.21 / 0.79 | 0.34 / 0.66 |
| comp51669_c0_seq1 (313) | C / T | 0.23 / 0.77 | 0.16 / 0.84 | 0.18 / 0.82 | 0.21 / 0.79 |
| comp51783_c0_seq1 (508) | T / G | 0.81 / 0.19 | 0.64 / 0.36 | 0.69 / 0.31 | 0.25 / 0.75 |
| comp52585_c0_seq1 (127) | C / G | 0.77 / 0.23 | 0.84 / 0.16 | 0.82 / 0.18 | 1 / 0 |
| comp52700_c0_seq1 (1559) | C / T | 0.85 / 0.15 | 0.84 / 0.16 | 0.85 / 0.15 | 0.84 / 0.16 |
| **comp52994_c0_seq1 (245)** | **T / C** | **0.12 / 0.88** | **0.3 / 0.7** | **0.24 / 0.76** | **0.9 / 0.1** |
| comp53206_c1_seq1 (1087) | C / T | 0.96 / 0.04 | 0.8 / 0.2 | 0.85 / 0.15 | 0.83 / 0.17 |
| comp53528_c0_seq1 (1313) | G / A | 0.97 / 0.03 | 0.96 / 0.04 | 0.96 / 0.04 | 0.87 / 0.13 |
| comp53591_c0_seq1 (320) | G / A | 0.13 / 0.87 | 0.13 / 0.87 | 0.13 / 0.87 | 0.14 / 0.86 |
| **comp53610_c0_seq1 (276)** | **A / G** | **0.31 / 0.69** | **0.39 / 0.61** | **0.37 / 0.63** | **0.51 / 0.49** |
| comp53749_c0_seq7 (2413) | T / C | 0.66 / 0.34 | 0.69 / 0.31 | 0.68 / 0.32 | 0.98 / 0.02 |
| comp53870_c0_seq1 (109) | G / A | 0.65 / 0.35 | 1 / 0 | 0.91 / 0.09 | 0.99 / 0.01 |
| comp54118_c0_seq1 (2585) | A / C | 0.71 / 0.29 | 0.47 / 0.53 | 0.54 / 0.46 | 0.3 / 0.7 |
| **comp54487_c0_seq1 (3338)** | **T / C** | **0.18 / 0.82** | **0.26 / 0.74** | **0.24 / 0.76** | **0.01 / 0.99** |
| comp55083_c0_seq1 (1074) | C / T | 0.72 / 0.28 | 0.55 / 0.45 | 0.6 / 0.4 | 0.99 / 0.01 |
| comp56793_c0_seq1 (670) | T / G | 1 / 0 | 0.98 / 0.02 | 0.98 / 0.02 | 0.86 / 0.14 |
| comp57453_c0_seq1 (2049) | A / G | 0.25 / 0.75 | 0.3 / 0.7 | 0.29 / 0.71 | 0.01 / 0.99 |
| comp58612_c0_seq1 (681) | A / G | 0.81 / 0.19 | 0.92 / 0.08 | 0.89 / 0.11 | 0.99 / 0.01 |
| comp59746_c0_seq1 (1253) | T / G | 0.78 / 0.22 | 0.64 / 0.36 | 0.68 / 0.32 | 0.78 / 0.22 |
| comp59967_c0_seq1 (1910) | G / A | 0.2 / 0.8 | 0.28 / 0.72 | 0.25 / 0.75 | 0.23 / 0.77 |
| comp60629_c0_seq1 (1021) | T / C | 0.64 / 0.36 | 0.83 / 0.17 | 0.78 / 0.22 | 0.99 / 0.01 |
| comp61109_c0_seq1 (1237) | C / A | 0.87 / 0.13 | 0.97 / 0.03 | 0.94 / 0.06 | 0.96 / 0.04 |
| comp63308_c0_seq1 (1097) | C / T | 0.82 / 0.18 | 0.7 / 0.3 | 0.74 / 0.26 | 0.85 / 0.15 |
| comp64225_c0_seq1 (1194) | G / T | 0.45 / 0.55 | 0.23 / 0.77 | 0.29 / 0.71 | 0.01 / 0.99 |
| comp71632_c0_seq1 (1208) | C / T | 0.03 / 0.97 | 0.09 / 0.91 | 0.07 / 0.93 | 0.09 / 0.91 |
| comp72132_c0_seq1 (584) | A / G | 0.89 / 0.11 | 0.93 / 0.07 | 0.92 / 0.08 | 0.93 / 0.08 |
| comp74829_c0_seq1 (165) | C / A | 0 / 1 | 0.1 / 0.9 | 0.07 / 0.93 | 0.04 / 0.96 |
| comp1178_c0_seq1 (146) | G / A | 0.01 / 0.99 | 0.03 / 0.97 | 0.02 / 0.98 | 0.09 / 0.91 |
| comp166107_c0_seq1 (81) | A / G | 0.75 / 0.25 | 0.54 / 0.46 | 0.6 / 0.4 | 0.13 / 0.87 |
| comp19558_c0_seq1 (688) | A / G | 0.87 / 0.13 | 0.89 / 0.11 | 0.89 / 0.11 | 0.99 / 0.01 |
| comp21559_c0_seq1 (165) | T / C | 0.28 / 0.72 | 0.1 / 0.9 | 0.15 / 0.85 | 0 / 1 |

| | | | | | |
|---|---|---|---|---|---|
| comp227564_c0_seq1 (556) | A / G | 0.96 / 0.04 | 0.77 / 0.23 | 0.83 / 0.17 | 0.48 / 0.52 |
| comp29071_c0_seq1 (277) | T / G | 0.53 / 0.47 | 0.35 / 0.65 | 0.4 / 0.6 | 0.07 / 0.93 |
| comp37643_c0_seq1 (1161) | G / A | 0.59 / 0.41 | 0.25 / 0.75 | 0.34 / 0.66 | 0.01 / 0.99 |
| comp39687_c0_seq1 (706) | C / T | 0.25 / 0.75 | 0.23 / 0.78 | 0.23 / 0.77 | 0.2 / 0.8 |
| comp40831_c0_seq1 (622) | C / T | 0.18 / 0.82 | 0.33 / 0.67 | 0.29 / 0.71 | 0.45 / 0.55 |
| comp43127_c0_seq1 (312) | A / C | 0.36 / 0.64 | 0.06 / 0.94 | 0.15 / 0.85 | 0.21 / 0.79 |
| comp45714_c0_seq1 (1064) | C / A | 0.56 / 0.44 | 0.84 / 0.16 | 0.75 / 0.25 | 1 / 0 |
| comp48574_c0_seq1 (1639) | G / A | 0.57 / 0.43 | 0.6 / 0.4 | 0.59 / 0.41 | 0.01 / 0.99 |
| comp49514_c0_seq1 (63) | A / G | 0.3 / 0.7 | 0.11 / 0.89 | 0.17 / 0.83 | 0.25 / 0.75 |
| comp49840_c0_seq1 (2135) | C / G | 0.72 / 0.28 | 0.81 / 0.19 | 0.79 / 0.21 | 0.97 / 0.03 |
| comp51626_c0_seq1 (1245) | C / T | 0.59 / 0.41 | 0.83 / 0.17 | 0.76 / 0.24 | 0.91 / 0.09 |
| comp52601_c0_seq1 (205) | A / G | 0.28 / 0.72 | 0.06 / 0.94 | 0.13 / 0.87 | 0.07 / 0.93 |
| comp58030_c0_seq1 (550) | A / G | 0.08 / 0.92 | 0.34 / 0.66 | 0.26 / 0.74 | 0.2 / 0.8 |
| comp66672_c0_seq1 (661) | C / G | 0.86 / 0.14 | 0.62 / 0.38 | 0.69 / 0.31 | 0.29 / 0.71 |
| 0_4394_01_1 (522) | G / A | 0.39 / 0.61 | 0.28 / 0.72 | 0.32 / 0.68 | 0.01 / 0.99 |
| CL12234_1_1 (626) | A / G | 0.01 / 0.99 | 0.01 / 0.99 | 0.01 / 0.99 | 0.33 / 0.67 |

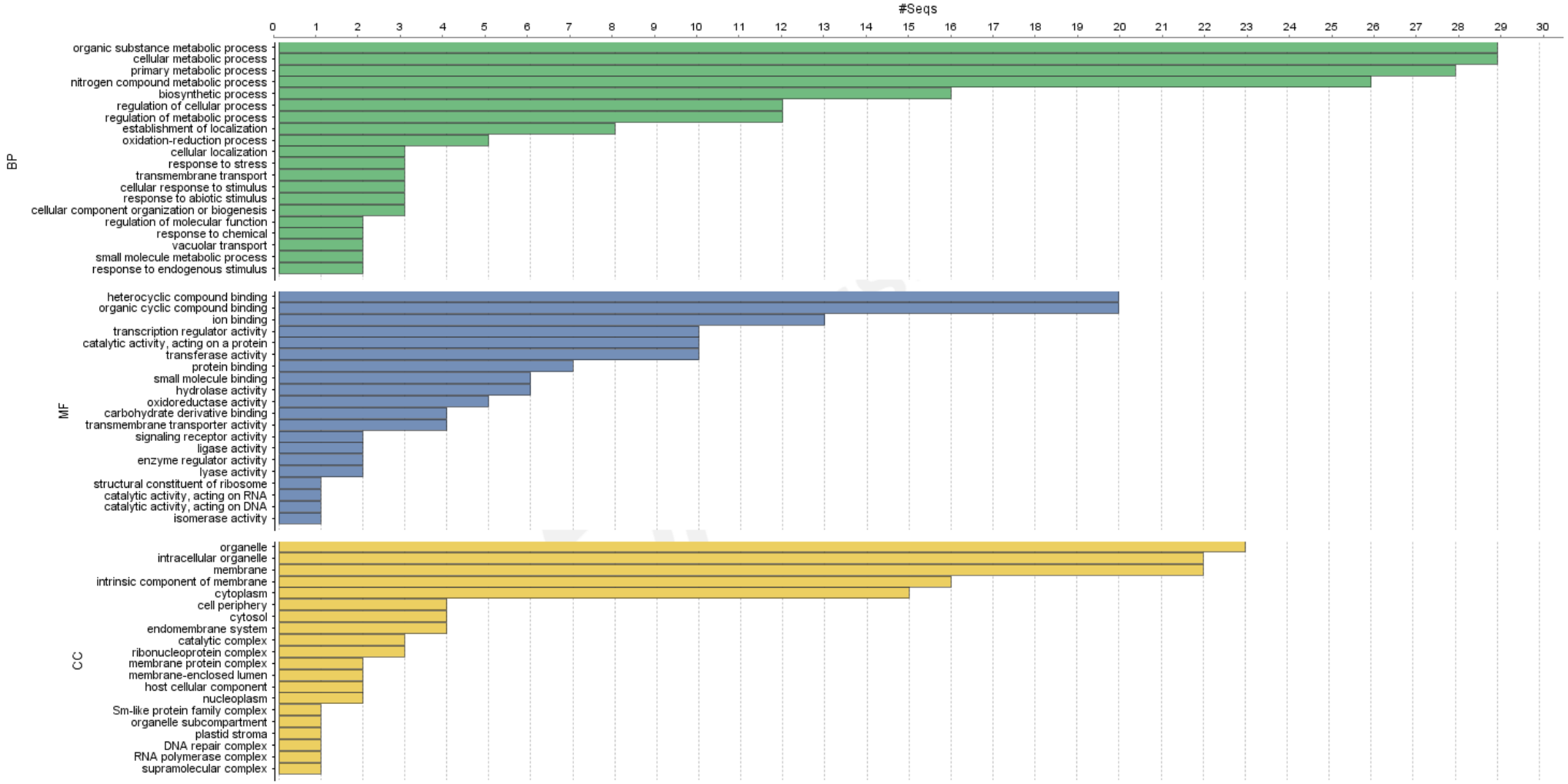



**Figure S7.** Distribution of top 20 third-level gene ontologies in each of the three main GO categories annotated for 64 sequences in which outlier SNPs were detected between dwarf and Pyrenean pines; BP - Biological Processes (annotated for 48 sequences), MF - Molecular Functions (50 sequences), CC - Cellular Components (38 sequences).

## 3.3. PUBLIKACJA TRZECIA

Julia Zaborowska, Annika Perry, Stephen Cavers & Witold Wachowiak

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

Journal of Systematics and Evolution [in press]

https://doi.org/10.1111/jse.12896

2022

JSE Journal of Systematics and Evolution

doi: 10.1111/jse.12896

Original Article

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

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

[1]Institute of Environmental Biology, Adam Mickiewicz University in Poznan, Uniwersytetu Poznanskiego 6, Poznan 61-614, Poland
[2]UK Centre for Ecology and Hydrology, Edinburgh Site, Bush Estate, Penicuik EH26 0QB, UK
*Author for correspondence.
E-mail: julia.zaborowska@amu.edu.pl.
Received 6 April 2022; Revised 21 May 2022; Accepted 26 May 2022; Article first published online 01 June 2022

**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*, *Pinus uncinata*, and *Pinus 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 *P. 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, the 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 selection within genomic regions, in the process of ecological divergence and adaptation to higher altitudes in pine taxa.

**Key words:** differential gene expression, high-altitude adaptations, mountain pines, *Pinus mugo* complex, transcriptome assembly.

## 1 Introduction

Plant growth, fitness, and survival are 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 in 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 variations have 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 expression patterns that may play important role in the 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, *Pinus uncinata* Ramond, and *Pinus sylvestris* L.—are morphologically diverged and adapted to different habitats, enforced by the disjunction of their geographical ranges and isolation during the Pleistocene glaciation (Christensen, 1987). *P. mugo*, the dwarf mountain pine, and *P. uncinata*, the Pyrenean pine, are sister taxa related to the 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 (Hamerník & Musil, 2007). Pyrenean pine is a typical coniferous tree (grows straight and may reach up to 25 m), which occurs from the Alps westward—in the Massif Central, Jura and Vosges, Pyrenees, and a few remote populations inside the Iberian Peninsula (Jalas & Suominen, 1973). Although they are distinguished by a number of other, more subtle traits, for example, phenology, cone size and shape, needle characteristics, and composition of volatiles and allozymes (please, see Table S1; Lewandowski et al., 2000; Monteleone et al., 2006; Boratyńska & Boratyński, 2007; Boratyńska et al., 2015; Wachowiak et al., 2018; Adams & Tashev, 2019), they are often aggregated under the *P. mugo* complex together with a few less defined groups (Christensen, 1987; Hamerník & Musil, 2007). For simplicity reasons, and following Businský & 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. The range of this pine spreads from the Mediterranean climate on the 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 (EUFORGEN, 2009). As a result of its large distribution, the species demonstrates high phenotypic differentiation with dozens of ecotypes described (Kew and Missouri Botanical Gardens, 2013). Pines from the *P. mugo* complex and *P. sylvestris* diverged about 5 million years BP (Wachowiak et al., 2011; Łabiszak & Wachowiak, 2021). However, they are genetically similar, and previous studies showed little differentiation between the taxa at karyotype (Bogunić et al., 2011), mitochondrial and plastid genomes (Heuertz et al., 2010; Dzialuk et al., 2017; Zaborowska et al., 2019; Sokołowska et al., 2021), 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 (Wachowiak et al., 2000; Kormut'ák et al., 2005; Żukowska & Wachowiak, 2017; Zaborowska et al., 2019). This close genetic similarity between the species is usually explained by relatively recent speciation in presence of gene flow (Christensen, 1987; Monteleone et al., 2006; Jasińska et al., 2010) and segregation of ancestral variation (Wachowiak et al., 2011, 2013). Comparative transcriptome analysis of these pines (Wachowiak et al., 2015) revealed thousands of polymorphic markers that were used for the development of SNP genotyping array (Perry et al., 2020), 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 the 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 a longer period of common history and a 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.

## 2 Material and Methods

### 2.1 Materials

Genetic material was obtained from needles of 11 two-year-old seedlings of three pine species (*Pinus mugo*—four samples, M; *Pinus uncinata*—three samples, UN; *Pinus sylvestris* L.—four 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,

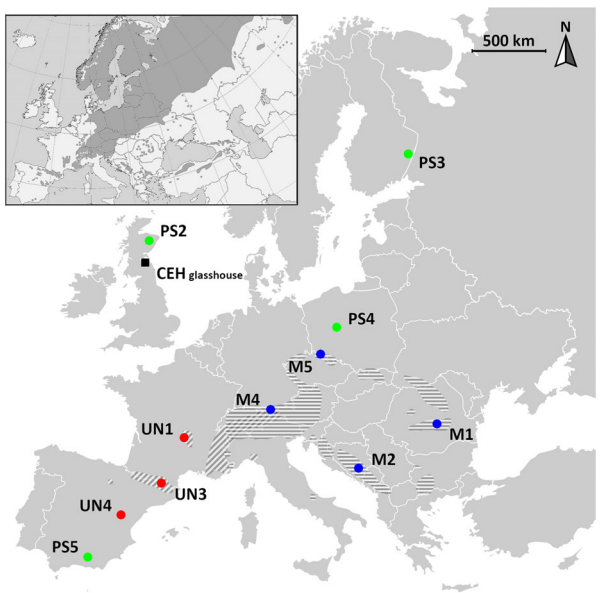



**Fig. 1.** Geographic distribution and sample collection sites of three investigated pine species. Blue dots and horizontal shading represent *Pinus mugo* (M) sites and range; red dots and diagonal shading correspond to *Pinus uncinata* (UN) locations; green points and dark shading on the inset represent *Pinus sylvestris* (PS) sample locations and the European part of its range. Black square marks the 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, accessed 4 October 2018) and ranges taken from Jalas and Suominen (1973). The inset map was prepared by EUFORGEN (2009) and adapted by the authors.

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

| Sp.[†] | Sample, ENA Acc.[‡] | Country code, population[§] | Latitude | Longitude | Alt.[¶] |
|---|---|---|---|---|---|
| *Pinus 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 |
| *Pinus 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 |
| | UN4, SAMEA2672724 | ES, Sierra de Gudar | 40°28′49″ N | 0°41′51″ W | 2000 |
| *Pinus 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 |

[†]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.

see Wachowiak et al., 2018). Raw paired-end Illumina reads from teams' previous RNA-sequencing attempts were used (the data are 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 (Andrews, 2021) 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. Afterward, reads were verified with FASTQC again. The resulting transcriptome, hereafter *MUS assembly*, was adopted as the reference in all subsequent examinations. 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 the 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 the TRINITY abundance estimation protocol (Haas, 2021). 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 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 indicates 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 (Haas, 2021). Transcripts that did not pass further expression filtering (FPKM > 0.5 for average across samples) were discarded.

### 2.3 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 a 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), 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, that is, a minimum 4-fold change in gene expression and *p*-value cutoff for false discovery rate set at 0.001 were thresholds for a significant result. Outputs from different software were compared and only the 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 the direct outcome from M-UN vs PS test ("board"), and its subset significant in three 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).

### 2.4 Functional annotation of the 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 steps 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 (Haas, 2018) using the following arguments: "runP = true," "transdecoder-m = 100," "complete = false" and "transdecoder-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 an *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 "contam = bacteria,opisthokonta" and "uninformative = conserved, predicted, unknown, unnamed, hypothetical, putative, unidentified, uncharacterized, uncultured, uninformative"); (4) Sequence functional analysis, the assignments of protein domains and gene ontology (GO) terms were performed with the EGGNOG-MAPPER v2.1.5 and eggNOG database v5.0 (Huerta-Cepas et al., 2019; Cantalapiedra et al., 2021). Single isoform was chosen to represent the predicted gene and the longest contig was chosen if there was more than one annotated. In the 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 script run_GOseq.pl from the 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 the first value might be exaggerated due to the generation of surplus isoforms during assembly, especially in longer transcripts. The new *MUS assembly* (submitted to the Dryad database under DOI: https://doi.org/10.5061/dryad.xsj3tx9j1), which was compared to the expected plant genes of two BUSCO collections, represented nearly complete sets of those sequences. In the case of the 425 "core" genes of Viridiplantae, 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 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 1939 gene models, was 1577 bp, while E90N50, measured on 17 830 gene models, was 2173 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). Exactly 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 — the 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: 7979, internal sequences: 9374). 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 the 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 orthologs 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, a total number of 43 573 isoforms, corresponding to 15 084 coding gene models (hereafter named genes), was left for final annotation steps in EGGNOG.

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 a 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, 9816 were unique, inclusive of 6163 in the biological process aspect, 2662 molecular functions, and 991 cellular components. Level 4 GO terms, which might be considered a 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. Among 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.

### 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 the 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 upregulated in the first plus 18 in the 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, 5). Nearly half of the gene models discriminating both mountain pines from Scots pine were confirmed in the "narrow" subset (Fig. 6). ORFs 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 gene models 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 in this subset (respectively: 85 and 24 unique), and 6433 GO terms were assigned: 3882 (1179 unique) in the category of biological processes, 1127

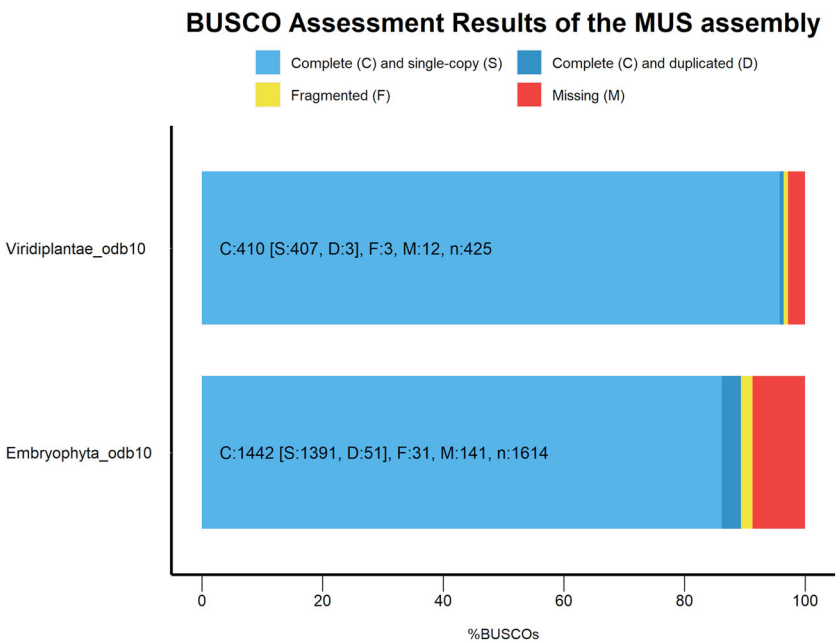



**Fig. 2.** Representation of 23 425 protein-coding genes from *MUS assembly* transcriptome compared to two sets of the BUSCOs—present in green plants (upper panel) and in land plants (lower panel). Only one isoform, the longest, per gene was involved in the test.

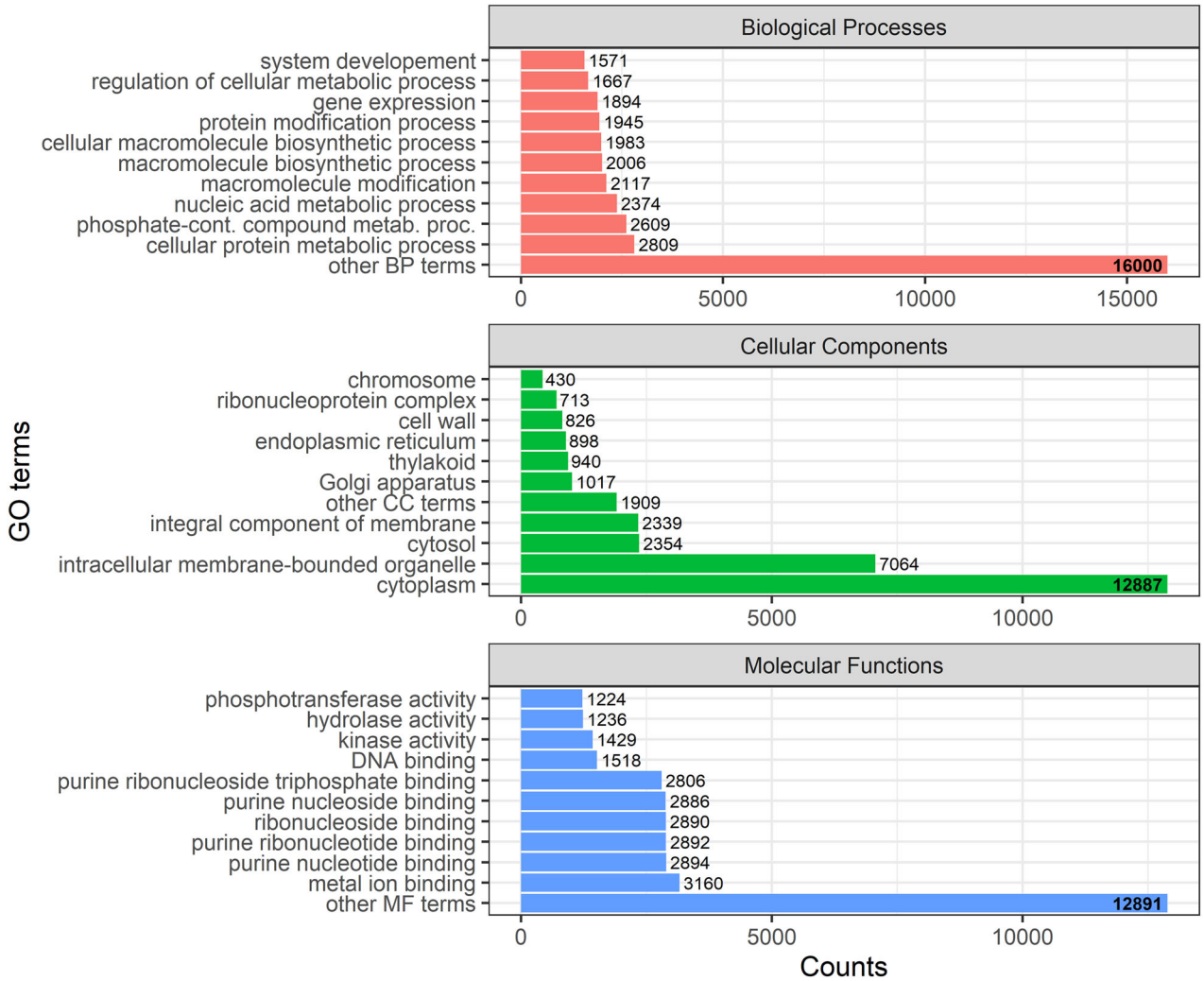



**Fig. 3.** The most frequently reported terms of level 4 in each main gene ontology (GO) aspect, as found in the annotated isoforms of the *MUS assembly*.

(244 unique) molecular functions, and 1424 (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 the 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 the 2OG-FE(II) oxygenase family, cis-zetain O-glucosylotransferase, protein disulfide isomerase, elongation factor, cullin 1 and photosystem II 10 kDa polypeptide. All of these were similarly upregulated in comparison to dwarf mountain pine vs Scots pine, while in the case of *Pinus uncinata*, the expression of the three last molecules did not differ from *Pinus sylvestris* levels. In the long list of DEGs differentiating dwarf from Scots pine, besides the few mentioned above, 30 others could be characterized (Tables S9, S11). There was one more O-metylotransferase, two additional peptides from the 2OG-Fe(II) oxygenase family, another component of photosystem II and an 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 3-beta hydroxysteroid dehydrogenase/isomerase activity (two genes; for the 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, although there were some genes concurrently and significantly upregulated in all three tests involving this taxon, none was annotated. Two aldehyde dehydrogenases were

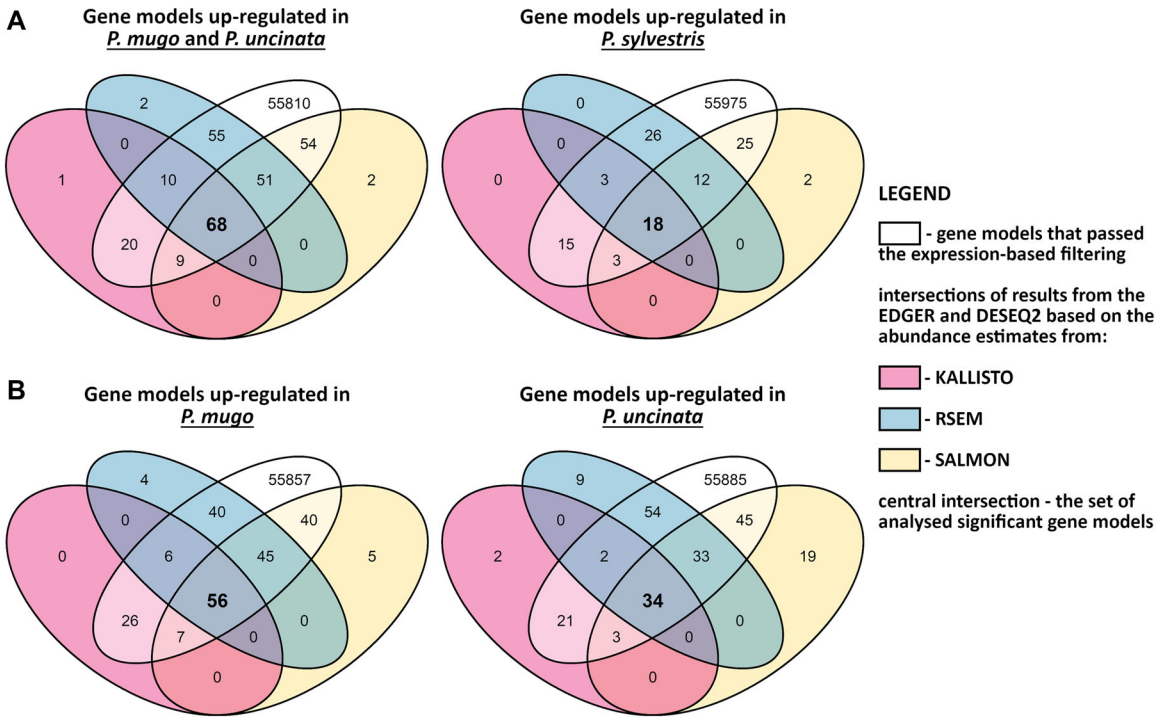



**Fig. 4.** Venn diagrams of results from different analysis tools used in search of differentially expressed gene models in the *MUS assembly*. **A,** Discriminating markers found between two mountain pines and Scots pine ("broad" M-UN vs PS comparison); **B,** Discriminating markers found between dwarf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

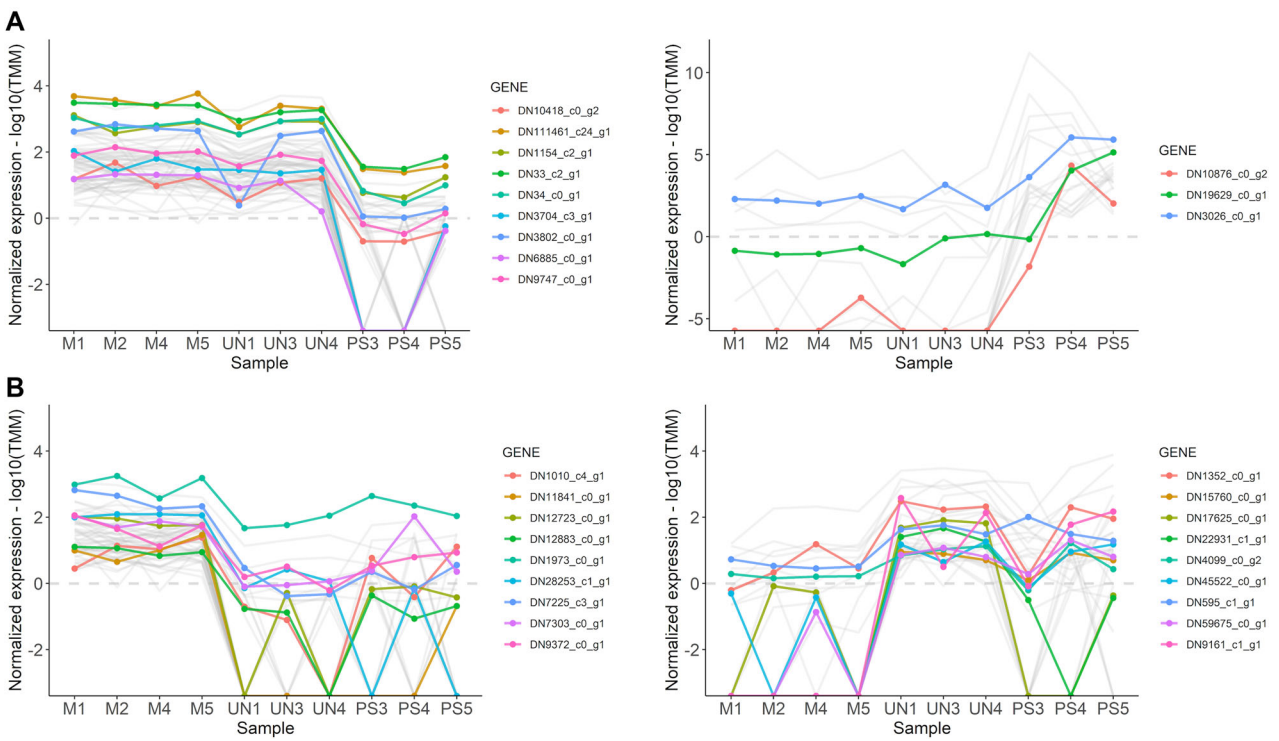



**Fig. 5.** Expression profiles of differentially expressed gene models, the successfully annotated ones are marked with colors. **A,** Significant markers found in M-UN vs PS pair, upregulated in both mountain pines (left) or in Scots pine (right); **B,** Markers significantly differentiating two mountain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

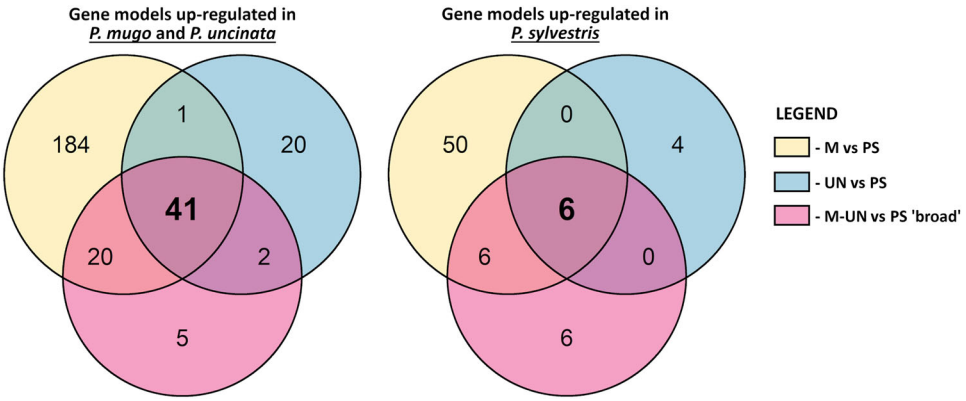



**Fig. 6.** Venn diagrams of results from three complementary comparisons of differentially expressed genes between two mountain pine taxa and Scots pine. The central intersection on left corresponds with the "narrow" M-UN vs PS set.

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. Fourteen more annotated DEGs distinguished Scots pine from the shrubby species (Table S11); besides 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*.

Among genes with satisfying annotation and variant expression between two focal mountain plants, we found nine sequences upregulated in dwarf pine and other nine 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 for 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 upregulated, at the same time, two coding for aldose 1-epimerase and for cysteine-rich receptor-like protein kinase were downregulated 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.

# 4 Discussion

## 4.1 New reference transcriptome

We examined differentiation in the gene expression between two mountain species from *Pinus mugo* complex and *Pinus 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 (Parchman et al., 2010; Pinosio et al., 2014; Duran et al., 2019). On the other hand, at 14 666, the number of 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., 2017; Proost et al., 2015). 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 (Cañas et al., 2017). Since the *MUS assembly* 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 *Pinus 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 a controlled glasshouse environment and uniform

**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 |
|---|---|---|
| *Upregulated in both mountain pines in relation to Scots pine* | | |
| 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_c0_g1 i1 | XP_024928927.1 *Ziziphus jujube* | O-methyltransferase |
| 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 |
| *Upregulated in Scots pine in relation to dwarf and Pyrenean pines* | | |
| DN10876_c0_g2 i1 | tr\|A9NV57_PICSI *Picea sitchensis* | *ALDH1A2*, aldehyde dehydrogenase |
| DN19629_c0_g1 i1 | XP_023522297.1 *Cucurbita pepo* | *ALDH11A3*, aldehyde dehydrogenase |
| DN3026_c0_g1 i2 | sp\|C72B1_PINTA *Pinus taeda* | Abietadienol/abietadienal oxidase |
| *Upregulated in dwarf pine in relation to Pyrenean pine* | | |
| 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 |
| 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 |
| *Upregulated in Pyrenean pine in relation to dwarf pine* | | |
| 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 |
| 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 |

Continued

**Table 2** Continued

| Gene, query isoform | Best hit sequence, species of origin | Predicted gene and/or function description |
|---|---|---|
| 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 i1 | tr\| A0A7J7C285_TRIWF *Tripterygium wilfordii* | *HSP90-1*, heat shock protein |

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 patterns showing similar 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), while triple that number was observed in the 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) $_{U}$ (U vs PS): 14) was not much larger than the numbers of DEGs specific to each of them ((M vs PS) $_{U}$ (M vs UN): 8 and (U vs PS) $_{U}$ (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. These results contrast with the pattern observed in a broad set of genome-wide SNP 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.

### 4.3 Putative signatures of mountain pine adaptations

The 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 upregulated in the *P. mugo* complex taxa (Table 2), and the O-methyltransferase activity was one of the main GO terms enriched in this 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 the 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* (González-Martínez et al., 2006; Wachowiak et al., 2022). 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 (Paasela et al., 2017; Lim et al., 2021), so we did not expect *COMT* expression in its close relatives. Another OMT was highly differentiated among *P. mugo* and *P. sylvestris* but showed 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, a 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 (Bičárová et al., 2019; Diaz-de-Quijano et al., 2019).

Another gene upregulated in mountain pines as compared to Scots pine encodes oxidoreductase from the

2-oxogluterate (2OG) Fe(II)-dependent oxygenase superfamily, which exhibits great diversity of metabolite biosynthesis, biodegradation, regulatory, or structural roles (Herr & Hausinger, 2018). Another transcript represented the nuclear *PsbR* gene for plastid photosystem II (PSII) 10 kDa polypeptide. The polypeptide is essential for the 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 UV-B radiation (Peng et al., 2021), and to have expression negatively correlated with the sun, temperature, and wind, but positively with precipitation and humidity (Sjödin et al., 2008). As the light intensity and UV-B radiation are greater at higher altitudes, upregulation in both mountain pines could reflect their adaptation and reaction to changed glasshouse conditions. Enrichment analysis indicated the regulation of circadian rhythm as an overrepresented term in *P. mugo* and *P. uncinata*, meaning that these plants might evolve to better synchronize with the light cycle, or other diurnal cycles altered in their environment. Another DEG encoded protein disulfide isomerase (*PDI*), which catalyzes the conversion of thiol-disulfide by formation and breakage of SS-bonds between cysteine residues, is responsible for the 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 (Zhang et al., 2018; Feldeverd et al., 2020), and may play a role in redox signaling (Wittenberg & Danon, 2008). An elongation factor was distinctly upregulated in the subalpine taxa too (better specified below).

Also strongly differentiated, but exhibiting the opposite direction of regulation (i.e., greater expression in Scots pine) was abietadienol/abiedienal oxidase, a cytochrome P450 monooxygenase unique to conifers. It is specialized in the synthesis of tricyclic diterpene resin acids (DRAs) (Ro et al., 2005; Bathe & Tissier, 2019), essential components of conifers' oleoresin—the defense blend against pathogens, pests and herbivores, and a chemotaxonomic marker effective in *Pinus* (Mitić 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.

### 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 the properties of fatty acids (FAs) chains by forming double bonds between adjacent carbon atoms. They are responsible for the 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 a 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 as 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 toward 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 others facilitating the elongation of synthesized peptides (Sasikumar et al., 2012). These specific two are related to the 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, which is related to the gibberellin metabolic processes (Li et al., 2018), and these hormones are known to affect 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 the adaptation of Pyrenean pine to the 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, and one cysteine-rich receptor-like protein kinase) were also found upregulated in *P. unicnata*. They belong to a big group of signaling, transmembrane proteins, which respond to diverse developmental and environmental prompts (Wrzaczek et al., 2010; Vaattovaara et al., 2019; Mou et al., 2021). Another differentially expressed transcript encoded a stem-specific protein that putative orthologs 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 the 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 a larger sample of populations, would be supportive. Finally, a complementary test, which would confirm how

actual the revealed expression patterns are, could be governed by individuals collected in their native conditions, across natural populations of the three pine taxa.

## 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 *Pinus 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 non-coding part of the genome, and epigenetic interactions.

## Acknowledgments

The authors would like to express their gratitude to institutions that 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.

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

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

Bičárová S, Sitková Z, Pavlendová H, Fleischer P Jr, Fleischer P Sr, Bytnerowicz A. 2019. The role of environmental factors in ozone uptake of *Pinus mugo* Turra. *Atmospheric Pollution Research* 10: 283–293.

Bogunić F, Siljak-Yakovlev S, Muratovic E, Pustahija F, Medjedović 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.

Boratyńska K, Boratyński 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.

Boratyńska K, Jasińska AK, Boratyński 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, 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.

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

Cañas RA, Li Z, Pascual MB, Castro-Rodríguez V, Ávila C, Sterck L, Van de Peer Y, Cánovas FM. 2017. The gene expression landscape of pine seedling tissues. *The Plant Journal* 91: 1064–1087.

Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021. eggNOG-mapper v2: Functional annotation, orthology assignments, and prediction at the metagenomic scale. *Molecular Biology and Evolution* 38: 5825–5829.

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.

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* CCoAOMT1 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 Á, Peñuelas 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.

Durn 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, Boratyńska K, Romo A, Boratyski A. 2017. Taxonomic and geographic variation of the *Pinus mugo* complex on chloroplast microsatellite markers. *Systematics and Biodiversity* 15: 464–479.

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

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: 610052.

Ford BA, Foo E, Sharwood R, Karafiatova M, Vrána J, MacMillan C, Nichols DS, Steuernagel B, Uauy C, Doležel 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.

Gonzalez-Ibeas D, Martinez-Garcia PJ, Famula RA, Delfino-Mix A, Stevens KA, Loopstra CA, Langley CH, Neale DB, Wegrzyn JL. 2016. Assessing the gene content of the megagenome: Sugar pine (*Pinus lambertiana*). *G3 Genes|Genomes|Genetics* 6: 3787–3802.

González-Martínez 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 B [online]. 2018. TransDecoder. Available from http://www.github.com/TransDecoder/TransDecoder/wiki [accessed 22 October 2018].

Haas B [online]. 2021. Trinity Transcript Quantification. Avaliable from http://www.github.com/trinityrnaseq/trinityrnaseq/wiki [accessed 20 April 2021].

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.

Hamerník 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.

Heuertz M, Teufel J, González-Martínez SC, Soto A, Fady B, Alía 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: 1473.

Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, 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.

Jasińska AK, Wachowiak W, Muchewicz E, Boratyska K, Montserrat JM, Boratyński A. 2010. Cryptic hybrids between *Pinus uncinata* and *P. sylvestris*. *Botanical Journal of the Linnean Society* 163: 473–485.

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

Kew and Missouri Botanical Gardes [online]. 2013. The Plant List: The Plant List: *Pinus sylvestris*. Version 1.1. Available from https://www.theplantlist.org [accessed 10 March 2021].

Kormut'ák A, Ostrolucká M, Vooková B, Pret'ová A, Fečková M. 2005. Artificial hybridization of *Pinus sylvestris* L. and *Pinus mugo* Turra. *Acta Biologica Cracoviensia Series 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.

Łabiszak 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, Glémin S, Savolainen O. 2016. *Local adaptation in plants*. 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. 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, Venäläinen M, Paulin L, Auvinen P, Kärkkäinen 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.

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, Simão 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*. Heidelberg: Springer. 64.

Mitic ZS, Jovanović SČ, Zlatković BK, Nikolić BM, Stojanović 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.

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, Martínez-García 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, Hällman J, Keech O, Klasson L, Koriabine M, Kucukoglu M, Käller M, Luthman J, Lysholm F, Niittylä T, Olson A, Rilakovic N, Ritland C, Rosselló JA, Sena J, Svensson T, Talavera-López C, Theißen G, Tuominen H, Vanneste K, Wu 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, 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, Pietiäinen 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.

Peng T, Wang Y, Yang T, Wang F, Luo J, Zhang Y. 2021. Physiological and biochemical responses, and comparative transcriptome profiling of two *Angelica sinensis* cultivars under enhanced ultraviolet-B radiation. *Frontiers in Plant Science* 12: 805407.

Perry A, Wachowiak W, Beaton J, Iason G, Cottrell J, Cavers S. 2022. Identifying and testing marker–trait associations for growth and phenology in three pine species: Implications for genomic prediction. *Evolutionary Applications* 15: 330–348.

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, González-Martínez SC, Bagnoli F, Cattonaro F, Grivet D, Marroni F, Lorenzo Z, Pausas JG, Verdú 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.

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

Proost S, VanBel M, Vaneechoutte D, Van de Peer Y, Inzé D, Mueller-Roeber B, Vandepoele K. 2015. PLAZA 3.0: An access point for plant comparative genomics. *Nucleic Acids Research* 43: D974–D981.

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.

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

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

Suorsa M, Sirpiö 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: 145–150.

Supek F, Bošnjak M, Škunca N, Šmuc T. 2011. REVIGO summarizes and visualizes long lists of Gene Ontology terms. *PLoS One* 6: e21800.

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

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

Wachowiak W, Boratyńska 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, Leśniewicz K, Odrzykoski IJ, Augustyniak H, Prus-Głowacki 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, Palmé 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.

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, González-Martínez SC, Cavers S. 2022. Admixture and selection patterns across the European distribution of Scots pine, *Pinus sylvestris* (Pinaceae). *Botanical Journal of the Linnean Society*, boac016.

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

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, Pédrono F, Pasquier E, Deluc LG, Marpeau AM, Aitzetmüller K. 2001. Fatty acid composition of Pinaceae as taxonomic markers. *Lipids* 36: 439–451.

Wrzaczek M, Brosché M, Salojärvi J, Kangasjärvi S, Idänheimo N, Mersmann S, Robatzek S, Karpiński S, Karpińska B, Kangasjärvi 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, Łabiszak 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, Łabiszak 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 *Eucalyptus* 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: 913.

Żukowska 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.

## Supplementary Material

The following supplementary material is available online for this article at http://onlinelibrary.wiley.com/doi/10.1111/jse.12896/suppinfo:

**Table S1.** Generalized phenotypic differentiation of the three analyzed pine species.

**Table S2.** Read abundances and metrics of sample mapping success to the reference *MUS assembly* transcriptome sequence.

**Table S3.** Basic statistics for the generated raw *MUS assembly* transcriptome and its expression-filtered subset.

**Table S4.** Results of *MUS assembly* isoform frame prediction and statistics for the resulting filtered transcriptome.

**Table S5.** Results of similarity search of translated *MUS assembly* isoforms against three protein databases.

**Table S6.** Similarity search results, most frequent species in best-hit, and contaminant alignments are listed.

**Table S7.** The most frequently reported protein domains form PFAM and SMART databases in filtered *MUS assembly*.

**Table S8.** The most frequently reported terms of level 3-5 in three main gene ontology (GO) aspects from the *MUS assembly*.

**Table S9.** Abundance estimations of 121 protein-coding genes with significantly differentiated expression patterns between three studied taxa. Markers significant in all six combinations of software dyads are presented, for simplification, solely fold change and *p*-values from the RSEM-EDGER pair are shown. Marked in bold are 74 successfully annotated sequences.

**Table S10.** Functional annotation - gene ontology (GO) and protein domains - of genes differentially expressed between mountain pine species and Scots pine, and between dwarf pine and Pyrenean pine. Only GO terms from the highest, most specific level of annotation are presented.

**Table S11.** Functional annotation - gene ontology (GO) and protein domains - of genes differentially expressed between each of the two mountain pine species and Scots pine. In each category, the most specific annotated GO terms are shown.

**Table S12.** Gene ontology (GO) terms represent enriched functional categories in sets of genes differentially expressed between analyzed pine taxa.

**Fig. S1.** ExN50 change plotted against the Ex value, comparison of outputs from three read abundance estimation methods used on *MUS assembly* gene models.

**Fig. S2.** Scatterplot of the GO terms from the Biological Process aspect enriched in differentially expressed genes upregulated in two mountain pines relative to Scots pine. Bubble color indicates the *p*-value for overrepresentation, size indicates the frequency of particular GO term in the underlying GO database (more general terms are larger), and group representatives are marked with blue asterisks.

**Fig. S3.** Scatterplot of the GO terms from the Biological Process aspect enriched in differentially expressed genes upregulated in Scots pine relative to two mountain pines.

**Fig. S4.** Scatterplot of the GO terms from the Biological Process aspect enriched in differentially expressed genes up-regulated in *P. mugo* compared to *P. uncinata*.

**Fig. S5.** Scatterplot of the GO terms from the Biological Process aspect enriched in differentially expressed genes upregulated in *P. uncinata* relative to *P. mugo*.

**Table S1.** Generalized phenotypic differentiation of the three analyzed pine species.

| Metric | *P. mugo* | *P. uncinata* | *P. sylvestris* | REF. [†] |
|---|---|---|---|---|
| Growth form | dwarfish – prostrate, polycormic shrub | monocormic, erect tree | monocormic, erect tree | 1, 2, 3 |
| Bark color and character | blackich or greyish brown, minutely scaly | blackich or greyish brown, with small plates curling up like shavings | orange-brown on branches and upper trunk, papery thin, flaking, later and on lower trunk forming thick grey ribs | 1 |
| Female cone characteristics: position; symmetry; color of fully matured cones and their shape on the exposed part of the cone | erect position; actinomorphic or subactinomorphic symmetry; apophyses glossy, light to chestnut brown, flat, slightly prominent, distally tumid or obliquely low pyramidal | erect position; conspicuously zygomorphic; rarely subactino-morphic symmetry; apophyses glossy, light to chestnut brown, elongated into high, oblique pyramid, rarely only slightly prominent | reflexed position; ± zygomorphic or subactinomorphic symmetry; apophyses dull, pale tawny with greyish tinge, slightly prominent to obliquely pyramidal in shape | 1, 3 |
| Needle characteristics: color; epidermis cells in cross section; sclerenchyma cells along and between vascular bundles; distance between vascular bundles; proportion of different sclerenchyma cell types (A-C – surrounding resin canals; AA-DD – between vascular bundles)[‡] | dark green; radially oblong; in thin layer or (almost) absent; 0.5-1.2 x vascular bundle width; small proportion of A type (8%), B and C types close to even in cell types surrounding resin canals (43% and 49%); Between the vascular bands predomination of the DD type (65%) and very small contribution of the AA type (< 2%) | dark green; radially oblong; in thin layer or (almost) absent; 0.6-1.4 x vascular bundle width; predomination of the B type among sclerenchyma cells surrounding resin canals, followed by A type (34%) and only few percent of C type; Among cells between vascular bundles mainly the CC and BB types present (43% and 31%, respectively), only minor component of DD type cells | greyish pale green; ± square; in multicell bundle of 1 triangular shape on cross section; 1.5-2.5 x vascular bundle width; the A type predominant (78%) among the cell types surrounding resin canals, almost no C type cells; The AA type (72%) most common between vascular bundles, the rest quite evenly distributed (BB - 13%, CC – 8%, DD - 7%) | 1, 4, 5 |
| Phenology and early growth rate (compared under glasshouse conditions) | earliest bud set, first to terminate dormancy, slowest growth rate | latest bud burst | fastest growth | 6 |

**Footnote**: [†] REF. – for more details, please see the publications: 1 – Businsky & Kirschner, 2010; 2 – Hamernik & Musil, 2007; 3 – Monteleone et al., 2006; 4 – Boratynska & Boratynski, 2007; 5 – Boratynska et al., 2015; 6 – Wachowiak et al., 2018; [‡] The cell type descriptions: A - fibre-like with restricted lumen, B - with thickened walls and intermediate lumen, C - with slightly thickened walls and distinct lumen, AA - fibre-like with highly thickened walls and restricted lumen, BB - with thickened walls and intermediate lumen, CC - with slightly thickened walls and distinct lumen, DD - with non-thickened walls and large lumen (after Boratynska & Boratynski, 2007).

**Table S2.** Read abundances and metrics of sample mapping success to the reference *MUS assembly* transcriptome sequence.

| Sample | Raw reads[†] | Cleaned reads[‡] (% of raw) | Duplicate reads (% of cleaned) | Unmapped (% of cleaned) | Mapped in proper pairs (% of cleaned) |
|---|---|---|---|---|---|
| **M1** | 21,907,358 | 20,903,206 (95.4) | 2,627,912 (12.6) | 1,568,764 (7.5) | 19,334,442 (92.5) |
| **M2** | 16,338,002 | 15,608,960 (95.5) | 2,029,470 (13.0) | 1,135,198 (7.3) | 14,473,762 (92.7) |
| **M4** | 26,548,182 | 25,346,548 (95.5) | 2,873,030 (11.3) | 2,174,986 (8.6) | 23,171,562 (91.4) |
| **M5** | 34,792,306 | 33,172,082 (95.3) | 2,627,912 (7.9) | 2,644,350 (8.0) | 30,527,732 (92.0) |
| **UN1** | 17,483,256 | 16,690,914 (95.5) | 1,253,694 (7.5) | 1,397,014 (8.4) | 15,293,900 (91.6) |
| **UN3** | 22,992,986 | 22,046,538 (95.9) | 2,537,826 (11.5) | 1,746,326 (7.9) | 20,086,382 (91.1) |
| **UN4** | 24,321,780 | 23,167,108 (95.3) | 2,537,826 (11.0) | 1,746,326 (7.5) | 21,420,782 (92.5) |
| **PS3** | 18,942,346 | 18,171,568 (95.9) | 3,529,774 (19.4) | 1,689,838 (9.3) | 16,481,730 (90.7) |
| **PS4** | 19,513,846 | 18,656,774 (95.6) | 1,735,108 (9.3) | 1,636,910 (8.8) | 17,019,864 (91.2) |
| **PS5** | 22,588,372 | 21,566,374 (95.5) | 1,971,162 (9.1) | 1,656,942 (7.7) | 19,909,432 (92.3) |
| **ALL** | 225,428,434 | 215,330,072 (95.5) | na[§] | 17,396,654 (8.1) | 197,719,588 (91.8) |
| **PS2**[¶] | 316,681,976 | 303,652,976 (95.9) | na | na | na |

**Footnote**: [†] Reads as might be downloaded from European Nucleotide Archive (sample IDs given in Table 1 of the main text); [‡] Reads after TRIMMOMATIC cleaning, only pairs left (trimming parameters used: "ILLUMINACLIP:anaconda3/envs/Trinity/share/adapters/TruSeq2-PE.fa:2:30:10:2:keepBothReads SLIDINGWINDOW:5:20 LEADING:3 TRAILING:3 MINLEN:30"; [§] Mapping was performed for each sample separately, duplicate reads could not be evaluated; [¶] Sample PS2 served only for the generation of the reference transcriptome; na – not applicable.

**Table S3.** Basic statistics for the generated raw *MUS assembly* transcriptome and its expression-filtered subset.

| Metric | Raw *MUS assembly* | Expression-filtered set |
|---|---|---|
| Number of contigs | 371,779 | 159,919 (43%) |
| Mean contig length [bp] | 716.8 | 1156.0 |
| Max. contig length [bp] | 24,660 | 24,660 |
| Total bases in contigs | 266,480,155 | 184,867,893 |
| GC content [%] | 42.17 | 42.13 |
| N50[†] | 1,290 | 1,955 |
| E90N50[‡] | 2,173 | 2,135 |
| Number of genes[§] | 241,804 | 56,077 |

**Footnote**: [†] N50 – the shortest length of contig that needs to be included for covering 50% of the transcriptome; [‡] E90N50 - same as N50 length but calculated only from the top most highly expressed contigs that represent 90% of the normalized expression data, here based on expression estimates from RSEM; [§] The number of gene models based on the initial TRINITY clustering of isoforms.

**Table S4.** Results of *MUS assembly* isoform frame prediction and statistics for the resulting filtered transcriptome.

| **Isoforms that passed expression filtering** | **159,919 (100%)** |
|---|---|
| Total sequences with predicted frame | 73,339 (45.9%) |
| - complete frames | 44,874 (28.1%) |
| - 5' partial frames | 11,112 (6.9%) |
| - 3' partial frames | 7,979 (5.0%) |
| - internal frames | 9,374 (5.9%) |
| Total length of filtered transcriptome [bp] | 83,393,445 |
| Average length [bp] | 1,137.1 |
| N50; N90 [bp] | 1,482; 522 |
| Longest sequence [bp] | 17,004 (TRINITY_DN7617_c0_g1_i3) |
| Shortest sequence [bp] | 327 (TRINITY_DN8003_c0_g1_i6) |

**Table S5.** Results of similarity search of translated *MUS assembly* isoforms against three protein databases.

| **Database** | **RefSeq plant proteins** | **Swiss-Prot** | **TrEMBL** | **OVERALL** |
|---|---|---|---|---|
| Total alignments | 1,123,311 | 71,501 | 614,855 | 1,809,667 |
| Total unique transcripts with an alignment | 55,358 | 34,387 | 59,096 | 59,368 |
| (pct. of total sequences with predicted frame) | (75.5%) | (46.9%) | (80.6%) | (81.0%) |
| Total unique informative alignments | 41,300 | 32,664 | 40,825 | 43,973 |
| (pct. of total transcripts with an alignment) | (74.6%) | (95.0%) | (69.1%) | (74.1%) |
| - complete frames | 27,803 | 25,100 | 27,100 | 28,963 |
| - 5' partial frames | 5,252 | 3,410 | 5,358 | 5,697 |
| - 3' partial frames | 4,208 | 2,581 | 4,125 | 4,608 |
| - internal frames | 4,037 | 1,573 | 4,242 | 4,705 |
| Total unique uninformative alignments | 14,058 | 1,723 | 18,271 | 15,395 |
| (pct. of total transcripts with an alignment) | (25.4%) | (5.0%) | (30.9%) | (25.9%) |
| - complete frames | 9,215 | 1,005 | 11,573 | 9,829 |
| - 5' partial frames | 1,695 | 247 | 2,385 | 2,104 |
| - 3' partial frames | 1,457 | 225 | 2,060 | 1,616 |
| - internal frames | 1,691 | 246 | 2,253 | 1,846 |
| Total unique contaminants | na | 4,683 | 1,119 | 837 |
| (pct. of total transcripts with an alignment) | | (13.6%) | (1.9%) | (1.4%) |
| - contaminants flagged as Bacteria | na | 980 | 191 | 122 |
| - contaminants flagged as Opisthokonta | na | 3,703 | 928 | 715 |
| Total unique transcripts without an alignment | 17,981 | 38,952 | 14,243 | 13,971 |
| (pct. of total sequences with predicted frame) | (24.5%) | (53.1%) | (19.4%) | (19.0%) |
| - complete frames | 7,856 | 18,769 | 6,201 | 6,082 |
| - 5' partial frames | 4,165 | 7,455 | 3,369 | 3,311 |
| - 3' partial frames | 2,314 | 5,173 | 1,794 | 1,755 |
| - internal frames | 3,646 | 7,555 | 2,879 | 2,823 |

**Footnote**: na - not applicable, in RefSeq database only plant sequences were screened

**Table S6**. Similarity search results, most frequent species in best-hit and contaminant alignments are listed.

| | | Best-hit taxon | # hits (%)[†] | Contaminant taxon | # hits (%)[‡] |
|---|---|---|---|---|---|
| **RefSeq plant prot.: 42,200, NA** | 1. | ***Amborella trichopoda*** | **7,526 (13.60)** | not applicable<br><br>**Footnote**: [†] Percentage of all unique transcripts with an alignment – corresponds with the first number after database name of left; [‡] Percentage of all unique transcripts successfully aligned but considered contaminants – corresponds with the second number after database name on left. | |
| | 2. | *Juglans microcarpa* x *J. regia* | 3,453 (6.24) | | |
| | 3. | *Nymphaea colorata* | 3,245 (5.86) | | |
| | 4. | *Dioscorea cayenensis* subsp. *rotundata* | 2,690 (4.86) | | |
| | 5. | *Nelumbo nucifera* | 1,801 (3.25) | | |
| | 6. | *Physcomitrium patens* | 1,778 (3.21) | | |
| | 7. | *Cucurbita pepo* subsp. *pepo* | 1,376 (2.49) | | |
| | 8. | *Elaeis guineensis* | 1,334 (2.41) | | |
| | 9. | *Selaginella moellendorffii* | 1,222 (2.21) | | |
| | 10. | *Phoenix dactylifera* | 1,209 (2.18) | | |
| **Swiss-Prot: 28,344, 3,787** | 1. | ***Arabidopsis thaliana*** | **21,071 (61.28)** | ***Homo sapiens*** | **722 (15.42)** |
| | 2. | *Oryza sativa* subsp. *japonica* | 2,364 (6.87) | *Mus musculus* | 552 (11.79) |
| | 3. | *Homo sapiens* | 722 (2.1) | *Schizosaccharomyces pombe* | 291 (6.21) |
| | 4. | *Mus musculus* | 552 (1.61) | *Xenopus laevis* | 249 (5.32) |
| | 5. | *Dictyostelium discoideum* | 459 (1.33) | *Bos taurus* | 240 (5.12) |
| | 6. | *Oryza sativa* subsp. *indica* | 405 (1.18) | *Danio rerio* | 215 (4.59) |
| | 7. | *Solanum lycopersicum* | 357 (1.04) | *Saccharomyces cerevisiae* | 145 (3.1) |
| | 8. | *Nicotiana tabacum* | 310 (0.9) | *Drosophila melanogaster* | 129 (2.75) |
| | 9. | *Schizosaccharomyces pombe* | 291 (0.85) | *Rattus norvegicus* | 121 (2.58) |
| | 10. | *Xenopus laevis* | 249 (0.72) | *Bacillus subtilis* | 111 (2.37) |
| **TrEMBL: 44,213 / 508** | 1. | ***Picea sitchensis*** | **19094 (32.31)** | ***Orchesella cincta*** | **178 (15.91)** |
| | 2. | *Araucaria cunninghamii* | 4605 (7.79) | *Pelagivirga sediminicola* | 68 (6.08) |
| | 3. | *Wollemia nobilis* | 3533 (5.98) | *Photinus pyralis* | 46 (4.11) |
| | 4. | *Cinnamomum micranthum* f. *kanehirae* | 2303 (3.9) | *Folsomia candida* | 26 (2.32) |
| | 5. | *Amborella trichopoda* | 1941 (3.28) | *Saitoella complicata*[x] | 24 (2.14) |
| | 6. | *Pinus taeda* | 1882 (3.18) | *Thalassorhabdomicrobium marinisediminis* | 24 (2.14) |
| | 7. | *Nelumbo nucifera* | 1149 (1.94) | *Heliconius melpomene cythera* | 23 (2.06) |
| | 8. | *Actinidia chinensis* var. *chinensis* | 840 (1.42) | *Bos indicus* x *B. taurus* | 21 (1.88) |
| | 9. | *Physcomitrium patens* | 634 (1.07) | *Jimgerdemannia flammicorona* | 14 (1.25) |
| | 10. | *Vitis vinifera* | 630 (1.07) | *Clavispora lusitaniae* | 14 (1.25) |
| **OVERALL: 44,341 / 837** | 1. | ***Picea sitchensis*** | **17,419 (29.34)** | ***Orchesella cincta*** | **107 (12.78)** |
| | 2. | *Araucaria cunninghamii* | 4,230 (7.13) | *Photinus pyralis* | 46 (5.50) |
| | 3. | *Wollemia nobilis* | 3,415 (5.75) | *Pelagivirga sediminicola* | 38 (4.54) |
| | 4. | *Amborella trichopoda* | 2,155 (3.63) | *Saitoella complicata* | 24 (2.87) |
| | 5. | *Cinnamomum micranthum* f. *kanehirae* | 1,915 (3.23) | *Folsomia candida* | 19 (2.27) |
| | 6. | *Pinus taeda* | 1,331 (2.24) | *Clavispora lusitaniae* | 14 (1.67) |
| | 7. | *Arabidopsis thaliana* | 1,017 (1.71) | *Tolypocladium ophioglossoides* | 11 (1.31) |
| | 8. | *Nymphaea colorata* | 875 (1.47) | *Jimgerdemannia flammicorona* | 11 (1.31) |
| | 9. | *Juglans microcarpa* x *J. regia* | 869 (1.46) | *Rhizophagus irregularis* | 10 (1.19) |
| | 10. | *Nelumbo nucifera* | 729 (1.23) | *Hortaea werneckii* | 8 (0.96) |

**Table S7.** The most frequently reported protein domains form PFAM and SMART databases in filtered *MUS assembly*.

| | | Domain ID, description | # |
|---|---|---|---|
| PFAM | 1. | **PPR**, PPR repeat | 1124 |
| | 2. | **Pkinase**, Protein kinase domain | 831 |
| | 3. | **Pkinase_Tyr**, Tyrosine kinase | 821 |
| | 4. | **LRR_1**, Leucine Rich Repeat | 638 |
| | 5. | **NB-ARC**, NB-ARC domain | 304 |
| | 6. | **NACHT**, NACHT domain | 261 |
| | 7. | **LRRNT_2**, Leucine rich repeat N-terminal domain | 248 |
| | 8. | **WD40**, WD domain, G-beta repeat | 231 |
| | 9. | **AAA**, ATPase family associated with various cellular activities | 203 |
| | 10. | **TIR**, TIR domain | 187 |
| SMART | 1. | **TRANS**, Transmembrane domain | 5217 |
| | 2. | **COIL**, Coiled coil | 3787 |
| | 3. | **SIGNAL**, Signal peptide | 2453 |
| | 1. | **STYKc**, Protein kinase; unclassified specificity | 743 |
| | 2. | **AAA**, ATPases associated with a variety of cellular activities | 604 |
| | 3. | **LRR**, Leucine-rich repeats, outliers | 566 |
| | 4. | **RING**, Ring finger | 289 |
| | 5. | **WD40**, WD40 repeats | 242 |
| | 6. | **ZnF_C2HC**, Zinc finger | 214 |
| | 7. | **LRR_TYP**, Leucine-rich repeats, typical (most populated) subfamily | 200 |
| | 8. | **RRM**, RNA recognition motif | 189 |
| | 9. | **TIR**, Toll - interleukin 1 - resistance | 187 |
| | 10. | **S_TKc**, Serine/Threonine protein kinases, catalytic domain | 152 |

**Table S8**. The most frequently reported terms of level 3-5 in three main gene ontology (GO) aspects from the *MUS assembly*.

| | | Level 3 (name, term, counts) | Level 4 (name, term, counts) | Level 5 (name, term, counts) |
|---|---|---|---|---|
| **Biological Process** | 1. | macromolecule metabolic process, GO:0043170, 5419 | cellular protein metabolic process, GO:0044267, 2809 | cellular protein modification process, GO:0006464, 1945 |
| | 2. | cellular macromolecule metabolic process, GO:0044260, 4872 | phosphate-containing compound metabolic process, GO:0006796, 2609 | phosphorylation, GO:0016310, 1465 |
| | 3. | organic cyclic compound metabolic process, GO:1901360, 3549 | nucleic acid metabolic process, GO:0090304, 2374 | regulation of gene expression, GO:0010468, 1401 |
| | 4. | cellular aromatic compound metabolic process, GO:0006725, 3442 | macromolecule modification, GO:0043412, 2117 | regulation of cellular biosynthetic process, GO:0031326, 1345 |
| | 5. | cellular nitrogen compound metabolic process, GO:0034641, 3365 | macromolecule biosynthetic process, GO:0009059, 2006 | regulation of macromolecule biosynthetic process, GO:0010556, 1296 |
| | 6. | heterocycle metabolic process, GO:0046483, 3343 | cellular macromolecule biosynthetic process, GO:0034645, 1983 | regul. of nucleobase-cont. compound metab. proc., GO:0019219, 1290 |
| | 7. | protein metabolic process, GO:0019538, 3261 | protein modification process, GO:0036211, 1945 | regulation of RNA metabolic process, GO:0051252, 1193 |
| | 8. | organic substance biosynthetic process, GO:1901576, 3218 | gene expression, GO:0010467, 1894 | carboxylic acid metabolic process, GO:0019752, 987 |
| | 9. | cellular biosynthetic process, GO:0044249, 3135 | regulation of cellular metabolic process, GO:0031323, 1667 | nucleotide metabolic process, GO:0009117, 931 |
| | 10. | nucleobase-containing compound metabolic process, GO:0006139, 3074 | system development, GO:0048731, 1571 | RNA biosynthetic process, GO:0032774, 880 |
| **Molecular Function** | 1. | nucleotide binding, GO:0000166, 3633 | metal ion binding, GO:0046872, 3160 | purine ribonucleoside binding, GO:0032550, 2886 |
| | 2. | nucleoside phosphate binding, GO:1901265, 3633 | purine nucleotide binding, GO:0017076, 2894 | adenyl nucleotide binding, GO:0030554, 2684 |
| | 3. | anion binding, GO:0043168, 3379 | purine ribonucleotide binding, GO:0032555, 2892 | transition metal ion binding, GO:0046914, 2089 |
| | 4. | cation binding, GO:0043169, 3271 | ribonucleoside binding, GO:0032549, 2890 | pyrophosphatase activity, GO:0016462, 1228 |
| | 5. | nucleic acid binding, GO:0003676, 2990 | purine nucleoside bind., GO:0001883, 2886 | protein kinase activity, GO:0004672, 1088 |
| | 6. | ribonucleotide binding, GO:0032553, 2957 | purine ribonucleoside triphosphate binding, GO:0035639, 2806 | cation transmembrane transporter activity, GO:0008324, 429 |
| | 7. | nucleoside bind., GO:0001882, 2898 | DNA binding, GO:0003677, 1518 | endopeptidase activity, GO:0004175, 328 |
| | 8. | transferase activity, transferring phosphorus-cont. groups, GO:0016772, 1864 | kinase activity, GO:0016301, 1429 | DNA polymerase activity, GO:0034061, 316 |
| | 9. | hydrolase activity, acting on acid anhydrides, GO:0016817, 1249 | hydrolase activ., acting on acid anhydrides, in phosphorus-cont. anhydrides, GO:0016818, 1236 | guanyl nucleotide binding, GO:0019001, 282 |
| | 10. | hydrolase activity, acting on ester bonds, GO:0016788, 928 | phosphotransferase activity, alcohol group as acceptor, GO:0016773, 1224 | sequence-specific DNA binding, GO:0043565, 277 |
| **Cellular Component** | 1. | intracellular anatomical structure, GO:0005622, 16532 | cytoplasm, GO:0005737, 12887 | plastid, GO:0009536, 4119 |
| | 2. | intracellular organelle, GO:0043229, 10852 | intracellular membrane-bounded organelle, GO:0043231, 7064 | nucleus, GO:0005634, 3909 |
| | 3. | cell periphery, GO:0071944, 3085 | cytosol, GO:0005829, 2354 | vacuole, GO:0005773, 1957 |
| | 4. | plasma membrane, GO:0005886, 2860 | integral component of membrane, GO:0016021, 2339 | mitochondrion, GO:0005739, 1648 |
| | 5. | intrinsic component of membrane, GO:0031224, 2385 | Golgi apparatus, GO:0005794, 1017 | cytoskeleton, GO:0005856, 628 |
| | 6. | organelle membr., GO:0031090, 1531 | thylakoid, GO:0009579, 940 | ribosome, GO:0005840, 540 |
| | 7. | intracellular non-membrane-bounded organelle, GO:0043232, 1317 | endoplasmic reticulum, GO:0005783, 898 | nucleolus, GO:0005730, 503 |
| | 8. | envelope, GO:0031975, 1141 | cell wall, GO:0005618, 826 | plastid thylakoid, GO:0031976, 457 |
| | 9. | organelle envelope, GO:0031967, 1106 | ribonucleoprot. compl., GO:0030529, 713 | nucleoplasm, GO:0005654, 425 |
| | 10. | plasmodesma, GO:0009506, 1068 | chromosome, GO:0005694, 430 | plastid thylak. membr., GO:0055035, 361 |

**Table S9.** Abundance estimations of 121 protein coding genes with significantly differentiated expression patterns between three studied taxa. Markers significant in all six combinations of software dyads are presented, for simplification, solely fold change and *p*-values from RSEM-EDGER pair are shown. Marked in bold are 74 successfully annotated sequences.

| GENE | Sample expression level [TMM cross-sample normalized counts] | | | | | | | | | | Log2(fold change) / *p*-value in taxa comparison | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | M1 | M2 | M4 | M5 | UN1 | UN3 | UN4 | PS3 | PS4 | PS5 | M-UN *vs* PS | M *vs* PS | UN *vs* PS | M *vs* UN |
| **DN34_c0_g1** | 1078.33 | 520.37 | 635.04 | 852.00 | 341.97 | 843.65 | 1000.49 | 6.69 | 2.85 | 9.88 | 6.9/ 1.6E-15 | 6.9/ 1.3E-14 | -6.8/ 4.6E-12 | ns |
| DN87290_c0_g1 | 98.18 | 60.04 | 64.41 | 85.53 | 47.22 | 39.48 | 36.70 | 0.00 | 0.00 | 0.54 | 8.2/ 3.1E-15 | 8.6/ 7.1E-14 | -7.7/ 5.2E-10 | ns |
| DN25212_c0_g1 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 9.43 | 3.76 | 19.52 | -9.8/ 4.0E-15 | -9.8/ 2.0E-10 | 9.8/ 9.3E-09 | ns |
| DN36898_c0_g1 | 8.33 | 12.96 | 20.64 | 14.86 | 9.31 | 10.07 | 5.44 | 0.00 | 0.00 | 0.00 | 9.9/ 1.5E-14 | 10.2/ 7.0E-15 | -9.4/ 1.3E-11 | ns |
| **DN9747_c0_g1** | 78.15 | 139.04 | 90.56 | 103.42 | 37.14 | 82.47 | 54.10 | 0.66 | 0.33 | 1.41 | 6.7/ 3.2E-14 | 7.0/ 1.7E-15 | -6.2/ 2.1E-11 | ns |
| DN19944_c0_g1 | 0.00 | 0.00 | 0.00 | 0.07 | 0.00 | 0.00 | 0.00 | 8.49 | 2.74 | 15.54 | -8.4/ 4.6E-13 | -7.9/ 7.2E-09 | 9.5/ 5.3E-08 | ns |
| DN14280_c0_g2 | 0.00 | 0.00 | 0.00 | 0.03 | 0.00 | 0.00 | 0.04 | 6.74 | 2.39 | 8.03 | -8.8/ 4.7E-13 | -8.8/ 2.6E-09 | 8.8/ 6.8E-08 | ns |
| DN22543_c0_g1 | 124.15 | 74.83 | 78.44 | 98.13 | 63.66 | 48.39 | 59.10 | 1.31 | 0.00 | 0.51 | 7.0/ 8.4E-13 | 7.3/ 1.6E-10 | -6.6/ 1.1E-07 | ns |
| **DN33_c2_g1** | 3098.63 | 2832.07 | 2655.37 | 2583.67 | 881.77 | 1591.25 | 1862.31 | 35.78 | 30.84 | 70.00 | 5.6/ 1.0E-12 | 6.0/ 2.8E-15 | -5.0/ 2.7E-10 | ns |
| DN7794_c0_g1 | 0.07 | 0.24 | 0.00 | 0.00 | 0.08 | 0.00 | 0.00 | 2.85 | 24.69 | 9.92 | -7.7/ 4.6E-12 | -7.4/ 6.9E-08 | ns | ns |
| DN15047_c0_g1 | 386.25 | 579.26 | 449.20 | 552.86 | 137.20 | 400.74 | 369.31 | 11.44 | 9.91 | 16.84 | 5.0/ 1.2E-11 | 5.3/ 1.5E-14 | -4.6/ 6.3E-09 | ns |
| **DN1154_c2_g1** | 1278.58 | 369.47 | 575.67 | 798.16 | 339.66 | 841.59 | 830.26 | 5.85 | 4.23 | 17.24 | 6.3/ 1.2E-11 | 6.4/ 3.8E-10 | -6.2/ 7.0E-10 | ns |
| **DN111461_c24_g1** | 4824.13 | 3723.25 | 2416.12 | 5859.51 | 570.20 | 2490.25 | 2027.80 | 30.70 | 24.22 | 37.85 | 6.7/ 2.0E-11 | 7.1/ 2.7E-18 | -5.8/ 2.1E-10 | ns |
| DN21971_c2_g1 | 0.49 | 0.00 | 0.37 | 0.33 | 0.00 | 0.42 | 0.42 | 5.03 | 7.50 | 22.69 | -5.3/ 3.0E-10 | -5.2/ 2.8E-07 | ns | ns |
| **DN3704_c3_g1** | 107.23 | 25.57 | 62.90 | 29.86 | 28.74 | 22.69 | 28.94 | 0.00 | 0.00 | 0.57 | 7.6/ 3.5E-09 | 8.0/ 1.4E-07 | ns | ns |
| DN882_c0_g1 | 486.44 | 277.55 | 131.03 | 706.40 | 60.54 | 208.30 | 113.96 | 2.84 | 2.02 | 3.91 | 6.6/ 6.4E-09 | 7.1/ 2.3E-12 | -5.4/ 2.0E-10 | ns |
| **DN3026_c0_g1** | 4.90 | 4.61 | 4.06 | 5.57 | 3.21 | 8.99 | 3.39 | 12.47 | 66.45 | 60.54 | -3.2/ 1.2E-08 | -3.2/ 4.4E-06 | ns | ns |
| DN30070_c0_g1 | 72.47 | 54.13 | 63.66 | 37.54 | 46.04 | 143.56 | 63.01 | 0.00 | 0.00 | 1.24 | 7.3/ 2.7E-08 | 7.0/ 8.4E-06 | ns | ns |
| **DN19629_c0_g1** | 0.55 | 0.47 | 0.49 | 0.62 | 0.31 | 0.94 | 1.12 | 0.90 | 16.45 | 35.50 | -4.7/ 2.8E-08 | ns | ns | ns |
| DN70256_c1_g1 | 51.13 | 32.46 | 30.31 | 37.54 | 21.35 | 28.05 | 12.67 | 0.00 | 1.09 | 0.00 | 6.3/ 8.5E-08 | 6.7/ 4.0E-06 | ns | ns |
| **DN3802_c0_g1** | 412.66 | 691.44 | 515.93 | 431.01 | 2.45 | 309.78 | 433.60 | 1.13 | 1.04 | 1.91 | 8.2/ 2.1E-07 | 8.6/ 7.0E-24 | ns | ns |
| **DN10418_c0_g2** | 15.06 | 47.88 | 9.47 | 17.59 | 3.10 | 11.67 | 15.89 | 0.20 | 0.20 | 0.41 | 5.8/ 2.2E-07 | 6.2/ 2.1E-09 | -5.0/ 1.8E-06 | ns |
| **DN10876_c0_g2** | 0.00 | 0.00 | 0.00 | 0.08 | 0.00 | 0.00 | 0.00 | 0.28 | 20.12 | 4.08 | -8.3/ 2.6E-07 | ns | ns | ns |
| DN13838_c0_g1 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.35 | 9.06 | 2.83 | 17.45 | -7.1/ 3.4E-07 | -9.6/ 1.5E-09 | ns | ns |
| **DN6885_c0_g1** | 15.30 | 21.24 | 20.49 | 19.63 | 8.26 | 13.60 | 1.61 | 0.00 | 0.00 | 0.41 | 6.4/ 5.4E-07 | 6.8/ 9.8E-13 | ns | ns |
| DN26628_c0_g1 | 59.57 | 40.83 | 34.81 | 43.61 | 16.51 | 29.54 | 15.15 | 4.80 | 2.77 | 2.95 | 3.3/ 1.9E-06 | 3.7/ 1.4E-09 | ns | ns |
| **DN14516_c0_g1** | 0.00 | 0.00 | 0.00 | 0.00 | 69.19 | 0.00 | 63.89 | 65.44 | 46.31 | 42.58 | ns | -12.0/ 1.9E-25 | ns | ns |
| **DN15289_c0_g1** | 93.71 | 49.05 | 78.16 | 73.49 | 0.00 | 36.92 | 62.00 | 0.00 | 0.00 | 0.00 | ns | 12.5/ 2.9E-23 | ns | ns |
| DN14751_c0_g1 | 62.25 | 47.49 | 26.22 | 41.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | ns | 11.8/ 5.3E-20 | ns | 11.8/ 2.4E-20 |
| **DN12723_c0_g1** | 103.09 | 91.44 | 54.89 | 58.46 | 0.00 | 0.51 | 0.00 | 0.67 | 0.81 | 0.38 | ns | 7.0/ 7.9E-18 | ns | 8.6/ 6.0E-13 |
| DN62591_c0_g1 | 30.16 | 19.03 | 12.10 | 29.14 | 0.53 | 0.00 | 16.95 | 0.00 | 0.00 | 0.00 | ns | 10.8/ 2.3E-16 | ns | ns |
| DN11689_c0_g1 | 91.38 | 30.13 | 143.75 | 124.36 | 78.57 | 0.57 | 106.83 | 0.00 | 0.00 | 0.00 | ns | 12.9/ 3.6E-16 | ns | ns |
| **DN2594_c0_g1** | 103.42 | 49.99 | 103.34 | 97.72 | 10.59 | 14.78 | 2.11 | 1.46 | 1.66 | 1.28 | ns | 5.9/ 3.9E-16 | ns | ns |
| **DN14974_c0_g1** | 1.30 | 1.30 | 0.85 | 1.02 | 40.76 | 1.79 | 1.06 | 46.30 | 108.01 | 34.96 | ns | -5.8/ 1.3E-15 | ns | ns |
| DN62550_c0_g1 | 54.39 | 60.35 | 117.09 | 72.37 | 1.66 | 52.38 | 42.61 | 0.90 | 1.06 | 1.50 | ns | 6.0/ 2.1E-15 | ns | ns |
| **DN244104_c1_g1** | 642.80 | 1278.74 | 1051.49 | 588.09 | 3.56 | 1268.41 | 695.92 | 2.16 | 3.45 | 7.69 | ns | 7.7/ 2.3E-15 | ns | ns |
| **DN16360_c2_g1** | 296.13 | 214.27 | 80.88 | 292.17 | 0.76 | 0.00 | 127.50 | 0.44 | 1.09 | 0.30 | ns | 8.6/ 6.7E-15 | ns | ns |
| **DN43514_c0_g1** | 29.22 | 16.89 | 11.60 | 12.18 | 10.05 | 0.40 | 0.00 | 0.00 | 0.00 | 0.00 | ns | 10.4/ 6.9E-15 | ns | ns |
| DN35757_c0_g1 | 74.44 | 14.58 | 48.85 | 38.04 | 0.00 | 0.57 | 0.00 | 0.00 | 0.00 | 0.00 | ns | 11.8/ 2.1E-14 | ns | 7.8/ 2.8E-08 |

| GENE | Sample expression level [TMM cross-sample normalized counts] | | | | | | | | | | Log2(fold change) / *p*-value in taxa comparison | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | M1 | M2 | M4 | M5 | UN1 | UN3 | UN4 | PS3 | PS4 | PS5 | M-UN *vs* PS | M *vs* PS | UN *vs* PS | M *vs* UN |
| DN38913_c0_g1 | 21.78 | 12.79 | 40.04 | 15.17 | 0.00 | 7.66 | 8.21 | 0.00 | 0.00 | 0.00 | ns | 10.8/ 2.3E-14 | ns | ns |
| **DN11048_c0_g1** | 96.74 | 156.40 | 198.02 | 319.09 | 2.97 | 63.55 | 135.60 | 2.38 | 1.89 | 0.89 | ns | 6.8/ 1.0E-13 | ns | ns |
| DN2063_c0_g1 | 1107.68 | 936.69 | 490.34 | 397.22 | 13.69 | 20.08 | 28.57 | 12.17 | 6.83 | 11.14 | ns | 6.2/ 1.9E-13 | ns | 5.1/ 1.4E-10 |
| **DN13150_c0_g1** | 38.70 | 8.66 | 24.83 | 37.96 | 0.16 | 30.11 | 28.24 | 0.00 | 0.00 | 0.00 | ns | 11.1/ 2.6E-13 | ns | ns |
| DN4576_c0_g1 | 0.00 | 0.00 | 0.14 | 0.12 | 9.44 | 7.65 | 1.79 | 9.78 | 17.34 | 8.53 | ns | -6.8/ 4.6E-13 | ns | ns |
| **DN2169_c0_g1** | 0.37 | 1.02 | 0.62 | 0.36 | 15.50 | 4.90 | 2.50 | 15.04 | 13.99 | 17.72 | ns | -4.8/ 9.2E-13 | ns | ns |
| **DN7054_c0_g1** | 0.61 | 0.00 | 0.12 | 0.00 | 20.91 | 18.87 | 0.00 | 60.06 | 38.02 | 36.17 | ns | -8.0/ 1.4E-12 | ns | ns |
| **DN28253_c1_g1** | 99.86 | 123.78 | 123.74 | 113.30 | 0.73 | 2.66 | 1.15 | 0.00 | 0.70 | 0.00 | ns | 8.9/ 1.5E-12 | ns | 6.3/ 2.5E-15 |
| DN8421_c0_g1 | 28.79 | 19.43 | 35.90 | 57.88 | 2.03 | 1.84 | 33.00 | 0.77 | 1.15 | 0.38 | ns | 5.6/ 2.4E-11 | ns | ns |
| **DN12883_c0_g1** | 12.71 | 11.60 | 6.86 | 8.82 | 0.17 | 0.13 | 0.00 | 0.43 | 0.09 | 0.21 | ns | 5.5/ 4.6E-11 | ns | 6.5/ 7.1E-13 |
| **DN13898_c1_g1** | 36.54 | 25.94 | 17.64 | 24.64 | 11.34 | 3.43 | 25.86 | 1.94 | 1.37 | 0.98 | ns | 4.2/ 1.3E-10 | ns | ns |
| **DN147_c0_g1** | 597.45 | 1373.96 | 414.95 | 467.60 | 67.78 | 252.34 | 154.90 | 21.12 | 11.24 | 10.85 | ns | 5.7/ 1.3E-10 | ns | ns |
| **DN1163_c2_g1** | 395.25 | 98.71 | 244.63 | 128.12 | 124.88 | 0.78 | 256.31 | 1.31 | 0.77 | 3.46 | ns | 6.9/ 2.2E-10 | ns | ns |
| DN21212_c0_g1 | 39.62 | 29.76 | 23.88 | 51.27 | 2.67 | 25.49 | 96.89 | 1.31 | 0.63 | 2.05 | ns | 4.8/ 3.5E-10 | ns | ns |
| DN9419_c0_g1 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 2.43 | 3.33 | 8.61 | 21.93 | ns | -9.8/ 1.0E-09 | ns | ns |
| **DN2019_c0_g1** | 35.77 | 24.80 | 13.68 | 28.27 | 6.72 | 6.17 | 28.91 | 0.69 | 0.45 | 1.54 | ns | 4.8/ 1.5E-09 | ns | ns |
| **DN20095_c0_g1** | 0.34 | 0.00 | 0.39 | 0.21 | 5.72 | 6.45 | 1.49 | 18.12 | 10.65 | 6.49 | ns | -5.4/ 2.0E-09 | ns | ns |
| **DN20956_c0_g1** | 70.01 | 60.46 | 64.14 | 178.66 | 1.12 | 142.41 | 0.00 | 0.73 | 0.90 | 0.00 | ns | 7.4/ 2.1E-09 | ns | ns |
| **DN22195_c1_g1** | 258.78 | 88.78 | 186.35 | 103.64 | 82.62 | 1.72 | 205.68 | 3.91 | 0.56 | 3.05 | ns | 6.0/ 4.4E-09 | ns | ns |
| DN13648_c1_g1 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.26 | 4.10 | 3.82 | 15.76 | ns | -9.3/ 5.1E-09 | ns | ns |
| **DN16426_c4_g1** | 320.85 | 536.07 | 21.02 | 267.13 | 1.01 | 103.11 | 39.84 | 0.42 | 1.03 | 1.13 | ns | 8.4/ 7.3E-09 | ns | ns |
| DN2089_c0_g2 | 2.05 | 5.88 | 5.40 | 3.92 | 0.00 | 3.42 | 0.00 | 0.00 | 0.00 | 0.00 | ns | 8.4/ 1.4E-08 | ns | ns |
| DN9776_c0_g1 | 27.68 | 30.75 | 116.07 | 114.96 | 1.00 | 40.51 | 34.43 | 0.55 | 0.89 | 1.84 | ns | 6.1/ 1.8E-08 | ns | ns |
| **DN1196_c0_g1** | 0.50 | 3.96 | 2.05 | 2.76 | 33.43 | 26.02 | 12.07 | 58.22 | 43.67 | 38.14 | ns | -4.3/ 3.3E-08 | ns | ns |
| **DN3741_c0_g1** | 299.39 | 665.79 | 19.38 | 269.60 | 1.07 | 128.13 | 35.57 | 0.57 | 1.45 | 0.36 | ns | 8.7/ 3.5E-08 | ns | ns |
| **DN49862_c0_g1** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.59 | 2.28 | 3.18 | 9.17 | ns | -8.6/ 6.1E-08 | ns | ns |
| **DN17420_c0_g1** | 0.40 | 0.37 | 1.03 | 0.44 | 92.45 | 0.49 | 177.05 | 4.08 | 72.63 | 171.42 | ns | -7.2/ 7.2E-08 | ns | ns |
| DN15406_c0_g1 | 38.66 | 92.85 | 56.68 | 59.40 | 25.26 | 25.88 | 34.48 | 3.00 | 0.33 | 1.52 | ns | 5.3/ 7.5E-08 | ns | ns |
| **DN36903_c0_g1** | 5.29 | 10.14 | 3.92 | 11.13 | 0.86 | 0.30 | 1.08 | 0.39 | 0.24 | 0.12 | ns | 5.1/ 9.4E-08 | ns | ns |
| **DN17004_c0_g1** | 12.51 | 2.97 | 2.59 | 7.66 | 0.00 | 0.42 | 2.43 | 0.00 | 0.00 | 0.00 | ns | 9.0/ 9.5E-08 | ns | ns |
| DN6929_c0_g1 | 1.02 | 2.74 | 1.24 | 1.05 | 36.42 | 2.87 | 18.21 | 10.84 | 22.21 | 21.49 | ns | -3.6/ 1.1E-07 | ns | ns |
| **DN142_c0_g1** | 17.17 | 4.17 | 15.84 | 8.40 | 45.55 | 172.90 | 185.17 | 215.78 | 100.24 | 292.24 | ns | -4.1/ 1.1E-07 | ns | ns |
| **DN1879_c0_g1** | 115.94 | 463.05 | 173.91 | 333.24 | 2.40 | 28.67 | 10.60 | 2.73 | 13.93 | 4.18 | ns | 5.3/ 2.6E-07 | ns | ns |
| DN15773_c0_g1 | 17.80 | 10.46 | 10.40 | 12.85 | 0.00 | 10.43 | 2.72 | 0.00 | 0.00 | 0.60 | ns | 5.8/ 2.7E-07 | ns | ns |
| **DN6288_c1_g1** | 16.37 | 12.29 | 34.80 | 33.23 | 15.26 | 83.67 | 5.45 | 0.12 | 0.00 | 0.84 | ns | 6.1/ 3.9E-07 | ns | ns |
| DN7990_c0_g1 | 127.45 | 54.33 | 344.00 | 154.27 | 74.50 | 94.31 | 125.72 | 8.93 | 4.26 | 6.85 | ns | 4.7/ 4.2E-07 | -3.9/ 1.7E-08 | ns |
| **DN5959_c0_g2** | 1.19 | 0.41 | 0.74 | 0.58 | 4.05 | 0.62 | 0.45 | 6.88 | 7.06 | 21.61 | ns | -4.0/ 4.4E-07 | ns | ns |
| DN3100_c0_g1 | 0.26 | 0.00 | 0.15 | 0.44 | 6.51 | 1.02 | 0.46 | 31.91 | 3.30 | 9.92 | ns | -6.0/ 5.0E-07 | ns | ns |
| **DN317_c17_g1** | 97.86 | 106.28 | 148.33 | 333.68 | 1.99 | 28.79 | 5.45 | 4.77 | 13.53 | 3.68 | ns | 4.6/ 6.8E-07 | ns | ns |
| **DN10386_c0_g1** | 6.91 | 39.08 | 8.97 | 27.51 | 7.52 | 1.92 | 1.18 | 0.76 | 0.36 | 0.77 | ns | 5.0/ 7.7E-07 | ns | ns |
| **DN595_c1_g1** | 5.35 | 3.37 | 2.85 | 3.23 | 42.69 | 57.59 | 30.60 | 102.04 | 31.11 | 19.38 | ns | -3.8/ 9.8E-07 | ns | -3.5/ 1.2E-10 |
| **DN8509_c0_g1** | 79.95 | 101.82 | 144.62 | 305.07 | 2.23 | 22.95 | 4.73 | 2.64 | 11.89 | 3.81 | ns | 4.7/ 1.1E-06 | ns | ns |
| **DN16054_c0_g1** | 0.19 | 1.49 | 2.59 | 1.35 | 2.76 | 13.72 | 7.21 | 14.01 | 79.08 | 27.92 | ns | -4.8/ 1.3E-06 | ns | ns |
| DN505_c0_g1 | 8.56 | 15.42 | 27.14 | 10.34 | 117.14 | 221.25 | 64.77 | 138.87 | 353.86 | 99.74 | ns | -3.6/ 1.5E-06 | ns | ns |

| GENE | Sample expression level [TMM cross-sample normalized counts] | | | | | | | | | | Log2(fold change) / *p*-value in taxa comparison | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | M1 | M2 | M4 | M5 | UN1 | UN3 | UN4 | PS3 | PS4 | PS5 | M-UN *vs* PS | M *vs* PS | UN *vs* PS | M *vs* UN |
| **DN21688_c1_g1** | 17.93 | 11.34 | 61.62 | 51.11 | 0.97 | 0.00 | 0.00 | 0.56 | 0.92 | 0.00 | ns | 6.3/ 1.7E-06 | ns | ns |
| **DN59675_c0_g1** | 0.00 | 0.00 | 0.14 | 0.00 | 7.27 | 11.99 | 6.27 | 1.91 | 20.26 | 6.39 | ns | -7.3/ 1.7E-06 | ns | -7.1/ 7.3E-12 |
| DN18417_c0_g1 | 5.59 | 15.34 | 1.81 | 10.63 | 0.00 | 0.08 | 1.16 | 0.13 | 0.19 | 0.08 | ns | 5.8/ 1.7E-06 | ns | ns |
| **DN11943_c1_g1** | 1.94 | 5.27 | 4.04 | 7.94 | 0.93 | 0.07 | 0.77 | 0.12 | 0.00 | 0.15 | ns | 5.4/ 1.9E-06 | ns | ns |
| **DN8457_c0_g1** | 145.30 | 193.37 | 207.99 | 112.09 | 28.77 | 198.67 | 105.80 | 27.77 | 24.02 | 15.26 | ns | 2.9/ 2.4E-06 | ns | ns |
| **DN3246_c0_g2** | 159.26 | 317.01 | 259.98 | 152.47 | 24.84 | 342.67 | 189.90 | 16.25 | 26.52 | 32.61 | ns | 3.2/ 2.4E-06 | ns | ns |
| DN7732_c0_g1 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.21 | 0.95 | 3.80 | 6.55 | ns | -8.2/ 2.4E-06 | ns | ns |
| **DN9000_c4_g1** | 96.77 | 172.48 | 75.24 | 0.39 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | ns | 12.7/ 8.6E-06 | ns | ns |
| **DN9183_c1_g1** | 1.47 | 0.29 | 0.24 | 0.12 | 15.61 | 0.89 | 3.24 | 9.24 | 17.15 | 5.95 | ns | -4.2/ 1.6E-05 | ns | ns |
| DN1988_c0_g1 | 0.53 | 369.80 | 393.51 | 0.58 | 207.38 | 579.94 | 482.37 | 1.85 | 2.55 | 3.23 | ns | ns | -7.4/ 1.0E-16 | ns |
| DN19247_c0_g1 | 195.74 | 72.17 | 0.83 | 0.00 | 0.24 | 0.21 | 0.00 | 102.38 | 51.50 | 56.36 | ns | ns | 8.9/ 1.8E-16 | ns |
| **DN17625_c0_g1** | 0.00 | 0.83 | 0.53 | 0.00 | 48.50 | 81.26 | 65.67 | 0.00 | 0.00 | 0.43 | ns | ns | -8.6/ 2.3E-12 | -7.8/ 3.0E-10 |
| **DN10565_c0_g1** | 0.00 | 0.35 | 0.34 | 0.00 | 101.37 | 57.35 | 25.94 | 0.00 | 0.28 | 0.10 | ns | ns | -8.7/ 8.5E-12 | ns |
| **DN22931_c1_g1** | 0.00 | 0.00 | 0.00 | 0.00 | 25.47 | 46.56 | 18.47 | 0.32 | 0.00 | 0.36 | ns | ns | -7.1/ 1.6E-11 | -11.2/ 5.5E-20 |
| **DN20781_c0_g1** | 10.83 | 0.39 | 0.19 | 8.44 | 20.45 | 45.40 | 41.57 | 0.47 | 0.00 | 0.21 | ns | ns | -7.3/ 8.0E-11 | ns |
| DN22141_c0_g1 | 9.10 | 0.00 | 0.00 | 0.08 | 9.77 | 5.50 | 4.16 | 0.00 | 0.00 | 0.00 | ns | ns | -9.0/ 2.3E-09 | ns |
| **DN12125_c0_g1** | 0.00 | 0.01 | 3.00 | 0.02 | 2.76 | 8.26 | 6.79 | 0.08 | 0.00 | 0.01 | ns | ns | -7.0/ 1.4E-07 | ns |
| **DN20325_c0_g1** | 0.00 | 0.00 | 28.12 | 0.74 | 31.50 | 27.66 | 31.88 | 0.67 | 0.00 | 0.00 | ns | ns | -7.2/ 2.1E-07 | ns |
| DN33656_c0_g1 | 0.00 | 0.00 | 0.00 | 0.00 | 47.97 | 4.07 | 139.30 | 0.00 | 0.00 | 0.00 | ns | ns | -12.3/ 2.9E-07 | ns |
| **DN32187_c0_g1** | 0.89 | 153.38 | 1.54 | 322.38 | 1.88 | 1.33 | 2.19 | 10.28 | 143.39 | 216.28 | ns | ns | 6.1/ 3.0E-06 | ns |
| **DN7303_c0_g1** | 109.00 | 49.09 | 75.43 | 51.71 | 0.81 | 0.89 | 1.15 | 2.45 | 105.62 | 2.27 | ns | ns | ns | 6.2/ 4.2E-16 |
| DN10255_c0_g1 | 12.26 | 21.70 | 11.76 | 9.21 | 0.00 | 0.00 | 0.00 | 19.89 | 23.54 | 0.00 | ns | ns | ns | 10.1/ 4.0E-15 |
| DN5464_c0_g1 | 0.00 | 0.00 | 0.00 | 0.00 | 11.27 | 9.21 | 25.36 | 34.79 | 10.81 | 0.00 | ns | ns | ns | -10.3/ 9.6E-15 |
| **DN15760_c0_g1** | 0.00 | 0.00 | 0.00 | 0.00 | 9.01 | 7.89 | 5.09 | 1.23 | 8.49 | 5.05 | ns | ns | ns | -9.2/ 9.3E-14 |
| DN18660_c0_g1 | 10.74 | 16.05 | 11.11 | 4.62 | 0.00 | 0.00 | 0.00 | 0.00 | 32.36 | 0.00 | ns | ns | ns | 9.7/ 3.1E-12 |
| **DN7225_c3_g1** | 661.95 | 444.52 | 181.77 | 213.19 | 2.92 | 0.42 | 0.47 | 2.26 | 0.65 | 3.58 | ns | ns | ns | 8.2/ 1.5E-10 |
| DN10556_c2_g1 | 0.00 | 0.00 | 0.07 | 0.03 | 5.19 | 10.19 | 3.24 | 2.66 | 5.03 | 0.03 | ns | ns | ns | -7.4/ 1.0E-09 |
| **DN11841_c0_g1** | 9.92 | 4.49 | 10.14 | 29.14 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.21 | ns | ns | ns | 10.1/ 1.2E-09 |
| DN56801_c0_g1 | 8.23 | 18.91 | 17.83 | 17.47 | 0.44 | 0.00 | 0.00 | 1.16 | 0.00 | 0.00 | ns | ns | ns | 6.7/ 1.7E-09 |
| **DN9161_c1_g1** | 0.00 | 0.00 | 0.00 | 0.00 | 376.43 | 3.18 | 135.06 | 0.83 | 59.70 | 147.56 | ns | ns | ns | -13.8/ 2.3E-08 |
| DN10393_c0_g1 | 0.00 | 0.89 | 0.00 | 0.36 | 35.01 | 33.18 | 39.07 | 0.00 | 6.57 | 35.66 | ns | ns | ns | -6.9/ 3.5E-08 |
| **DN1010_c4_g1** | 2.81 | 13.62 | 10.78 | 24.32 | 0.20 | 0.08 | 0.00 | 5.84 | 0.39 | 12.87 | ns | ns | ns | 6.8/ 1.7E-07 |
| **DN4099_c0_g2** | 1.94 | 1.44 | 1.61 | 1.65 | 6.97 | 10.76 | 13.34 | 0.62 | 16.49 | 2.69 | ns | ns | ns | -2.6/ 1.1E-06 |
| **DN45522_c0_g1** | 0.50 | 0.00 | 0.38 | 0.00 | 14.94 | 4.46 | 18.40 | 0.74 | 9.06 | 15.08 | ns | ns | ns | -5.7/ 1.7E-06 |
| **DN1973_c0_g1** | 965.35 | 1758.51 | 371.62 | 1523.20 | 46.66 | 58.10 | 111.35 | 436.99 | 224.19 | 108.82 | ns | ns | ns | 4.0/ 3.5E-06 |
| DN12894_c0_g1 | 24.16 | 0.77 | 32.07 | 85.37 | 0.00 | 0.00 | 0.00 | 0.59 | 0.49 | 0.00 | ns | ns | ns | 11.5/ 3.6E-06 |
| **DN1352_c0_g1** | 0.62 | 2.13 | 15.23 | 2.84 | 305.26 | 171.52 | 209.21 | 1.81 | 199.18 | 89.96 | ns | ns | ns | -5.4/ 3.6E-06 |
| **DN9372_c0_g1** | 113.11 | 44.65 | 13.26 | 56.87 | 1.57 | 3.20 | 0.62 | 3.42 | 6.25 | 8.56 | ns | ns | ns | 5.0/ 3.8E-06 |

**Table S10**. Functional annotation - gene ontology (GO) and protein domains - of genes differentially expressed between mountain pine species and Scots pine, and between dwarf pine and Pyrenean pine. Only GO terms from the highest, most specific level of annotation are presented.

| Gene, isoform | PFAM / SMART protein domains | Biological Processes, top level GO terms | Molecular Functions, top level GO terms | Cellular Components, top level GO terms |
|---|---|---|---|---|
| **Up-regulated in *P. mugo* and *P. uncinata* in relation to *P. sylvestris* levels** | | | | |
| DN10418_c0_g2, i1 | UDPGT / - | lvl3: GO:0006810 | lvl6: GO:0050502 | lvl5: GO:0005634, GO:0005643, GO:0044428 |
| DN111461_c24_g1, i1 | 2OG-FeII_Oxy / - | lvl3: GO:0055114, GO:0080167 | lvl6: GO:0005506 | - |
| DN1154_c2_g1, i1 | Methyltransf_2, Dimerisation / - | lvl6: GO:0009809 | lvl6: GO:0030744, GO:0030755, GO:0033799, GO:0047763 | lvl5: GO:0005634 |
| DN33_c2_g1, i6 | Methyltransf_3 / - | lvl6: GO:0009809, GO:0009809 | lvl6: GO:0042409, GO:0042409 | - |
| DN34_c0_g1, i1 | Methyltransf_2, Dimerisation / - | lvl2: GO:0032259 | lvl5: GO:0008171 | - |
| DN3704_c3_g1, i1 | EF1_GNE, GST_C / EF1_GNE | lvl6: GO:0006414 | lvl6: GO:0003746 | lvl5: GO:0005773, GO:0005853 |
| DN3802_c0_g1, i1 | PsbR / TRANS | lvl8: GO:0010270 | - | lvl6: GO:0009507, GO:0009523, GO:0009534, GO:0009535, GO:0042651, GO:0044434 |
| DN6885_c0_g1, i1 | Cullin_Nedd8, Cullin / Cullin_Nedd8, Cullin | lvl8: GO:0006511, GO:0016567 | lvl2: GO:0016491 | lvl7: GO:0009941 |
| DN9747_c0_g1, i1 | Thioredoxin, ERp29_N, Calsequestrin, AhpC-TSA / SIGNAL, TRANS | lvl7: GO:0010154 | lvl5: GO:0003756, GO:0015035 | lvl7: GO:0000326, GO:0000327 |
| **Up-regulated in *P. sylvestris* in relation to *P. mugo* and *P. uncinata* levels** | | | | |
| DN10876_c0_g2, i1 | Aldedh, LuxC, Formyl_trans_C, Formyl_trans_N / - | lvl8: GO:0001947, GO:0006776, GO:0032291, GO:0032292, GO:0033121, GO:0035115, GO:0039023, GO:0042552, GO:0042904, GO:0043010, GO:0051289 | lvl7: GO:0005524 | lvl6: GO:0005759, GO:0009507 |
| DN19629_c0_g1, i1 | Aldedh, LuxC, Formyl_trans_C, Formyl_trans_N / - | lvl3: GO:0055114 | lvl5: GO:0008886 | lvl4: GO:0005737 |
| DN3026_c0_g1, i2 | p450 / TRANS | lvl3: GO:0055114 | lvl6: GO:0005506, GO:0005506 | lvl4: GO:0016021 |
| **Up-regulated in *P. mugo* in relation to *P. uncinata* levels** | | | | |
| DN1010_c4_g1, i2 | Pkinase_Tyr, Pkinase, LRR_1 / TRANS, STYKc, SIGNAL | lvl6: GO:0006468 | lvl7: GO:0005524 | lvl3: GO:0005886, GO:0009506, GO:0071944 |
| DN11841_c0_g1, i2 | Aldedh, LuxC / - | lvl4: GO:0009699, GO:0019438, GO:0044550, GO:1901362 | lvl6: GO:0050269 | lvl4: GO:0005737, GO:0005829, GO:0044444 |
| DN12723_c0_g1, i1 | GTP_EFTU_D3, GTP_EFTU, GTP_EFTU_D2 / - | lvl9: GO:0006409 | lvl8: GO:0015078 | lvl7: GO:0045335 |
| DN12883_c0_g1, i3 | Terpene_synth_C, Terpene_synth / COIL, TRANS | lvl7: GO:0043693 | lvl5: GO:0000287, GO:0010333, GO:0000287, GO:0010333 | lvl6: GO:0009507 |

| | | | | |
|---|---|---|---|---|
| DN1973_c0_g1, i8 | Peroxidase / SIGNAL, TRAN | lvl6: GO:0070301 | lvl4: GO:0020037, GO:0046872 | lvl5: GO:0005773, GO:0009505 |
| DN28253_c1_g1, i1 | FA_desaturase, Cyt-b5 / - | lvl6: GO:0032787, GO:0072330 | lvl6: GO:0005506 | - |
| DN7225_c3_g1, i1 | GTP_EFTU_D3, GTP_EFTU, GTP_EFTU_D2 / - | lvl9: GO:0006409 | lvl8: GO:0015078 | lvl7: GO:0045335 |
| DN7303_c0_g1, i3 | E1-E2_ATPase, Hydrolase, Cation_ATPase_N / TRANS, Cation_ATPase_N, COIL | lvl8: GO:0006754, GO:0009168, GO:0009169, GO:0009206, GO:0009207, GO:0015991 | lvl9: GO:0043492 | lvl6: GO:0005774 |
| DN9372_c0_g1, i10 | Terpene_synth, Terpene_synth_C / COIL | lvl9: GO:0051504 | lvl5: GO:0000287, GO:0010333, GO:0000287, GO:0010333, GO:0050554 | lvl6: GO:0009507, GO:0009507 |
| **Up-regulated in *P. uncinata* in relation to *P. mugo* levels** | | | | |
| DN1352_c0_g1, i10 | DUF26 / SIGNAL, TRANS | lvl5: GO:0016310 | lvl4: GO:0016301 | lvl1: GO:0005576 |
| DN15760_c0_g1, i1 | DUF26 / SIGNAL, TRANS | lvl5: GO:0016310 | lvl4: GO:0016301 | lvl1: GO:0005576 |
| DN17625_c0_g1, i1 | HSP90, HATPase_c / COIL, HATPase_c | lvl8: GO:0009169, GO:0009207 | lvl7: GO:0005524, GO:0016887 | lvl7: GO:0009570 |
| DN22931_c1_g1, i1 | DUF3700, DUF3711 / DUF3700 | - | - | lvl3: GO:0005886, GO:0071944 |
| DN4099_c0_g2, i1 | Pkinase_Tyr, Pkinase, LRR_1 / TRANS, STYKc, SIGNAL | lvl6: GO:0006468 | lvl7: GO:0005524 | lvl3: GO:0005886, GO:0009506, GO:0071944 |
| DN45522_c0_g1, i4 | Ribosomal_L18e / - | lvl5: GO:0006412 | lvl2: GO:0003735 | lvl6: GO:0005774, GO:0009507, GO:0015934, GO:0022625 |
| DN595_c1_g1, i1 | Aldose_epim / TRANS | lvl5: GO:0019318 | lvl5: GO:0004034 | lvl4: GO:0005737, GO:0005829, GO:0044444 |
| DN59675_c0_g1, i5 | DUF26, Pkinase_Tyr, Pkinase / TRANS, STYKc, SIGNAL | lvl7: GO:0015992 | lvl7: GO:0005524 | lvl6: GO:0009507 |
| DN9161_c1_g1, i1 | HSP90, HATPase_c / COIL, HATPase_c | lvl8: GO:0009169, GO:0009207 | lvl7: GO:0005524, GO:0016887 | lvl7: GO:0009570 |

**Table S11**. Functional annotation - gene ontology (GO) and protein domains - of genes differentially expressed between each of two mountain pine species and Scots pine. In each category the most specific of annotated GO terms are shown.

| Gene, isoform | Diamond best hit | EGGNOG predicted gene ID and/or function description | | |
|---|---|---|---|---|
| **PFAM / SMART protein domains** | | **Biological Processes, top level GO terms** | **Molecular Functions, top level GO terms** | **Cellular Components, top level GO terms** |
| **Up-regulated in *P. mugo* pine in relation to *P. sylvestris*** | | | | |
| **DN10386_c0_g1, i15** | XP_006844819.1 | **NAS2**, sensor for the physiological iron status within the plant, might be involved in the transport of iron | | |
| NAS / - | | lvl8: GO:0006826 | lvl4: GO:0030410 | - |
| **DN11048_c0_g1, i3** | tr\| A0A151R6I0_CAJCA | **PSBC**, one of the components of the core antenna complex of photosystem II; binds chlorophyll and helps catalyze the primary light-induced photochemical processes of photosystem II (by similarity) | | |
| PSII / TRANS | | lvl6: GO:0009772, GO:0018298 | lvl4: GO:0016168 | lvl7: GO:0009570 |
| **DN1163_c2_g1, i1** | tr\| A0A0C9QNY1_9SPER | 3-beta hydroxysteroid dehydrogenase/isomerase family | | |
| 3Beta_HSD, NmrA, Epimerase / TRANS | | lvl5: GO:0009617, GO:0042742 | lvl6: GO:0005507 | lvl7: GO:0009570 |
| **DN11943_c1_g1, i3** | XP_028804495.1 | receptor-like protein kinase | | |
| Pkinase_Tyr, Pkinase, Glyco_hydro_18 / TRANS, STYKc, SIGNAL | | lvl6: GO:0006468, GO:0007178 | lvl7: GO:0005524 | lvl4: GO:0016021 |
| **DN13150_c0_g1, i1** | tr\| A0A2K3LWF7_TRIPR | domain of unknown function DUF221 | | |
| DUF221 / TRANS, COIL | | - | - | lvl1: GO:0016020 |
| **DN13898_c1_g1, i1** | XP_020244917.1 | **BGAL7**, beta-galactosidase | | |
| Glyco_hydro_35, Gal_Lectin, Glyco_hydro_42 / SIGNAL, TRANS | | lvl3: GO:0005975 | lvl6: GO:0004565 | lvl2: GO:0048046 |
| **DN147_c0_g1, i2** | tr\| Q9SYU3_PINRA | O-methyltransferase | | |
| Methyltransf_2, Dimerisation /- | | lvl2: GO:0032259 | lvl5: GO:0008171 | - |
| **DN15289_c0_g1, i1** | tr\| A0A453NWV4_AEGTS | ABC1 family | | |
| ABC1 / TRANS, COIL | | lvl5: GO:0016310 | lvl6: GO:0004190 | lvl7: GO:0009941 |
| **DN16360_c2_g1, i1** | XP_039823216.1 | **ACTA2**, actin | | |
| Actin / ACTIN, ANK | | lvl9: GO:0018393, GO:0051532 | lvl8: GO:0000146 | lvl10: GO:0035267 |
| **DN16426_c4_g1, i1** | XP_034696707.1 | hydroxycinnamoyl-coenzyme A shikimate/quinate hydroxycinnamoyltransferase-like | | |
| Transferase / - | | lvl1: GO:0008152 | lvl4: GO:0016747 | - |
| **DN17004_c0_g1, i1** | tr\| A9P1C4_PICSI | 2-oxoacid-dependent dioxygenase | | |
| 2OG-FeII_Oxy / - | | lvl3: GO:0055114 | lvl6: GO:0005506 | - |
| **DN1879_c0_g1, i3** | XP_020242727.1 | **LOX**, lipoxygenase | | |
| Lipoxygenase, PLAT / LH2 | | lvl8: GO:0048527 | lvl6: GO:0005506 | lvl6: GO:0009507 |
| **DN2019_c0_g1, i10** | tr\| C0PST5_PICSI | F-box protein | | |
| F-box / FBOX | | lvl8: GO:0016567 | lvl2: GO:0030246 | lvl5: GO:0005634 |
| **DN20956_c0_g1, i1** | XP_031282903.1 | ketol-acid reductoisomerase | | |
| IlvC, IlvN / - | | lvl6: GO:0009097, GO:0009099 | lvl6: GO:0005507 | lvl7: GO:0009570, GO:0009941 |
| **DN21688_c1_g1, i1** | tr\| I6QL29_PRUPE | **HSP90C**, heat shock protein | | |
| HSP90, HATPase_c / HATPase_c, COIL | | lvl7: GO:0006278, GO:0044743, GO:0045036, GO:0045037, GO:0072596 | lvl7: GO:0005524 | lvl7: GO:0009570, GO:0009941 |
| **DN22195_c1_g1, i1** | tr\| A9NPJ0_PICSI | 3-beta hydroxysteroid dehydrogenase/isomerase family | | |

| | | | | |
|---|---|---|---|---|
| 3Beta_HSD, Epimerase, NmrA / - | | lvl4: GO:0009737, GO:0033993, GO:0097305 | lvl3: GO:0000166, GO:1901265 | lvl7: GO:0009570 |
| **DN244104_c1_g1, i1** | tr\| A0A392NMB0_9FABA | **RCA**, ribulose bisphosphate carboxylase oxygenase activase | | |
| AAA / - | | lvl7: GO:0048366 | lvl7: GO:0005524 | lvl7: GO:0009570, GO:0009941 |
| **DN2594_c0_g1, i2** | tr\| B8LRP0_PICSI | zinc finger A20 and AN1 domain-containing stress-associated protein | | |
| zf-AN1, zf-A20 / ZnF_AN1, ZnF_A20 | | lvl2: GO:0006950 | lvl6: GO:0008270 | - |
| **DN317_c17_g1, i1** | XP_020242721.1 | **LOX**, lipoxygenase | | |
| Lipoxygenase, PLAT / LH2 | | lvl8: GO:0048527 | lvl6: GO:0005506 | lvl6: GO:0009507 |
| **DN3246_c0_g2, i1** | XP_025828258.1 | **RCA**, ribulose bisphosphate carboxylase oxygenase activase | | |
| AAA / - | | lvl7: GO:0048366 | lvl7: GO:0005524 | lvl7: GO:0009570, GO:0009941 |
| **DN36903_c0_g1, i1** | tr\| B8LL97_PICSI | bark storage protein A-like | | |
| PNP_UDP_1, DNA_pol_delta_4 / SIGNAL, TRANS | | lvl5: GO:0006260 | lvl1: GO:0003824 | lvl5: GO:0005634 |
| **DN3741_c0_g1, i1** | tr\| A9NZE2_PICSI | NmrA-like family | | |
| 3Beta_HSD, NmrA / - | | - | - | - |
| **DN43514_c0_g1, i1** | tr\| C1K4P6_ELAOL | **ACX5**, acyl-coenzyme A oxidase | | |
| ACOX, Acyl-CoA_dh_M, Acyl-CoA_dh_1 / - | | lvl7: GO:0009694, GO:0009695 | lvl5: GO:0003997 | lvl6: GO:0005777 |
| **DN6288_c1_g1, i1** | tr\| C0PTH7_PICSI | 2-oxoglutarate-iron(II)-dependent oxygenase | | |
| 2OG-FeII_Oxy / - | | lvl7: GO:0010421 | lvl6: GO:0005506 | - |
| **DN8457_c0_g1, i8** | tr\| A9NVS6_PICSI | **FLS1**, flavonol synthase | | |
| 2OG-FeII_Oxy / - | | lvl5: GO:0051555 | lvl6: GO:0005506 | lvl5: GO:0005634 |
| **DN8509_c0_g1, i1** | tr\| A0A1S3CTT5_CUCME | lipoxygenase | | |
| Lipoxygenase, PLAT / LH2, COIL | | lvl7: GO:0031408 | lvl6: GO:0005506 | lvl6: GO:0009507 |
| **DN9000_c4_g1, i1** | tr\| Q4LAW9_CAPCH | malate dehydrogenase | | |
| Ldh_1_N, Ldh_1_C / TRANS | | lvl7: GO:0006099, GO:0006108 | lvl5: GO:0030060 | lvl7: GO:0009570 |
| For information on the remaining 12 annotated DEGs in this comparison: **DN10418_c0_g2_i1**, **DN111461_c24_g1_i1**, **DN1154_c2_g1_i1**, **DN12723_c0_g1_i1**, **DN12883_c0_g1_i3**, **DN28253_c1_g1_i1**, **DN33_c2_g1_i6**, **DN34_c0_g1_i1**, **DN3704_c3_g1_i1**, **DN3802_c0_g1_i1**, **DN6885_c0_g1_i1**, **DN9747_c0_g1_i1**, please see their duplicates in Table 2, in Table S8 or above. | | | | |
| **Up-regulated in *P. sylvestris* in relation to *P. mugo* levels** | | | | |
| **DN1196_c0_g1, i7** | tr\| A9NWC5_PICSI | **UGT73B5**, flavonol 3-O-glucosyltransferase activity | | |
| UDPGT / TRANS, COIL | | lvl5: GO:0051555 | lvl6: GO:0010294, GO:0047893, GO:0080043, GO:0080044 | lvl4: GO:0005737, GO:0005829, GO:0031410, GO:0043231, GO:0044444 |
| **DN142_c0_g1, i1** | tr\| Q944B7_PINPS | destroys radicals which are normally produced within the cells and which are toxic to biological systems (by similarity) | | |
| Sod_Cu / SIGNAL | | lvl8: GO:0019430, GO:0035195 | lvl6: GO:0005507, GO:0008270 | lvl7: GO:0009570 |
| **DN14516_c0_g1, i1** | tr\| A6N1R4_ORYSI | serine-glyoxylate | | |
| Aminotran_5, Aminotran_1_2 / - | | lvl5: GO:0009853 | lvl5: GO:0004760, GO:0008453, GO:0050281 | lvl7: GO:0009570 |
| **DN14974_c0_g1, i1** | tr\| A0A0D5X1V8_9POAL | transcription factor | | |
| AP2 / AP2 | | lvl7: GO:0006355 | lvl4: GO:0003677 | lvl5: GO:0005634 |
| **DN16054_c0_g1, i26** | XP_031496404.1 | **SANT** | | |
| Myb_DNA-binding / SANT | | lvl7: GO:0006355, GO:0010154 | lvl4: GO:0003677 | lvl5: GO:0005634 |
| **DN17420_c0_g1, i7** | tr\| A9NP17_PICSI | ribulose-phosphate 3-epimerase | | |
| Ribul_P_3_epim / - | | lvl10: GO:0002098 | lvl5: GO:0004750 | lvl7: GO:0009570, GO:0009941 |

| | | | | |
|---|---|---|---|---|
| **DN20095_c0_g1, i1** | NP_001290022.1 | **HB**, may not function as an oxygen storage or transport protein, but might act as an oxygen sensor or play a role in electron transfer, possibly to a bound oxygen molecule | | |
| Globin / - | | lvl6: GO:0015671 | lvl6: GO:0005506 | lvl4: GO:0005618, GO:0005737, GO:0005829, GO:0044444 |
| **DN2169_c0_g1, i1** | tr\| A0A023ZYJ7_PINMS | **AP2** | | |
| AP2 / AP2 | | lvl7: GO:0006355 | lvl4: GO:0003677 | lvl5: GO:0005634 |
| **DN49862_c0_g1, i5** | XP_030526856.1 | **ARP1**, K05692 actin beta gamma 1 | | |
| Actin / ACTIN | | lvl6: GO:0002064, GO:0007283, GO:0048515 | lvl7: GO:0005524 | lvl7: GO:0097346 |
| **DN5959_c0_g2, i4** | tr\| A0A4U5R5P0_POPAL | homeobox-leucine zipper protein | | |
| START, Homeobox / HOX, START, COIL | | lvl7: GO:0006355, GO:0010154, GO:0048366, GO:0048825 | lvl7: GO:0000976 | lvl5: GO:0005634 |
| **DN7054_c0_g1, i2** | tr\| A9NPJ0_PICSI | 3-beta hydroxysteroid dehydrogenase/isomerase family | | |
| 3Beta_HSD, NmrA, Epimerase / TRANS | | lvl5: GO:0009617, GO:0042742 | lvl6: GO:0005507 | lvl7: GO:0009570 |
| **DN9183_c1_g1, i2** | tr\| A0A3G6JA47_PINTA | transcription factor | | |
| K-box, SRF-TF / COIL, MADS, SIGNAL, TRANS | | lvl7: GO:0006355 | lvl4: GO:0003677 | lvl5: GO:0005634 |
| For information on the remaining 3 annotated DEGs in this comparison: **DN3026_c0_g1_i2**, **DN595_c1_g1_i1**, **DN59675_c0_g1_i5**, please see their duplicates in other compared pairs in other tables or above. | | | | |
| **Up-regulated in *P. uncinata* in relation to *P. sylvestris* levels** | | | | |
| **DN10565_c0_g1, i2** | XP_039687868.1 | RING finger and CHY zinc finger domain-containing protein | | |
| zf-CHY, zf-C3HC4 / RING | | - | lvl6: GO:0008270 | - |
| **DN12125_c0_g1, i4** | XP_034685795.1 | reverse transcriptase (RNA-dependent DNA polymerase) | | |
| RVT_1 / - | | lvl7: GO:0006278 | lvl6: GO:0003964 | - |
| **DN20325_c0_g1, i1** | tr\| A0A078HA49_BRANA | 26S proteasome non-ATPase regulatory subunit | | |
| RPN7, PCI / PINT, COIL | | lvl4: GO:0009057, GO:0030163 | - | lvl5: GO:0005634 |
| **DN20781_c0_g1, i1** | tr\| A0A0C9RKK7_9SPER | **RPL23**, ribosomal protein | | |
| Ribosomal_L14 / - | | lvl5: GO:0006412 | lvl2: GO:0003735 | lvl6: GO:0015934, GO:0022625, GO:0031981 |
| For information on the remaining 8 annotated DEGs in this comparison: **DN10418_c0_g2_i1**, **DN111461_c24_g1_i1**, **DN1154_c2_g1_i1**, **DN17625_c0_g1_i1**, **DN22931_c1_g1_i1**, **DN33_c2_g1_i6**, **DN34_c0_g1_i1**, **DN9747_c0_g1_i1**, see duplicates in other tables or above. | | | | |
| **Up-regulated in *P. sylvestris* in relation to *P. uncinata* levels** | | | | |
| **DN32187_c0_g1, i3** | tr\| A0A2Z4HIC2_PASED | **TUFA**, promotes the GTP-dependent binding of aminoacyl-tRNA to the A-site of ribosomes during protein biosynthesis (by similarity) | | |
| GTP_EFTU, GTP_EFTU_D2, GTP_EFTU_D3 / - | | lvl8: GO:0009207, GO:1901069 | lvl7: GO:0003924, GO:0005525 | lvl7: GO:0009570, GO:0009941 |

**Table S12**. Gene ontology (GO) terms representing enriched functional categories in sets of genes differentially expressed between analyzed pine taxa.

| Term ID | Aspect | Term name | *P*-value for over-representation† | Value | LogSize | Freq. | Uniq. | Disp. |
|---|---|---|---|---|---|---|---|---|
| **Enriched in DEGs up-regulated in *P. mugo* and *P. uncinata* in relation to *P. sylvestris*** | | | | | | | | |
| GO:0008150 | BP | biological process | 0.0275 | -1.5611 | 4.3379 | 100.0000 | 1.0000 | 0.0000 |
| GO:0008152 | BP | metabolic process | 0.0052 | -2.2831 | 4.0076 | 46.7365 | 1.0000 | 0.0000 |
| GO:0032259 | BP | methylation | 0.0008 | -3.0863 | 2.6551 | 2.0716 | 0.9236 | 0.0000 |
| GO:0042752 | BP | regulation of circadian rhythm | 0.0253 | -1.5970 | 1.7324 | 0.2434 | 0.9899 | 0.0000 |
| GO:0048316 | BP | seed development | 0.0179 | -1.7477 | 2.7672 | 2.6825 | 0.9201 | 0.0000 |
| GO:0065003 | BP | protein-containing complex assembly | 0.0118 | -2.6787 | 2.7505 | 2.5814 | 0.7761 | 0.0000 |
| GO:0034975 | BP | protein folding in endoplasmic reticulum | 0.0028 | -2.5576 | 0.4771 | 0.0092 | 0.9792 | 0.0239 |
| GO:0034976 | BP | response to endoplasmic reticulum stress | 0.0100 | -2.0007 | 2.0492 | 0.5099 | 0.9290 | 0.0347 |
| GO:0000327 | CC | lytic vacuole within protein storage vacuole | 0.0014 | -2.8557 | 0.3010 | 0.0048 | 0.7897 | 0.0000 |
| GO:0032991 | CC | protein-containing complex | 0.0153 | -2.4895 | 3.5085 | 15.3905 | 1.0000 | 0.0000 |
| GO:0000793 | CC | condensed chromosome | 0.0179 | -1.7468 | 1.8751 | 0.3533 | 0.7684 | 0.0236 |
| GO:0009579 | CC | thylakoid | 0.0453 | -1.3442 | 2.7752 | 2.8404 | 0.9623 | 0.0253 |
| GO:0005853 | CC | eukaryotic translation elongation factor 1 complex | 0.0146 | -1.8356 | 0.6990 | 0.0191 | 0.8532 | 0.0371 |
| GO:0005788 | CC | endoplasmic reticulum lumen | 0.0157 | -1.8034 | 1.6335 | 0.2005 | 0.8140 | 0.0465 |
| GO:0003746 | MF | translation elongation factor activity | 0.0450 | -1.3466 | 1.7853 | 0.2797 | 1.0000 | 0.0000 |
| GO:0008171 | MF | O-methyltransferase activity | 0.0000 | -5.5720 | 1.6628 | 0.2098 | 0.4367 | 0.0000 |
| **Enriched in DEGs up-regulated in *P. sylvestris* in relation to *P. mugo* and *P. uncinata*** | | | | | | | | |
| GO:0008152 | BP | metabolic process | na | -1.9097 | 4.0076 | 46.7365 | 1.0000 | 0.0000 |
| GO:0009699 | BP | phenylpropanoid biosynthetic process | 0.0225 | -1.6486 | 2.0334 | 0.4915 | 0.3350 | 0.7954 |
| GO:0016903 | MF | oxidoreductase activity, acting on the aldehyde or oxo group of donor | 0.0002 | -3.7592 | 1.8195 | 0.3030 | 0.4452 | 0.0000 |
| **Enriched in DEGs up-regulated in *P. mugo* in relation to *P. uncinata*** | | | | | | | | |
| GO:0008150 | BP | biological process | 0.0248 | -1.6064 | 4.3379 | 100.0000 | 1.0000 | 0.0000 |
| GO:0008152 | BP | metabolic process | 0.0050 | -2.9781 | 4.0076 | 46.7365 | 1.0000 | 0.0000 |
| GO:0009987 | BP | cellular process | 0.0151 | -1.8214 | 4.1479 | 64.5630 | 1.0000 | 0.0000 |
| GO:0010035 | BP | response to inorganic substance | 0.0075 | -2.1222 | 2.8657 | 3.3669 | 0.9673 | 0.0000 |
| GO:0010259 | BP | multicellular organism aging | 0.0009 | -3.0537 | 2.6385 | 1.9958 | 0.9331 | 0.0000 |
| GO:0040007 | BP | growth | 0.0177 | -1.7510 | 2.5832 | 1.7546 | 1.0000 | 0.0000 |
| GO:0040011 | BP | locomotion | 0.0020 | -2.6917 | 1.6532 | 0.2021 | 1.0000 | 0.0000 |
| GO:0043066 | BP | negative regulation of apoptotic process | 0.0007 | -3.1671 | 0.9031 | 0.0322 | 0.9406 | 0.0000 |
| GO:0051017 | BP | actin filament bundle assembly | 0.0001 | -4.0771 | 1.3010 | 0.0873 | 0.9236 | 0.0000 |
| GO:0051031 | BP | tRNA transport | 0.0001 | -4.2936 | 0.6990 | 0.0184 | 0.8961 | 0.0000 |
| GO:0002182 | BP | cytoplasmic translational elongation | 0.0001 | -3.9356 | 0.7782 | 0.0230 | 0.7843 | 0.0201 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| GO:0030175 | CC | filopodium | 0.0001 | -4.2936 | 1.1761 | 4.0050 | 0.9314 | 0.0000 |
| GO:0005853 | CC | eukaryotic translation elongation factor 1 complex | 0.0002 | -3.8029 | 0.6990 | 0.0191 | 0.9473 | 0.0008 |
| GO:0042995 | CC | cell projection | 0.0018 | -2.7385 | 2.0828 | 0.5728 | 0.9979 | 0.0011 |
| GO:0005615 | CC | extracellular space | 0.0019 | -2.7137 | 2.2833 | 0.9118 | 0.9978 | 0.0012 |
| GO:0031252 | CC | cell leading edge | 0.0001 | -3.8458 | 2.9661 | 4.4077 | 0.9973 | 0.0015 |
| GO:0009506 | CC | plasmodesma | 0.0002 | -3.8233 | 2.9279 | 4.0386 | 0.9419 | 0.0015 |
| GO:0030054 | CC | cell junction | 0.0001 | -3.8292 | 2.9279 | 4.0386 | 0.9974 | 0.0015 |
| GO:0055044 | CC | symplast | 0.0001 | -3.8371 | 2.9284 | 4.0433 | 0.9974 | 0.0015 |
| GO:0043025 | CC | neuronal cell body‡ | 0.0003 | -3.5540 | 2.9661 | 4.4002 | 0.9973 | 0.0015 |
| GO:0044297 | CC | cell body | 0.0003 | -3.5540 | 2.9661 | 4.4002 | 0.9973 | 0.0015 |
| GO:0071944 | CC | cell periphery | 0.0155 | -1.8084 | 3.5219 | 15.8726 | 0.9967 | 0.0019 |
| GO:0005576 | CC | extracellular region | 0.0336 | -1.4733 | 3.2815 | 9.1226 | 0.9970 | 0.0022 |
| GO:0005811 | CC | lipid droplet | 0.0009 | -3.0390 | 1.7709 | 0.2769 | 0.8729 | 0.0161 |
| GO:0031982 | CC | vesicle | 0.0081 | -2.0920 | 2.9504 | 4.2534 | 0.9441 | 0.0401 |
| GO:0003674 | MF | molecular function | 0.0142 | -1.8474 | 4.3315 | 100.0000 | 1.0000 | 0.0000 |
| GO:0003824 | MF | catalytic activity | 0.0009 | -3.0692 | 3.9787 | 44.3787 | 1.0000 | 0.0000 |
| GO:0004781 | MF | sulfate adenylyltransferase (ATP) activity | 0.0001 | -4.1098 | 0.6990 | 0.0186 | 0.8912 | 0.0000 |
| GO:0005215 | MF | transporter activity | 0.0117 | -1.9313 | 3.1242 | 6.1993 | 1.0000 | 0.0000 |
| GO:0005488 | MF | binding | 0.0058 | -2.2354 | 4.1108 | 60.1473 | 1.0000 | 0.0000 |
| GO:0019003 | MF | GDP binding | 0.0001 | -4.1904 | 0.6990 | 0.0186 | 0.8376 | 0.0000 |
| GO:0022890 | MF | inorganic cation transmembrane transporter activity | 0.0005 | -3.8297 | 2.5658 | 1.7106 | 0.8436 | 0.0000 |
| **Enriched in DEGs up-regulated in *P. uncinata* in relation to *P. mugo*** | | | | | | | | |
| GO:0002376 | BP | immune system process | 0.0021 | -2.6742 | 2.3820 | 1.1024 | 1.0000 | 0.0000 |
| GO:0009908 | BP | flower development | 0.0046 | -2.3405 | 2.6532 | 2.0624 | 0.9077 | 0.0000 |
| GO:0046034 | BP | ATP metabolic process | 0.0320 | -1.4947 | 2.2625 | 0.8360 | 0.9304 | 0.0000 |
| GO:0050821 | BP | protein stabilization | 0.0001 | -3.8423 | 1.5185 | 0.1470 | 0.9738 | 0.0000 |
| GO:0071277 | BP | cellular response to calcium ion | 0.0001 | -4.2785 | 0.9031 | 0.0322 | 0.7819 | 0.0000 |
| GO:0006457 | BP | protein folding | 0.0062 | -2.2044 | 2.3222 | 0.9600 | 0.9714 | 0.0244 |
| GO:0005575 | CC | cellular component | 0.0476 | -1.3222 | 4.3212 | 100.0000 | 1.0000 | 0.0000 |
| GO:0005774 | CC | vacuolar membrane | 0.0089 | -2.0487 | 2.4713 | 1.4082 | 0.8761 | 0.0000 |
| GO:0048046 | CC | apoplast | 0.0119 | -1.9246 | 2.6551 | 2.1530 | 0.9972 | 0.0012 |
| GO:0071944 | CC | cell periphery | 0.0023 | -2.6416 | 3.5219 | 15.8726 | 0.9960 | 0.0016 |
| GO:0005576 | CC | extracellular region | 0.0044 | -2.3574 | 3.2815 | 9.1226 | 0.9964 | 0.0022 |
| GO:0016020 | CC | membrane | 0.0170 | -1.7691 | 3.9194 | 39.6506 | 0.9950 | 0.0034 |
| GO:0004034 | MF | aldose 1-epimerase activity | 0.0051 | -2.2966 | 0.6990 | 0.0186 | 0.8330 | 0.0000 |
| GO:0009678 | MF | pyrophosphate hydrolysis-driven proton transmembrane transporter activity | 0.0359 | -1.4444 | 1.6990 | 0.2284 | 1.0000 | 0.0000 |
| GO:0051082 | MF | unfolded protein binding | 0.0002 | -3.6087 | 2.0569 | 0.5267 | 0.9544 | 0.0000 |
| GO:0032550 | MF | purine ribonucleoside binding | 0.0296 | -1.5294 | 0.8451 | 0.0280 | 0.6447 | 0.0302 |

**Footnote**: † - *p*-value from run_GOseq.pl of TRINITY, all the other from REVIGO (run settings: large list (0.9), *p*-value provided, obsolete terms removed, *Arabidopsis thaliana* as reference species, SimRel measure; accessed 18 May 2022); ‡ - unknown reason of this appearance; Freq. – Frequency; Uniq. – Uniqueness; Disp. – Dispensability; na - not applicable

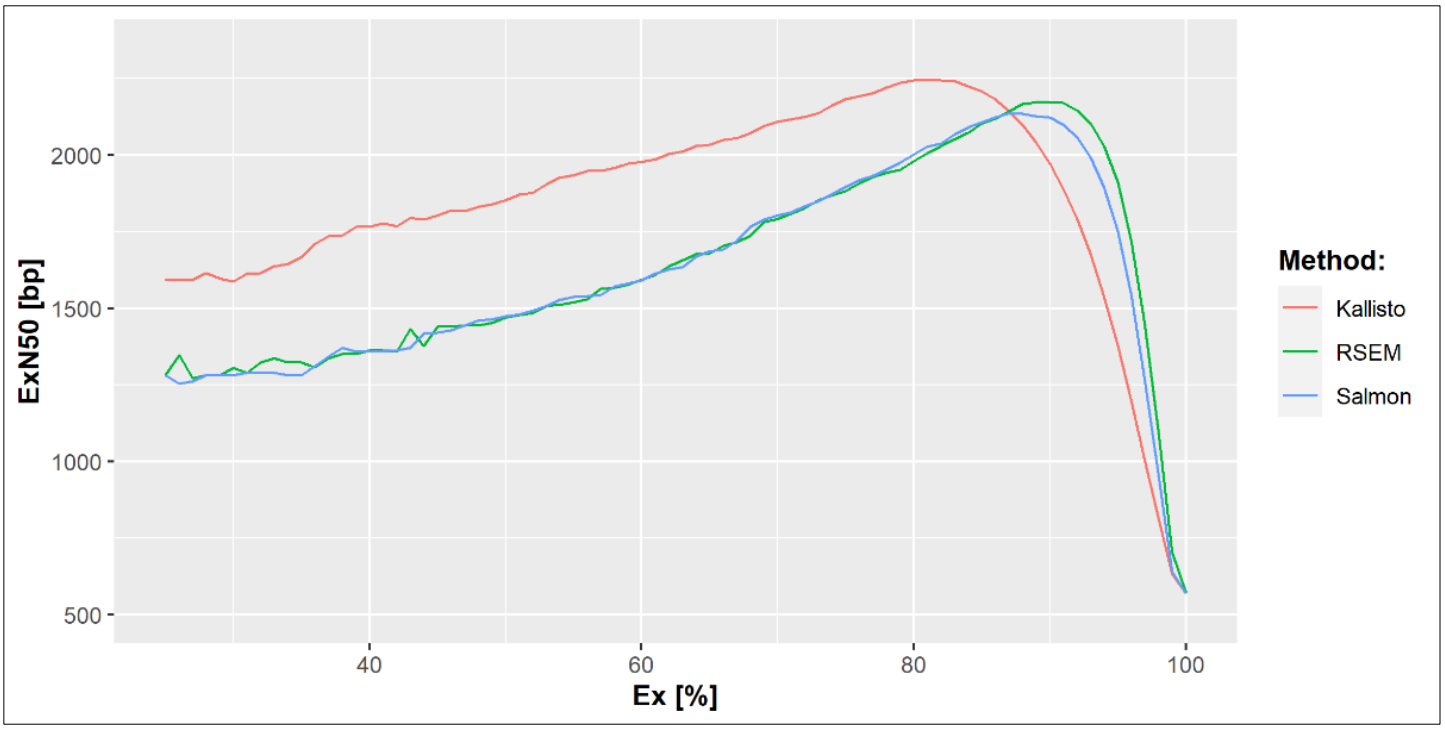



**Fig. S1** ExN50 change plotted against the Ex value, comparison of outputs from three read abundance estimation methods used on *MUS assembly* gene models.

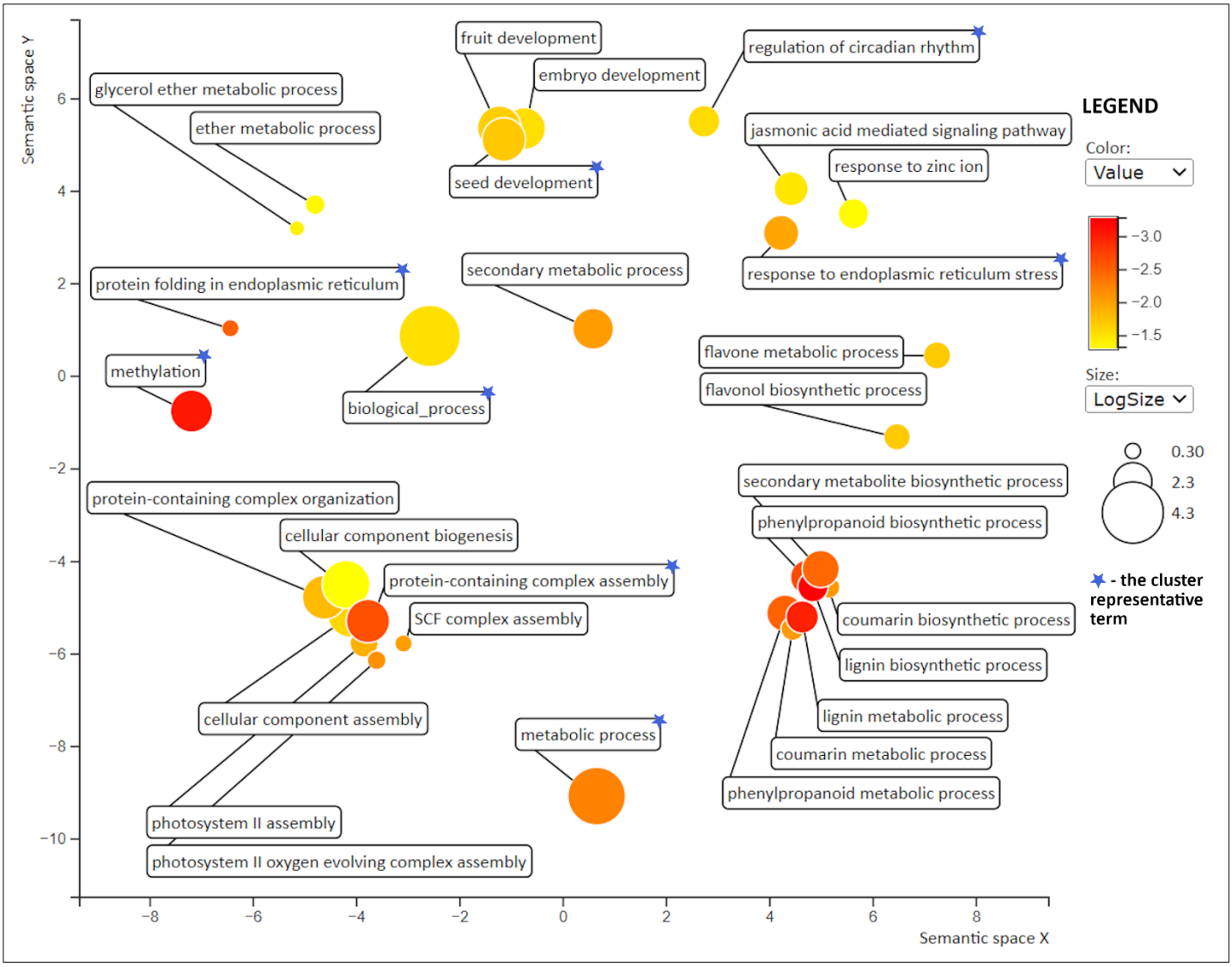



**Fig. S2** Scatterplot of the GO terms from the Biological Process aspect enriched in differentially expressed genes up-regulated in two mountain pines relative to Scots pine. Bubble color indicates the *p*-value for overrepresentation, size indicates the frequency of particular GO term in the underlying GO database (more general terms are larger), group representatives are marked with blue asterisks.

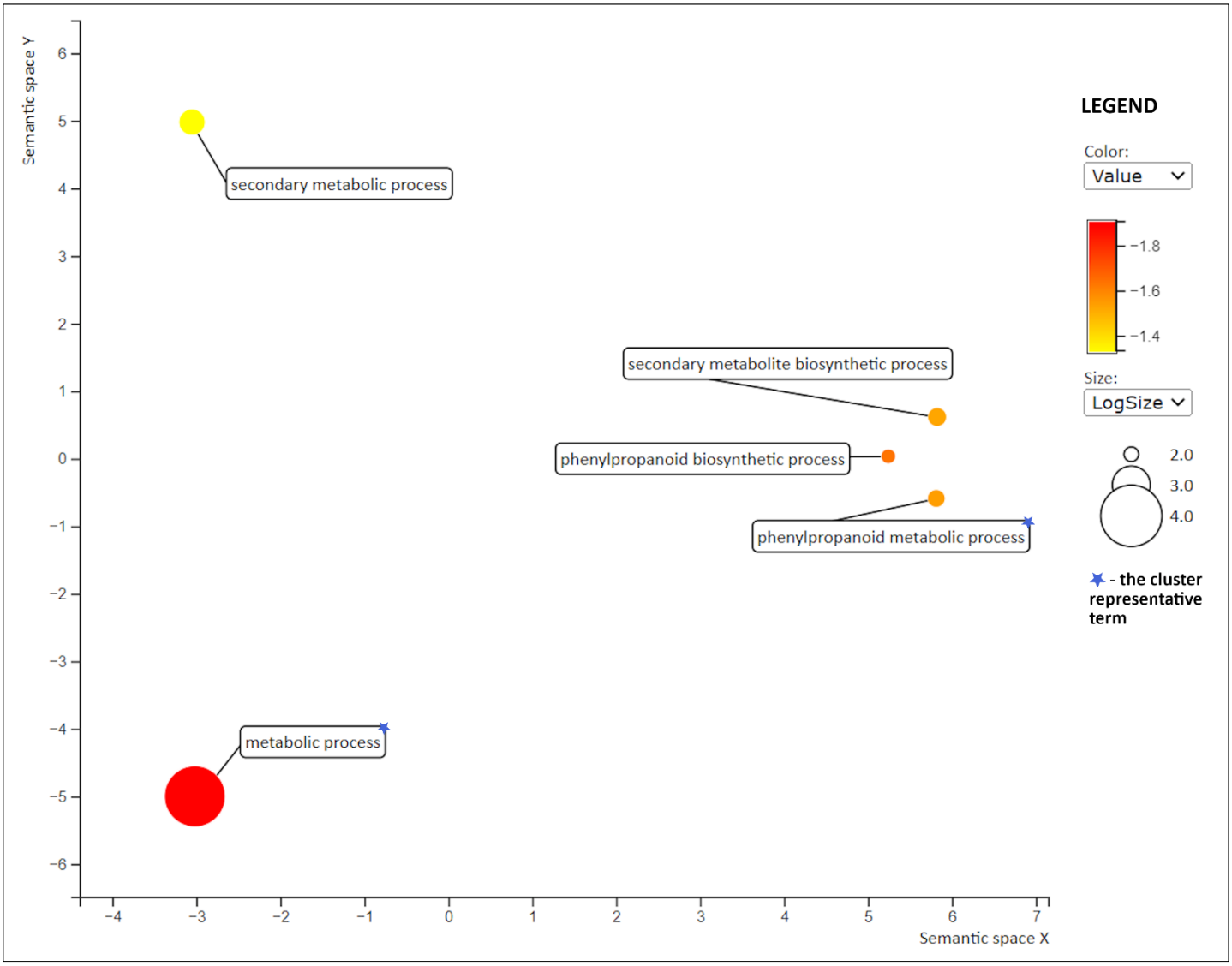



**Fig. S3** Scatterplot of the GO terms from the Biological Process aspect enriched in differentially expressed genes up-regulated in Scots pine relative to two mountain pines.

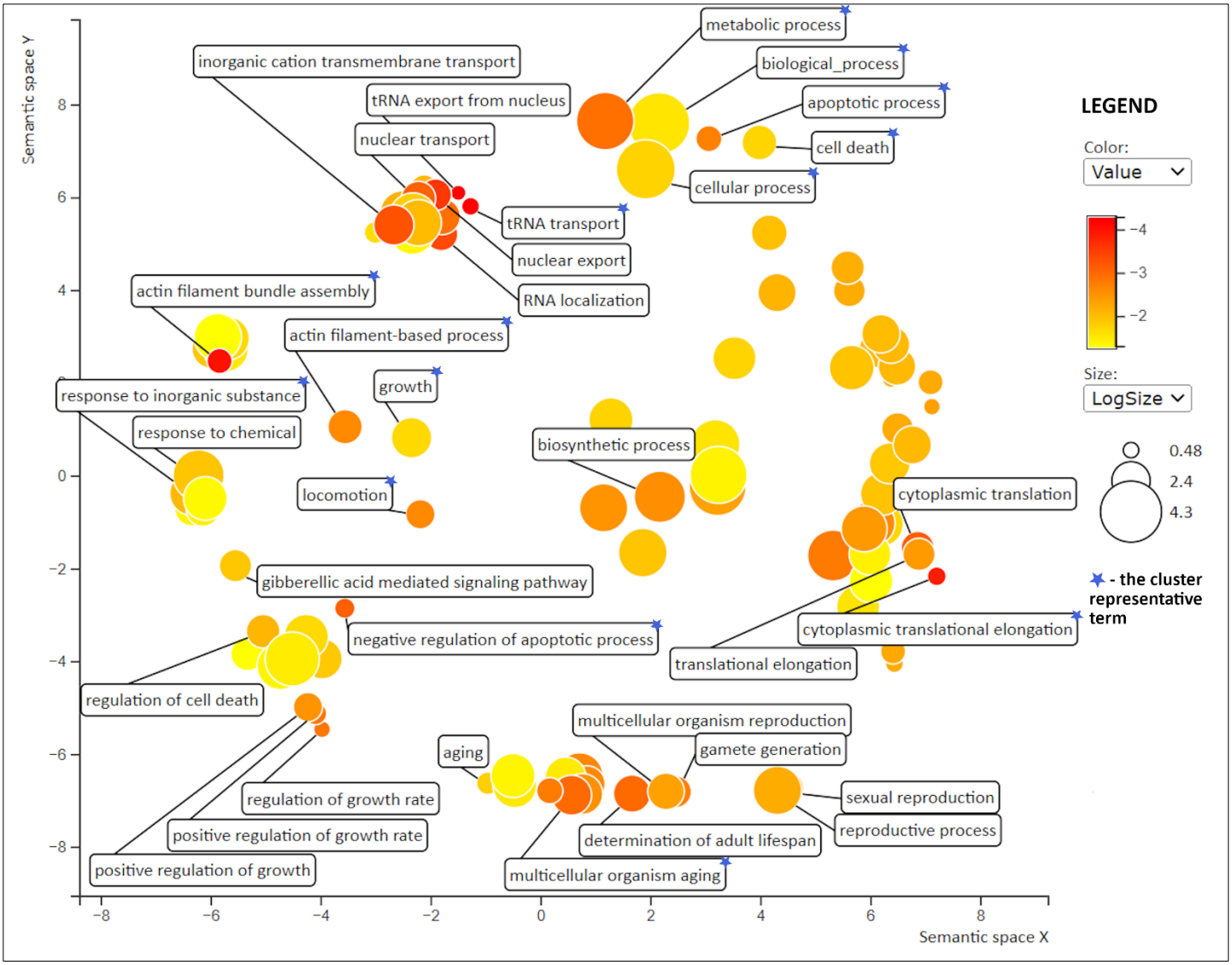



**Fig. S4** Scatterplot of the GO terms from the Biological Process aspect enriched in differentially expressed genes up-regulated in *P. mugo* compared to *P. uncinata*.

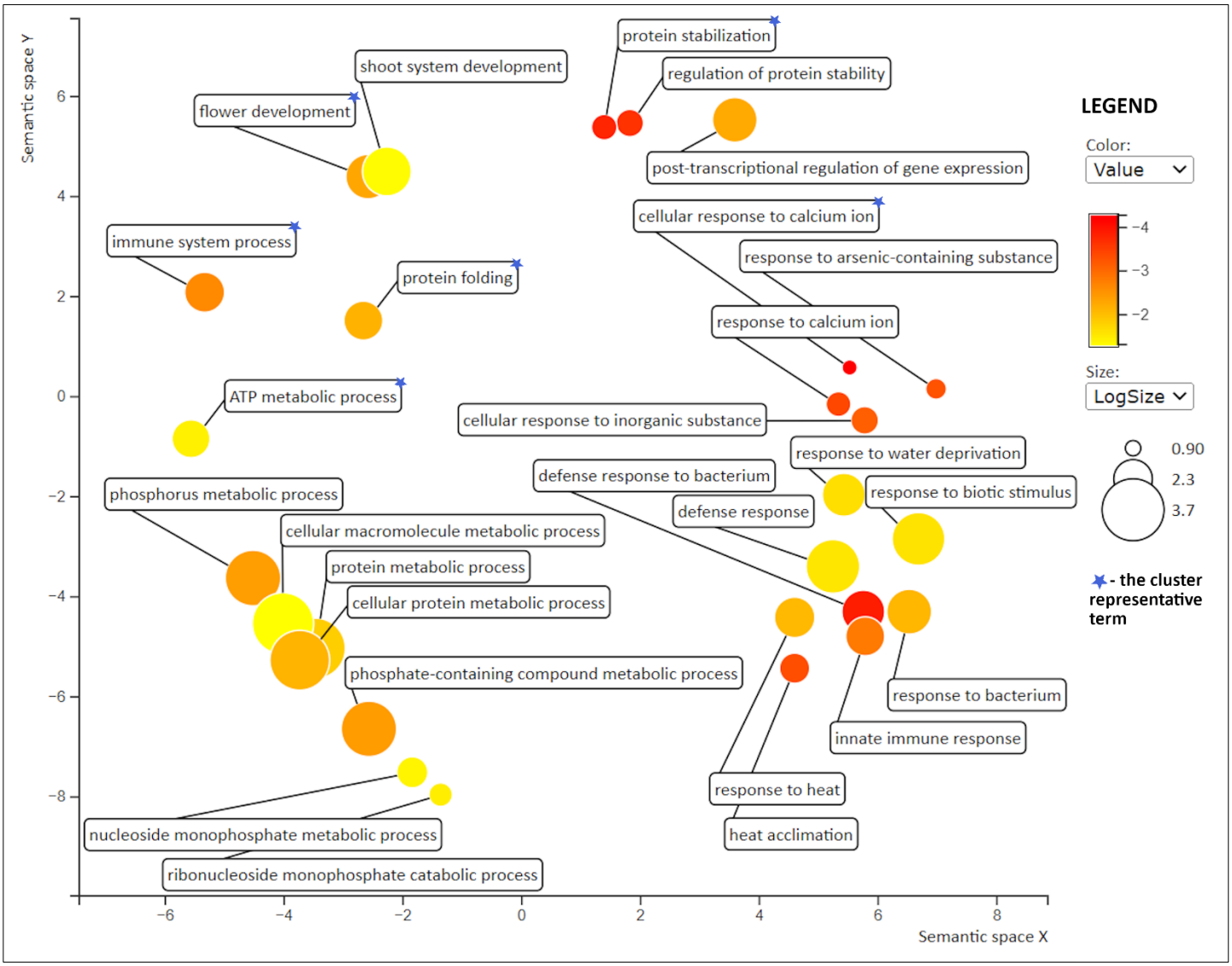



**Fig. S5** Scatterplot of the GO terms from the Biological Process aspect enriched in differentially expressed genes up-regulated in *P. uncinata* relative to *P. mugo*.

# 4. OŚWIADCZENIA

## oświadczenia autorów o ich wkładzie w powstanie publikacji wchodzących w skład dysertacji

## 4.1. OŚWIADCZENIE KANDYDATKI O JEJ WKŁADZIE W POWSTANIE KAŻDEJ Z TRZECH PUBLIKACJI

Poznań, 23.06.2022

**Mgr Julia Zaborowska**
Zakład Ekologii Roślin i Ochrony Środowiska
Instytut Biologii Środowiska, Wydział Biologii
Uniwersytet im. Adama Mickiewicza w Poznaniu
julia.zaborowska@amu.edu.pl

**Oświadczenie określające wkład w powstanie publikacji**

Niniejszym oświadczam, że jestem współautorką, pierwszą autorką, zarazem autorką korespondencyjną trzech niżej wymienionych ubikacji:

I. Zaborowska J, Łabiszak 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

Mój udział w powstawaniu publikacji pierwszej polegał na współtworzeniu koncepcji badania, a następnie przeprowadzeniu większości procedur molekularnych związanych z genotypowaniem prób (namnożenie fragmentów *mt*DNA, trawienie enzymatyczne i elektroforezą, przygotowanie prób do sekwencjonowania metodą Sangera), a także ich wstępnym filtrowaniu. Przeprowadziłam całość przedstawionych analiz z zakresu genetyki populacyjnej i filogeografii (wyliczenie statystyk zmienności i zróżnicowania genetycznego populacji, opracowanie wszystkich zamieszczonych tabel i rycin oraz całości dołączonych materiałów uzupełniających). Byłam główną osobą odpowiedzialną za interpretację rezultatów oraz przygotowanie manuskryptu i późniejsze jego korekty. Swój wkład w powstanie tego artykułu oceniam na 75%.

II. Zaborowska J, Łabiszak 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

Do powstania publikacji drugiej przyczyniłam się poprzez współtworzenie planu badań, a później ocenę i selekcję danych wejściowych, analizę materiału (ocenę zmienności i zróżnicowania genetycznego populacji a także ich grupowania, identyfikację 'outlierów' i przypisanie im potencjalnych funkcji) wraz z interpretacją wyników, napisanie pierwotnej wersji artykułu i jej późniejsze dopracowywanie. Wszystkie zamieszczone w artykule tabele i ryciny zostały wykonane przeze mnie, podobnie całość materiałów uzupełniających. Swój ostateczny wkład w jej powstanie oceniam na 65%.

III. Zaborowska J, Perry A, Cavers S, Wachowaik W. 2022 [in press]. **Evolutionary targets of gene expression divergence in a complex of closely related pine species**. *Journal of Systematics and Evolution* DOI: https://doi.org/10.1111/jse.12896

W procesie powstawania publikacji trzeciej uczestniczyłam biorąc udział w kształtowaniu planu badań i w selekcji danych wejściowych, byłam odpowiedzialna za całość przedstawionych analiz bioinformatycznych, za opracowanie wyników włącznie z przygotowaniem rycin i tabel, a także za przygotowanie zrębu maszynopisu i późniejsze jego modyfikacje - po uwzględnieniu sugestii współautorów i recenzentów. Jestem autorką wszystkich materiałów dodatkowych zamieszczonych w suplemencie. Mój całkowity wkład w powstanie publikacji trzeciej to 75%.

## 4.2. OŚWIADCZENIA WSPÓŁAUTORÓW O ICH WKŁADZIE W POWSTANIE PUBLIKACJI PIERWSZEJ

Zaborowska J, Łabiszak 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

Poznań, 05.07.2022

**Dr Bartosz Łabiszak**
Zakład Ekologii Roślin i Ochrony Środowiska
Instytut Biologii Środowiska, Wydział Biologii
Uniwersytet im. Adama Mickiewicza w Poznaniu
ul. Uniwersytetu Poznańskiego 6, 61-614 Poznań
bartosz.labiszak@amu.edu.pl

**Oświadczenie określające wkład w powstanie publikacji**

Niniejszym oświadczam, że jestem współautorem publikacji:

Zaborowska J, Łabiszak 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

Mój wkład w jej powstanie polegał na wygenerowaniu części danych mitochondrialnych, pomocy przy ich analizie, a także korekcie manuskryptu. Swój wkład oceniam na 10%.

Poznań, 24.06.2022

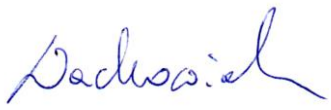
**Prof. dr hab. Witold Wachowiak**
Zakład Ekologii Roślin i Ochrony Środowiska
Instytut Biologii Środowiska, Wydział Biologii
Uniwersytet im. Adama Mickiewicza w Poznaniu
ul. Uniwersytetu Poznańskiego 6, 61-614 Poznań
witwac@amu.edu.pl

**Oświadczenie określające wkład w powstanie publikacji**

Niniejszym oświadczam, że jestem współautorem publikacji:

Zaborowska J, Łabiszak 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

Mój wkład w powstanie publikacji polegał na opracowaniu planu badań, dostarczeniu materiału do analiz, pomocy przy opracowaniu i interpretacji wyników, a także udziale w opracowaniu maszynopisu i korekcie publikacji. Swój wkład oceniam na 15%.



## 4.3. OŚWIADCZENIA WSPÓŁAUTORÓW O ICH WKŁADZIE W POWSTANIE PUBLIKACJI DRUGIEJ

Zaborowska J, Łabiszak 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

Poznań, 05.07.2022

**Dr Bartosz Łabiszak**
Zakład Ekologii Roślin i Ochrony Środowiska
Instytut Biologii Środowiska, Wydział Biologii
Uniwersytet im. Adama Mickiewicza w Poznaniu
ul. Uniwersytetu Poznańskiego 6, 61-614 Poznań
bartosz.labiszak@amu.edu.pl

**Oświadczenie określające wkład w powstanie publikacji**

Niniejszym oświadczam, że jestem współautorem publikacji:

Zaborowska J, Łabiszak 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

Mój wkład w powstanie publikacji polegał na pomocy przy analizie SNP w programie STRUCTURE oraz pakietach środowiska R, a także na konsultacji treści manuskryptu. Swój wkład oceniam na 5%.

Łabiszak

Edinburgh, 14.06.2022

**Dr. Annika Perry**
UK Centre for Ecology & Hydrology
Bush Estate
Penicuik, Midlothian
EH26 0QB, UK
annt@ceh.ac.uk

**Declaration of contribution to the publication**

I hereby confirm that I am a co-author of the publication:

Zaborowska J, Łabiszak 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

My contribution was the participation in development of the SNP genotyping array used in the study and genotyping of the tested samples. I estimate my total contribution to this article at 10%.

Edinburgh, 14.06.2022

**Dr. Stephen Cavers**
UK Centre for Ecology & Hydrology
Bush Estate
Penicuik, Midlothian
EH26 0QB, UK
scav@ceh.ac.uk

**Declaration of contribution to the publication**

I hereby confirm that I am a co-author of the publication:

Zaborowska J, Łabiszak 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

My contribution was related to development of the Axiom SNP array used in the study and to critical revision of the manuscript. I estimate my total contribution to this article at 5%.

Dr Stephen Cavers.

Poznań, 24.06.2022

**Prof. dr hab. Witold Wachowiak**
Zakład Ekologii Roślin i Ochrony Środowiska
Instytut Biologii Środowiska, Wydział Biologii
Uniwersytet im. Adama Mickiewicza w Poznaniu
ul. Uniwersytetu Poznańskiego 6, 61-614 Poznań
witwac@amu.edu.pl

**Oświadczenie określające wkład w powstanie publikacji**

Niniejszym oświadczam, że jestem współautorem publikacji:

Zaborowska J, Łabiszak 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

Mój wkład w powstanie publikacji polegał na współtworzeniu koncepcji badań, nadzorowaniu i udziale w analizie danych SNP oraz na interpretacji wyników i krytycznej recenzji maszynopisu artykułu. Swój wkład oceniam na 15%.

## 4.4. OŚWIADCZENIA WSPÓŁAUTORÓW O ICH WKŁADZIE W POWSTANIE PUBLIKACJI TRZECIEJ

Zaborowska J, Perry A, Cavers S, Wachowiak W. 2022. **Evolutionary targets of gene expression divergence in a complex of closely related pine species**. *Journal of Systematics and Evolution* in press (DOI: https://doi.org/10.1111/jse.12896).

Edinburgh, 14.06.2022

**Dr. Annika Perry**
UK Centre for Ecology & Hydrology
Bush Estate
Penicuik, Midlothian
EH26 0QB, UK
annt@ceh.ac.uk

**Declaration of contribution to the publication**

I hereby confirm that I am a co-author of the publication:

Zaborowska J, Perry A, Cavers S, Wachowiak W. 2022. **Evolutionary targets of gene expression divergence in a complex of closely related pine species**. *Journal of Systematics and Evolution* **https://doi.org/10.1111/jse.12896**

My contribution involves molecular laboratory procedures related to acquisition of the transcriptome used in the study. I have also participated in revision of the final article. I estimate my total contribution to this article at 5%.

Edinburgh, 14.06.2022

**Dr. Stephen Cavers**
UK Centre for Ecology & Hydrology
Bush Estate
Penicuik, Midlothian
EH26 0QB, UK
scav@ceh.ac.uk

**Declaration of contribution to the publication**

I hereby confirm that I am a co-author of the publication:

Zaborowska J, Perry A, Cavers S, Wachowiak W. 2022. **Evolutionary targets of gene expression divergence in a complex of closely related pine species**. *Journal of Systematics and Evolution* **https://doi.org/10.1111/jse.12896**

My contribution was related to generation of the transcriptome sequences used in the study and critical revision of the manuscript. I estimate my total contribution to this article at 5 %.

Dr Stephen Cavers.

Poznań, 24.06.2022

**Prof. dr hab. Witold Wachowiak**
Zakład Ekologii Roślin i Ochrony Środowiska
Instytut Biologii Środowiska, Wydział Biologii
Uniwersytet im. Adama Mickiewicza w Poznaniu
ul. Uniwersytetu Poznańskiego 6, 61-614 Poznań
witwac@amu.edu.pl

**Oświadczenie określające wkład w powstanie publikacji**

Niniejszym oświadczam, że jestem współautorem publikacji:

Zaborowska J, Perry A, Cavers S, Wachowaik W. 2022. **Evolutionary targets of gene expression divergence in a complex of closely related pine species**. *Journal of Systematics and Evolution* (DOI: https://doi.org/10.1111/jse.12896)

Mój wkład w powstanie publikacji polegał na współtworzeniu koncepcji analiz, dostarczeniu surowych danych transkryptomowych, konsultacji analiz i ich wyników. udziału w opracowaniu i rewizji maszynopisu artykułu. Swój wkład oceniam na 15%.