Wydział Biologii
Instytut Biologii Eksperymentalnej
Zakład Genetyki

JOANNA SIKORA

Zastosowanie metody przeczesywania genomu do charakterystyki różnorodności genetycznej blisko spokrewnionych taksonów z kompleksu *Pinus mugo*

Application of genome skimming to characterize the genetic diversity of closely related taxa in the *Pinus mugo* complex

ROZPRAWA DOKTORSKA

Promotor

prof. UAM dr hab. Konrad Celiński

Poznań, 2024

PODZIĘKOWANIA

Pragnę złożyć serdeczne podziękowania:

promotorowi **prof. UAM dr hab. Konradowi Celińskiemu** za wyrozumiałość, wsparcie merytoryczne i nieocenioną pomoc w trakcie realizacji i przygotowywania rozprawy doktorskiej.

Dziękuję również:

Pracownikom Zakładu Genetyki, Wydziału Biologii, Uniwersytetu im. Adama Mickiewicza w Poznaniu

dr Hannie Fuchs z Instytutu Dendrologii Polskiej Akademii Nauk w Kórniku

Pracownikom Laboratorium Technik Biologii Molekularnej, w szczególności prof. UAM dr hab. Mirosławie Dabert, mgr Julii Olechnowicz i Urszuli Sobczyńskiej

dr Pascalowi von Sengbusch z Biura raportów ekologicznych i nauk o torfowiskach, Kandern, Niemcy

mgr Ladislavowi Rektorisowi z Agencji Ochrony Przyrody i Krajobrazu Republiki Czeskiej, Obszar Chronionego Krajobrazu Třeboňsko, Czechy

Szczególne podziękowania składam moim najbliższym:

Rodzicom, Mężowi oraz Przyjaciołom

za motywację i wsparcie okazywane na każdym etapie mojego rozwoju naukowego.

FINANSOWANIE

Publikacje wchodzące w skład niniejszej rozprawy doktorskiej powstały w wyniku realizacji kierowanego przeze mnie projektu Diamentowy Grant Nr DI2017003147, pt.: „Rekonstrukcja relacji filogenetycznych w kompleksie *Pinus mugo* z zastosowaniem sekwencjonowania nowej generacji (ang. *next generation sequencing*) i metody przeczesywania genomu (ang. *genome skimming*)", finansowanego przez Ministerstwo Nauki i Szkolnictwa Wyższego w latach 2018-2022.

SPIS TREŚCI

STRESZCZENIE

Kompleks *Pinus mugo* to duża grupa blisko spokrewnionych europejskich sosen, które są składową ekosystemów najważniejszych pasm górskich, takich jak Alpy, Karpaty i Pireneje. Duża zmienność morfologiczna poszczególnych taksonów, występowanie sympatryczne, procesy hybrydyzacji i introgresji, obecność mieszańców oraz wielu nazw synonimicznych, a także złożona historia rekolonizacji powoduje, że kompleks *Pinus mugo* stanowi ciekawy, ale jednocześnie bardzo skomplikowany i wymagający obiekt badawczy. Relacje taksonomiczne i genetyczne między taksonami w tym kompleksie są przedmiotem badań i dyskusji od wielu lat. Pomimo zastosowania wielu różnych podejść badawczych, wciąż wymagają one wyjaśnienia i uzupełnienia o nowe dane i wyniki, zwłaszcza te uzyskane metodami wysokoprzepustowymi i narzędziami bioinformatycznymi.

W niniejszej rozprawie doktorskiej zastosowano metodę przeczesywania genomu (ang. *genome skimming*) i sekwencjonowanie nowej generacji (ang. *next-generation sequencing*, NGS) do: 1) charakterystyki zróżnicowania genetycznego sosen z kompleksu *Pinus mugo* i blisko z nimi spokrewnionych taksonów, tj. *Pinus sylvestris* i *Pinus* × *rhaetica*, 2) wytypowania regionów potencjalnie przydatnych do ich identyfikacji i dalszej analizy oraz 3) wnioskowania filogenetycznego w oparciu o sekwencje pochodzące z genomu chloroplastowego, mitochondrialnego i jądrowego.

Metoda przeczesywania genomu to nowatorskie podejście, które umożliwia uzyskanie frakcji DNA, które występują w komórce w dużej liczbie kopii, tj. jądrowego rybosomalnego DNA (nrDNA) oraz częściowych lub kompletnych sekwencji plastomów i mitogenomów.

Na tej podstawie po raz pierwszy przeprowadzono szczegółową analizę porównawczą genomów chloroplastowych trzech głównych przedstawicieli kompleksu *Pinus mugo*, tj. *Pinus mugo* Turra, *Pinus uncinata* Ramond i *Pinus uliginosa* Neumann, jak również blisko spokrewnionej *Pinus sylvestris*. Uzyskane wyniki wykazały wysokie podobieństwo genomów chloroplastowych trzech analizowanych taksonów, pod względem długości, struktury i liczby genów, i potwierdziły ich bliskie relacje filogenetyczne. Dodatkowo, przeprowadzona również po raz pierwszy ocena zasadności stosowania jądrowego regionu ITS2 (ang. *internal transcribed spacer* 2) do identyfikacji taksonów w rodzinie Pinaceae, z uwzględnieniem taksonów z kompleksu *Pinus mugo*, wykazała, że użyteczność regionu ITS2 do tego celu jest dość mocno ograniczona. Za pomocą metody przeczesywania genomu wygenerowano również siedem zestawów danych złożonych z sekwencji nukleotydowych pochodzących z trzech genomów organellowych i wykorzystano je do oceny różnorodności genetycznej

taksonów i populacji należących do kompleksu *Pinus mugo* oraz określono ich przydatność do identyfikacji analizowanych taksonów z zastosowaniem trzech różnych metod delimitacji. Najwyższy procent sukcesu w rozróżnianiu taksonów z tego kompleksu osiągnięto przy użyciu metody opartej na odległościach i zestawu złożonego z regionów wysoce zmiennych (tzw. *hotspotów*).

Uzyskane wyniki dostarczają nowych i obszernych zasobów genomicznych do dalszych badań nad kompleksem *Pinus mugo*, a także potwierdzają skuteczność i zasadność stosowania metody przeczesywania genomu do analizy różnorodności genetycznej blisko spokrewnionych taksonów.

Słowa kluczowe: kompleks *Pinus mugo*, przeczesywanie genomu, sekwencjonowanie nowej generacji, identyfikacja taksonów

ABSTRACT

The *Pinus mugo* complex is a large group of closely related European pines that are a component of the ecosystems of the most important mountain ranges, such as the Alps, the Carpathians and the Pyrenees. The high morphological variability of individual taxa, sympatric occurrence, hybridization and introgression processes, the presence of hybrids and many synonymous names, as well as the convoluted history of recolonization, make the *Pinus mugo* complex an interesting, but at the same time very complicated and demanding research object. Taxonomic and genetic relationships between taxa in this complex have been the subject of research and discussion for many years. Despite the use of many different research approaches, they still require clarification and extension with new data and results, especially those obtained using high-throughput methods and bioinformatic tools.

In this doctoral dissertation, genome skimming and next-generation sequencing were used to: 1) characterize the genetic diversity of the *Pinus mugo* complex pines and closely related taxa, i.e. *Pinus sylvestris* and *Pinus* × *rhaetica*, 2) select regions potentially useful for their identification and further analysis, and 3) make phylogenetic inferences based on sequences from the chloroplast, mitochondrial, and nuclear genomes.

Genome skimming is a novel approach that allows obtaining DNA fractions that occur in a cell in a high copy number, i.e. nuclear ribosomal DNA (nrDNA) and partial or complete sequences of plastomes and mitogenomes.

On this basis, a detailed comparative analysis of the chloroplast genomes of the three main representatives of the *Pinus mugo* complex, i.e. *Pinus mugo* Turra, *P. uncinata* Ramond and *Pinus uliginosa* Neumann, as well as the closely related *Pinus sylvestris*, was performed for the first time. The obtained results showed a high similarity of the chloroplast genomes of the three analyzed taxa, in terms of length, structure and number of genes, and confirmed their close phylogenetic relationships. Additionally, a first-ever assessment of the validity of using the nuclear internal transcribed spacer 2 (ITS2) region for identifying taxa in the Pinaceae family, including taxa from the *Pinus mugo* complex, showed that the usefulness of the ITS2 region for this purpose is quite limited. Seven data sets composed of nucleotide sequences from three organellar genomes were also generated using genome skimming and used to assess the genetic diversity of taxa and populations belonging to the *Pinus mugo* complex and to determine their suitability for identifying the analyzed taxa using three different

delimitation methods. The highest percentage of success in distinguishing taxa from this complex was achieved using the distance-based method and a set composed of highly variable regions (so-called hotspots).

The obtained results provide new and extensive genomic resources for further studies of the *Pinus mugo* complex, and confirm the effectiveness and validity of using the genome skimming method to analyze the genetic diversity of closely related taxa.

Keywords: *Pinus mugo* complex, genome skimming, next-generation sequencing, taxa identification

WYKAZ PRAC NAUKOWYCH WCHODZĄCYCH W SKŁAD CYKLU

1. **Sokołowska, J.;** Fuchs, H.; Celiński, K. New Insight into Taxonomy of European Mountain Pines, *Pinus mugo* Complex, Based on Complete Chloroplast Genomes Sequencing. Plants 2021, 10, 1331. DOI:10.3390/plants10071331.

   Q1(WoS), 71% (SCOPUS), IF = 2,632; (IF_5 = 4,4), MNiSW = 70 pkt

2. **Sokołowska, J.**; Fuchs, H.; Celiński, K. Assessment of ITS2 Region Relevance for Taxa Discrimination and Phylogenetic Inference among Pinaceae. Plants 2022, 11, 1078. DOI:10.3390/plants11081078.

   Q1(WoS), 83% (SCOPUS), IF = 2,632; (IF_5 = 4,4), MNiSW = 70 pkt

3. **Sikora, J.**; Celiński, K. Exploring taxonomic and genetic relationships in the *Pinus mugo* complex using genome skimming data. International Journal of Molecular Sciences 2024; 25(18):10178. DOI:10.3390/ijms251810178.

   Q1(WoS), 79% (SCOPUS), IF = 4,9; (IF_5 = 5,6), MNiSW = 140 pkt

Sumaryczny IF: **10,164**

Sumaryczna punktacja MNiSW/MEiN: **280**

1. Celiński K., **Sokołowska J.**, Fuchs, H., Maděra P., Wiland-Szymańska J.E. Characterization of the Complete Chloroplast Genome Sequence of the Socotra Dragon's Blood Tree (*Dracaena cinnabari* Balf.). Forests 2022; 13 (6), 1-12. DOI:10.3390/f13060932.

   Q1(WoS), 79%, IF = 2,116; (IF_5 = 2,453), MNiSW = 100 pkt

2. Celiński, K., **Sokołowska, J.**, Zemleduch-Barylska, A., Kuna, R., Kijak, H., Staszak, A.M., Wojnicka-Półtorak, A., Chudzińska, E. Seed Total Protein Profiling in Discrimination of Closely Related Pines: Evidence from the *Pinus mugo* Complex. Plants 2020; 9, 872. DOI:10.3390/plants9070872.

   Q1(WoS), 56%, IF =2,632; (IF_5 =4,4), MNiSW = 70 pkt

3. Celiński K., Kijak H., Wojnicka-Półtorak A., Buczkowska-Chmielewska K., **Sokołowska J.**, Chudzińska E. Effectiveness of the DNA barcoding approach for closely related conifers discrimination: A case study of the *Pinus mugo* complex. Comptes Rendus Biologies 2017; 340: 339–348. DOI:10.1016/j.crvi.2017.06.002.

   Q3(WoS), 73%, IF = 1,313; (IF5=1,330), MNiSW = 30 pkt

1. WSTĘP

Definiowanie i wyznaczanie kategorii taksonomicznych, w tym gatunków, było domeną taksonomii od czasów Karola Linneusza. Identyfikacja, delimitacja i opis gatunków stanowi podstawę badań taksonomicznych oraz filogenetycznych i jest kluczowa dla ochrony rzadkich i zagrożonych gatunków, wspierając działania na rzecz zachowania różnorodności biologicznej. Tradycyjne metody identyfikacji gatunkowej opierają się na cechach morfologicznych, co może stanowić problem w przypadku gatunków kryptycznych i zjawiska plastyczności fenotypowej (Herbert i in. 2003; Jarman i Elliott 2000; Ragupathy i in. 2009). Ponadto klucze oparte na cechach morfologicznych mogą okazać się skuteczne jedynie w odniesieniu do określonego etapu rozwojowego lub płci. Z kolei nowoczesne markery molekularne są stabilne, wykrywalne we wszystkich tkankach, niezależnie od etapu rozwojowego i nie są podatne na działanie warunków środowiskowych (Agarwal i in. 2008), a także umożliwiają identyfikację gatunków na podstawie ich śladowych fragmentów.

Jednakże, granice międzygatunkowe mogą być trudne do określenia, w przypadku gatunków, które powstały w wyniku niedawnej dywergencji, ze względu na niewielkie zróżnicowanie genetyczne, jak również występowanie niezgodności między zestawami cech lub topologiami drzew filogenetycznych opartych na różnych zestawach genów (Harrison i in. 2014). Niezgodność ta może wynikać zarówno z trwającej wymiany genów, jak i długiego czasu potrzebnego do osiągnięcia monofiletyczności w większości *loci* (Hudson i Coyne 2002). Główna trudność w wyznaczaniu granic międzygatunkowych występuje na początku formowania się gatunków. Jest to szczególnie skomplikowane, gdy gatunki wykazują dużą zmienność morfologiczną, przy jednocześnie niewielkiej zmienności genetycznej, co powszechnie występuje w procesie radiacji ewolucyjnej (Shaffer i Thomson 2007). Ponadto, dodatkową trudność w delimitacji gatunków, potęguje brak jasności i jednomyślności co do samej definicji gatunku (de Queiroz 2007).

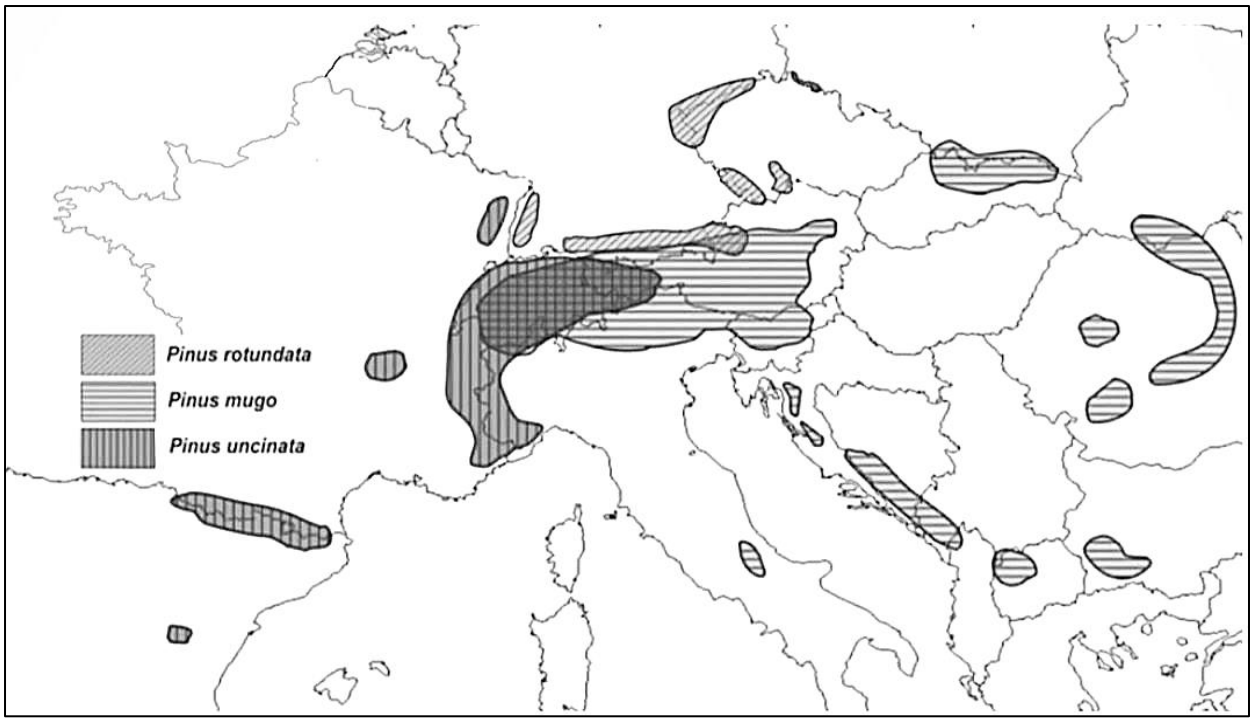
W rodzinie *Pinaceae* występuje wiele blisko spokrewnionych taksonów o bardzo podobnej morfologii igieł lub szyszek, zajmujących te same nisze ekologiczne, mające podobne zasięgi geograficzne i tworzące populacje sympatryczne. W takich populacjach dochodzi do przepływu genów, co prowadzi do powstawania osobników o fenotypie pośrednim w stosunku do taksonów macierzystych, a to dodatkowo utrudnia proste i skuteczne przyporządkowanie ich do konkretnych gatunków. A bez takiej jednoznacznej

identyfikacji trudno prowadzić efektywnie działania ochronne, szczególnie w przypadku taksonów zagrożonych.

Blisko spokrewnione taksony, które charakteryzuje bardzo zbliżona morfologia i brak jasnych determinant gatunkowych grupuje się w większe jednostki taksonomiczne zwane kompleksami. Kompleksy gatunków to grupy, w których wyznaczniki gatunkowe i liczebność są często niejasne. Niejednoznaczność ta zwykle wynika z wielu czynników, w tym m.in. ukrytej zmienności morfologicznej, niewystarczająco szerokiego próbkowania, rozległych zasięgów geograficznych, hybrydyzacji, introgresji czy niedawnej dywergencji (Grube i Kroken 2000). Badania delimitacji taksonów zgrupowanych w ramach takich kompleksów gatunkowych jest niezwykle ważne, ponieważ umożliwia lepsze zrozumienie pierwszych etapów formowania się gatunków. Ponadto niepełna wiedza na temat stanu bioróżnorodności w kompleksach gatunków może spowodować nieprawidłową ocenę bioróżnorodności, biogeografii i procesów specjacji całej grupy (Heath i in. 2008).

Jednym z najlepszych przykładów takiego złożonego systemu obejmującego wszystkie wyżej wymienione czynniki (problemy) jest europejski kompleks sosen górskich – *Pinus mugo* Turra *sensu lato*, złożony z grupy blisko spokrewnionych taksonów z rodziny Pinaceae, podrodzaju *Pinus*, sekcji *Pinus*, podsekcji *Pinus* (Gernandt i in. 2005). Jego naturalny zasięg występowania obejmuje góry południowej i środkowej Europy, w tym Pireneje, Sierra Cebollera, Sierra de Gudar, Masyw Centralny, Wogezy, Schwarzwald, Jura, Alpy, Karpaty, Rodopy, Abruzzy, Sudety (Franco 1986; Christensen 1987b; Jalas i Suominen 1973; Hamerník i Musil 2007; Rycina 1).

Na przestrzeni lat, w obrębie tego kompleksu opisano 16 gatunków, 91 odmian i 19 form (Christensen 1987b). Taksony wchodzące w skład kompleksu są zróżnicowane pod względem cech morfologicznych oraz wymagań ekologicznych (Hamernik i Musil 2007). Wyszczególniając, kosodrzewina (*Pinus mugo* Turra *sensu stricto*) ma typowy polikormiczny (krzewiasty) pokrój i występuje powszechnie w pasie subalpejskim w Europie Środkowej i Południowo-Wschodniej, tj. Alpach, Sudetach, Karpatach (Jalas i Suominen 1973; Rycina 1). Z kolei sosna hakowata (*Pinus uncinata* Ramond *sensu stricto*) charakteryzuje się monokormicznym pokrojem, osiągając między 12 a 20 m wysokości i występuje głównie na stanowiskach górskich Europy Zachodniej, tj. w Pirenejach, Alpach Zachodnich (1400–2700 m n.p.m.), Sierra de Gudar, Sierra Cebollera, Masywie Centralnym, Jurze i Wogezach (Jalas i Souminen 1973; Franco 1986; Rycina 1).



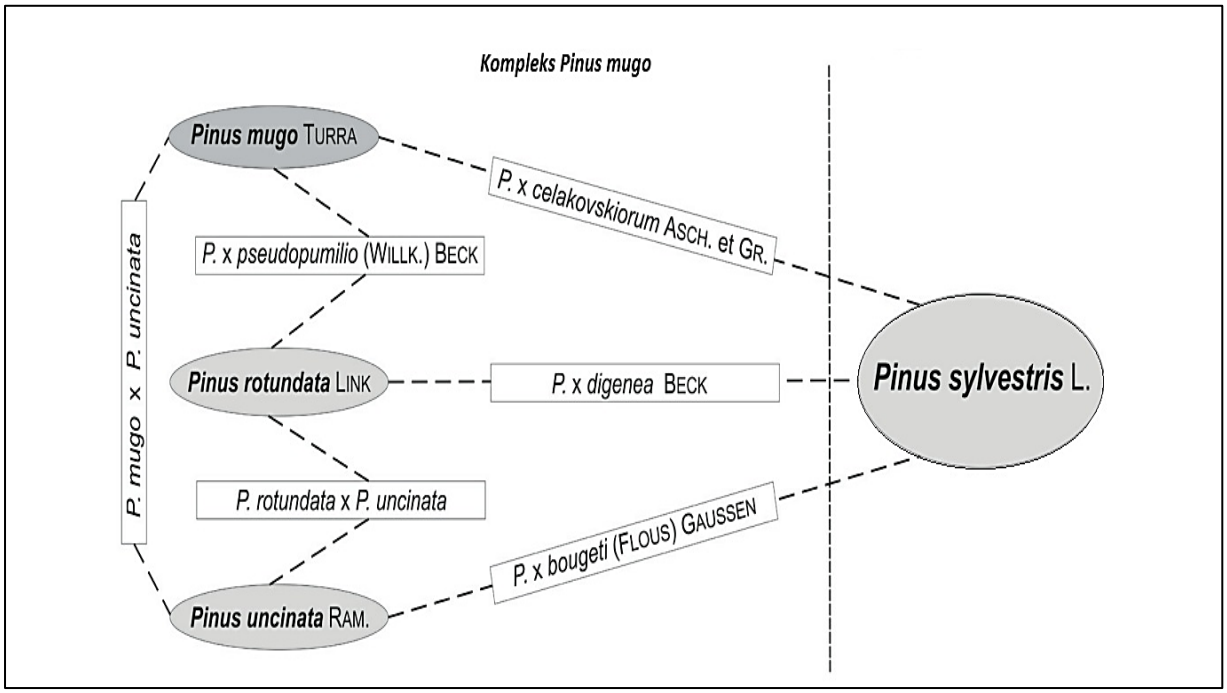
Rycina 1. Zasięg geograficzny *Pinus rotundata*, *Pinus uncinata* i *Pinus mugo*, zaadaptowany z Jalas i Suominen, 1973 (za: Bastl 2008, zmienione).

Natomiast *Pinus rotundata* Link jest endemicznym taksonem o pokroju drzewiastym spotykanym na podgórskich torfowiskach w Europie Środkowej (Jalas i Suominen 1973; Rycina 1). Takson ten krzyżuje się z kosodrzewiną, tworząc mieszańca znanego jako *P.* × *pseudopumilio* (Willk.) Beck, jak również z sosną zwyczajną tworząc wtedy *P.* × *digenea* Beck (Skalický 1988, Rycina 2).

Niektórzy badacze traktują kompleks *Pinus mugo* jako jeden, polimorficzny gatunek bez wyróżniania w nim taksonów wewnątrzgatunkowych (np. Schlechtendal 1857; Christ 1863; Goppert 1864; Shaw 1914). Wg Christensena (1987 a,b) w obrębie kompleksu można wyróżnić natomiast dwa podgatunki, tj. *Pinus mugo* Turra ssp. *mugo* i *Pinus mugo* Turra ssp. *uncinata* (Ram.) Domin, oraz jeden takson mieszańcowy, powstały w wyniku krzyżowania się wyżej wymienionych dwóch taksonów – *Pinus mugo* Turra nothossp. *rotundata* (Link) Janchen & Neumayer. Z kolei Businský (Businský 1999) wyodrębnił trzy osobne taksony w randze gatunku, tj. *Pinus mugo* Turra, *Pinus uncinata* Ramond i *Pinus rotundata* Link (Rycina 2).

Identyfikacja oraz delimitacja taksonów z kompleksu *Pinus mugo* stanowiła i nadal stanowi duże wyzwanie ze względu na ich dużą zmienność morfologiczną (Christensen 1987a; Christensen 1987b; Boratyńska i Boratyński 2007; Marcysiak i Boratyński, 2007) oraz

naturalną hybrydyzację w strefach kontaktu z sosną zwyczajną, *Pinus sylvestris* L. (Christensen 1987a, b; Wachowiak i Prus-Głowacki 2008; Lewandowski i Wiśniewska 2006), Między innymi przez to w literaturze naukowej często pojawiały się nieścisłości dotyczące interpretacji cech fenotypowych lub nazw poszczególnych taksonów (Businský 1999; Hamerník i Musil 2007).



Rycina 2. Zależności taksonomiczne w kompleksie *Pinus mugo*, z uwzględnieniem hybrydyzacji w obrębie kompleksu oraz z *Pinus sylvestris* L. (za: Hamernik i Musil 2007, zmienione).

Tytułem przykładu, taksonem posiadającym różne nazwy oraz rangi taksonomiczne jest sosna błotna, która znana jest m.in. jako *Pinus uliginosa* Neumann ex Wimmer (Neumann 1837), *Pinus rotundata* Link (Jalas i Suominen 1973; Skalická i Skalický 1988), *Pinus mugo* subsp. *rotundata* (Link) Janch. et Neumayer (Domin 1936, Jasicova 1966) czy *Pinus* × *rhaetica* Brügger (Staszkiewicz 1993). Po raz pierwszy została opisana w 1837 roku w Parku Narodowym Gór Stołowych, na „Wielkim Torfowisku Batorowskim" (tzw. *locus classicus*) (Neumann 1837). Ze względu na zmniejszającą się liczebność populacji i brak naturalnej regeneracji (Gołąb 1999), a także ryzyko zanieczyszczenie puli genowej poprzez wymianę genów z sosną zwyczajną i kosodrzewiną, sosna błotna jest gatunkiem zagrożonym na wyginięcie i objętym ścisłą ochroną w Polsce pod nazwą *Pinus* × *rhaetica* (Dz. U. z 2014 r. poz. 1409).

Wspólnym synonimem używanym dla sosny błotnej w czeskich i niemieckich populacjach jest *Pinus rotundata* Link (Christensen 1987; Businsky´ 1999), chociaż w późniejszym czasie sugerowano, że *Pinus rotundata* Link i *Pinus uliginosa* Neumann ex Wimmer mogą być odrębnymi gatunkami (Businsky´i Kirschner 2006).

Pochodzenie *Pinus uliginosa* Neumann ex Wimmer nie jest do końca ustalone, jednakże na podstawie badanych cech biochemicznych (allozymy) i biometrycznych (typy komórek sklerenchymatycznych) postawiono hipotezę, że *P. uliginosa* powstała w wyniku hybrydyzacji pomiędzy *Pinus sylvestris* oraz *Pinus mugo* i/lub taksonów z kompleksu *Pinus mugo* (Lewandowski i in. 2000; Boratyńska i Boratyński 2007).

Określenie wzajemnych relacji taksonomicznych w obrębie kompleksu *Pinus mugo* było przedmiotem wielu badań prowadzonych różnymi metodami i technikami. Jednakże, uzyskane w nich wyniki nie są jednoznaczne. W badaniach tych wykorzystano, m.in.: cechy morfologiczne i anatomiczne igieł (np. Christensen 1987a, Boratyńska i Bobowicz 2001; Boratyńska 2004; Boratyńska i Boratyński 2007; Boratyńska i in. 2015); cechy morfologiczne szyszek (np. Christensen 1987a; Monteleone i in. 2006; Marcysiak i Boratyński 2007); metody serologiczne (Prus-Głowacki i Szweykowski 1979; Prus-Głowacki i in. 1985); izoenzymy (Prus-Głowacki i Szweykowski 1983; Siedlewska i Prus-Głowacki 1995; Prus-Głowacki i in. 1998; Bobowicz i in. 2000); allozymy (Lewandowski i in. 2000); markery RAPD (ang. Randomly Amplified Polymorphic DNA) (Monteleone i in. 2006); metody cytogenetyki molekularnej i cytometrii przepływowej (Bogunić i in. 2011; Celiński i in. 2019); badania transkryptomiczne (Wachowiak i in. 2015); analizy chemotaksonomiczne (Bonikowski i in. 2015; Celiński i in. 2015); profilowanie białek nasion (Celiński i in. 2020) czy barkoding DNA (Celiński i in. 2017).

Na podstawie przeprowadzonych badań, wykazano że taksony wchodzące w skład kompleksu *Pinus mugo,* charakteryzują się dużą zmiennością morfologiczną, a jednocześnie niewspółmiernie niskim zróżnicowaniem genetycznym. Wysokie podobieństwo genetyczne zaobserwowano z wykorzystaniem 64 polimorficznych *loci* RAPD (ang. *Randomly Amplified Polymorphic* DNA, RAPD) (Monteleone i in. 2006) pomiędzy alpejskimi populacjami *Pinus mugo* subsp. *mugo* i *Pinus mugo* subsp. *uncinata*, jak również chloroplastowych *loci* mikrosatelitarnych, tj. Pt15169, Pt41093 i Pt71936 (ang. *chloroplast Simple Sequence Repeats*, cpSSRs) u *Pinus uncinata* Ram. i zachodnich populacji *Pinus mugo* Turra *s.s.* oraz *P. rotundata* Link/*P. × pseudopumilio* (Willk.) Beck. (Heuertz i in.

2010) czy 12 *loci* jądrowych potencjalnie związanych ze stresem odwodnieniowym i/lub tolerancją na zimno u *Pinus mugo* Turra s.l., *P. uncinata* Ramond i *P. uliginosa* (Wachowiak i in. 2013).

Analiza porównawcza z wykorzystaniem ośmiu chloroplastowych regionów barkodowych (tj. *matK, rbcL, trnH-psbA, trnL-trnF, rpl20-rps18, trnV, ycf1, ycf2*) o łącznej długości 5361 nukleotydów, również nie wykazała istnienia różnic w analizowanych sekwencjach pomiędzy blisko spokrewnionymi taksonami – *P. mugo, P. uncinata* i *P. uliginosa* (Celiński i in. 2017a). Co więcej, w przypadku *Pinus mugo* i *Pinus uncinata* potwierdzono brak różnic w organizacji i wielkości genomu (Bogunić i in. 2011), a w przypadku *P. uliginosa* i *Pinus* × *rhaetica* nie zaobserwowano różnic w zawartości DNA względem postulowanych gatunków rodzicielskich, tj. *P. sylvestris* i *P. mugo* (Celiński i in. 2019). Wykazano również, że niektóre taksony *P. mugo s.l.* i *P. sylvestris* posiadają wspólne haplotypy chloroplastów (Heuertz i in. 2010). Z drugiej strony, wykazano, że *Pinus uncinata* posiada unikalne mitotypy (Łabiszak i in. 2019). Przyczyną niskiego zróżnicowania genetycznego taksonów w obrębie kompleksu *Pinus mugo* może być krótki czas dywergencji, co wyjaśnia brak stałych różnic w analizowanych regionach DNA (Wachowiak i in. 2011).

Ostatnio prowadzone badania chemotaksonomiczne ujawniły specyficzne gatunkowo różnice w składzie olejków eterycznych i związków lotnych wyekstrahowanych z igieł niektórych przedstawicielach kompleksu *Pinus mugo,* tj. *Pinus uncinata, Pinus uliginosa* i *Pinus mugo* (Celiński i in. 2015). Z kolei analiza densytometryczna profili białkowych nasion (ang. *Seed Total Protein*, STP) ujawniła obecność charakterystycznych i diagnostycznych prążków dla *Pinus uliginosa*, *Pinus mugo* i *P.* × *rhaetica* (Celiński i in. 2020), co umożliwiło rozróżnienie tych blisko spokrewnionych taksonów. Uzyskane wyniki były zgodne z przeprowadzoną wcześniej analizą cytogenetyczną, która wykazała istnienie różnic między *Pinus* × *rhaetica* a *P. uliginosa* na podstawie wzorów pasm C na ich chromosomach (Celiński i in. 2019). Taki wynik sugeruje, że obie nazwy nie są tożsame i nie powinny być używane wymiennie jako synonimy.

Punktem wyjścia do przeprowadzenia moich badań były wyniki poprzednio prowadzonych prac, w których z jednej strony wykazano dużą zmienność morfologiczną i anatomiczną taksonów z kompleksu *Pinus mugo,* a z drugiej strony wykazano brak różnic specyficznych gatunkowo na poziomie sekwencji DNA w analizowanych wybranych rejonach barkodowych. Ponadto, dotychczas prowadzone badania nad kompleksem *Pinus*

mugo z zastosowaniem rozmaitych metod i technik nie dały pełnego obrazu złożoności tego kompleksu, gdyż nie uwzględniały wielu ważnych taksonów i populacji, np. z terenu Czech i Niemiec.

W związku z powyższym w niniejszej rozprawie doktorskiej po raz pierwszy zastosowano wysokoprzepustowe sekwencjonowanie nowej generacji (ang. *next-generation sequencing*, NGS) i metodę przeczesywania genomu (ang. *genome skimming*) do eksploracji nowych obszarów trzech genomów organellowych o znacznej długości w celu uchwycenia oraz odzwierciedlenia różnorodności i złożoności kompleksu *Pinus mugo*, na terenie Polski, Czech i Niemiec.

2. CEL PRACY

Głównymi celami badawczymi mojej rozprawy doktorskiej było: 1) charakterystyka różnorodności genetycznej wybranych taksonów z europejskiego kompleksu górskich sosen – *Pinus mugo* i blisko z nim spokrewnionych taksonów, tj. *Pinus sylvestris* i *Pinus* × *rhaetica*; 2) wytypowanie regionów potencjalnie przydatnych do ich identyfikacji i dalszej analizy; oraz 3) wnioskowanie filogenetyczne w oparciu o wyniki uzyskane dzięki sekwencjonowaniu nowej generacji i metodzie przeczesywania genomu.

Szczegółowymi celami badawczymi było:

1) sekwencjonowanie, analiza porównawcza i charakterystyka kompletnych genomów chloroplastowych *P. mugo, P. uliginosa* i *P. uncinata*;

2) identyfikacja i selekcja regionów wysoce zmiennych (tzw. *hotspotów*) w genomach chloroplastowych, potencjalnie przydatnych w identyfikacji taksonów z kompleksu *Pinus mugo*;

3) przeprowadzenie wnioskowania filogenetycznego na temat pokrewieństwa taksonów z kompleksu *Pinus mugo* w oparciu o kompletne genomy chloroplastowe i wybrane regiony DNA;

4) określenie przydatności regionu ITS2 do rozróżniania, identyfikacji i wnioskowania filogenetycznego u odległych i blisko spokrewnionych taksonów z rodziny *Pinaceae*;

5) wygenerowanie zróżnicowanych zestawów danych genomicznych, reprezentujących regiony pochodzące z genomu mitochondrialnego, chloroplastowego i jądrowego;

6) określenie zmienności genetycznej oraz przydatności zestawów danych genomicznych, z genomu mitochondrialnego, chloroplastowego i jądrowego w badaniach blisko spokrewnionych taksonów sosny;

7) określenie skuteczności rozróżniania analizowanych taksonów przy użyciu różnych zestawów danych i metod delimitacji;

8) ujawnienie powiązań genetycznych między *P.* × *rhaetica, P. rotundata* i *P. ulignosa*

9) zbadanie zróżnicowania genetycznego taksonów i populacji wchodzących w skład kompleksu *Pinus mugo.*

3. MATERIAŁY I METODY

W celu jak najlepszego odzwierciedlenia różnorodności i złożoności kompleksu *Pinus mugo* pozyskano materiał roślinny w postaci igieł, z terenu Polski, Czech i Niemiec dla następujących taksonów: *Pinus uliginosa*, *Pinus rotundata*, *Pinus* × *rhaetica*, *Pinus mugo*, *Pinus uncinata* i *Pinus sylvestris*. Zbiór materiału roślinnego obejmował: obszar ochronny Gąsiennicowa na terenie Tatrzańskiego Parku Narodowego, Rezerwat Przyrody „Wielkie Torfowisko Batorowskie", Rezerwat Przyrody „Bór nad Czerwonem", Rezerwat przyrody „Torfowisko pod Węglińcem", Ogród Dendrologiczny Uniwersytetu Przyrodniczego w Poznaniu, obszar Moraska, Poznań (Polska); Park Narodowy Szumawa (lokalizacja: „Novohůrecká slat"), Rezerwat Przyrody „Červené blato" (Czechy); Rezerwat Przyrody „Rotmeer", Rezerwat przyrody „Kirchspielwald-Ibacher Moos" oraz Rezerwat Przyrody „Steerenmoss", będące częścią obszaru Natura 2000 w południowym Schwarzwaldzie (Niemcy).

Genomowy DNA (ang. *genomic* DNA, gDNA), wyizolowany z igieł poszczególnych osobników, został wykorzystany do przygotowania bibliotek genomowych typu shotgun o długości 200–300 pz dla platformy Ion Torrent. W celu późniejszego zmultipleksowania (zmieszania) poszczególnych bibliotek genomowych przeprowadzono ligację indeksowanych adapterów (ang. *barcoded libraries*) przy użyciu zestawu Ion Xpress™ Barcode Adapters Kit. Analizę jakości i długości przygotowanych bibliotek przeprowadzono przy użyciu 2200 Tapestation Bioanalyzer (Agilent Technologies Waldbronn, Niemcy).

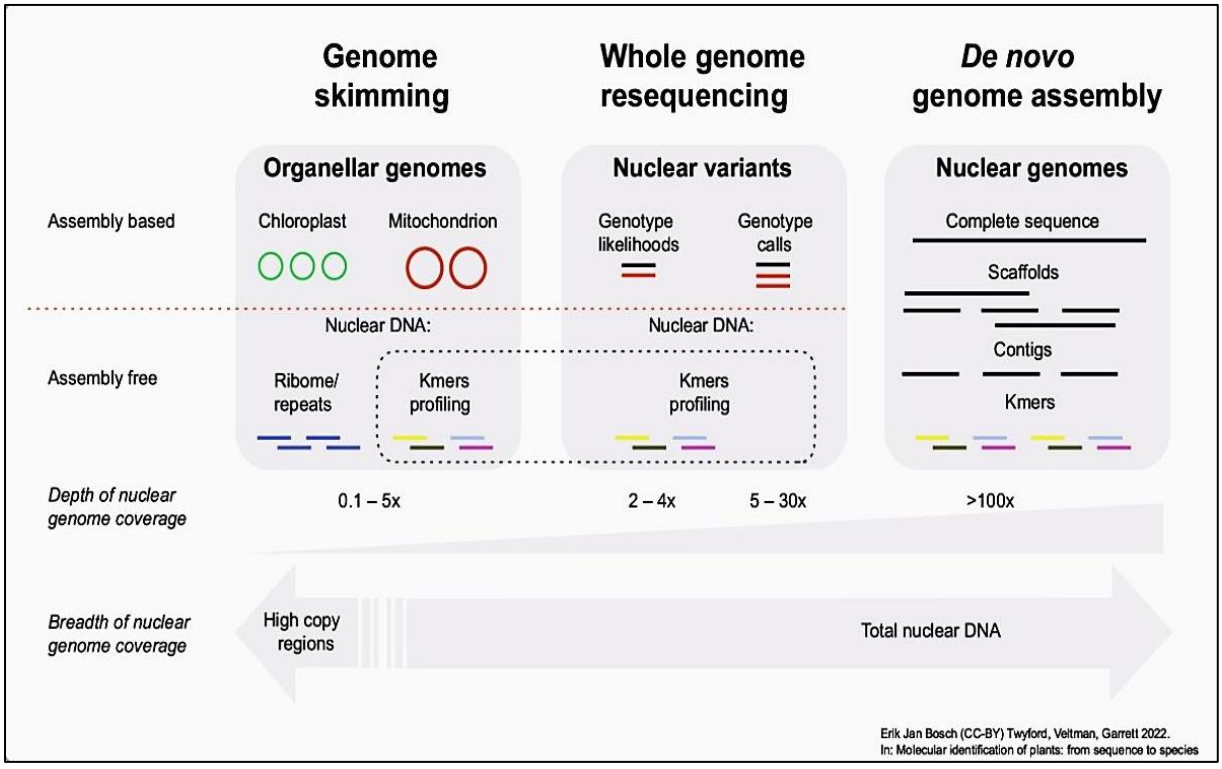
Sekwencjonowanie nowej generacji (ang. *next-generation sequencing*, NGS) uzyskanych bibliotek genomowych przeprowadzono w oparciu o chip Ion 318™ Chip v2 BC z wykorzystaniem systemu Ion Personal Genome Machine™ (PGM™) (Thermo Fisher Scientific, Waltham, USA), jak również w oparciu o chip Ion 540™ z wykorzystaniem Ion GeneStudio™ S5 System (Thermo Fisher Scientific, Waltham, USA).

Surowe dane uzyskane w postaci odczytów poddano kontroli jakości i filtrowaniu w celu usunięciu adapterów i odczytów o niskiej jakości, a następnie wykorzystano metodę przeczesywania genomu z zastosowaniem pakietu programów bioinformatycznych Geneious Prime 2020.2.5 (Kearse i in. 2012).

Metoda przeczesywania genomu (ang. *genome skimming)* umożliwia uzyskanie danych genomowych na dużą skalę w oparciu o frakcje DNA, pochodzące z płytkiego sekwencjonowania całego genomu (Straub i in. 2012; Dodsworth 2015, Rycina 3) i polega na ekstrakcji, składaniu oraz adnotacji frakcji genomowego DNA (ang. *genomic* DNA, gDNA)

występującego w komórce w wielu kopiach, tj.: dużych powtórzeń tandemowych rybosomalnego cistronu jądrowego (ang. *nuclear ribosomal* DNA, nrDNA), jak również chloroplastowego (ang. *chloroplast* DNA, cpDNA) i mitochondrialnego DNA (ang. *mitochondrial* DNA, mtDNA) przy pokryciu ok. 0,1–5x (Dodsworth i in. 2019).

W ostatnich latach podejście to znalazło zastosowanie w identyfikacji i delimitacji gatunków roślin (np. Nevill i in. 2020; Mascarello i in. 2021; Ji i in. 2022) czy poszukiwaniu nowych specyficznych markerów (np. Liu i in. 2021; Dobral i in. 2021; Mascarello i in. 2021). Przeczesywanie genomu zostało również wykorzystane do wnioskowania filogenetycznego na różnych poziomach taksonomicznych (np. Parks i in. 2009; Whittall i in.; 2010; Straub i in. 2012; Malé i in. 2014; Fonseca i Lohmann 2020; Simmonds i in. 2021; He i in. 2021). Ogromną zaletą tej metody jest prostota i stosunkowo niskie koszty analizy, w tym brak konieczności stosowania specjalnie zaprojektowanych par starterów ukierunkowanych na określone regiony genomu, jak również brak potrzeby posiadania genomu referencyjnego. Dodatkowo, metoda sprawdza się w przypadku sekwencjonowania zdegradowanego DNA pochodzącego z okazów muzealnych (Zeng i in. 2018; Liu i in. 2019; Nevill i in. 2020; Wang i in. 2020).



Rycina 3. Graficzne przedstawienie metody przeczesywania genomu (ang. *genome skimming*). (Za: de Boer i in. 2022, zmienione).

W związku z powyższym, zastosowanie takiego nowoczesnego i wysokoprzepustowego podejścia gwarantuje uzyskanie bardzo dużej ilości danych o podłożu genetycznym, które następnie dzięki zaawansowanym analizom bioinformatycznym i filogenetycznym pozwolą na wieloregionową i wielogenomową analizę różnorodności genetycznej taksonów należących do kompleksu *Pinus mugo* i odpowiedź na pytania o wzajemne relacje i status taksonomiczny rodzimej sosny błotnej *Pinus uliginosa* w stosunku do *P. mugo, P. rotundata* i *P.* × *rhaetica* oraz poszukiwanie markerów DNA specyficznych dla tych taksonów.

4. KOMENTARZE DO PUBLIKACJI

4.1. Publikacja 1

Ze względu na ograniczenia tradycyjnych regionów barkodowych DNA, które charakteryzują się ograniczoną zmiennością genetyczną i stosunkowo niską mocą dyskryminacyjną w przypadku identyfikacji blisko spokrewnionych taksonów (Li i in. 2014), zaistniała potrzeba znalezienia nowej metody do identyfikacji blisko spokrewnionych gatunków roślin. W ostatnich latach zaproponowano wykorzystanie genomu chloroplastowego jako „superbarkodu" na poziomie gatunku, który może znacząco zwiększyć rozdzielczość w analizach filogenetycznych, filogeograficznych oraz genetycznych (Parks i in. 2009).

W pierwszej części mojej rozprawy doktorskiej po raz pierwszy przeprowadziłam szczegółową, analizę porównawczą opartą na *de novo* zsekwencjonowanych kompletnych genomach chloroplastowych taksonów *Pinus mugo subsp. mugo*, tj. *Pinus mugo* (MZ3334), *Pinus mugo subsp. rotundata*, tj. *P. uliginosa* (MZ333465), *Pinus mugo subsp. uncinata* voucher 1347, tj. *P. uncinata* (MZ333464) oraz genomie referencyjnym blisko spokrewnionej *Pinus sylvestris* L. (KR476379) (Sokołowska i in. 2021).

Uzyskane wyniki wykazały wysoką konserwatywność genomów chloroplastowych, zarówno pod względem ich długości, struktury jak i liczby genów. Całkowita długość kompletnych plastomów *P. mugo*, *P. uliginosa* i *Pinus uncinata* wynosi odpowiednio 119 765 pz, 119 759 pz i 119 780 pz. Genomy te zawierają 121 genów, w tym: 73 geny kodujące białka, 36 genów tRNA i 4 geny rybosomalnego RNA. Podobnie, nie ma różnic w długości sekwencji kodujących białka (60 339 pz) oraz procentowej zawartości GC (38,5%).

Dopasowanie kompletnych sekwencji genomów chloroplastowych taksonów z kompleksu *Pinus mugo* oraz *Pinus sylvestris* z wykorzystaniem algorytmu MAUVE, nie ujawniło żadnych zdarzeń rearanżacji bądź inwersji i potwierdziło bliskie powiązania ewolucyjne między wszystkimi analizowanymi taksonami. Wyniki te były zgodne z wcześniejszymi badaniami, w których potwierdzono podobieństwo w organizacji genomu chloroplastowego *Pinus densiflora*, *P. sylvestris*, *P. thunbergii*, *P. tabuliformis* i *P. taeda* (Kang i in. 2019). Konserwatywność struktury genomów chloroplastowych sosen

potwierdzono również w późniejszej analizie porównawczej z udziałem 33 gatunków z rodzaju *Pinus* (Xia i in. 2023).

Analiza porównawcza przeprowadzona na podstawie metody "*sliding window*", ujawniła występowanie pięciu regionów wysoce zmiennych, tzw. *hotspotów* (tj. *trnG, atpI-rps2, trnE-clpP, clpP-rps12* i *rrn4.5-rrn5*) o wartości różnorodności nukleotydowej (ang. *nucleotide diversity*, Pi) większej niż 0,00238 (Pi > 0,00238) dla taksonów z kompleksu *Pinus mugo* oraz występowanie dziewięciu *hotspotów* (tj. *trnS-psaM, trnE-clpP, psaJ-trnP, psaM - trnS, petB-petD, ycf3-psaA, rrn4.5-rrn5, ycf1* i *ycf2*) dla taksonów z kompleksu *Pinus mugo* i *Pinus sylvestris* o wartości różnorodności nukleotydowej większej niż 0,00589 (Pi > 0,00589). Ponadto, w genomach chloroplastowych *Pinus mugo, Pinus uliginosa* i *Pinus uncinata* wykryto łącznie pięćdziesiąt dziewięć *loci* mikrosatelitarnych (ang. *Simple Sequence Repeats*, SSR) o długości co najmniej 10 pz. Ich liczba wahała się od dziewiętnastu u *P. uncinata* do dwudziestu u *P. mugo* oraz *Pinus uliginosa* i była podobna do *P. sylvestris* (tj. 22 mikrosatelity). Zidentyfikowane różnice w liczbie i rozmieszczeniu sekwencji mikrosatelitarnych pomiędzy tymi czterema taksonami pozwalają na ich hipotetyczne rozróżnienie.

Analizy filogenetyczne przeprowadzone z wykorzystaniem metody największej wiarygodności (ang. *Maximum Likelihood*, ML) oraz wnioskowania bayesowskiego (ang. *Bayesian Inference*, BI) u szesnastu przedstawicieli z rodziny Pinaceae, jak również z rodziny Podocarpaceae (tj. *Podocarpus latifolius*) jako grupy zewnętrznej (ang. *outgroup*) na podstawie dwóch zestawów danych, tj.: kompletnych plastomów oraz wysoce zmiennego rejonu *ycf1* (Hernández-León i in. 2013; Celiński i in., 2017; Olsson i in. 2018), potwierdziły bliskie pokrewieństwo między *P. uncinata, P. mugo* i *Pinus uliginosa* oraz wykazały, że sosny te tworzą odrębny klad w obrębie rodzaju i podrodzaju *Pinus.*

PUBLIKACJA 1

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.

https://doi.org/10.3390/plants10071331

plants

MDPI

Article

New Insight into Taxonomy of European Mountain Pines, *Pinus mugo* Complex, Based on Complete Chloroplast Genomes Sequencing

Joanna Sokołowska [1], Hanna Fuchs [2] and Konrad Celiński [1,*]

[1] Department of Genetics, Institute of Experimental Biology, Faculty of Biology, School of Natural Sciences, Adam Mickiewicz University, Poznań, Uniwersytetu Poznańskiego 6, 61-614 Poznań, Poland; joanna.sokolowska@amu.edu.pl

[2] Institute of Dendrology, Polish Academy of Sciences, Parkowa 5, 62-035 Kórnik, Poland; hkijak@man.poznan.pl

* Correspondence: konrad.celinski@amu.edu.pl

Abstract: The *Pinus mugo* complex is a large group of closely related mountain pines, which are an important component of the ecosystems of the most important mountain ranges, such as the Alps, Carpathians and Pyrenees. The phylogenetic relationships between taxa in this complex have been under discussion for many years. Despite the use of many different approaches, they still need to be clarified and supplemented with new data, especially those obtained with high-throughput methods. Therefore, in this study, the complete sequences of the chloroplast genomes of the three most recognized members of the *Pinus mugo* complex, i.e., *Pinus mugo*, *Pinus rotundata* and *Pinus uncinata*, were sequenced and analyzed to gain new insight into their phylogenetic relationships. Comparative analysis of their complete chloroplast genome sequences revealed several mutational hotspots potentially useful for the genetic identification of taxa from the *Pinus mugo* complex. Phylogenetic inference based on sixteen complete chloroplast genomes of different coniferous representatives showed that pines from the *Pinus mugo* complex form one distinct monophyletic group. The results obtained in this study provide new and valuable omics data for further research within the European mountain pine complex. They also indicate which regions may be useful in the search for diagnostic DNA markers for the members of *Pinus mugo* complex and set the baseline in the conservation of genetic resources of its endangered taxa.

Keywords: Pinaceae; European mountain pines; closely related taxa; next-generation sequencing

Citation: Sokołowska, J.; Fuchs, H.; Celiński, K. New Insight into Taxonomy of European Mountain Pines, *Pinus mugo* Complex, Based on Complete Chloroplast Genomes Sequencing. *Plants* **2021**, *10*, 1331. https://doi.org/10.3390/plants10071331

Academic Editor: Jung Sung Kim

Received: 3 June 2021
Accepted: 27 June 2021
Published: 29 June 2021

1. Introduction

The *Pinus mugo* complex is a large and polymorphic complex of closely related pines native to the main mountains of Europe, including the Pyrenees, the Alps and the Carpathians [1,2]. Some researchers indicate that in this group there may be even more than a hundred endemic forms classified into various taxonomic ranks, i.e., species, subspecies or varieties [1]. However, among them only three taxa, i.e., *Pinus mugo* subsp. *mugo*, *Pinus mugo* subsp. *rotundata* and *Pinus mugo* subsp. *uncinata*, are more widely known and thoroughly studied. These taxa differ in some phenotypic features, geographical distribution or preferred habitat. *Pinus mugo* subsp. *mugo*, also known as *Pinus mugo* Turra (dwarf mountain pine) or *Pinus mugo sensu stricto*, is a shrub with long, curved branches, reaching up to 3.5 m in height. The taxa has a wide geographical range, including the Alps, Pyrenees, Carpathians and Balkans [3], but most often occur in the higher parts of the mountains at an altitude of 1600–2200 m.a.s.l. [4]. *Pinus mugo* subsp. *rotundata*, identified by some researchers as a synonym for *Pinus uliginosa* Neumann (peat-bog pine), is usually a tree-shaped form, with a geographical range limited to peat bog areas of Poland, the Czech Republic and Germany [5]. *Pinus mugo* subsp. *uncinata* known as *Pinus uncinata*

Rammond (mountain pine) is a tree with a height of 12–20 m, occurring in the Pyrenees and the western Alps, as well as the Central Massif and the Iberian System [1,2].

These three taxa are considered to be independent species or subspecies inside *Pinus mugo* complex known also *Pinus mugo* Turra *sensu lato* [1,2,6]. The International Union for Conservation of Nature (IUCN) has defined the status of *Pinus mugo* subsp. *mugo* and *Pinus mugo* subsp. *uncinata* as least concern (LC), while *Pinus mugo* subsp. *rotundata* is identified as endangered (EN) [7]. However, conservation of these taxa can be difficult for a number of reasons. One of them is the problematic identification and classification of atypical individuals to specific taxa, especially in sympatric populations. In such populations, natural and uncontrolled gene flow is observed, as well as the formation of hybrid individuals with a phenotype intermediate between those of parent taxa [8–11]. Another serious problem is the functioning of synonyms in the scientific literature, which probably (but not for sure) refer to the same taxon, e.g., *Pinus mugo* subsp. *rotundata* also appears in the literature as *Pinus uliginosa* but can also be understood as *Pinus* × *rhaetica* (as a hybrid of *Pinus sylvestris* × *Pinus mugo*) [2]. The relations between these synonyms require urgent and detailed analyses, especially since *Pinus uliginosa* is the most endangered pine in Poland, as the number of individuals is gradually declining [12,13].

Until now, representatives of the *Pinus mugo* complex have been the subject of many different studies, including needle biometric analyses [14–16], characteristics of allozyme variability [17,18], patterns of genetic diversity distribution in the geographical aspect [19–21], gene flow and hybridization [22–24], molecular cytogenetics or flow cytometric analyses [25,26]. Some of the most important aspects so far undertaken by researchers were also attempts to establish relations between taxa in this complex [27,28] or searching for diagnostic features or additional determinants allowing for their unambiguous and simple differentiation [29–31]. However, based on the results obtained so far, it is extremely difficult to draw consistent and unambiguous conclusions, especially far-reaching ones. On the one hand, numerous studies indicate differences in gene expression products such as volatiles [32], essential oils [33] or seed protein patterns [31] between *P. mugo*, *P. uliginosa* and *P. uncinata*. On the other hand, other studies indicate that taxa from the *Pinus mugo* complex share common chloro- and mitotypes [20,34], have a complex genetic background and are generally characterized by a conservative organization of genomes [25,26].

Despite many studies on *P. mugo*, *P. rotundata* and *P. uncinata*, their origin, species distinctiveness and taxonomic status within the *Pinus mugo* complex as well as the identification of additional diagnostic determinants for them require further analysis.

The use of complete chloroplast genome sequences obtained by high throughput techniques could greatly help in this regard, by significantly increasing the phylogenetic resolution and providing new insight into the taxonomic relationships within this complex. This approach is particularly recommended in the case of closely related taxa, where the use of whole chloroplast genomes as one super-barcode should bring better resolution effects than the use of one or even several universal or specific DNA barcodes, which may be too little variable in a given group of plants [35]. This approach seems to be particularly relevant in the case of the *Pinus mugo* complex, where the analysis of several core, supplementary and candidate barcode regions failed to distinguish these taxa at the DNA level [30]. A detailed comparative analysis of the complete sequences of chloroplast genomes was successfully used in the research, among others, in *Pseudolarix* and *Tsuga* [36], *Corylus* [37], *Magnolia* [38] or *Quercus* [39] as well as many others plant taxa.

Therefore, the main objectives of our research were: (1) sequencing, analysis and characterization of the entire genomes of *P. mugo*, *P. rotundata* and *P. uncinata* chloroplasts; (2) comparative analysis of the obtained complete chloroplast genome sequences with previously published data for other members of the *Pinus* genus, especially those for *P. sylvestris*; (3) identifying and selecting mutation regions (hot spots) in chloroplast genomes potentially useful in identifying *Pinus mugo* taxa; and (4) performing a phylogenetic inference about the relatedness of three closely related taxa of the *Pinus mugo* complex based on the complete sequences of the chloroplast genomes as well as selected regions.

Our results gain new insight into the taxonomy of this highly polymorphic group of closely related taxa, significantly increasing phylogenetic resolution and providing new genomic resources for further taxonomic research and as a baseline to take conservation measures for this ecologically important group of European mountain pines.

2. Results and Discussion

2.1. General Features of P. mugo, P. rotundata and P. uncinata Chloroplast Genomes

Chloroplast genomes are typically about 150 kb in length and have a fairly distinctive quadripartite structure consisting of a large single copy (LSC), a small single copy (SSC) regions and two inverted repeats (IR) that separate them. Usually these repeats (IRa and IRb) are about 20-30 kb long, although in the case of the Pinaceae they are extremely reduced-to fragments sometimes even within 400 bp. The number of genes annotated in chloroplast genomes is variable, ranging from 63 to even 209 genes, although usually it does not exceed the range of 110 and 130 [39–41].

The length of complete chloroplast genomes of three closely related *P. mugo*, *P. rotundata* and *P. uncinata* analyzed in this study is comparable and amounts to 119,765 bp, 119,759 bp and 119,780 bp, respectively for these taxa (Figure 1 and Table 1). Chloroplast genomes of representatives of the *Pinus mugo* complex are circular molecules with a typical quadripartite structure consisting of a large single copy (LSC), a small single copy (SSC) and two very short inverted repeated IRs (IRa and IRb). The length of the LSC region ranges from 65,879 bp for *P. rotundata* to 65,899 bp for *P. uncinata* and *P. mugo* while the length of the SSC region ranges from 53,164 bp for *P. mugo*, 53,168 bp for *P. rotundata* to 53,169 bp for *P. uncinata*. The IR regions, on the other hand, are strongly reduced and, in the case of the representatives of the *Pinus mugo* complex, they are only 365 bp, which is one of the shortest so far described in the Pinaceae family. For comparison, the IR length in *Pinus taeda* (KC427273) is 485 bp, and for *Pinus sylvestris* (KR476379), *Pinus densiflora* (MK285358) or *Pinus yunnanensis* (MK007968) is exactly 495 bp [42,43]. For other species, differences in IR lengths are also observed, and several studies report that contraction and expansion of IR regions are quite common phenomena in plants [44]. Moreover, it happens that in some species these regions are completely lost [45–47]. It is postulated that the contraction and expansion of the IR regions play a major role in evolution and are responsible for altering the length of genomic sequences.

Table 1. Basic features of chloroplast genomes among the seven taxa of Pinaceae.

Genome Features	*Pinus mugo*	*Pinus rotundata*	*Pinus uncinata*	*Pinus sylvestris*	*Pinus densiflora*	*Larix decidua*	*Abies alba*
Genome size (bp)	119,765	119,759	119,780	119,758	119,875	122,747	121,243
Total coding length (bp)	67,592	67,593	67,592	67,625	67,684	68,621	67,983
Protein coding length (bp)	60,339	60,339	60,339	60,384	60,444	61,524	60,810
rRNA coding length (bp)	4517	4518	4518	4518	4518	4520	4522
tRNA coding length (bp)	2736	2736	2735	2654	2723	2577	2651
Total GC content (%)	38.5	38.5	38.5	38.5	38.5	38.8	38.3
Total number of genes	121	121	121	116	118	110	113
Number of protein-coding genes	73	73	73	73	73	72	74
Number of rRNA genes	4	4	4	4	4	4	4
Number of tRNA genes	36	36	36	35	36	34	35
GenBank Acc. No.	MZ333466	MZ333465	MZ333464	KR476379	MK285358	AB501189	NC_042410

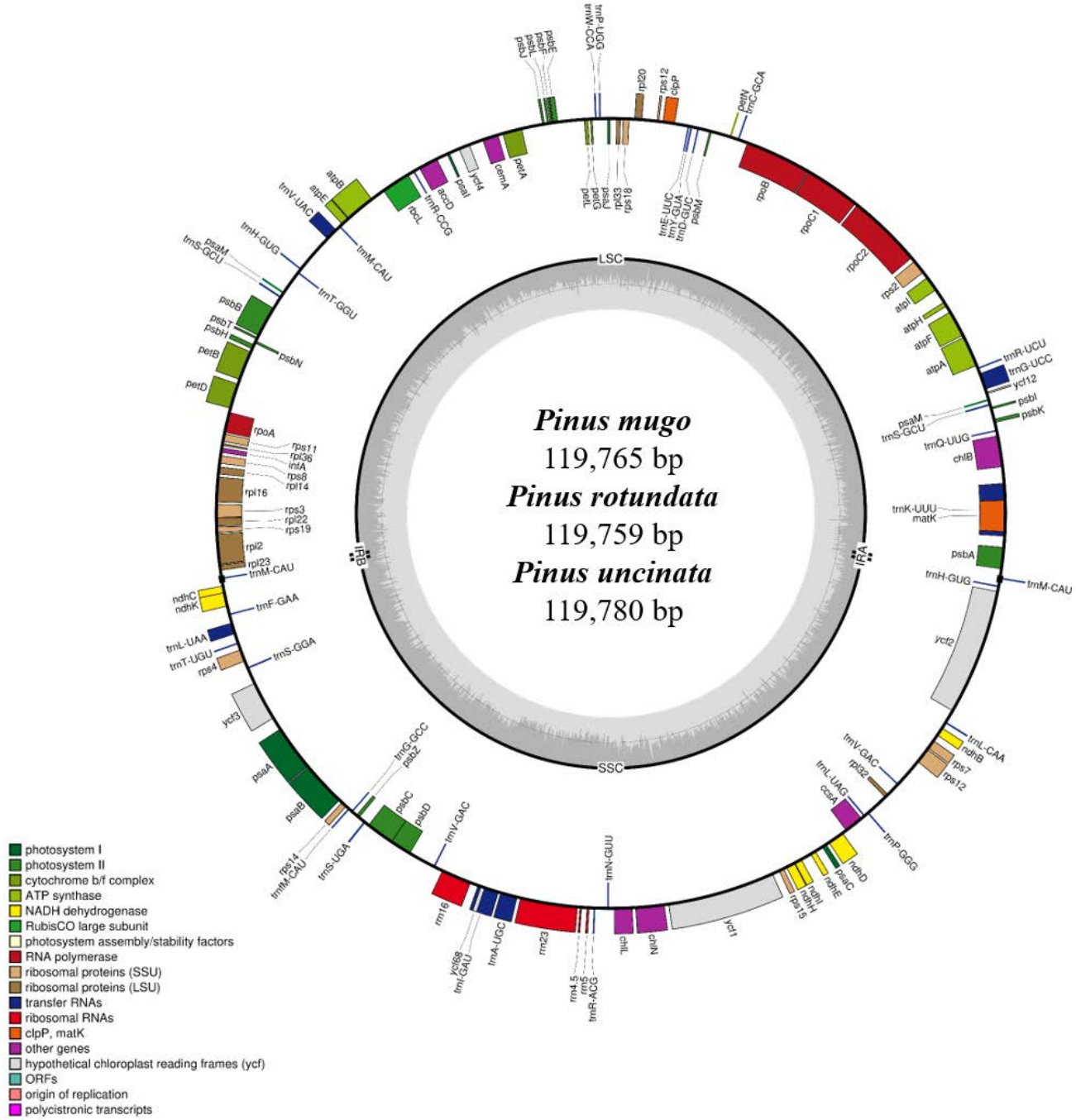


Figure 1. Gene maps of the three *Pinus* chloroplast genomes. The genes inside the circle are transcribed clockwise, and those outside are transcribed counterclockwise. Genes of different functions are color coded. The darker gray in the inner circle shows the GC content, while the lighter gray shows the AT content. IRA, IRB, inverted repeats; LSC, large single copy region; SSC, small single copy region.

The plastomes of *P. mugo*, *P. rotundata* and *P. uncinata* contain 121 genes, including 115 unique genes (excluding duplicate ones), 73 protein-coding genes, 36 transfer RNA genes, and four ribosomal RNA genes (Figure 1, Table 1). Five genes are duplicated, i.e., *psaM* (x2), *trnH-GUG* (x2), *trnM-CAU* (x3), *trnS-GCU* (x2) and *trnV-GAC* (x2). The functional classification of these genes is presented in Supplementary Table S1. The total content of GC is 38.5% and there are no differences in this parameter between the analyzed taxa. Likewise, there are no significant differences in the length of the protein coding sequences (60,339 bp), the total and unique number of genes (121 and 115, respectively) or

the number of rRNA and tRNA genes (4 and 36, respectively). Due to the uniform gene number, order and their names, annotated chloroplast genomes of these three taxa from the *Pinus mugo* complex are presented on one circular map (Figure 1).

Our results obtained in this study are fully consistent with those previously published for other *Pinus* representatives, i.e., *P. sylvestris* (KR476379) or *P. densiflora* (MK285358) [42] in terms of genome size, total coding length, and protein coding length, as well as number of predicted genes or GC content (Table 1). There are only slight differences in genomic features between *Pinus* taxa and *Larix* or *Abies* taxa. They mainly concern the size of the genome and the number of genes. Taxa of the genera *Larix* and *Abies* have slightly longer genomes and fewer genes than representatives of the genus *Pinus*.

2.2. Genome Comparative Analysis and Identification of Divergent Hotspots

The complete sequences of the *P. mugo*, *P. rotundata* and *P. uncinata* chloroplast genomes were aligned with the complete *P. sylvestris* chloroplast genome (KR476379) to compare the organization of their genomes (Figure 2). *Pinus sylvestris* was chosen as the reference taxon closest to this complex but not belonging to it. Figure 2 shows only one locally collinear block (LCB) between all analyzed chloroplast genomes, which suggests a high level of similarity in genome organization between the analyzed *Pinus* taxa.

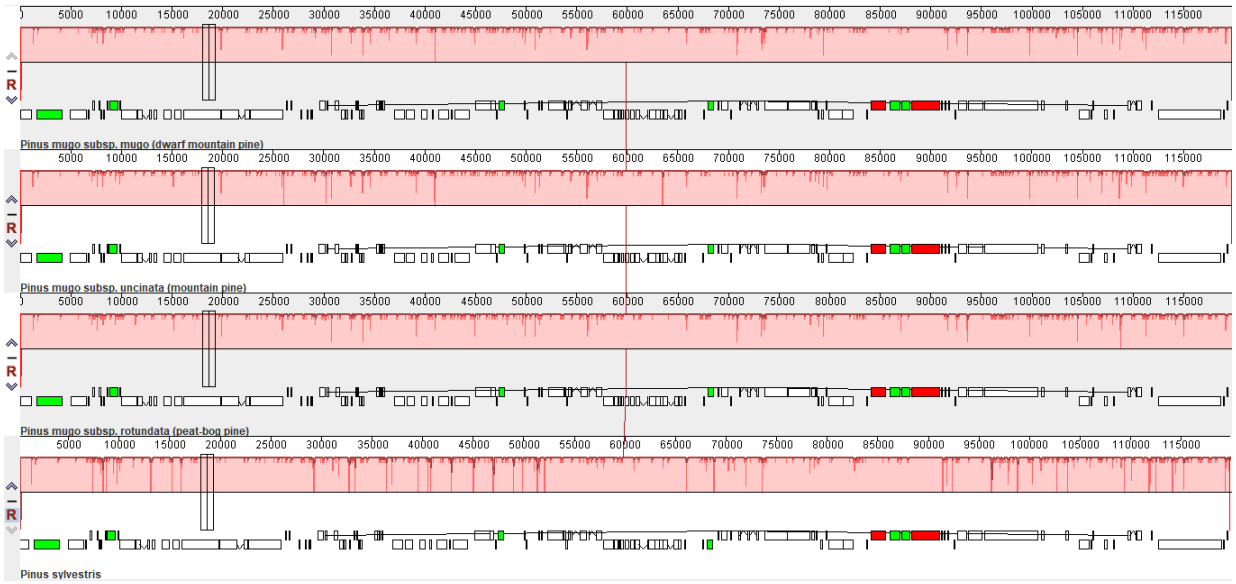


Figure 2. MAUVE alignment of three *Pinus mugo* complex representatives; *P. mugo* subsp. *mugo*, *P. mugo* subsp. *uncinata*, *P. mugo* subsp. *rotundata*. The *Pinus sylvestris* chloroplast genome is shown at bottom as a reference. Within each of the alignments, local collinear blocks are represented by blocks of the same color connected by lines.

In summary, whole-genome alignment of the chloroplast sequences did not reveal any rearrangement or inversion events among *Pinus* chloroplast genomes, and confirmed the close evolutionary relationships between all analyzed taxa (both those belonging to the *Pinus mugo* complex and not). Our results are fully consistent with earlier studies on *Pinus* species [42], in which the gene content and order of the *P. densiflora* chloroplast genome were similar to four other pines, i.e., *P. sylvestris*, *P. thunbergii*, *P. tabuliformis* and *P. taeda* [42].

The K2p distance values calculated as an estimator of evolutionary divergence (Table 2) differ between *Pinus* taxa from 0.000259 in a pair of *P. mugo* and *P. uncinata* to 0.00318 in a pair of *P. uncinata* and *P. sylvestris*, with an average of 0.001741 for all four analyzed *Pinus* taxa.

Table 2. Estimates of evolutionary divergence between four *Pinus* species. The number of base differences per site from between sequences are shown. Standard error estimate(s) are shown above the diagonal and were obtained by a bootstrap procedure (1000 replicates). This analysis involved four nucleotide sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 120279 positions in the final dataset. Evolutionary analyses were conducted in MEGA X [48].

	P. mugo	*P. rotundata*	*P. uncinata*	*P. sylvestris*
P. mugo	-	0.000044	0.000055	0.000158
P. rotundata	0.000259	-	0.000061	0.000158
P. uncinata	0.000351	0.000409	-	0.000158
P. sylvestris	0.003117	0.003126	0.003184	-

DnaSP was used to perform two sliding window analyses in order to identify mutational regions. One analysis concerned only three taxa from the *Pinus mugo* complex (Figure 3A), while the other, apart from *P. mugo*, *P. rotundata* and *P. uncinata*, also included *P. sylvestris* (Figure 3B).

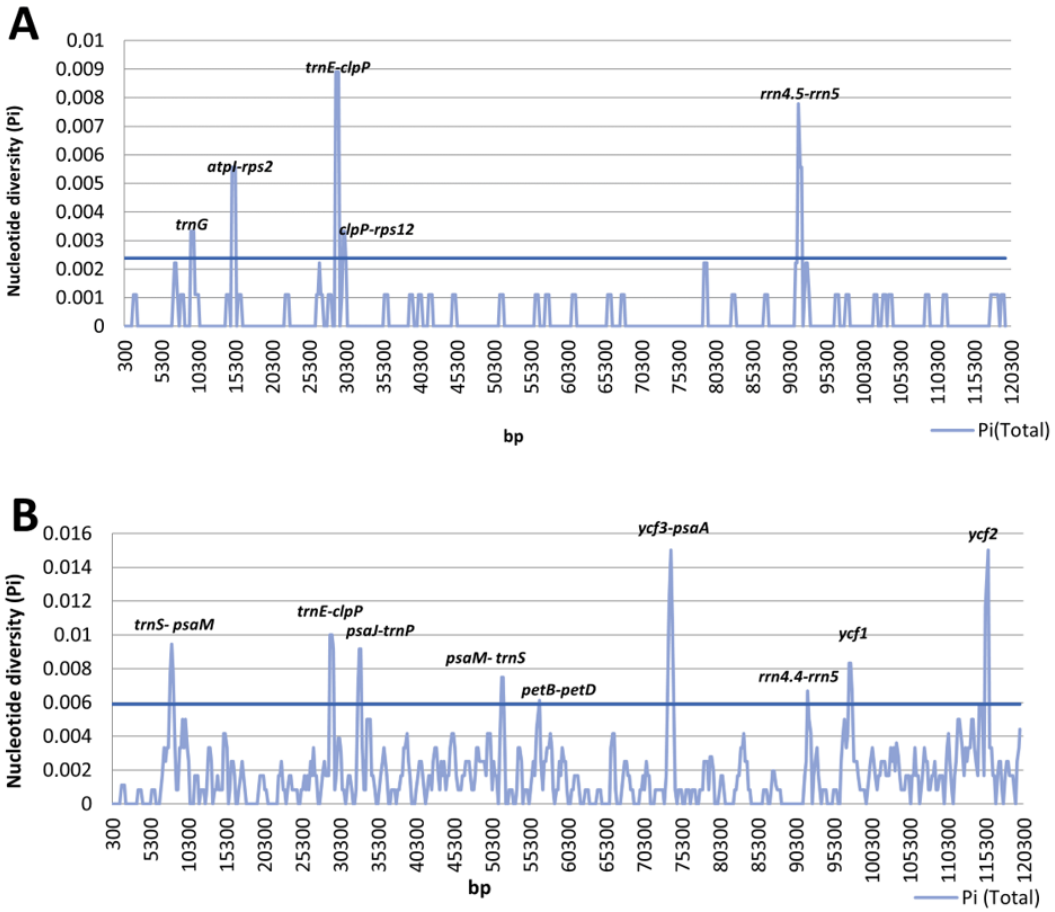


Figure 3. Sliding window analysis of the whole chloroplast genomes. Window length: 600 bp; step size: 200 bp. X-axis: position of the midpoint of a window. Y-axis: nucleotide diversity of each window. (**A**) Pi between three the *Pinus mugo* complex representatives. (**B**) Pi among the *Pinus mugo* complex representatives and *Pinus sylvestris*. The horizontal line on the graph sets the threshold separately for the *Pinus mugo* complex representatives (0.00238) and separately for the *Pinus mugo* complex representatives and *Pinus sylvestris* (0.00589).

The results in Figure 3A clearly show that for the *Pinus mugo* complex taxa there were five divergent hotspots with a high Pi value (>0.00238), i.e., *trnG*, *atpI-rps2*, *trnE-*

clpP, *clpP-rps12*, and *rrn4.5-rrn5*. For the second combination, taxa from the *Pinus mugo* complex and *P. sylvestris*, a total of nine unique mutational regions with a high Pi value (>0.00589) were detected, i.e., *trnS-psaM*, *trnE-clpP*, *psaJ-trnP*, *psaM-trnS*, *petB-petD*, *ycf3-psaA*, *rrn4.5-rrn5*, *ycf1* and *ycf2* (Figure 3B). The average value of nucleotide diversity (Pi) was 0.00036 and 0.00174 for the *Pinus mugo* complex taxa and for the *Pinus mugo* taxa together with *P. sylvestris*, respectively. This result is in line with expectations because the second combination included more distant pines, not just three closely related taxa. A similar relationship was found also in the case of other species [49].

Pairwise distance analysis for the highly variable regions (Figure 4A,B) showed that the highest K2p distance between taxa from the *Pinus mugo* complex is between *P. mugo* and *P. rotundata* (0.01239) in the *trnE-clpP* region (Figure 4A). In turn, the highest K2p distance between *P. sylvestris* and any taxon from the *Pinus mugo* complex (Figure 4B) is 0.03298 and concerns the *trnS-psaM* region and the *P. sylvestris* vs. *P. rotundata*. Overall, a detailed pairwise distance analysis revealed what values of discrepancy and in which regions of the chloroplast genome sequence can be expected between the analyzed taxa pairs.

Chloroplast DNA regions selected in this study can be preferentially used as specific barcodes for further studies of *Pinus mugo* taxonomy. A species-specific barcode is defined as a fragment of a DNA sequence with a sufficiently high mutation rate to enable the species to be identified within a given taxonomic group [35]. The *ycf1* and *ycf2* regions seem of particular interest in this regard for the genus *Pinus*. Several studies show that the *ycf1* region in particular has extremely high discriminatory power in some genera and much greater potential than the commonly used universal core barcodes [30,50,51].

2.3. Simple Sequence Repeats Analysis

Simple sequence repeats (SSRs or microsatellites) are very often used in population, ecological and conservation genetics as effective molecular markers. Their most important advantages are the high level of genetic polymorphism detected by them and wide distribution throughout the genome of chloroplasts, as well as trouble-free amplification, fast electrophoretic separation or objective and simple statistical analysis [52–55].

In this study, a total of fifty-nine SSRs with a length of at least 10 bp were detected in the chloroplast genomes of three members of the *Pinus mugo* complex. The number of detected SSR loci ranged slightly from nineteen in *P. uncinata* to twenty in *P. mugo* and *P. rotundata* and was similar to *P. sylvestris* (22 microsatellites) but much lower than that found recently with another pine, *Pinus taeda* (151) [56].

Interestingly, the identified differences in the number of SSRs between the four analyzed taxa hypothetically allow these taxa to be distinguished using microsatellite loci. A detailed analysis of the number and distribution of SSRs brings very interesting results. For *P. mugo* and *P. uncinata*, we found a microsatellite between 54,429 and 54,438 bp and between 54,428 and 54,437 bp, respectively, which was not observed in the genomes of *P. rotundata* or *P. sylvestris*. Similarly, in the case of *P. mugo* and *P. rotundata*, we found the presence of a microsatellite repeat between 44,949 and 44,961 bp and 44,940 and 44,954 bp, respectively, which is not present in the chloroplast genome of *P. uncinata*. A comparison of the 100,883-100,892 bp region in *P. rotundata* with the 100,844-100,853 region in *P. sylvestris* reveals that these taxa differ in the repeat motif; *P. sylvestris* has an A repeat, and *P. rotundata* has a T repeat. Moreover, to a similar extent, no microsatellite repetitions were found in the other two taxa, i.e., *P. mugo* and *P. uncinata*. Most of the SSRs identified in this study (47/59) were located in the intergenic distance region (IGS) (Table 3). The most common microsatellite repeat motif was mononucleotide (84.75%), followed by dinucleotide (10.17%) and compound (5.08%). Our results are fully consistent with the observations from other previously conducted studies in which SSRs in chloroplast genomes have a motif composed mainly of short polyadenine (polyA) or polythymine (polyT) repeats and much less often contain guanidine (G) or cytosine (C) tandem repeats [38,56].

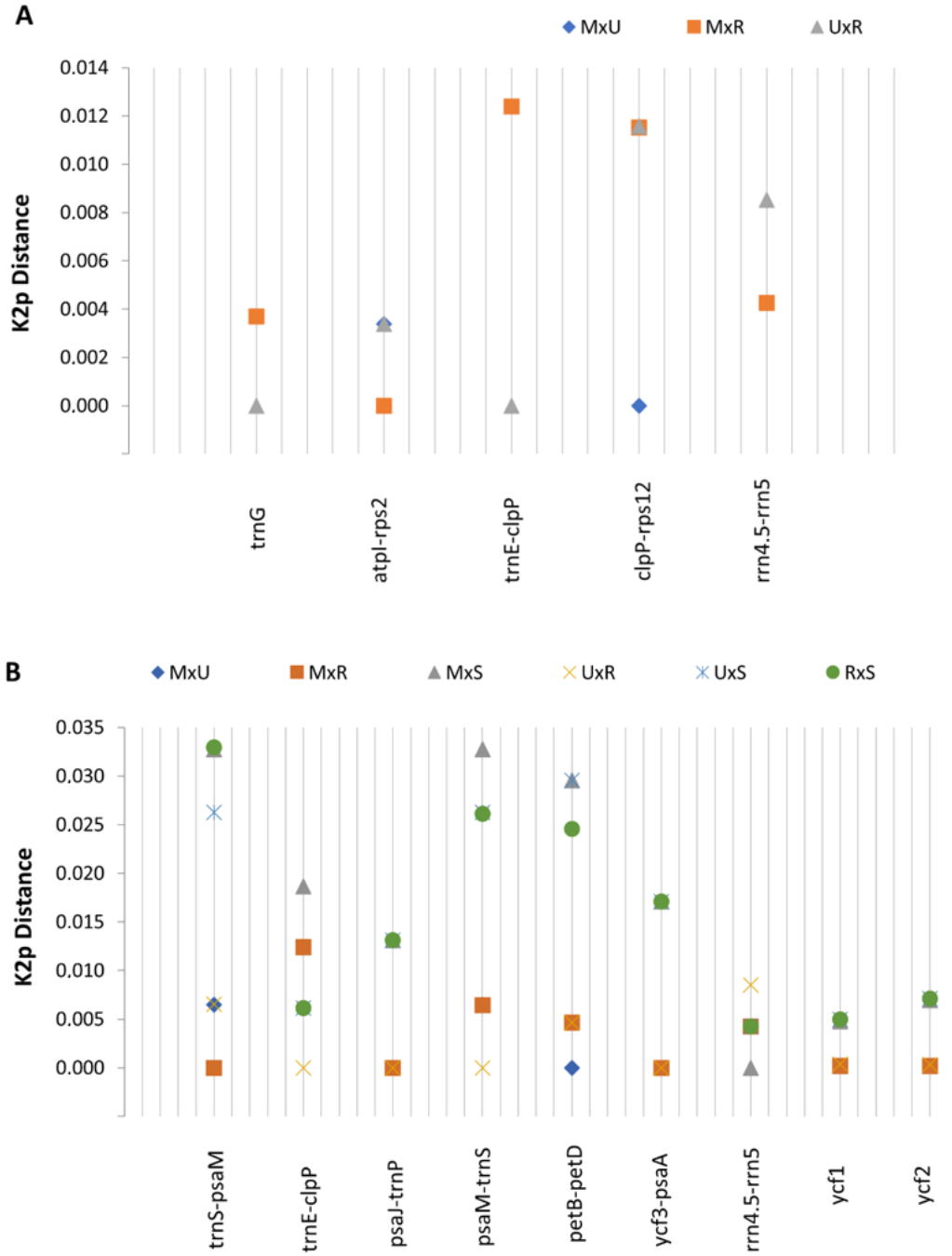


Figure 4. K2p distance for selected hotspot regions separately for the *Pinus mugo* complex representatives (**A**) and taxa for the *Pinus mugo* complex and *P. sylvestris* (**B**). Abbreviations: M, *Pinus mugo*; R, *Pinus rotundata*; U, *Pinus uncinata*; S, *Pinus sylvestris*.

The SSRs identified in this study can be used for further research on the representatives of the *Pinus mugo* complex, i.e., *P. mugo*, *P. rotundata* and *P. uncinata*, and to characterize their genetic resources. The SSRs described in this study can potentially be used to distinguish taxa in the *Pinus mugo* complex and also complement other microsatellite loci used so far for this purpose [57,58].

2.4. Phylogenetic Inference

The phylogenesis of many different groups of plants was determined by analyzing the sequences of both the complete genome of chloroplasts and selected regions [59–62]. In this study, we were particularly interested in the relationships within the *Pinus mugo*

complex between three closely related taxa, as the phylogeny of the genus *Pinus* is well known. Therefore, phylogenetic trees were constructed using the ML and Bayes algorithms using the nucleotide sequences of the chloroplast genomes of sixteen taxa representing the two main conifer families, Pinaceae and Podocarpaceae (Table 4). We used two datasets. The first involved alignment of entire chloroplast genome sequences, while the second was based on alignment of the highly variable *ycf1* gene only. In many previous studies, researchers indicate its very high level of genetic diversity, useful in phylogenic analyses [30,51,63].

As shown in Figure 5, both obtained ML and Bayesian phylogenetic trees clearly indicated that *P. mugo*, *P. rotundata* and *P. uncinata* belonging to the *Pinus mugo* complex formed a separate cluster within the *Pinus* genus. Although phylogenetic reconstruction was not the main focus of this work, the overall topology of the trees obtained here (regardless of the data set and analysis methods used) was not surprising, and is consistent with the well-known and widely accepted division of the Pinaceae family into basic genera, i.e., *Picea*, *Larix*, *Abies* and *Pinus*. Additionally, in the genus *Pinus*, the analyzed pine taxa formed two separate clades. One clade consisted of *Pinus strobus* and *Pinus cembra* belonging to the subgenus *Strobus*, while the other clade consisted of taxa included in the subgenus *Pinus*, i.e., *Pinus taeda*, *Pinus pinea*, *Pinus densiflora*, *P. sylvestris* as well as three closely related taxa from of the *Pinus mugo* complex; *P. mugo*, *P. rotundata* and *P. uncinata*. It is worth noting that in the ML and BI trees, most of the nodes had 100% bootstrap support and 1.0 Bayesian posterior probability (Figure 5). *Podocarpus latifolius* from the Podocarpaceae family, as predicted, was outside the main group of taxa from the Pinaceae family.

Table 3. Estimates Simple sequence repeats (SSRs) identified in the *P. mugo*, *P. rotundata*, *P. uncinata* and *P. sylvestris* chloroplast genomes.

Taxon	ID	Type	Repeat Motif	Length (bp)	Start	End	Location	ID	Type	Repeat Motif	Length (bp)	Start	End	Location
P. mugo	1	p1	(C)12	12	15142	15153	IGS	11	p1	(A)11	11	79749	79759	IGS
	2	p1	(A)12	12	26050	26061	IGS	12	p1	(T)10	10	87077	87086	IGS
	3	c	(A)10(G)10	20	30198	30217	IGS	13	p1	(A)10	10	100605	100614	IGS
	4	p1	(T)23	23	40994	41016	IGS	14	p1	(T)10	10	103575	103584	IGS
	5	p1	(T)13	13	44949	44961	IGS	15	p1	(G)11	11	104142	104152	CDS (*ndhD*)
	6	p1	(T)10	10	48132	48141	IGS	16	p1	(A)13	13	106928	106940	IGS
	7	p1	(A)10	10	54429	54438	IGS	17	p1	(T)11	11	107335	107345	CDS (*rpl32*)
	8	p1	(A)10	10	67826	67835	IGS	18	p1	(A)10	10	109379	109388	IGS
	9	p1	(T)11	11	71751	71761	CDS (*ycf3*)	19	p1	(A)10	10	109840	109849	CDS (*rps12*)
	10	p2	(AT)6	12	73254	73265	IGS	20	p2	(AT)6	12	111752	111763	IGS
P. rotundata	1	p1	(C)13	13	15141	15153	IGS	11	p1	(T)11	11	87069	87079	IGS
	2	p1	(A)12	12	26050	26061	IGS	12	p1	(A)10	10	100597	100606	IGS
	3	c	(A)11(G)10	21	30197	30217	IGS	13	p1	(T)10	10	100883	100892	IGS
	4	p1	(T)15	15	40993	41007	IGS	14	p1	(T)10	10	103568	103577	IGS
	5	p1	(T)15	15	44940	44954	IGS	15	p1	(G)10	10	104135	104144	CDS (*ndhD*)
	6	p1	(T)11	11	48122	48132	IGS	16	p1	(A)13	13	106920	106932	IGS
	7	p1	(A)10	10	67816	67825	IGS	17	p1	(T)11	11	107327	107337	CDS (rpl32)
	8	p1	(T)12	12	71741	71752	CDS (*ycf3*)	18	p1	(A)10	10	109375	109384	IGS
	9	p2	(AT)6	12	73245	73256	IGS	19	p1	(A)10	10	109836	109845	CDS (*rps12*)
	10	p1	(A)11	11	79740	79750	IGS	20	p2	(AT)6	12	111746	111757	IGS
P. uncinata	1	p1	(C)13	13	15142	15154	IGS	11	p1	(T)11	11	87091	87101	IGS
	2	p1	(A)15	15	26051	26065	IGS	12	p1	(A)10	10	100620	100629	IGS
	3	c	(A)11(G)10	21	30203	30223	IGS	13	p1	(T)10	10	103590	103599	IGS
	4	p1	(T)23	23	40998	41020	IGS	14	p1	(G)10	10	104157	104166	CDS (*ndhD*)
	5	p1	(T)11	11	48131	48141	IGS	15	p1	(A)13	13	106942	106954	IGS

Table 3. *Cont.*

Taxon	ID	Type	Repeat Motif	Length (bp)	Start	End	Location	ID	Type	Repeat Motif	Length (bp)	Start	End	Location
	6	p1	(A)10	10	54428	54437	IGS	16	p1	(T)11	11	107349	107359	CDS (*rpl32*)
	7	p1	(A)10	10	67836	67845	IGS	17	p1	(A)11	11	109393	109403	IGS
	8	p1	(T)13	13	71760	71772	CDS (*ycf3*)	18	p1	(A)11	11	109855	109865	CDS (*rps12*)
	9	p2	(AT)6	12	73265	73276	IGS	19	p2	(AT)6	12	111767	111778	IGS
	10	p1	(A)12	12	79761	79772	IGS							
P. sylvestris	1	p1	(T)11	11	1376	1386	IGS	12	p1	(A)10	10	79947	79956	IGS
	2	p1	(A)10	10	9837	9846	IGS	13	p1	(T)10	10	87277	87286	IGS
	3	c	(C)10(T)11	21	15195	15215	IGS	14	p1	(A)10	10	100844	100853	IGS
	4	p1	(A)12	12	26112	26123	IGS	15	p1	(T)11	11	101130	101140	IGS
	5	c	(A)11(G)10	21	30269	30289	IGS	16	p1	(T)10	10	101833	101842	CDS (*ndhH*)
	6	p1	(T)11	11	41059	41069	IGS	17	p1	(T)10	10	102658	102667	IGS
	7	p1	(T)19	19	45043	45061	IGS	18	p1	(G)11	11	104388	104398	CDS (*ndhD*)
	8	p1	(A)12	12	68030	68041	IGS	19	p1	(T)11	11	107567	107577	CDS (*rpl32*)
	9	p1	(T)14	14	71957	71970	CDS (*ycf3*)	20	p1	(A)10	10	109610	109619	IGS
	10	p2	(AT)6	12	73462	73473	IGS	21	p1	(A)12	12	110071	110082	CDS (*rps12*)
	11	p2	(AT)6	12	79134	79145	IGS	22	p2	(AT)7	14	111984	111997	IGS

c, compound SSR; p1, mono-nucleotide SSR; p2, di-nucleotide SSR.

Table 4. GenBank information on complete chloroplast genomes of conifer taxa used in phylogenetic analyses in this study.

GenBank Accession	Taxon	Common Name	Family
NC_042410	*Abies alba*	silver fir	Pinaceae
KP742350	*Abies koreana*	Korean fir	Pinaceae
AB501189	*Larix decidua*	common larch	Pinaceae
NC_036811	*Larix sibirica*	Siberian larch	Pinaceae
NC_021456	*Picea abies*	Norway spruce	Pinaceae
NC_032367	*Picea asperata*	dragon spruce	Pinaceae
MN536531	*Pinus cembra*	Swiss stone pine	Pinaceae
MK285358	*Pinus densiflora*	Japanese red pine	Pinaceae
MZ333466	*Pinus mugo* subsp. *mugo*	dwarf mountain pine	Pinaceae
MZ333465	*Pinus mugo* subsp. *rotundata*	peat-bog pine	Pinaceae
MZ333464	*Pinus mugo* subsp. *uncinata*	mountain pine	Pinaceae
NC_039585	*Pinus pinea*	Italian stone pine	Pinaceae
NC_026302	*Pinus strobus*	Eastern white pine	Pinaceae
KR476379	*Pinus sylvestris*	Scots pine	Pinaceae
KY964286	*Pinus taeda*	loblolly pine	Pinaceae
MH536745	*Podocarpus latifolius*	broad-leaved yellowwood	Podocarpaceae

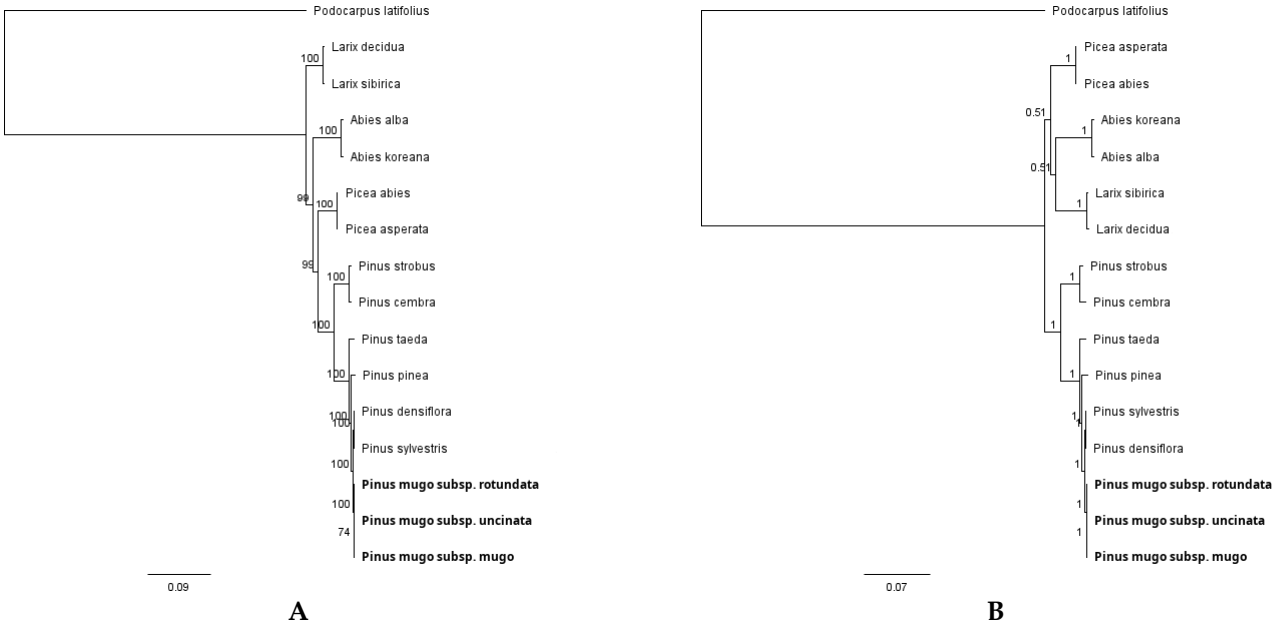


Figure 5. *Cont*.

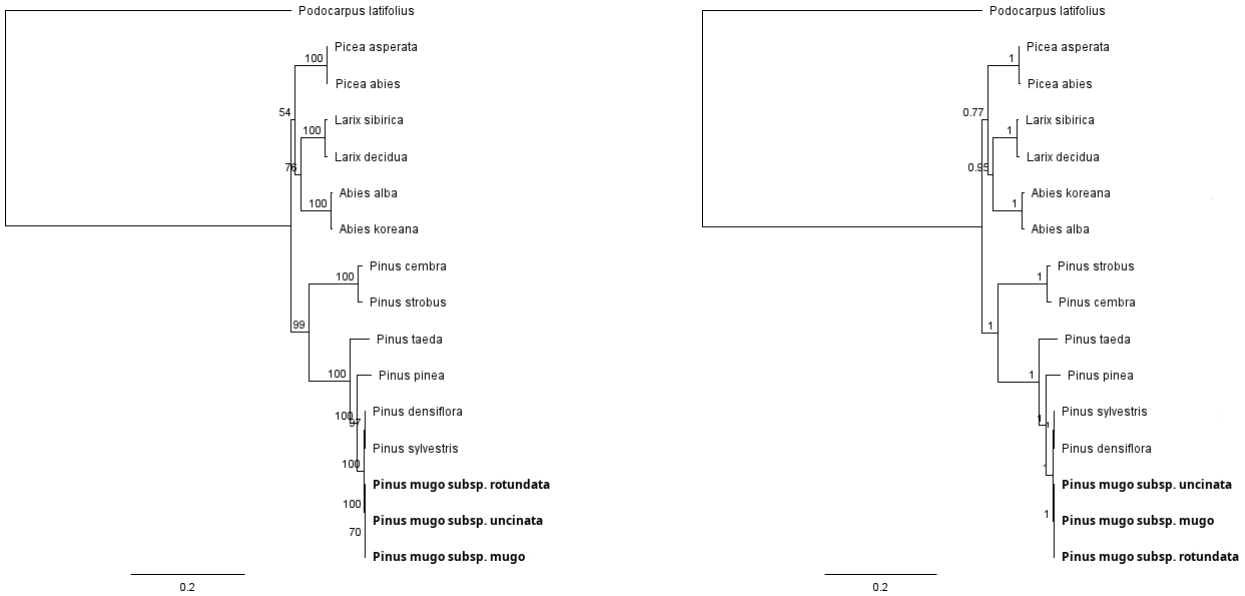


Figure 5. Phylogenetic relationships between sixteen conifers taxa based on complete sequences of chloroplast genomes (**A**,**B**) inferred from ML and BI analyses, respectively, and only on *ycf1* (**C**,**D**) also inferred from ML and BI analyses, respectively.

3. Materials and Methods

3.1. Sampling, DNA Extraction and Genomic Library Preparation

Fresh and healthy needles of the three most recognized members of the *Pinus mugo* complex were collected as follows: *Pinus mugo* subsp. *uncinata* (hereinafter referred to for short as *Pinus uncinata)* (collection number 1347) from the Dendrological Garden of University of Life Sciences, Poznań, Poland (52°25′37″ N, 16°53′48″ E); *Pinus mugo* subsp. *rotundata* (hereinafter referred to for short as *Pinus rotundata*) from the Great Peat Bog of Batorów located in Stołowe Mountains National Park, Poland (50°15′42.48″ N, 16°8′31.92″ E) and finally *Pinus mugo* subsp. *mugo* (hereinafter referred to for short as *Pinus mugo*) from the Tatra National Park (UNESCO Biosphere Reserve), Poland (49°10′0″ N, 19°55′0″ E). The collected needles were stored at 4 °C, until DNA extraction. Genomic DNA was isolated using the CTAB method [64]. The quality and integrity of isolated DNA were determined using agarose gel electrophoresis and measurement on a NanoDrop spectrophotometer (Thermo Fisher Scientific, Carlsbad, CA, USA). The genomic library was prepared according to the manufacturer's recommendations with protocol: Ion Xpress™ Plus gDNA Fragment Library Preparation, using Ion Xpress Plus Fragment Library Kit (Pub. No. MAN0009847) (ThermoFisher Scientific, Waltham, MA, USA). The 100 ng of total genomic DNA was fragmented using Ion Shear Plus Reagents with 8 min incubation time at 37 °C, targeting fragments length of 200–300 bp. Then, the fragmented DNA was purified using 1.8× sample volume of Agencourt™ AMPure™ XP Reagent. The fragment size was checked by 2200 Tapestation Bioanalyzer and Agilent™ High Sensitivity DNA Kit (Agilent Technologies, Waldbronn, Germany), according to protocol: Agilent HS D1000 ScreenTape System Quick Guide. For *Pinus uncinata*, the adapters ligation was conducted for reaction setup for non-barcoded libraries using Ion Plus Fragment Library Kit Adapters. For *P. mugo* and *P. rotundata,* the adapters ligation was conducted for reaction setup for barcoded libraries using the Ion Xpress™ Barcode Adapters Kit. AMPure purification was performed after ligation using a 1.2× sample volume of Agencourt™ AMPure™ XP Reagent (ThermoFisher Scientific, Waltham, MA) for 200–300-base-read library size. The size selection procedure was performed on the E-Gel™ SizeSelect™ 2% Agarose Gel, then the libraries were amplified and purified using a 1.2× sample volume of Agencourt™ AMPure™ XP Reagent (ThermoFisher Scientific, Waltham, MA, USA). Quality and length analysis was conducted using 2200 Tapestation Bioanalyzer (Agilent

Technologies Waldbronn, Germany). Chloroplast genomes are typically about 150 kb in length and have a fairly distinctive quadripartite

3.2. Next Generation Sequencing

The genomic library was diluted to 100 pM. The concentration was measured on the Qubit™ 2.0 Fluorometer using Qubit™ dsDNA HS Assay Kit (Pub. No. MAN0002326 Revision: B.0) (Life Technologies). The *P. uncinata* template preparation was performed according to protocol: Ion PGM™ Hi-Q™ View OT2 Kit (Cat. No. A29900, Pub. No. MAN0014580 Rev. C.0). *P. mugo* and *P. rotundata* templates preparation were performed according to protocol: Ion 540™ Kit – OT2 (Cat. No A27753 Pub. No. MAN0010852 Rev. E.0). Evaluation of the templated Ion Sphere™ Particles (ISPs) was conducted using Ion Sphere™ Quality Control Kit (Cat.No. 4468656), according to protocol Ion Sphere™ Assay on the Qubit ™ 2.0 Fluorometer (Pub. No. MAN0016387 Revision A.0) (ThermoFisher Scientific, Waltham, MA, USA). *P. uncinata* genome sequencing was conducted on Ion 318™ Chip v2 BC by Ion Personal Genome Machine™ (PGM™) System (Thermo Fisher Scientific, Waltham, MA, USA) according to manufacturer's recommendations using protocol: Ion PGM™ Hi-Q™ View Sequencing Kit user guide (Cat. No. A30044, Pub. No. MAN0014583). Then, *P. mugo* and *P. rotundata* genome sequencing was conducted on Ion 540™ Chip by GeneStudio™ S5 System (Thermo Fisher Scientific, Waltham, USA) according to manufacturer's recommendations using protocol: Ion 540™ Kit – OT2 User Guide (Cat. No A27753, Pub. No MAN0010850, Rev. D).

3.3. Chloroplast Genomes Assembly and Gene Annotation

BBDuk Adapter/Quality Trimming V. 35.82 available in Geneious Prime 2020.2.5 [65] was used to filter low quality reads and trim low quality ends and adapters. The filtered reads were de novo assembled into contigs using Geneious Assembler on default options with merging homopolymer variants. Contigs were mapped to the reference genome *Pinus sylvestris* (NC_035069.1) using Geneious Mapper with minimum mapping quality: 30. Reads, which mapped to the reference genome, were used to assemble de novo the complete chloroplast genome sequences of *P. mugo*, *P. rotundata* and *P. uncinata*. Assembled genomes were initially annotated using CPGAVAS2, an integrated plastome sequence annotator [66], and GeSeq [67] and compared to the *Pinus sylvestris* (RefSeq: NC_035069.1) reference sequence. Location of large single copy region (LSC) and small single copy region (SSC) as well as calculation of GC content was carried out in Geneious Prime 2020.2.5 [65] by comparison with homologous sequences available to other *Pinus* representatives. Transfer RNAs were also checked with tRNAscan-RE v2.0.3. [68] incorporated in GeSeq [67] using default settings. OrganellarGenomeDRAW (OGDRAW) version 1.3.1 [69] was used to draw a circular map chloroplast genome of *P. mugo*, *P. rotundata* and *P. uncinata*. The complete sequences of the chloroplast genomes of these three taxa mentioned above have been deposited in GenBank under the following accession numbers: MZ333466 for *Pinus mugo* subsp. *mugo*; MZ333465 for *Pinus mugo* subsp. *rotundata* and MZ333464 for *Pinus mugo* subsp. *uncinata*.

3.4. Genome Comparative Analysis and Identification of Divergent Hotspots

In order to study genome-wide evolutionary dynamics among *P. mugo*, *P. rotundata* and *P. uncinata* from the *Pinus mugo* complex and to search evolutionary events such as gene loss, duplication, rearrangements and translocations, multiple alignments were made using progressive MAUVE algorithm with default settings via MAUVE [70] plugin v1.1.1 available in Geneious Prime 2020.2.5 [65]. The complete sequences of the *P. mugo*, *P. rotundata* and *P. uncinata* chloroplast genomes were compared with this previously published sequence for *Pinus sylvestris* (KR476379), which is the nearest taxa to the *Pinus mugo* complex, but does not belong to it (Table 1). Evolutionary divergence between the three representatives of the *Pinus mugo* complex and *P. sylvestris* was estimated by

calculating genetic distances using the Kimura 2-parameters (K2p) evolution model [46,71] implemented in MEGA X [48].

Identification of divergent hotspots was performed separately only for the representatives of the *Pinus mugo* complex and for those representatives and *P. sylvestris* on the basis of three and four complete sequences of chloroplast genomes, respectively. The relevant chloroplast genomes were aligned using MAFFT v7.450 with default options [72], and then nucleotide diversity (Pi) was calculated through sliding window analysis using DnaSP version 6 [73]. The window length was set to 600 bp, with a step size 200 bp. The diversity thresholds for the *Pinus mugo* complex (0.00238) and for the *Pinus mugo* complex and together with *P. sylvestris* (0.00589) were calculated by sum of the average and double the standard deviation [74]. Regions with levels of nucleotide diversity higher than these thresholds were recommended as highly variable regions. Pairwise distance was also determined for these regions using the Kimura 2-parameters (K2p) evolution model [46,71] implemented in MEGA X [48].

3.5. Identification of Simple Sequence Repeats

Simple sequence repeats (SSRs) in chloroplast genomes of *Pinus mugo* complex representatives and *Pinus sylvestris* were detected by MIcroSAtellite (MISA) [75], with the following parameters set at ≥10 for mononucleotides, 6≥ for dinucleotides and ≥5 for tri-, tetra-, penta- and hexanucleotides, respectively.

3.6. Phylogenetic Inference

Phylogenetic inferences were constructed by maximum likelihood (ML) and Bayesian inference (BI) were constructed by maximum likelihood (ML) analysis using sixteen complete sequences of chloroplast genomes of various conifers representatives (including data obtained in this study for *P. mugo*, *P. rotundata* and *P. uncinata*). The list of taxa included in the study, along with GenBank accession numbers, is given in Table 4. In order to better explain the topology of the tree, both closely related taxa from the Pinaceae family, such as *Pinus*, and more distant taxa from the genus *Abies*, *Larix* and *Picea*, were selected. The outgroup was *Podocarpus latifolius* from the Podocarpaceae family.

Complete chloroplast genomes were aligned with MAFFT v7.450 using default settings [51]. A General Time Reversible + Gamma nucleotide substitution model (GTR + G) was selected according to Akaike's information criterion (AIC) [76] with MEGA X [48], as the best substitution model for the ML and BI analyses. The ML analyses were conducted in RaxML v8.2.11 [77], with 1000 rapid bootstrap replicates along with a search for the best-scoring ML tree in every run and parsimony random seed set to 10.

BI analyses were conducted using MrBayes v 3.2.6 [78,79]. The Markov Chain Monte Carlo (MCMC) algorithm was run for 100,000 generations and the trees were sampled every 100 generations. The first 25% of the trees were discarded as a burn-in, and remaining trees were used to generate the consensus tree, including clade posterior probability (PP). Convergence was determined by examining the average standard deviation of the split frequencies (<0.01).

4. Conclusions

In this study, we aimed to increase the phylogenetic resolution within the European mountain pine complex using, for the first time, a detailed comprehensive comparative analysis of the complete chloroplast genome sequences of the three main representatives of this complex, i.e., *Pinus mugo*, *P. rotundata* and *P. uncinata*. The obtained results revealed a high conservation of their chloroplast genomes in terms of length, structure and number of genes. We confirmed very close relationships between these three taxa using inference and phylogenetic trees topology in which *P. mugo*, *P. rotundata* and *P. uncinata* form one distinct clade within the genus *Pinus* with strong support. Highly variable regions and distinct microsatellite loci patterns have been identified in the genomes of chloroplast members of the *Pinus mugo* complex that could potentially be used in the future to discriminate and

identify these taxa. Our analyses increase the knowledge of the *Pinus mugo* complex phylogeny and provide a valuable genomic baseline for future research into the evolutionary history and conservation of this highly polymorphic and enigmatic group, as well as the Pinaceae family in general.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/plants10071331/s1, Table S1: List of genes annotated in the chloroplast genomes of P. mugo, P. rotundata and P. uncinata sequenced in this study.

Author Contributions: J.S. and K.C. conceived of and designed the research framework; J.S. performed most of the experiments and data analysis; K.C. participated in the data analysis; J.S. and K.C. wrote the original draft manuscript as well as reviewing and editing the final manuscript; J.S. and H.F. assembled and annotated the genome; J.S. and K.C. collected the samples; K.C. supervised the project. All authors have read and agreed to the published version of the manuscript.

Funding: Praca naukowa finansowana ze środków budżetowych na naukę w latach 2018-2020, jako projekt badawczy w ramach programu "Diamentowy Grant" Nr DI2017003147. Research paper financed from the budget for science in 2018-2020, as a research project under the "Diamond Grant" program No. DI2017003147.

Data Availability Statement: Data is contained within the article.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Christensen, K.L. Taxonomic revision of the *Pinus mugo* complex and *P. rhaetica* (*P. mugo* sylvestris) (Pinaceae). *Nord. J. Bot.* **1987**, *7*, 383–408. [CrossRef]
2. Hamernik, J.; Musil, I. The *Pinus mugo* complex—its structuring and general overview of the used nomenclature. *J. For. Sci.* **2008**, *53*, 253–266. [CrossRef]
3. Critchfield, W.B.; Little, E.L. *Geographic Distribution of the Pines of the World*; Departament of Agriculture, Forest Service: Washington, DC, USA, 1966. [CrossRef]
4. Ballian, D.; Ravazzi, C.; de Rigo, D.C. Pinus mugo in Europe: Distribution, Habitat, Usage and Threats. In *European Atlas of Forest Tree Species*; San-Miguel-Ayanz, J., De Rigo, D., Caudullo, G., Houston Durrant, T., Mauri, A., Eds.; Publication Office of the European Union: Luxembourg, 2016; pp. 124–125.
5. Jalas, J.; Suominen, J. *Atlas Florae Europaeae: Distribution of Vascular Plants in Europe. Gymnospermae, Volume 2*; The Commitee for Mapping the Flora of Europe and Society Biology Fenn: Helsinki, Finland, 1973; p. 40.
6. Businský, R.; Kirschner, J. Nomenclatural Notes on the *Pinus mugo* Complex in Central Europe. *Phyt. Ann. Rei Bot.* **2006**, *46*, 129–139.
7. IUCN. The IUCN Red List of Threatened Species. Version 2020-3. Available online: https://www.iucnredlist.org (accessed on 12 January 2021).
8. Lewandowski, A.; Wiśniewska, M. Short Note: Crossability between Pinus Uliginosa and Its Putative Parental Species *Pinus sylvestris* and *Pinus mugo*. *Silvae Genet.* **2006**, *55*, 52–54. [CrossRef]
9. Wachowiak, W.; Celiński, K.; Prus-Głowacki, W. Evidence of Natural Reciprocal Hybridisation between Pinus Uliginosa and *P. sylvestris* in the Sympatric Population of the Species. *Flora Morphol. Distrib. Funct. Ecol. Plants* **2005**, *200*, 563–568. [CrossRef]
10. Wachowiak, W.; Prus-Głowacki, W. Hybridisation Processes in Sympatric Populations of Pines *Pinus sylvestris* L., *P. mugo* Turra and *P. uliginosa* Neumann. *Plant Syst. Evol.* **2008**, *271*, 29–40. [CrossRef]
11. Wachowiak, W.; Żukowska, W.B.; Wójkiewicz, B.; Cavers, S.; Litkowiec, M. Hybridization in Contact Zone between Temperate European Pine Species. *Tree Genet. Genomes* **2016**, *12*. [CrossRef]
12. Danielewicz, W.Z.J. Ochrona Sosny Błotnej Pinus Uliginosa A. Neumann Na Terenie Borów Dolnośląskich. *Przegląd Przyr.* **2000**, *11*, 113–124.
13. Gołąb, Z. Sosna Błotna (Pinus Uliginosa Neumann) Na Wielkim Torfowisku Batorowskim w Górach Stołowych. *Szczeliniec* **1999**, *3*, 41–48.
14. Boratyńska, K.; Boratyński, A. Taxonomic Differences among Closely Related Pines *Pinus sylvestris*, *P. mugo*, *P. uncinata*, *P. rotundata* and *P. uliginosa* as Revealed in Needle Sclerenchyma Cells. *Flora Morphol. Distrib. Funct. Ecol. Plants* **2007**, *202*, 555–569. [CrossRef]
15. Boratyńska, K.M. Pinus Uncinata Ramond Taxonomy Basedon Needle Characters. *Plant Syst. Evol.* **2001**, *227*, 183–194.
16. Boratyńska, K.; Boratyński, A.; Lewandowski, A. Morphology of Pinus Uliginosa (Pinaceae) Needles from Populations Exposed to and Isolated from the Direct Influence of *Pinus sylvestris*. *Bot. J. Linn. Soc.* **2003**, *124*, 83–91. [CrossRef]

17. Siedlewska, A.; Prus-Głowacki, W. Genetic Structure and Taxonomic Position of Pinus Uliginosa Neumann Population from Wielkie Torfowisko Batorowskie in Stołowe Mts. Locus Classicus. *Acta Soc. Bot. Pol.* **1995**, *64*, 51–58. [CrossRef]
18. Lewandowski, A.; Burczyk, J.; Wachowiak, W.; Boratyński, A.; Prus-Głowacki, W. Genetic Evaluation of Seeds of Highly Endangered Pinus Uliginosa Neumann from Węgliniec Reserve for Ex-Situ Conservation Program. *Acta Soc. Bot. Pol.* **2005**, *74*, 237–242. [CrossRef]
19. Celiński, K.; Pawlaczyk, E.M.; Wojnicka-Półtorak, A.; Chudzińska, E.; Prus-Głowacki, W. Cross-Species Amplification and Characterization of Microsatellite Loci in *Pinus mugo* Turra. *Biology* **2013**, *68*, 621–626. [CrossRef]
20. Heuertz, M.; Teufel, J.; González-Martínez, S.C.; Soto, A.; Fady, B.; Alía, 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]
21. Celiński, K.; Zbránková, V.; Wojnicka-Półtorak, A.; Chudzińska, E. Biogeography and evolutionary factors determine genetic differentiation of *Pinus mugo* (Turra) in the Tatra Mountains (Central Europe). *J. Mt. Sci.* **2015**, *12*, 549–557. [CrossRef]
22. Danusevičius, D.; Marozas, V.; Brazaitis, G.; Petrokas, R.; Christensen, K.I. Spontaneous Hybridization between *Pinus mugo* and *Pinus sylvestris* at the Lithuanian Seaside: A Morphological Survey. *Sci. World J.* **2012**, *2012*, 1–11. [CrossRef] [PubMed]
23. Kormutak, A.; Demankova, B.; Gömöry, D. Spontaneous Hybridization between *Pinus sylvestris* L. and *P. mugo* Turra in Slovakia. *Silvae Genet.* **2008**, *57*, 76–82. [CrossRef]
24. Kormutak, A.; Galgoci, M.; Bolecek, P.; Gőmőry, D.; Libantova, J. Reinforced evidence on partial compatibility between *Pinus sylvestris* and *Pinus mugo* and on maternal inheritance of chloroplast DNA in the *Pinus mugo* × *Pinus sylvestris* cross. *Silvae Genet.* **2020**, *69*, 108–115. [CrossRef]
25. Bogunić, F.; Siljak-Yakovlev, S.; Muratovic, E.; Pustahija, F.; Medjedović, S. Molecular cytogenetics and flow cytometry reveal conserved genome organization in *Pinus mugo* and *P. uncinata*. *Ann. For. Sci.* **2011**, *68*, 179–187. [CrossRef]
26. Celiński, K.; Chudzińska, E.; Gmur, A.; Piosik, Ł.; Wojnicka-Półtorak, A. Cytological characterization of three closely related pines-*Pinus mugo*, *P. uliginosa* and *P.* × *rhaetica* from the *Pinus mugo* complex (Pinaceae). *Biology* **2019**, *74*, 751–756. [CrossRef]
27. Lewandowski, A.; Boratyński, 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]
28. Prus-Głowacki, W.; Bujas, E.; Ratyńska, H. Taxonomic Position of Pinus Uliginosa Neumann as Related to Other Taxa of *Pinus mugo* Complex. *Acta Soc. Bot. Pol.* **1998**, *67*, 269–274. [CrossRef]
29. Celiński, K.; Kijak, H.; Barylski, J.; Grabsztunowicz, M.; Wojnicka-Półtorak, A.; Chudzińska, E. Characterization of the complete chloroplast genome of *Pinus uliginosa* (Neumann) from the *Pinus mugo* complex. *Conserv. Genet. Resour.* **2016**, *9*, 209–212. [CrossRef]
30. Celiński, K.; Kijak, H.; Wojnicka-Półtorak, A.; Buczkowska-Chmielewska, K.; Sokołowska, J.; Chudzińska, E. Effectiveness of the DNA barcoding approach for closely related conifers discrimination: A case study of the *Pinus mugo* complex. *Comptes Rendus Biol.* **2017**, *340*, 339–348. [CrossRef]
31. Celiński, K.; Sokołowska, J.; Zemleduch-Barylska, A.; Kuna, R.; Kijak, H.; Staszak, A.M.; Wojnicka-Półtorak, A.; Chudzińska, E. Seed Total Protein Profiling in Discrimination of Closely Related Pines: Evidence from the *Pinus mugo* Complex. *Plants* **2020**, *9*, 872. [CrossRef]
32. Celiński, K.; Bonikowski, R.; Wojnicka-Półtorak, A.; Chudzińska, E.; Maliński, T. Volatiles as Chemosystematic Markers for Distinguishing Closely Related Species within the *Pinus mugo* Complex. *Chem. Biodivers.* **2015**, *12*, 1208–1213. [CrossRef]
33. Bonikowski, R.; Celinski, K.; Wojnicka-Półtorak, A.; Maliński, T. Composition of Essential Oils Isolated from the Needles of *Pinus uncinata* and *P. uliginosa* Grown in Poland. *Nat. Prod. Commun.* **2015**, *10*, 371–373. [CrossRef]
34. Cavers, S.; Wachowiak, W.; Boratyńska, K. 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.
35. Li, X.; Yang, Y.; Henry, R.; Rossetto, M.; Wang, Y.; Chen, S. Plant DNA barcoding: From gene to genome. *Biol. Rev.* **2015**, *90*, 157–166. [CrossRef] [PubMed]
36. Sudianto, E.; Wu, C.-S.; Lin, C.-P.; Chaw, S.-M. Revisiting the Plastid Phylogenomics of Pinaceae with Two Complete Plastomes of Pseudolarix and Tsuga. *Genome Biol. Evol.* **2016**, *8*, 1804–1811. [CrossRef]
37. Yang, Z.; Zhao, T.; Ma, Q.; Liang, L.; Wang, G. Comparative Genomics and Phylogenetic Analysis Revealed the Chloroplast Genome Variation and Interspecific Relationships of Corylus (Betulaceae) Species. *Front. Plant Sci.* **2018**, *9*, 927. [CrossRef] [PubMed]
38. Li, Y.; Sylvester, S.P.; Li, M.; Zhang, C.; Li, X.; Duan, Y.; Wang, X. The Complete Plastid Genome of *Magnolia zenii* and Genetic Comparison to Magnoliaceae species. *Molecules* **2019**, *24*, 261. [CrossRef] [PubMed]
39. Li, X.; Li, Y.; Zang, M.; Li, M.; Fang, Y. Complete Chloroplast Genome Sequence and Phylogenetic Analysis of Quercus acutissima. *Int. J. Mol. Sci.* **2018**, *19*, 2443. [CrossRef]
40. Jansen, R.K.; Ruhlman, T.A. Plastid Genomes of Seed Plants. In *Genomics of Chloroplasts and Mitochondria. Advances in Photosynthesis and Respiration (Including Bioenergy and Related Processes)*; Bock, R., Knoop, V., Eds.; Springer: Dordrecht, The Netherlands, 2012; Volume 35, pp. 103–126.
41. Lin, M.; Qi, X.; Chen, J.; Sun, L.; Zhong, Y.; Fang, J.; Hu, C. The complete chloroplast genome sequence of Actinidia arguta using the PacBio RS II platform. *PLoS ONE* **2018**, *13*, e0197393. [CrossRef] [PubMed]

42. Kang, H.-I.; Lee, H.O.; Lee, I.H.; Kim, I.S.; Lee, S.-W.; Yang, T.J.; Shim, D. Complete Chloroplast Genome of Pinus densiflora Siebold & Zucc. and Comparative Analysis with Five Pine Trees. *Forests* **2019**, *10*, 600. [CrossRef]
43. Qiu, J.; Chen, L.; Yi, X.; Li, M. The complete chloroplast genome of Pinus yunnanensis Franchet (Pinaceae). *Mitochondrial DNA Part B* **2019**, *4*, 2600–2601. [CrossRef]
44. Kim, K.-J. Complete Chloroplast Genome Sequences from Korean Ginseng (Panax schinseng Nees) and Comparative Analysis of Sequence Evolution among 17 Vascular Plants. *DNA Res.* **2004**, *11*, 247–261. [CrossRef]
45. Perry, A.; Wolfe, K.H. Nucleotide Substitution Rates in Legume Chloroplast DNA Depend on the Presence of the Inverted Repeat. *J. Mol. Evol.* **2002**, *55*, 501–508. [CrossRef] [PubMed]
46. Jiang, M.; Chen, H.; He, S.; Wang, L.; Chen, A.J.; Liu, C. Sequencing, Characterization, and Comparative Analyses of the Plastome of Caragana rosea var. rosea. *Int. J. Mol. Sci.* **2018**, *19*, 1419. [CrossRef]
47. Yi, X.; Gao, L.; Wang, B.; Su, Y.-J.; Wang, T. The Complete Chloroplast Genome Sequence of Cephalotaxus oliveri (Cephalotaxaceae): Evolutionary Comparison of Cephalotaxus Chloroplast DNAs and Insights into the Loss of Inverted Repeat Copies in Gymnosperms. *Genome Biol. Evol.* **2013**, *5*, 688–698. [CrossRef]
48. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [CrossRef]
49. Celiński, K.; Kijak, H.; Wiland-Szymańska, J. Complete Chloroplast Genome Sequence and Phylogenetic Inference of the Canary Islands Dragon Tree (*Dracaena draco* L.). *Forests* **2020**, *11*, 309. [CrossRef]
50. Dong, W.; Xu, C.; Li, C.; Sun, J.; Zuo, Y.; Shi, S.; Cheng, T.; Guo, J.; Zhou, S. ycf1, the most promising plastid DNA barcode of land plants. *Sci. Rep.* **2015**, *5*, 8348. [CrossRef]
51. Olsson, S.; Grivet, D.; Cid-Vian, J. Species-diagnostic markers in the genus Pinus: Evaluation of the chloroplast regions matK and ycf. *For. Syst.* **2018**, *27*, e016. [CrossRef]
52. Ellegren, H. Microsatellites: Simple sequences with complex evolution. *Nat. Rev. Genet.* **2004**, *5*, 435–445. [CrossRef] [PubMed]
53. Urbaniak, L.; Wojnicka-Półtorak, A.; Celinski, K.; Lesiczka, P.; Pawlaczyk, E.; Aučina, A. Genetic resources of relict populations of *Pinus sylvestris* (L.) in Western Carpathians assessed by chloroplast microsatellites. *Biologia* **2019**, *74*, 1077–1086. [CrossRef]
54. Oliveira, E.; Pádua, J.G.; Zucchi, M.I.; Vencovsky, R.; Vieira, M.L.C. Origin, evolution and genome distribution of microsatellites. *Genet. Mol. Biol.* **2006**, *29*, 294–307. [CrossRef]
55. Gómez, A.; González-Martínez, S.C.; Collada, C.; Climent, J.; Gil, L. Complex population genetic structure in the endemic Canary Island pine revealed using chloroplast microsatellite markers. *Theor. Appl. Genet.* **2003**, *107*, 1123–1131. [CrossRef] [PubMed]
56. Asaf, S.; Khan, A.L.; Khan, M.A.; Shahzad, R.; Lubna; Kang, S.M.; Al-Harrasi, A.; Al-Rawahi, A.; Lee, I.-J. Complete chloroplast genome sequence and comparative analysis of loblolly pine (Pinus taeda L.) with related species. *PLoS ONE* **2018**, *13*, e0192966. [CrossRef]
57. Dzialuk, A.; Boratyńska, K.; Romo, A.; Boratynski, A. Taxonomic and geographic variation of the *Pinus mugo* complex on chloroplast microsatellite markers. *Syst. Biodivers.* **2016**, *15*, 464–479. [CrossRef]
58. Dzialuk, A.; Muchewicz, E.; Boratyński, A.; Montserrat, J.M.; Boratyńska, K.; Burczyk, J. Genetic variation of Pinus uncinata (Pinaceae) in the Pyrenees determined with cpSSR markers. *Plant Syst. Evol.* **2009**, *277*, 197–205. [CrossRef]
59. Li, D.-M.; Zhao, C.-Y.; Liu, X.-F. Complete Chloroplast Genome Sequences of Kaempferia Galanga and Kaempferia Elegans: Molecular Structures and Comparative Analysis. *Molecules* **2019**, *24*, 474. [CrossRef] [PubMed]
60. Vu, H.-T.; Tran, N.; Nguyen, T.-D.; Vu, Q.-L.; Bui, M.-H.; Le, M.-T.; Le, L. Complete Chloroplast Genome of Paphiopedilum delenatii and Phylogenetic Relationships among Orchidaceae. *Plants* **2020**, *9*, 61. [CrossRef] [PubMed]
61. Zhang, Z.; Zhang, Y.; Song, M.; Guan, Y.; Ma, X. Species Identification of Dracaena Using the Complete Chloroplast Genome as a Super-Barcode. *Front. Pharmacol.* **2019**, *10*, 1441. [CrossRef]
62. Ge, J.; Cai, L.; Bi, G.-Q.; Chen, G.; Sun, W. Characterization of the Complete Chloroplast Genomes of Buddleja colvilei and B. sessilifolia: Implications for the Taxonomy of Buddleja L. *Molecules* **2018**, *23*, 1248. [CrossRef]
63. Hernández-León, S.; Gernandt, D.S.; De La Rosa, J.A.P.; Barbolla, L.J. Phylogenetic Relationships and Species Delimitation in Pinus Section Trifoliae Inferrred from Plastid DNA. *PLoS ONE* **2013**, *8*, e70501. [CrossRef]
64. Doyle, J.J.; Doyle, J.L. Isolation of Plants DNA from Fresh Tissue. *Focus* **1990**, *12*, 13–15.
65. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649. [CrossRef] [PubMed]
66. Shi, L.; Chen, H.; Jiang, M.; Wang, L.; Wu, X.; Huang, L.; Liu, C. CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Res.* **2019**, *47*, W65–W73. [CrossRef] [PubMed]
67. Tillich, M.; Lehwark, P.; Pellizzer, T.; Ulbricht-Jones, E.S.; Fischer, A.; Bock, R.; Greiner, S. GeSeq–versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* **2017**, *45*, W6–W11. [CrossRef]
68. Chan, P.P.; Lowe, T.M. tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. *Methods Mol. Biol.* **2019**, *1962*, 1–14. [CrossRef]
69. Greiner, S.; Lehwark, P.; Bock, R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* **2019**, *47*, W59–W64. [CrossRef]
70. Darling, A.E.; Mau, B.; Perna, N.T. progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. *PLoS ONE* **2010**, *5*, e11147. [CrossRef]

71. Kimura, M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **1980**, *16*, 111–120. [CrossRef] [PubMed]
72. Katoh, K.; Standley, D.M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef]
73. Rozas, J.; Ferrer-Mata, A.; Sánchez-DelBarrio, J.C.; Guirao-Rico, S.; Librado, P.; Ramos-Onsins, S.; Sánchez-Gracia, A. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Mol. Biol. Evol.* **2017**, *34*, 3299–3302. [CrossRef]
74. Bi, Y.; Zhang, M.-F.; Xue, J.; Dong, R.; Du, Y.-P.; Zhang, X.-H. Chloroplast genomic resources for phylogeny and DNA barcoding: A case study on Fritillaria. *Sci. Rep.* **2018**, *8*, 1–12. [CrossRef] [PubMed]
75. Beier, S.; Thiel, T.; Münch, T.; Scholz, U.; Mascher, M. MISA-web: A web server for microsatellite prediction. *Bioinformatics* **2017**, *33*, 2583–2585. [CrossRef]
76. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control.* **1974**, *19*, 716–723. [CrossRef]
77. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [CrossRef]
78. Huelsenbeck, J.P.; Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **2001**, *17*, 754–755. [CrossRef] [PubMed]
79. Ronquist, F.; Huelsenbeck, J.P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **2003**, *19*, 1572–1574. [CrossRef] [PubMed]

Supplementary Table S1

Supplementary Table S1. List of genes annotated in the chloroplast genomes of *P. mugo*, *P. rotundata* and *P. uncinata* sequenced in this study.

No.	Classification of genes	Name of genes	Number
1	Photosystem I	*psaA, psaB, psaC, psaI, psaJ, psaM(x2), ycf3, ycf4*	9
2	Photosystem II	*psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ*	15
3	Cytochrome b/f complex	*petA, petB, petD, petG, petL, petN*	6
4	ATP synthase	*atpA, atpB, atpE, atpF, atpH, atpI*	6
5	NADH dehydrogenase	*ndhB, ndhC, ndhD, ndhE, ndhH, ndhI, ndhK*	7
6	RubisCO large subunit	*rbcL*	1
7	RNA polymerase	*rpoA, rpoB, rpoC1, rpoC2*	4
8	Ribosomal proteins – small units (SSU)	*rps2, rps3, rps4, rps7, rps8, rps11, rps12, rps14, rps15, rps18, rps19*	11
9	Ribosomal proteins – large units (LSU)	*rpl2, rpl14, rpl16, rpl20, rpl22, rpl23, rpl32, rpl33, rpl36*	9
10	Other genes/Miscellaneous	*accD, ccsA, cemA, chlB, chlL, chlN, clpP, infA, matK*	9
11	Protein of unknown function / Hypothetical chloroplast reading frames	*ycf1, ycf2, ycf12, ycf68*	4
12	Transfer RNAs	*trnA-UGC, trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnfM-CAU, trnG-GCC, trnG-UCC, trnH-GUG* (x2), *trnI-GAU, trnK-UUU, trnL-CAA, trnL-UAG, trnL-UAA, trnM-CAU* (x3), *trnN-GUU, trnP-GGG, trnP-UGG, trnQ-UUG, trnR-ACG, trnR-CCG, trnR-UCU, trnS-UGA, trnS-GGA, trnS-GCU* (x2), *trnT-UGU, trnT-GGU, trnV-GAC* (x2), *trnV-UAC, trnW-CCA, trnY-GUA*	36
13	Ribosomal RNAs	*rrn4.5, rrn5, rrn16, rrn23*	4
	Total		**121**

(x2) indicates that the number of the repeat unit is 2

4.2. Publikacja 2

Według China Plant BOL Group (China Plant BOL Group i in. 2011) region jądrowego DNA rybosomalnego – ITS (ang. *internal transcribed spacer*), a w szczególności ITS2, posiada znacznie lepszą zdolność do rozróżniania gatunków w porównaniu z regionami plastydowymi, ponieważ charakteryzuje się wyższym tempem mutacji niż regiony kodujące i jest dziedziczony od obojga rodziców, co pozwala na jego wykorzystanie do identyfikacji nawet blisko spokrewnionych gatunków (Álvarez i Wendel 2003). Wobec powyższego, tego typu marker DNA byłby bardzo przydatny do weryfikacji hipotezy o mieszańcowym pochodzeniu sosny błotnej.

W związku z tym w **drugiej części mojej rozprawy doktorskiej dokonałam oceny skuteczności stosowania regionu jądrowego cistronu rybosomalnego (ang. *nuclear ribosomal* DNA, nrDNA) ITS2 do dyskryminacji bliżej i dalej spokrewnionych taksonów oraz wnioskowania filogenetycznego w rodzinie *Pinaceae*, w tym także w kompleksie *Pinus mugo* (Sokołowska i in. 2022).** W tym celu przeanalizowanych zostało 368 sekwencji ITS2 reprezentujących 71 taksonów z siedmiu rodzajów (tj. *Abies, Keteleeria, Tsuga, Larix, Pseudotsuga, Picea* i *Pinus*), należących do trzech podrodzin (*Abietoideae, Laricoideae, Pinoideae*) w obrębie rodziny *Pinaceae*. Do analiz wykorzystano uzyskane poprzez sekwencjonowanie nowej generacji, 22 sekwencje ITS2 dla przedstawicieli kompleksu *Pinus mugo* oraz 346 sekwencje pobrane z GenBank. Jak dotąd nie przeprowadzono tak kompleksowego badania podsumowującego zasadność wykorzystania regionu ITS2 do rozróżniania i identyfikacji taksonów, a także wnioskowania filogenetycznego w tej rodzinie.

Przeprowadzone analizy obejmowały ocenę przydatności sekwencji ITS2 jako potencjalnego regionu barkodowego, w tym: oszacowanie zmienności wewnątrz- i międzygatunkowej pomiędzy analizowanymi taksonami, jak również identyfikację oraz dyskryminację analizowanych taksonów na podstawie sekwencji ITS2 na różnych poziomach taksonomicznych, tj. kompleksu gatunkowego (kompleks *Pinus mugo*), rodzaju, podrodziny i rodziny. Do tego celu wykorzystano trzy metody, a mianowicie: 1) metodę opartą na odległości genetycznej (ang. *Pairwise Genetic Distance*, PWG-distance); 2) metodę opartą na podobieństwie sekwencji DNA (ang. *sequence similarity-based method*) z wykorzystaniem programów TaxonDNA i BLAST oraz 3) wnioskowanie filogenetyczne z wykorzystaniem metody największej wiarygodności (ang. *Maximum Likelihood*, ML).

Na podstawie uzyskanych wyników wykazano, że analizowane sekwencje ITS2 różnią się znacznie pod względem wartości podstawowych parametrów opisujących zmienność genetyczną, w zależności od rozpatrywanego poziomu taksonomicznego. Całkowita średnia

odległość genetyczna dla rodziny Pinaceae wyniosła 0,342. Oszacowane wartości podstawowych parametrów charakteryzujących zmienność genetyczną sekwencji ITS2 były najwyższe na poziomie rodziny, a najniższe na poziomie kompleksu gatunkowego, tj. kompleksu *Pinus mugo*, u którego nie zaobserwowano żadnej zmienności sekwencji. Dodatkowo na podstawie metody opartej na podobieństwie sekwencji, wykazano że 100% analizowanych sekwencji ITS2 zostało niejednoznacznie zidentyfikowanych przy użyciu opcji BM (ang. *Best Match*) i BCM (ang. *Best Close Match*). Żadna sekwencja ITS2 w kompleksie *Pinus mugo* nie została zidentyfikowana jako poprawna (0%) lub niepoprawna (0%). Z kolei ogólna topologia otrzymanego drzewa filogenetycznego na podstawie metody największej wiarygodności (ang. Maximum Likelihood, ML) oparta na sekwencjach ITS2 jest zgodna z przyjętym podziałem rodziny Pinaceae, a wszystkie analizowane taksony zostały przypisane do odpowiednich rodzajów zgodnie z powszechnie przyjętą taksonomią, jednakże ich rozróżnienie w ramach tych rodzajów nie było jednoznaczne.

Podsumowując, na podstawie uzyskanych wyników można stwierdzić, że wykorzystanie sekwencji ITS2 do analiz filogenetycznych blisko spokrewnionych taksonów z rodziny Pinaceae, w tym kompleksu *Pinus mugo* nie wydaje się być zasadne ze względu na zbyt niską zmienność genetyczną tego regionu.

PUBLIKACJA 2

Sokołowska, J.; Fuchs, H.; Celiński, K. Assessment of ITS2 Region Relevance for Taxa Discrimination and Phylogenetic Inference among Pinaceae. Plants 2022, 11, 1078.

https://doi.org/10.3390/plants11081078.

plants

MDPI

Article

Assessment of ITS2 Region Relevance for Taxa Discrimination and Phylogenetic Inference among Pinaceae

Joanna Sokołowska [1], Hanna Fuchs [2] and Konrad Celiński [1,*]

[1] Department of Genetics, Institute of Experimental Biology, Faculty of Biology, School of Natural Sciences, Adam Mickiewicz University, Poznań, Uniwersytetu Poznańskiego 6, 61-614 Poznań, Poland; joanna.sokolowska@amu.edu.pl

[2] Institute of Dendrology, Polish Academy of Sciences, Parkowa 5, 62-035 Kórnik, Poland; hkijak@man.poznan.pl

* Correspondence: konrad.celinski@amu.edu.pl

Abstract: The internal transcribed spacer 2 (ITS2) is one of the best-known universal DNA barcode regions. This short nuclear region is commonly used not only to discriminate taxa, but also to reconstruct phylogenetic relationships. However, the efficiency of using ITS2 in these applications depends on many factors, including the family under study. Pinaceae represents the largest family of extant gymnosperms, with many species of great ecological, economic, and medical importance. Moreover, many members of this family are representatives of rare, protected, or endangered species. A simple method for the identification of Pinaceae species based on DNA is necessary for their effective protection, authentication of products containing Pinaceae representatives, or phylogenetic inference. In this study, for the first time, we conducted a comprehensive study summarizing the legitimacy of using the ITS2 region for these purposes. A total of 368 sequences representing 71 closely and distantly related taxa of the seven genera and three subfamilies of Pinaceae were characterized for genetic variability and divergence. Intra- and interspecies distances of ITS2 sequences as well as rates of sequence identification and taxa discrimination among Pinaceae at various taxonomic levels, i.e., the species complex, genus, subfamily, and family, were also determined. Our study provides a critical assessment of the suitability of the ITS2 nuclear DNA region for taxa discrimination among Pinaceae. The obtained results clearly show that its usefulness for this purpose is limited.

Keywords: Pinaceae; internal transcribed spacer; ITS2; DNA barcoding; taxa identification

Citation: Sokołowska, J.; Fuchs, H.; Celiński, K. Assessment of ITS2 Region Relevance for Taxa Discrimination and Phylogenetic Inference among Pinaceae. *Plants* **2022**, *11*, 1078. https://doi.org/10.3390/plants11081078

Academic Editor: Igor Bartish

Received: 4 March 2022
Accepted: 12 April 2022
Published: 15 April 2022

1. Introduction

The Pinaceae family is the largest extant family of all gymnosperms [1]. Within this family, 225 species have been distinguished, which are grouped in three subfamilies (Pinoideae Pilg., Laricoideae Melchior et Werdermann, Abietoideae Pilg.) and eleven genera, i.e., *Abies* Mill., *Cathaya* Chun & Kuang, *Cedrus* Belon ex Trew, *Keteleeria* Carriere., *Larix* Mill., *Nothotsuga* Hu ex C. N. Page, *Picea* A. Dietr., *Pinus* L., *Pseudolarix* Gordon, *Pseudotsuga* Carriere, and *Tsuga* Carriere [2]. Representatives of the Pinaceae are an extremely important component of temperate, boreal, subalpine, and subtropical forests in the northern hemisphere. Moreover, conifers from the Pinaceae family are of great economic importance, especially the genus *Pinus*, which has been used for many years as a valuable source of wood, resins, essential oils, and seeds [3]. Many studies also show that members of this family represent a valuable source of active substances with medicinal properties [4–8].

The high economic and medical value of members of this family may be a strong temptation to over-exploit particular species, obtain them illegally, or even falsify products containing them. Since many species of the Pinaceae family are rare, endangered, or protected, this can pose a serious threat. Another equally serious risk is the misidentification of certain species. There are many closely related taxa in the Pinaceae family with a very similar needle or cone morphology. Moreover, they have the same ecological niches, similar

geographic ranges, and often grow together in sympatric populations. In such populations, gene flow occurs, which leads to the formation of hybrid individuals with a phenotype intermediate to the parent taxa. Thus, the processes of hybridization and introgression additionally hinder the simple and effective assignment of such individuals to particular species. One of the best examples of such a complex system involving all of the above-mentioned factors is the European mountain pine association, *Pinus mugo* Turra *sensu lato*. This complex has been the subject of research using various techniques and methods for several decades [9–15]. However, the main problem in this complex remains the same, i.e., finding easy-to-use specific diagnostic determinants for individual taxa, preferably based on DNA markers.

The unambiguous identification of plant material or distinguishing taxa plays an important role not only in the authentication of Pinaceae-containing products, but also in establishing genetic relationships in this family, characterizing the genetic resources of its taxa, determining their geographical distribution, developing conservation plans for individual taxa, monitoring gene flow, and finally analyzing various evolutionary processes, such as hybridization or introgression.

DNA barcoding is a popular method of species identification using short and standardized gene fragments as molecular markers [16]. This rapid approach has commonly been used in taxonomic research for species-level identification in different domains of the tree of life for several years. The undeniable features of good DNA barcode regions are easy PCR amplification and sequencing, and most importantly, ensuring a high level of species discrimination [17]. The short length of the PCR amplicons facilitates work with very degraded DNA, e.g., from herbarium specimens. However, the effectiveness of identifying the analyzed biological material with DNA barcodes and assigning it to a specific family, genus, or species depends on many factors (including, among others, the analyzed family, DNA region or abundance of the reference base) and is not always satisfactorily high [18–21]. Therefore, in recent years, many studies have been carried out to circumvent these limitations by selecting appropriate regions of DNA that guarantee a high percentage of satisfactory identifications, both for universal barcoding and for species identification within individual families or genera [22–26]. One of the best known and widely used universal DNA barcode regions is the internal transcribed spacer 2 (ITS2) [27]. This short nuclear region, due to its properties, is also used to reconstruct phylogenetic relationships at the species and genus level [28–32]. So far, ITS2 has been used to analyze taxa from the Pinaceae family rather sporadically, while evaluating the usefulness of this region as barcode DNA for various groups of plants and animals [26,27].To the best of the authors' knowledge, so far, there is no comprehensive study summarizing the validity of using the ITS2 region for the discrimination and identification of taxa as well as phylogenetic inference in this family carried out on a large number of species and sequences.

Therefore, the main objective of this study is to determine whether ITS2: (1) is a useful DNA marker for the discrimination and identification of both distant and closely related taxa in the Pinaceae family; (2) has a sufficiently high potential for the authentication of products containing Pinaceae representatives; and (3) is valuable nuclear region for inferring phylogenetic relationships of conifers in this family.

2. Results

2.1. Genetic Variation of ITS2 Sequences in Pinaceae

A total of 368 sequences representing 71 Pinaceae conifer taxa were analyzed (Table S1), of which 346 sequences were downloaded from GenBank and the remaining 22 sequences were obtained in this study and deposited in the same genetic database (Table S2). The accession numbers of the newly obtained sequences are shown in Table S3.

The genetic variability of the ITS2 sequence in the Pinaceae family was precisely characterized. All analyses and calculations were performed separately for four different taxonomic levels, i.e., species complex: *Pinus mugo* complex; genus: *Abies*, *Keteleeria*, *Tsuga*,

Larix, *Pseudotsuga*, *Picea,* and *Pinus*; subfamily: Abietoideae, Laricoideae, Pinoideae, and family: Pinaceae.

The obtained results indicate that the analyzed ITS2 sequences in Pinaceae differ significantly in the values of the basic parameters describing genetic variation depending on the considered taxonomic level (Table 1). At the species complex level, represented by the closely related European mountain pine complex, *Pinus mugo* complex (PMC), no sequence variation was observed. At the genera level, the alignment length varied only slightly. The shortest alignment length was noticed for the genus *Pseudotsuga* (232 bp), while the longest for the genus *Pinus* (252 bp). The number of conserved sites (CS) ranged from 70 to 237. The highest percentage of calculated conserved sites (CS) was observed for the genus *Pseudotsuga* (100%) and the lowest for the *Pinus* genus (27.78%). The number of variable sites (VS) ranged from 0 (for the genus *Pseudotsuga*) to 177 (for the *Pinus* genus), which gives 0% and 70.24% respectively. The genus *Pseudotsuga* was also characterized by zero parsimony informative sites (PIS) and singleton sites (SS), and consequently also zero overall mean distance (OMD). On the other hand, the highest values of these coefficients were recorded for the genus *Pinus* (28.97%) and *Keteleeria* (48.5%), for PIS and SS, respectively. The highest OMD value (0.236) was found for the genus *Keteleeria*. At the subfamily level, the alignment length varied from 233 bp to 254 bp. The number of conserved sites ranged from 61 to 148. The highest percentage of calculated conserved sites (CS) was observed for the Laricoideae (63.52%) and the lowest for the Pinoideae subfamily (24.21%). The number of variable sites varied substantially from 85 to 187 within the subfamilies. The lowest proportion of variable sites (VS) was observed for the Laricoideae (36.48%) and the highest for the Pinoideae (74.21%) followed by Abietoideae (73.23%). The number of parsimony informative sites (PIS) ranged from 48 (Laricoideae) to 107 (Pinoideae), which is 20.60% and 42.4%, respectively. The values of the number of singleton sites (SS) and overall mean distance (OMD) were also the lowest for the members of the Laricoideae subfamily and amounted to 15.88% and 0.023, respectively. The highest percentage of calculated singleton sites was observed for the Abietoideae subfamily (38.19%), and the highest overall mean distance (OMD) was noticed for the Pinoideae (0.149). At the family level, the estimated alignment length was equal to 270 bp. The proportion of calculated conserved sites (CS), variable sites (VS), parsimony informative sites (PIS), singleton sites (SS) was: 17 (6.30%), 250 (92.59%), 196 (72.5%), and 49 (18.15%), respectively.

Table 1. Values of basic parameters characterizing genetic variation in ITS2 sequences in Pinaceae. AL—alignment length; CS—conserved sites; VS-variable sites; PIS—parsimony informative sites; SS—singleton sites; OMD—overall mean distance. Values in brackets are given as percentages.

Taxonomic Group/Level	Name	AL	CS	VS	PIS	SS	OMD
Species complex	*P. mugo complex*	243	243 (100.00)	0 (0.00)	0 (0.00)	0 (0.00)	0.000
Genera	*Abies*	246	169 (68.70)	75 (30.49)	27 (10.98)	48 (19.51)	0.016
	Keteleeria	249	118 (47.39)	122 (48.99)	0 (0.00)	121 (48.5)	0.236
	Tsuga	247	237 (95.95)	10 (4.05)	7 (2.83)	3 (1.21)	0.026
	Larix	233	172 (73.82)	61 (26.18)	15 (6.44)	46 (19.74)	0.013
	Pseudotsuga	232	232 (100.00)	0 (0.00)	0 (0.00)	0 (0.00)	0.000
	Picea	236	232 (98.31)	4 (1.69)	3 (1.27)	1 (0.42)	0.004
	Pinus	252	70 (27.78)	177 (70.24)	73 (28.97)	101 (40.0)	0.077
Subfamily	Abietoideae	254	66 (25.98)	186 (73.23)	88 (34.65)	97 (38.19)	0.074
	Laricoideae	233	148 (63.52)	85 (36.48)	48 (20.60)	37 (15.88)	0.023
	Pinoideae	252	61 (24.21)	187 (74.21)	107 (42.4)	77 (30.56)	0.149
Family	Pinaceae	270	17 (6.30)	250 (92.59)	196 (72.5)	49 (18.15)	0.342

The overall mean distance (OMD) for the Pinaceae is 0.342. In general, the highest values of the basic parameters characterizing the genetic variation of the ITS2 sequence in Pinaceae were found at the family level, and the lowest at the species complex level.

2.2. Genetic Divergence within and between Pinaceae Taxa

MEGA version X was used to calculate the genetic divergence of ITS2 sequences in the Pinaceae family. Table 2 summarizes in detail the values of five genetic divergence parameters, i.e., all interspecific distance, minimum interspecific distances, theta, all intraspecific distance, and coalescence depth obtained in this study at the species complex, genus, subfamily, and family levels.

Table 2. Values of genetic divergence indices in ITS2 sequences in Pinaceae.

Taxonomic Group/Level	Name	All Interspecific Distance	Minimum Interspecific Distance	All Intraspecific Distance	Theta	Coalescent Depth
Species complex	*P. mugo complex*	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000
Genera	*Abies*	0.0175 ± 0.0006	0.0000 ± 0.0000	0.0085 ± 0.0012	0.0176 ± 0.0039	0.0190 ± 0.0046
	Keteleeria	0.1998 ± 0.0973	0.0000 ± 0.0000	0.1972 ± 0.1919	0.1995 ± 0.0250	0.2958 ± 0.2874
	Tsuga	0.0356 ± 0.0119	0.0291 ± 0.0109	0.0061 ± 0.0061	0.0258 ± 00820	0.0062±0.0061
	Larix	0.0137 ± 0.0006	0.0000 ± 0.0000	0.0067 ± 0.0010	0.0130 ± 0.0033	0.0271 ± 0.0074
	Pseudotsuga	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000
	Picea	0.0048 ± 0.0009	0.0000 ± 0.0000	0.0017 ± 0.0012	0.0043 ± 0.0023	0.0024 ± 0.0018
	Pinus	0.0392 ± 0.0257	0.0000 ± 0.0000	0.0959 ± 0.0050	0.0768 ± 0.0102	0.0486 ± 0.0259
Subfamily	Abietoideae	0.1089±0.0082	0.0000 ± 0.0000	0.0370 ± 0.0291	0.0665 ± 0.0065	0.0562 ± 0.0389
	Laricoideae	0.0436 ± 0.0083	0.0000 ± 0.0000	0.0063 ± 0.0011	0.0222 ± 0.0037	0.0249 ± 0.0071
	Pinoideae	0.1702 ± 0.0059	0.0000 ± 0.0000	0.0342 ± 0.0229	0.1492 ± 0.0157	0.0396 ± 0.0210
Family	Pinaceae	0.3586 ± 0.0041	0.0000 ± 0.0000	0.0294 ± 0.0141	0.3417 ± 0.0287	0.0423 ± 0.0162

At the *P. mugo* complex level, all genetic divergence indices were zero. At the genera level, the highest all interspecific and all intraspecific variation was observed in the genus *Keteleeria* (0.1998 ± 0.0973 and 0.1972 ± 0.1919, respectively). This genus was also characterized by the highest value of theta (0.1995 ± 0.0250) and coalescent depth (0.2958 ± 0.2874). The lowest values of the above-mentioned coefficients were observed for the genus *Picea*. The genus *Pseudotsuga* was the only one represented by only one taxon. Therefore, the value of all five coefficients for this type was zero. At the subfamily level, the lowest all interspecific and all intraspecific distances were revealed for Laricoideae (0.0436 ± 0.0083 and 0.0063 ± 0.0011). In turn, the highest all interspecific distance was found in Pinoideae (0.1702 ± 0.0059), and the highest all intraspecific distance in Abietoideae (0.0370 ± 0.0291). For all analyzed subfamilies, the minimum interspecific distance was equal to 0.0000. The theta value ranged from 0.0222 for Laricoideae to 0.1492 for Pinoideae. The highest value of coalescent depth was observed for Abietoideae (0.0562) while the lowest (0.0249) for Laricoideae. At the family level, all interspecific and all intraspecific distances parameters were 0.3586 and 0.0294, respectively. The theta value was equal 0.3417 and coalescent depth value was 0.0423.

The pairwise genetic distance-based method was used to determine the genetic divergence of the analyzed ITS2 sequences. Figure 1 shows the relative distribution of all intraspecific and all interspecific distances obtained for the analyzed samples from the Pinaceae family. The value of intraspecific distance in the range from 0.0% to 1.0% constituted 46.04% of the total observation, while the values of interspecific distance in the range of 0.0 to 1.0 and 1.0 to 2.0 were respectively 14.74% and 14.39%. Overall, the vast majority of intraspecific distance values (95.06%) are in the range from 0% to 8%, with 83.93% in the range from 0.0% to 2.0%. The interspecific value of the distance in the range from 0% to 14% constitutes 90.49% of the calculations, and the range of distance from 0% to 6% constitutes 67.75% (Table S4).

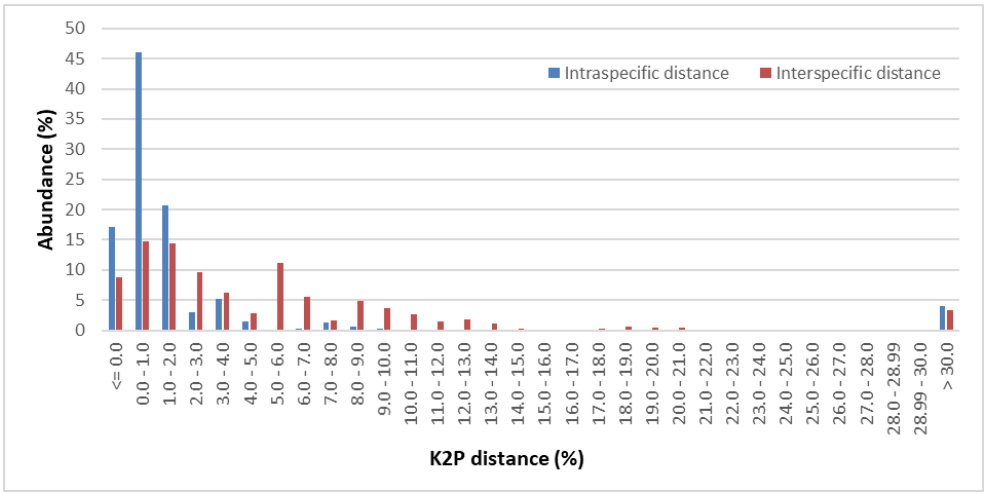


Figure 1. Abundance of all intraspecific and all interspecific K2P pairwise distance in Pinaceae.

2.3. Rates of Sequences Identification and Taxa Discrimination among Pinaceae

BLAST1 and TAXONDNA/Species Identifier 1.8 were used to determine the percentage of correct, ambiguous, and incorrect ITS2 sequence identification and taxa discrimination among Pinaceae.

Using the BLAST1 method, the correct ITS2 sequence identification rate was the highest for the *Tsuga* genus (100%) and the lowest for *Pseudotsuga* (0%). More than 40% of ITS2 sequences were correctly identified in the genus *Abies* and just over 30% in the genus *Pinus*. In contrast, in the genus *Picea*, only 8.33% of the ITS2 sequences were correctly identified. Ambiguous ITS2 sequences identification was highest for *Pseudotsuga* (100%) and lowest (0%) for *Tsuga*. Generally, for six out of seven analyzed Pinaceae genera, the value of ambiguous identification rate was more than 50%. Incorrect identification rates were highest for *Picea* (25%), followed by *Pinus* (13.51%) and *Abies* (4.76%). Overall success in correct species discrimination based on the BLAST1 method and ITS2 sequences was moderate (32.88%). Figure 2 and Table S5 summarize the results obtained with BLAST1 method.

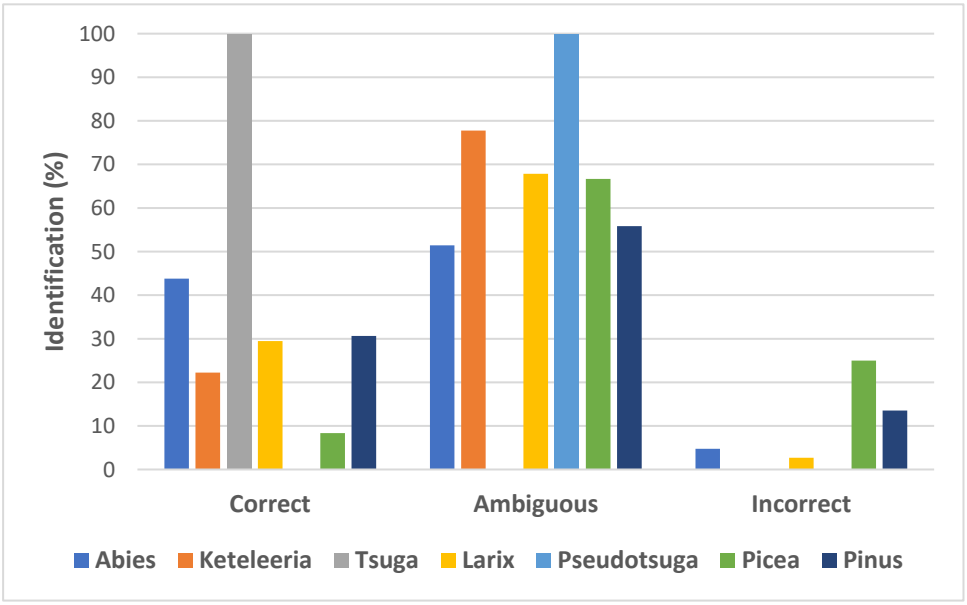


Figure 2. Species discrimination rates based on BLAST1 method.

TAXONDNA/Species Identifier 1.8 program, using "Best Match" (BM) and "Best Close Match" (BCM) options, was used to calculate the percentage of correct ITS2 sequence

identification at the species complex, genus, subfamily, and family level for taxa represented in the analysis by at least two sequences. The summary of the success of taxa identification for analyzed genera among the Pinaceae family using the TaxonDNA method is shown in Table 3. At the species complex level, represented by the closely related European mountain pine complex, *Pinus mugo* complex (PMC), 100% of the analyzed ITS2 sequences were ambiguously identified using the BM and BCM options. None of the ITS2 sequence in this species complex were identified as correct (0%) or incorrect (0%). At the genera level, *Pseudotsuga* and *Tsuga* were the most successfully discriminated (100%), while *Keteleeria* had the lowest discriminatory success. At the subfamily level, the highest success of identification was noticed for Laricoideae (23.48%), and the lowest success for the subfamily Abietoideae (17.80%). At the family level, the success rate of species discrimination among the Pinaceae was relatively low. Only 71 of 368 sequences were correctly identified to species level (19.2%) based on best match (BM) analysis. It is worth emphasizing that the ambiguous identification was over three times higher than the correct identification (74.1%). Incorrect identification concerned 7.5% of the sequences. Similar results were obtained based on the best close match (BCM) analysis.

Table 3. Percentage of correct, incorrect and ambiguous sequence identifications based on the 'best match' and 'best close match' with TaxonDNA software.

Taxonomic Group/Level	Name	Best Match (BM)			Best Close Match (BCM)			
		Correct	Incorrect	Ambiguous	Correct	Incorrect	Ambiguous	No Match
Species complex	*P. mugo complex*	0.00	0.00	100.00	0.00	0.00	100.00	0.00
Genera	*Abies*	16.19	8.57	75.24	16.19	8.57	75.24	0.00
	Keteleeria	11.11	0.00	88.89	11.10	0.00	88.89	0.00
	Tsuga	100.00	0.00	0.00	100.00	0.00	0.00	0.00
	Larix	21.43	3.57	75.00	21.43	2.68	75.00	0.89
	Pseudotsuga	100.00	0.00	0.00	100.00	0.00	0.00	0.00
	Picea	20.83	0.00	79.17	20.83	0.00	79.17	0.00
	Pinus	16.22	9.01	74.77	15.32	7.21	72.07	5.40
Subfamily	Abietoideae	17.80	8.47	73.73	17.80	7.63	73.73	0.84
	Laricoideae	23.48	3.48	73.04	23.40	2.61	75.00	0.87
	Pinoideae	17.04	7.40	75.50	16.30	5.93	73.33	4.44
Family	Pinaceae	19.20	6.52	74.10	19.00	5.43	73.38	2.17

2.4. Phylogenetic Inference

Phylogenetic inference was performed in the RaxML v8.2.11 [33] using 371 ITS2 sequences and the maximum likelihood (ML) method. The resulting phylogenetic tree sorted all analyzed Pinaceae taxa into the appropriate genera according to the commonly accepted taxonomy, most with high bootstrap support. Although the ITS2 sequences allowed the assignment of taxa to the different Pinaceae genera, their distinction within these genera was not so obvious. These observations are fully consistent with the results obtained by us with BLAST1 and Taxon DNA in terms of the correct sequence identification and discrimination of Pinaceae taxa, where the species level also turned out to be the most critical.

The overall topology of the obtained phylogenetic ML tree based on ITS2 sequences is consistent with the commonly accepted division of Pinaceae and strongly supports the concept of monophyly of particular Pinaceae genera (Figure S1).

3. Discussion

The main aim of the article was to assess the suitability of the ITS2 region for the discrimination and identification of taxa in Pinaceae and phylogenetic inference in this large family of conifers. Detailed analysis of the genetic variation of the ITS2 sequence was performed for the first time in one comprehensive study. Particular attention has been paid

to precisely characterize the level of genetic variation in the ITS2 sequence at four different taxonomic levels and to determine the success of sequence and taxa discrimination using both phylogenetically distant and closely related conifers.

In contrast to some reports in the literature [26,27] our observations clearly showed that although ITS2 is easy to analyze, it has severe limitations (mainly low level of genetic variability) and is not the best choice for identification, authentication, or phylogenetic inference in Pinaceae. The finding was quite unexpected as there are quite a few ITS2 sequences in the databases. Moreover, many studies show that ITS2 is a very good DNA barcode region [22,23,25,34,35]. According to China Plant BOL Group [36], the nuclear ribosomal DNA region, ITS, is characterized by a much better ability to discriminate species compared to plastid regions. Internal transcribed spacer (ITS1, 5.8S, ITS2) is characterized by a higher rate of evolution than the coding regions, is inherited biparentally, and shows a high level of divergence at the species level, which allows it to be used to identify even closely related species [37]. This is one of the reasons why the assessment of the ability of ITS2 to discriminate between closely related taxa in the Pinaceae family was one of the goals of our work.

In seed plants, the length of ITS varies usually from 500 to 700 bp [38], while in gymnosperms it is much longer, especially in Pinaceae (1500–3700 bp) [37,39], except that half the length of ITS1 consists of subrepeats [40–43], which results in some difficulty in PCR amplification and typical Sanger sequencing reads. Hence, the availability of complete ITS sequences in genetic databases for different species is still very limited. Therefore, China Plant BOL Group has suggested using only the ITS2 sequence as an alternative to the complete ITS region, due to the corresponding variability of the primary and secondary structure [23,27,44].

There are many reports in the literature on the assessment of the identification potential of various genomic regions postulated for distinguishing and analyzing different groups of organisms that would be ideal DNA barcodes. For animals, the cytochrome c oxidase 1 subunit 1 has been proposed as the main bar code [16]. In the case of plant identification, this region will not work well due to the insufficient rate of nucleotide substitution in the plant's mitochondrial genome [45]. To solve this problem, several promising highly variable plastid DNA regions have been proposed, including both coding and non-coding loci, which can be used singly or together in various combinations. In this way, among others, *rbcL* [45–47], *matK* [48–52], combination *rbcL* + *matK* [53], intergenic spacer region-*trnH-psbA* [46,49,50,54–57], *rpoB* [45,49,51], *rpoC1* [49,51], *atpF-atpH* [45], *ndhJ*, *ycf*5, or *accD* [51], were selected as valuable. Recently, the *ycf*1 region was indicated as very promising due to its very high variability, and even recommended as the main barcode in terrestrial plants [58]. Although the usefulness of *ycf*1 as a species-diagnostic marker has been demonstrated in the case of the Pinaceae family [59], it should be used in the analyses with caution as it was confirmed that this region is under positive selection in all *Pinus* plastomes [60], as well as frequent hybridizations within the genus *Pinus*, and finally the complex model of inheritance of the chloroplast genome [61]. Some difficulty in the widespread use of the complete *ycf*1 region may also arise from its considerable length, which makes it difficult to apply conventional Sanger sequencing with readings of 600–800 bp.

In this respect, nuclear markers, especially short ITS2, seem to be a better solution than chloroplast regions, especially given that the ITS2 region has so far been used quite successfully for the discrimination and identification of taxa in many angiosperm families. ITS2 turned out to be the best marker differentiating species from the Araliaceae family [24]. Several universal and popular DNA barcode regions (*matK*, *rbcL*, ITS2, *psbA-trnH*, and *ycf*5) were assessed by Liu [24] for their ability to identify 1113 sequences derived from 276 species from 23 genera. ITS2 correctly identified 85.23% and 97.29% of the sequences at the species and genus levels, respectively. Additionally, it was suggested that the *psbA-trnH* region could be an additional candidate DNA barcode for the identification of the Araliaceae family. Similarly, the high efficiency of the ITS2 region in discriminating taxa

(>90% and 100% at the species and genus levels, respectively) was demonstrated in the Euphorbiaceae family based on the analysis of 1183 samples representing 871 species and 66 genera [34]. Similar results were obtained in the study of the Rutaceae family [62], where ITS2 was shown to be superior to the other six barcodes (*psbA-trnH*, *matK*, *ycf*5, *rpo*C1, *rbcL*, ITS). It was characterized by the highest interspecific divergence with regard to intraspecific divergence. Moreover, it has also been proven to be effective in distinguishing between closely related species. However, in our research, the effectiveness of the ITS2 sequence is zero at the *P. mugo* complex level. Our research clearly shows that the higher the taxonomic level, the higher the percentage of success in discriminating Pinaceae taxa. However, it seems that this percentage is decreasing not only with decreasing phylogenetic distance of analyzed samples, but also with increasing number of available sequences and individuals in the database.

ITS2 was also used with greater or lesser success in the analysis of the Apocynaceae family [22] or Asteraceae [35]. In the latter family, ITS2 was considered as a suitable, but not ideal, barcode for identifying species of high medical importance, belonging to the largest family of flowering plants. Compared to other barcodes, ITS2 was characterized by the greatest universality, specific divergence, and discrimination, which makes it a promising marker for Asteraceae authentication. In the Fabaceae family, ITS2 has been shown to be effective in identifying medicinal plants. It is worth noting that ITS2 also turned out to be an appropriate phylogenetic marker, but it did not solve all taxonomic problems [23]. This suggests that effective DNA barcoding using ITS2 varies from family to family and can be unreliable in complex taxonomic groups. In the Brassicaceae and Roasaceae families, ITS2 proved to be an imperfect barcode due to its low resolution [63,64].

As generally accepted, the ideal DNA barcode region should have a higher interspecies than intraspecies variation. In the case of the current study based on the pairwise genetic distance-based method (PWG-distance) for 368 ITS2 sequences from the Pinaceae family, the estimated inter-specific divergence parameter is higher than the intra-specific distances for all analyzed groups except for the *Pinus* genus. The DNA sequence similarity-based method, on the other hand, showed that in the case of the Pinaceae family, the success of taxa discrimination using ITS2 was relatively low. This shows that internal transcribed spacer 2 has some problems in identifying both subspecies and varieties, so closely related species are not properly distinguished.

Our results are fully consistent with those obtained for other gymnosperm families, including Podocarpaceae, the second largest family among gymnosperms, in which ITS2 was not characterized by too high an index of taxa discrimination [20]. Moreover, Yao's [26] studies on several families of gymnosperms showed that ITS 2 had the least discriminatory success at the species level in comparison to other plant groups (mosses, ferns, monocotyledons, dicotyledons) as well as animals.

4. Materials and Methods

4.1. Sampling and Plant Materials

In our study, a total of 371 ITS2 sequences were analyzed, of which 368 sequences belonged to 71 Pinaceae taxa of 7 genera, and three *Podocarpus* sp. Sequences were outgroups only for phylogenetic analyzes (Tables S1 and S2). Of the 368 ITS2 Pinaceae sequences, 346 samples were downloaded from GenBank, and 22 sequences were obtained in this study by analyzing 20 individuals belonging to the *Pinus mugo* complex and 2 individuals of Scots pine (*Pinus sylvestris*) as the closest related taxa to the *Pinus mugo* complex. The analyzed specimens were collected in Poland, the Czech Republic and Germany (Table S3). The suitability of the ITS2 region to discriminate against taxa of the Pinaceae family was assessed by analyzing both phylogenetically distant and closely related taxa.

4.2. DNA Extraction and Next-Generation Sequencing

The 100 mg of fresh plant tissue was used to extract genomic DNA using the Genomic Mini AX Plant Spin kit (A&A Biotechnology, Gdańsk, Poland). The quality and

concentration of the extracted DNA were verified using gel electrophoresis and a NanoDrop spectrophotometer (Thermo Fisher Scientific, Carlsbad, CA, United States). The genomic libraries were constructed with protocol: Ion Xpress™ Plus gDNA Fragment Library Preparation, using Ion Xpress Plus Fragment Library Kit (Pub. No. MAN0009847) (ThermoFisher Scientific, Waltham, MA, USA). Next-generation sequencing was performed on the GeneStudio™ S5 System (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's instructions using the protocol: Ion 540 ™ Kit—OT2 User Guide (Cat. No. A27753, Publication No. MAN0010850, Revision D). The reads generated from next-generation sequencing were assembled and annotated using the Geneious Prime 2020.2.5 package. Reads were mapped to MT735327 sequence from genbank using Genious mapper with default settings and minimum mapping quality 30%. Sequences that were mapped were subsequently assembled de novo using Geneious algorithm on default settings and annotated based on MT735327 sequence. The obtained 22 complete ITS2 sequences were submitted to GenBank. Each sequence was assigned an accession number (Table S2).

4.3. Data Validation

The sequences of internal transcribed spacer 2 from GenBank were downloaded using query "internal transcribed spacer 2 Pinaceae" (in July 2021). The nuclear ribosomal ITS2 sequences with ambiguous bases were discarded from further analyzes. In order to extract ITS2 sequence, the ITS2 database (Internal Transcribed Spacer 2 Ribosomal DNA Database) (version 3.0.13) available online (http://its2.bioapps.biozentrum.uni-wuerzburg.de (accessed on 20 August 2021) was used [65–67]. Single sequence species were excluded from the analysis.

4.4. Data Analysis

In order to carry out a deeper and more precise genetic characterization, i four taxonomic levels were considered, namely species complex, genus, subfamily, and family. ITS2 sequences were aligned using the ClustalW with default parameters available in MEGA version X [68]. Then, the length of the alignment was estimated and the percentage of conserved sites (CS), variable sites (VS), parsimony-informative sites (PIS), singleton sites (SS), and overall mean distance (OMD) were calculated. To assess the suitability of the ITS2 sequence as a potential barcode at the level of genus, subfamily, and family of Pinaceae, selected methods were used, namely the pairwise genetic distance-based method (PWG-distance), the DNA sequence similarity-based method (TaxonDNA, BLAST), and phylogenetic tree method (maximum likelihood).

The pairwise genetic distance-based method was used to determine the genetic divergence of the obtained ITS2 sequences. The five parameters were calculated in MEGA version X [68] using the Kimura two-parameter distance model (K2P) [69] with pairwise deletion option to define the interspecific and intraspecific variability. The interspecific divergence has been characterized by all interspecific distance and minimum interspecific distance parameters [27,70,71]. Intraspecific variability was calculated using the K2P distance matrix by applying three parameters: all intraspecific distance, theta(θ) and coalescent depth [50,70]. Using the "Pairwise summary" option based on the K2P model available in the TaxonDNA/SpeciesIdentifier 1.8 software, the frequency of the distribution of interspecific distance and intraspecific variability was obtained. Plots were made to illustrate the barcode gap for each genus and for the entire Pinaceae family.

Sequence identification and taxa discrimination rates among Pinaceae were calculated using the two different methods. The first was the method based on DNA sequence similarity-based method implemented in TAXONDNA/Species Identifier 1.8 program. "Best Match" (BM) and "Best Close Match" (BCM) options were used to verify the percentage of correct ITS2 sequence identification at the species complex, genus, subfamily and family level for taxa represented in at least two sequences [72]. The second was BLAST1. In this method, performed with the BLAST program (http://blast.ncbi.nlm.nih.gov/Blast.cgi, accessed on 15 February 2022), all ITS2 Pinaceae sequences were used as query sequences to

search the reference database. Correct identification means that the best BLAST hits of the query sequence are from the expected species, while ambiguous identification means that the best BLAST hits for the query sequence turned out to be those of several species, including the expected species. Incorrect identification in turn means that the query sequence's best BLAST hit is not from the expected species.

Phylogenetic inference was constructed by maximum likelihood (ML) analysis using 371 ITS2 Pinaceae sequences in RaxML v8.2.11 [33], with 1000 rapid bootstrap replicates along with a search for the best-scoring ML tree in every run and parsimony random seed set to 10. *Podocarpus longefoliolatus* (AY083065), *Podocarpus macrophyllus* voucher HZ20070103 (EF660588), and *Podocarpus neriifolius* isolate VNMN000814 (KR674120) were used as an outgroup.

5. Conclusions

Our study provides a critical assessment of nuclear DNA ITS2 region relevance for taxa discrimination among Pinaceae. Based on the results obtained, it can be concluded that although ITS2 fulfills some of the important features of an ideal DNA barcode region, its usefulness for distinguishing taxa among Pinaceae is severely limited. It seems that the correct and successful identification of taxa is reserved only for those that are phylogenetically distant and represent different genera rather than species, and for those for which few sequences are available in genetic databases. The closely related conifers at the species complex level are indistinguishable using ITS2. The application of this region for study relationships in the Pinaceae family does not seem justified due to its low genetic variability, which results low phylogenetic resolution. Nevertheless, further research on ITS2 is needed, especially in terms of extending the available genetic databases with new records, especially for species from the genera *Tsuga* or *Pseudotsuga*. A wider database would determine whether the high success in distinguishing taxa in these two genera is due to the low number of deposited samples in these databases or to the high efficiency of ITS2. Another interesting direction for further research could also be to determine the effectiveness of complete internal transcribed spacer 1 (ITS1) in discriminating Pinaceae taxa and studying the phylogeny of this large and important family of conifers.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/plants11081078/s1, Figure S1: Evolutionary relationships analyzed 371 ITS2 sequences, Table S1: Summary of the Pinaceae samples analyzed in this study. Table S2: List of taxa and records with accession numbers analyzed in this study, Table S3: List of individuals representing *Pinus mugo* complex taxa sequenced in this study, Table S4: Abundance of all intraspecific and all interspecific K2P pairwise distance in Pinaceae, Table S5: The results of species discrimination in seven genera of the Pinaceae family based on the BLAST1 method.

Author Contributions: J.S. and K.C. conceived of and designed the research framework; J.S. performed the experiments; J.S. and K.C. wrote the original draft manuscript, reviewed and edited the final manuscript; J.S. and H.F. performed data analysis; K.C. participated in the data analysis; J.S. and K.C. collected the samples; K.C. supervised the project. All authors have read and agreed to the published version of the manuscript.

Funding: Research paper financed from the budget for science in 2018–2020, as a research project under the "Diamond Grant" program No. DI2017003147.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data are available within the article and Supplementary Materials.

Acknowledgments: The authors would like to thank Pascal von Sengbusch—Büro für ökologische Gutachten/Moorkundeand Ladislav Rektoris—Agentura ochrany přírody a krajiny České Republiky for their help in collecting the plant material. We also thank four anonymous reviewers for reading our manuscript carefully and for their many insightful comments and suggestions.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Liston, A.; Vining, T.F.; Campbell, C.S.; Gernandt, D.S.; Piñero, D. Molecular phylogeny of pinaceae and pinus. *Acta Hortic.* **2003**, *615*, 107–114. [CrossRef]
2. Farjon, A. *A Natural History of Conifers*; Timber Press: Portland, OR, USA, 2008.
3. Le Maitre, D.C. Pines in cultivation: A global view. In *Ecology and Biogeography of Pinus*; David, M.R., Ed.; Cambridge University Press: Cambridge, UK, 1998; pp. 407–431.
4. Leduc, C.; Coonishish, J.; Haddad, P.; Cuerrier, A. Plants used by the cree nation of eeyou istchee (Quebec, Canada) for the treatment of diabetes: A novel approach in quantitative ethnobotany. *J. Ethnopharmacol.* **2006**, *105*, 55–63. [CrossRef] [PubMed]
5. Yasin, M.; Hussain Janbaz, K.; Imran, I.; Gilani, A.U.H.; Bashir, S. Pharmacological Studies on the antispasmodic, bronchodilator and anti-platelet activities of abies webbiana. *Phyther. Res.* **2014**, *28*, 1182–1187. [CrossRef]
6. Kizilarslan, Ç.; Sevg, E. Ethnobotanical uses of *genus pinus*, L. (Pinaceae) in Turkey. *Indian J. Tradit. Knowl.* **2013**, *12*, 209–220.
7. Hussain, K.; Nisar, M.F.; Majeed, A.; Nawaz, K.; Bhatti, K.H. Ethnomedicinal survey for important plants of Jalalpur Jattan, District Gujrat, Punjab, Pakistan. *Ethnobot. Leafl.* **2010**, *14*, 807–825.
8. Kaushik, P.; Lal, S.; Rana, A.C.; Kaushik, D. GC-MS analysis of bioactive constituents of pinus roxburghii sarg. (Pinaceae) from Northern India. *Res. J. Phytochem.* **2014**, *8*, 42–46. [CrossRef]
9. Celiński, K.; Kijak, H.; Wojnicka-Półtorak, A.; Buczkowska-Chmielewska, K.; Sokołowska, J.; Chudzińska, E. Effectiveness of the DNA barcoding approach for closely related conifers discrimination: A case study of the *pinus mugo* complex. *Comptes Rendus-Biol.* **2017**, *340*, 339–348. [CrossRef]
10. Celiński, K.; Bonikowski, R.; Wojnicka-Półtorak, A.; Chudzińska, E.; Maliński, T. Volatiles as chemosystematic markers for distinguishing closely related species within the *pinus mugo* complex. *Chem. Biodivers.* **2015**, *12*, 1208–1213. [CrossRef]
11. Bonikowski, R.; Celiński, K.; Wojnicka-Półtorak, A.; Maliński, T. Composition of essential oils isolated from the needles of Pinus Uncinata and P. Uliginosa Grown in Poland. *Nat. Prod. Commun.* **2015**, *10*, 371–373. [CrossRef]
12. Celiński, K.; Kijak, H.; Barylski, J.; Grabsztunowicz, M.; Wojnicka-Półtorak, A.; Chudzińska, E. Characterization of the complete chloroplast genome of *pinus uliginosa* (neumann) from the *pinus mugo* complex. *Conserv. Genet. Resour.* **2017**, *9*, 209–212. [CrossRef]
13. Celiński, K.; Chudzińska, E.; Gmur, A.; Piosik, Ł.; Wojnicka-Półtorak, A. Cytological characterization of three closely related pines—*Pinus Mugo*, *P. Uliginosa* and P. × Rhaetica from the Pinus Mugo complex (Pinaceae). *Biologia* **2019**, *74*, 751–756. [CrossRef]
14. Wachowiak, W.; Celiński, K.; Prus-Głowacki, W. Evidence of natural reciprocal hybridisation between Pinus Uliginosa and P. Sylvestris in the sympatric population of the species. *Flora Morphol. Distrib. Funct. Ecol. Plants* **2005**, *200*, 563–568. [CrossRef]
15. Sokołowska, J.; Fuchs, H.; Celiński, K. New insight into taxonomy of european mountain pines, *pinus mugo* complex, based on complete chloroplast genomes sequencing. *Plants* **2021**, *10*, 1331. [CrossRef] [PubMed]
16. Hebert, P.D.N.; Cywinska, A.; Ball, S.L.; DeWaard, J.R. Biological Identifications through DNA Barcodes. *Proc. R. Soc. B Biol. Sci.* **2003**, *270*, 313–321. [CrossRef]
17. Hollingsworth, P.M.; Graham, S.W.; Little, D.P. Choosing and using a plant DNA barcode. *PLoS ONE* **2011**, *6*, e19254. [CrossRef]
18. Wang, X.C.; Liu, C.; Huang, L.; Bengtsson-Palme, J.; Chen, H.; Zhang, J.H.; Cai, D.; Li, J.Q. ITS1: A DNA barcode better than ITS2 in eukaryotes? *Mol. Ecol. Resour.* **2015**, *15*, 573–586. [CrossRef]
19. Ran, J.H.; Wang, P.P.; Zhao, H.J.; Wang, X.Q. A test of seven candidate barcode regions from the Plastome in Picea (Pinaceae). *J. Integr. Plant Biol.* **2010**, *52*, 1109–1126. [CrossRef]
20. Little, D.P.; Knopf, P.; Schulz, C. DNA barcode identification of Podocarpaceae—The second largest conifer family. *PLoS ONE* **2013**, *8*, e81008. [CrossRef]
21. Maia, V.H.; da Mata, C.S.; Franco, L.O.; Cardoso, M.A.; Cardoso, S.R.S.; Hemerly, A.S.; Ferreira, P.C.G. DNA barcoding bromeliaceae: Achievements and pitfalls. *PLoS ONE* **2012**, *7*, e29877. [CrossRef]
22. LV, Y.N.; YANG, C.Y.; SHI, L.C.; ZHANG, Z.L.; XU, A.S.; ZHANG, L.X.; LI, X.L.; LI, H.T. Identification of Medicinal plants within the apocynaceae family using ITS2 and PsbA-TrnH barcodes. *Chin. J. Nat. Med.* **2020**, *18*, 594–605. [CrossRef]
23. Gao, T.; Yao, H.; Song, J.; Liu, C.; Zhu, Y.; Ma, X.; Pang, X.; Xu, H.; Chen, S. Identification of Medicinal plants in the family fabaceae using a potential DNA barcode ITS2. *J. Ethnopharmacol.* **2010**, *130*, 116–121. [CrossRef] [PubMed]
24. Liu, Z.; Zeng, X.; Yang, D.; Chu, G.; Yuan, Z.; Chen, S. Applying DNA barcodes for identification of plant species in the family araliaceae. *Gene* **2012**, *499*, 76–80. [CrossRef] [PubMed]
25. Feng, S.; Jiang, M.; Shi, Y.; Jiao, K.; Shen, C.; Lu, J.; Ying, Q.; Wang, H. Application of the ribosomal DNA ITS2 region of physalis (Solanaceae): DNA barcoding and phylogenetic study. *Front. Plant Sci.* **2016**, *7*, 1047. [CrossRef] [PubMed]
26. Yao, H.; Song, J.; Liu, C.; Luo, K.; Han, J.; Li, Y.; Pang, X.; Xu, H.; Zhu, Y.; Xiao, P.; et al. Use of ITS2 region as the universal DNA barcode for plants and animals. *PLoS ONE* **2010**, *5*, e13102. [CrossRef] [PubMed]
27. Chen, S.; Yao, H.; Han, J.; Liu, C.; Song, J.; Shi, L.; Zhu, Y.; Ma, X.; Gao, T.; Pang, X.; et al. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS ONE* **2010**, *5*, e8613. [CrossRef]
28. Coleman, A.W. ITS2 is a double-edged tool for eukaryote evolutionary comparisons. *Trends Genet.* **2003**, *19*, 370–375. [CrossRef]

29. Schultz, J.; Maisel, S.; Gerlach, D.; Müller, T.; Wolf, M. A common core of secondary structure of the internal transcribed spacer 2 (ITS2) throughout the Eukaryota. *RNA* **2005**, *11*, 361–364. [CrossRef]
30. Miao, M.; Warren, A.; Song, W.; Wang, S.; Shang, H.; Chen, Z. Analysis of the internal transcribed spacer 2 (ITS2) region of scuticociliates and related taxa (ciliophora, oligohymenophorea) to infer their evolution and phylogeny. *Protist* **2008**, *159*, 519–533. [CrossRef]
31. Coleman, A.W. Pan-eukaryote ITS2 homologies revealed by RNA secondary structure. *Nucleic Acids Res.* **2007**, *35*, 3322–3329. [CrossRef]
32. Schultz, J.; Wolf, M. ITS2 sequence-structure analysis in phylogenetics: A how-to manual for molecular systematics. *Mol. Phylogenetics Evol.* **2009**, *52*, 520–523. [CrossRef]
33. Stamatakis, A. RAxML Version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [CrossRef] [PubMed]
34. Pang, X.; Song, J.; Zhu, Y.; Xie, C.; Chen, S. Using DNA barcoding to identify species within euphorbiaceae. *Planta Med.* **2010**, *76*, 1784–1786. [CrossRef] [PubMed]
35. Gao, T.; Yao, H.; Song, J.; Zhu, Y.; Liu, C.; Chen, S. Evaluating the feasibility of using candidate DNA barcodes in discriminating species of the large Asteraceae family. *BMC Evol. Biol.* **2010**, *10*, 324. [CrossRef] [PubMed]
36. China Plant BOL Group; Li, D.Z.; Gao, L.M.; Li, H.T.; Wang, H.; Ge, X.J.; Liu, J.Q.; Chen, Z.D.; Zhou, S.L.; Chen, S.L.; et al. Comparative analysis of a large dataset indicates that Internal Transcribed Spacer (ITS) Should be incorporated into the core barcode for seed plants. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 19641–19646. [CrossRef] [PubMed]
37. Álvarez, I.; Wendel, J.F. Ribosomal ITS sequences and plant phylogenetic inference. *Mol. Phylogenet. Evol.* **2003**, *29*, 417–434. [CrossRef]
38. Baldwin, B.G.; Sanderson, M.J.; Porter, J.M.; Wojciechowski, M.F.; Campbell, C.S.; Donoghue, M.J. The ITS Region of nuclear ribosomal DNA: A valuable source of evidence on angiosperm phylogeny. *Ann. Missouri Bot. Gard.* **1995**, *85*, 247–277. [CrossRef]
39. Liston, A.; Robinson, W.A.; Oliphant, J.M.; Alvarez-Buylla, E.R. Length variation in the nuclear ribosomal DNA internal transcribed spacer region of non-flowering seed plants. *Syst. Bot.* **1996**, *21*, 109–120. [CrossRef]
40. Marrocco, R.; Gelati, M.T.; Maggini, F. Nucleotide sequence of the internal transcribed spacers and 5.8 s region of ribosomal DNA in *Phus pinea* L. *Mitochondrial DNA* **1996**, *6*, 175–177. [CrossRef]
41. Vining, T.F.; Campbell, C.S. Phylogenetic signal in sequence repeats within nuclear ribosomal DNA internal transcribed spacer 1 in Tsuga. *Am. J. Bot.* **1997**, *84*, 702.
42. Gernandt, D.S.; Liston, A. Internal transcribed spacer region evolution in Larix and Pseudotsuga (Pinaceae). *Am. J. Bot.* **1998**, *86*, 711–723. [CrossRef]
43. Maggini, F.; Frediani, M.; Gelati, M.T. Nucleotide sequence of the internal transcribed spacers of ribosomal DNA in Picea Abies Karst. *Mitochondrial DNA* **2000**, *11*, 87–89. [CrossRef]
44. Han, J.; Zhu, Y.; Chen, X.; Liao, B.; Yao, H.; Song, J.; Chen, S.; Meng, F. The short ITS2 sequence serves as an efficient taxonomic sequence tag in comparison with the full-length ITS. *Biomed Res. Int.* **2013**, *2013*, 3–10. [CrossRef] [PubMed]
45. Fazekas, A.J.; Burgess, K.S.; Kesanakurti, P.R.; Graham, S.W.; Newmaster, S.G.; Husband, B.C.; Percy, D.M.; Hajibabaei, M.; Barrett, S.C.H. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLoS ONE* **2008**, *3*, e2802. [CrossRef] [PubMed]
46. Kress, W.J.; Erickson, D.L. A Two-locus global DNA barcode for land plants: The coding *rbcL* gene complements the non-coding *trnH-PsbA* spacer region. *PLoS ONE* **2007**, *2*, e508. [CrossRef] [PubMed]
47. Newmaster, S.G.; Fazekas, A.J.; Ragupathy, S. DNA barcoding in land plants: Evaluation of RbcL in a multigene tiered approach. *Can. J. Bot.* **2006**, *84*, 335–341. [CrossRef]
48. Pennisi, E. Taxonomy. Wanted: A barcode for plants. *Science* **2007**, *318*, 190–191. [CrossRef]
49. Chase, M.W.; Cowan, R.S.; Hollingsworth, P.M.; Van Den Berg, C.; Madriñán, S.; Petersen, G.; Seberg, O.; Jørgsensen, T.; Cameron, K.M.; Carine, M.; et al. A Proposal for a standardised protocol to barcode all land plants. *Taxon* **2007**, *56*, 295–299. [CrossRef]
50. Lahaye, R.; Van Der Bank, M.; Bogarin, D.; Warner, J.; Pupulin, F.; Gigot, G.; Maurin, O.; Duthoit, S.; Barraclough, T.G.; Savolainen, V. DNA barcoding the floras of biodiversity hotspots. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 2923–2928. [CrossRef]
51. Ford, C.S.; Ayres, K.L.; Toomey, N.; Haider, N.; Stahl, J.V.A.; Kelly, L.J.; Wikström, N.; Hollingsworth, P.M.; Duff, R.J.; Hoot, S.B.; et al. Selection of candidate coding DNA barcoding regions for use on land plants. *Bot. J. Linn. Soc.* **2009**, *159*, 1–11. [CrossRef]
52. Starr, J.R.; Naczi, R.F.C.; Chouinard, B.N. Plant DNA barcodes and species resolution in sedges (Carex, Cyperaceae). *Mol. Ecol. Resour.* **2009**, *9* (Suppl. S1), 151–163. [CrossRef]
53. CBOL Plant Working Group. A DNA barcode for land plants. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 12794–12797. [CrossRef] [PubMed]
54. Kress, W.J.; Wurdack, K.J.; Zimmer, E.A.; Weigt, L.A.; Janzen, D.H. Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. USA.* **2005**, *102*, 8369–8374. [CrossRef] [PubMed]
55. Newmaster, S.G.; Ragupathy, S. Esting plant barcoding in a sister species complex of pantropical acacia (Mimosoideae, Fabaceae). *Mol. Ecol. Resour.* **2009**, *9*, 172–180.
56. Pang, X.; Liu, C.; Shi, L.; Liu, R.; Liang, D.; Li, H.; Cherny, S.S.; Chen, S. Utility of the TrnH-PsbA intergenic spacer region and its combinations as plant DNA barcodes: A meta-analysis. *PLoS ONE* **2012**, *7*, e48833. [CrossRef]

57. Bolson, M.; De Camargo Smidt, E.; Brotto, M.L.; Silva-Pereira, V. ITS and TrnH-PsbA as efficient DNA barcodes to identify threatened commercial woody angiosperms from Southern Brazilian atlantic rainforests. *PLoS ONE* **2015**, *10*, e0143049. [CrossRef]
58. Dong, W.; Xu, C.; Li, C.; Sun, J.; Zuo, Y.; Shi, S.; Cheng, T.; Guo, J.; Zhou, S. Ycf1, the most promising plastid DNA barcode of land plants. *Sci. Rep.* **2015**, *5*, 8348. [CrossRef]
59. Olsson, S.; Grivet, D.; Cid Vian, J. Species-diagnostic markers in the genus pinus: Evaluation of the chloroplast regions Matk and Ycf1. *For. Syst.* **2018**, *27*, e016. [CrossRef]
60. Zeb, U.; Dong, W.L.; Zhang, T.T.; Wang, R.N.; Shahzad, K.; Ma, X.F.; Li, Z.H. Comparative plastid genomics of pinus species: Insights into sequence variations and phylogenetic relationships. *J. Syst. Evol.* **2020**, *58*, 118–132. [CrossRef]
61. Kormutak, A.; Galgoci, M.; Bolecek, P.; Gomory, D.; Libantova, J. Reinforced evidence on partial compatibility between *pinus sylvestris* and *pinus mugo* and on maternal inheritance of chloroplast DNA in the *Pinus Mugo* × *Pinus Sylvestris* cross. *Silvae Genet.* **2020**, *69*, 108–115. [CrossRef]
62. Luo, K.; Chen, S.L.; Chen, K.L.; Song, J.Y.; Yao, H.; Ma, X.; Zhu, Y.J.; Pang, X.H.; Yu, H.; Li, X.W.; et al. Assessment of candidate plant DNA barcodes using the rutaceae family. *Sci. China Life Sci.* **2010**, *53*, 701–708. [CrossRef]
63. Pang, X.; Song, J.; Zhu, Y.; Xu, H.; Huang, L.; Chen, S. Applying plant DNA barcodes for rosaceae species identification. *Cladistics* **2011**, *27*, 165–170. [CrossRef] [PubMed]
64. Sun, X.Q.; Qu, Y.Q.; Yao, H.; Zhang, Y.M.; Yan, Q.Q.; Hang, Y.Y. Applying DNA barcodes for identification of economically important species in Brassicaceae. *Genet. Mol. Res.* **2015**, *14*, 15050–15061. [CrossRef] [PubMed]
65. Koetschan, C.; Förster, F.; Keller, A.; Schleicher, T.; Ruderisch, B.; Schwarz, R.; Müller, T.; Wolf, M.; Schultz, J. The ITS2 database III sequences and structures for phylogeny. *Nucleic Acids Res.* **2009**, *38* (Suppl. S1), 275–279. [CrossRef] [PubMed]
66. Selig, C.; Wolf, M.; Müller, T.; Dandekar, T.; Schultz, J. The ITS2 database II: Homology modelling RNA Structure for molecular systematics. *Nucleic Acids Res.* **2008**, *36* (Suppl. S1), 377–380. [CrossRef] [PubMed]
67. Schultz, J.; Müller, T.; Achtziger, M.; Seibel, P.N.; Dandekar, T.; Wolf, M. The internal transcribed spacer 2 database—A web server for (not only) low level phylogenetic analyses. *Nucleic Acids Res.* **2006**, *34*, 704–707. [CrossRef]
68. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [CrossRef]
69. Kimura, M. A Simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **1980**, *16*, 111–120. [CrossRef]
70. Meyer, C.P.; Paulay, G. DNA barcoding: Error rates based on comprehensive sampling. *PLoS Biol.* **2005**, *3*, e422. [CrossRef]
71. Meier, R.; Zhang, G.; Ali, F. The use of mean instead of smallest interspecific distances exaggerates the size of the "barcoding gap" and leads to misidentification. *Syst. Biol.* **2008**, *57*, 809–813. [CrossRef]
72. Meier, R.; Shiyang, K.; Vaidya, G.; Ng, P.K.L. DNA barcoding and taxonomy in Diptera: A tale of high intraspecific variability and low identification success. *Syst. Biol.* **2006**, *55*, 715–728. [CrossRef]

Figure S1. Phylogenetic inference constructed from 371 ITS2 sequences from the Pinaceae family using Maximum Likelihood (ML). The numbers in the nodes in the tree indicate the bootstrap values (> 50%).

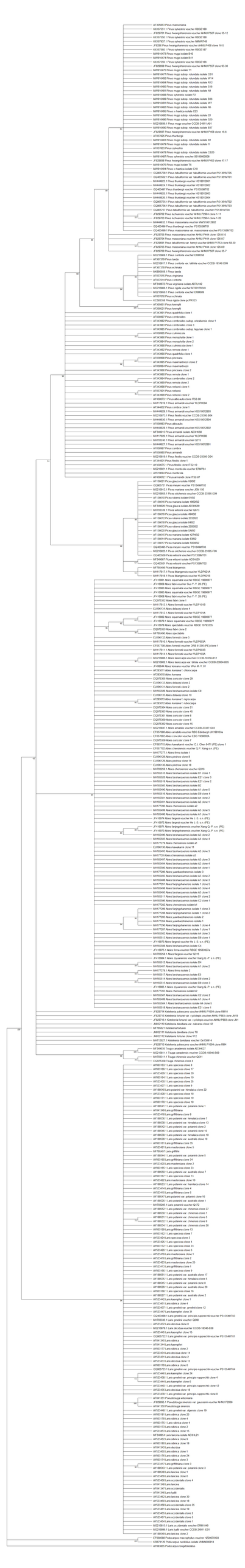

AF393900 Pinus massoniana
KX167551 1 Pinus sylvestris voucher RBSC169
JF828751 Pinus hwangshanensis voucher AHNU P557 clone 35-12
KX167550 1 Pinus sylvestris voucher RBSC168
KX167547 1 Pinus sylvestris voucher MANVI748
JF8296 Pinus hwangshanensis voucher AHNU P458 clone 16-5
KX167549 1 Pinus sylvestris voucher RBSC167
MW816473 Pinus mugo isolate B40
MW816474 Pinus mugo isolate B41
KX167558 1 Pinus sylvestris voucher RBSC180
JF828699 Pinus hwangshanensis voucher AHNU P357 clone 83-30
MW816475 Pinus mugo isolate T1
MW816477 Pinus mugo subsp. rotundata isolate C81
MW816482 Pinus mugo subsp. rotundata isolate W14
MW816484 Pinus mugo subsp. rotundata isolate W13
MW816485 Pinus mugo subsp. rotundata isolate S18
MW816481 Pinus mugo subsp. rotundata isolate M4
MW816468 Pinus sylvestris isolate P2
MW816469 Pinus mugo subsp. rotundata isolate B36
MW816491 Pinus mugo subsp. rotundata isolate W7
MW816492 Pinus mugo subsp. rotundata isolate W5
MW816495 Pinus × rhaetica isolate C23
MW816496 Pinus mugo subsp. rotundata isolate G1
MW816498 Pinus mugo subsp. rotundata isolate S20
MW816500 X Pinus mugo voucher CCDB-29817-A01
MW816490 Pinus mugo subsp. rotundata isolate B37
JF828662 Pinus hwangshanensis voucher AHNU P464 clone 16-6
KF233253 Pinus thunbergii
MW816483 Pinus mugo subsp. rotundata isolate R1
MW816476 Pinus mugo subsp. rotundata isolate I1
KF037003 Pinus sylvestris
MW816478 Pinus mugo subsp. rotundata isolate C820
MW816487 Pinus sylvestris voucher SR19000008
JF828752 Pinus hwangshanensis voucher AHNU P453 clone 47-17
MW816476 Pinus mugo isolate T6
MW816494 Pinus × rhaetica isolate C16
OQ805728.1 Pinus taiwanensis var. takasihumo voucher PS13016735
OQ805502.1 Pinus taiwanensis var. takasihumo voucher PS13016731
MH444823.1 Pinus thunbergii voucher HS1801260
MH444824.1 Pinus thunbergii voucher HS1801262
OQ405467 Pinus thunbergii voucher PS13536702
MH444825.1 Pinus thunbergii voucher HS1801263
MH444826.1 Pinus thunbergii voucher HS1801264
OQ805725.1 Pinus taiwanensis var. takasihumo voucher PS13016732
OQ805726.1 Pinus taiwanensis var. takasihumo voucher PS13016733
OQ805727 Pinus taiwanensis var. takasihumo voucher PS13016734
JF828752 Pinus luchuensis voucher AHNU P564 clone 1-11
JF828703 Pinus luchuensis voucher AHNU P564 clone 1-3
MH444822.1 Pinus luchuensis voucher MWS1801260
OQ405466 Pinus thunbergii voucher PS13536701
OQ405469.1 Pinus massoniana var. massoniana voucher PS13536732
JF828704 Pinus massoniana voucher AHNU P444 clone 126-10
JF828754 Pinus massoniana voucher AHNU P444 clone 126-47
JF828753 Pinus massoniana voucher AHNU P170 clone 58-50
JF828705 Pinus massoniana voucher AHNU P444 clone 126-9
JF828706 Pinus hwangshanensis voucher AHNU P587 clone 55-7
MG210998.1 Pinus contorta voucher ERM558
AF367079 Pinus banksiana
MG210817.1 Pinus contorta var. latifolia voucher CCDB-18346-D08
AF367078 Pinus echinata
MK890658.1 Pinus taeda
AF037010 Pinus virginiana
AF037014 Pinus contorta
MF348837 Pinus virginiana isolate ADTLH42
MG210996.1 Pinus rigida voucher MT897764B
MG210650.1 Pinus contorta voucher ERM609
AF037018 Pinus echinata
KC563358 Pinus rigida clone pa.PR123
AF393601 Pinus krempfii
AF393721 Pinus krempfii
AF343861 Pinus quadrifolia clone 1
AF039267 Pinus cembroides
AF343862 Pinus cembroides subsp. orizabensis clone 1
AF343863 Pinus cembroides clone 3
AF343864 Pinus cembroides subsp. lagunae clone 1
AF039266 Pinus culminicola
AF343867 Pinus monophylla clone 1
AF343868 Pinus monophylla clone 2
AF343869 Pinus culminicola clone 1
AF343860 Pinus remota clone 1
AF343866 Pinus quadrifolia clone 1
AF039268 Pinus pinceana
AF343870 Pinus pinceana clone 2
AF039269 Pinus maximartinezii
AF343896 Pinus pinceana clone 2
AF343895 Pinus remota clone 1
AF343894 Pinus cembroides clone 2
AF343893 Pinus lagunae clone 2
AF343898 Pinus nelsonii clone 1
AF039270 Pinus nelsonii
AF343899 Pinus nelsonii clone 2
KY438573.1 Pinus albicaulis clone IT22-08
MH178193.1 Pinus armandii voucher YLDP008A
AF344852 Pinus cembra clone 1
MH444829.1 Pinus armandii voucher HS1801260I
MG210973.1 Pinus flexilis voucher CCDB-23395-E04
MH444830.1 Pinus armandii voucher HS1801260H
AF039903 Pinus albicaulis
MH444828.1 Pinus armandii voucher HS1801260Z
MF348415 Pinus armandii isolate ACFH408
MH178321.1 Pinus armandii voucher YLDP080
MH178146.1 Pinus armandii voucher QZ15
MH444827.1 Pinus armandii voucher HS1801260I
AF039901 Pinus cembra
AF039902 Pinus armandi
MG210819.1 Pinus flexilis voucher CCDB-23395-D04
AF344851 Pinus flexilis clone 1
KY438573.1 Pinus flexilis clone IT32-10
MG210821.1 Pinus monticola voucher ERM764
KY438574 Pinus monticola
KY438572.1 Pinus armandii clone IT32-07
MG210995.1 Pinus lambertiana voucher 05052
JQ865721 Pinus mayeri voucher PSU-MM2102
MG210912.1 Picea maritae voucher JEM-150
MG210955.1 Picea sitchensis voucher CCDB-23395-E09
AY139010 Picea rubens isolate 61552
AY139014 Picea rubens isolate 49E2552
MT349520 Picea glauca isolate AC3W439
MH703339.1 Picea sitchensis voucher QZ73
AY139019 Picea glauca isolate 49H552
MT349521 Picea rubens isolate 3333552
AY139018 Picea glauca isolate 6465Z
AY139013 Picea rubens isolate 2505552
AY139020 Picea glauca isolate 5AB5Z
AY139015 Picea rubens isolate 4274552
AY139074 Picea mariana isolate 6350Z
AY139077 Picea mariana isolate 5064552
OQ405495 Picea morrisonicola PS13496703
MG210955.1 Picea sitchensis voucher CCDB-23395-F09
OQ405500 Picea wilsonii voucher PS13496701
MF349687 Picea wilsonii isolate AC3N129
OQ405501 Picea wilsonii PS13496702
MF795498 Picea likiangensis
MH178167.1 Picea likiangensis voucher YLDP021A
MH178169.1 Picea likiangensis voucher YLDP023B
AY189681 Abies squamata voucher RBGE 19890977
AY189683 Abies fabri voucher Sun Y. F. 28 (PE)
AY189681 Abies squamata voucher RBGE 19890977
AY189682 Abies squamata voucher RBGE 19890977
AY189684 Abies fabri voucher Sun Y. F. 28 (PE)
AY189681 Abies squamata
MH178151.1 Abies forrestii voucher YLDP1551
EU196124 Abies delavayi clone 4
MH178152.1 Abies forrestii voucher YLDP1554
AY189682 Abies squamata voucher RBGE 19890977
JF418978.1 Abies squamata voucher RBGE 19890977
AY189678 Abies spectabilis voucher RBGE 19793335
DQ915353 Abies fabri clone 2
MF795496 Abies spectabilis
EU196122 Abies forrestii clone 3
MH178153.1 Abies forrestii voucher YLDP0534
EF657768 Abies forrestii voucher Rushan YLDP(2018) clone 1
MH178151.1 Abies forrestii voucher YLDP0538
MH178154.1 Abies forrestii voucher YLDP1534
MG210686.1 Abies beaicarpa voucher CCDB-16858-B12
MG210682.1 Abies beaicarpa var. tibetica voucher CCDB-23964-B05
JF499944 Abies koreana voucher Won M. Y. 01
AF263971 Abies koreana f. chlorocarpa
AF263970 Abies koreana
DQ915360 Abies delavayi clone 19
EU196133 Abies delavayi clone 2
EU196131 Abies forrestii clone 2
MH105505 Abies beshanzuensis isolate C8
EU196135 Abies delavayi clone 10
AF263973 Abies koreana f. nigrocarpa
AF263972 Abies koreana f. rubrocarpa
DQ915364 Abies concolor clone 21
DQ915363 Abies concolor clone 20
DQ915361 Abies delavayi clone 6
DQ915365 Abies concolor clone 15
DQ915362 Abies concolor clone 13
MG210687.1 Abies arnabilis voucher CCDB-23327-G03
EF657666 Abies arnabilis voucher RBG Edinburgh 19780153x
EF657682 Abies concolor voucher EBG 1938002A
DQ915358 Abies concolor clone 7
EF657715 Abies kawakamii voucher C.J. Chen 9471 (PE) clone 1
EF657742 Abies chensiensis voucher Q.F. Xiang s.n. (PE)
MH772377.1 Abies ferreana isolate 1
EU196128 Abies yerdoe clone 8
EU196129 Abies yerdoe clone 14
EU196130 Abies yerdoe clone 18
MH703259.1 Abies chensiensis voucher QZ10
MH105510 Abies beshanzuensis isolate E1 clone 1
MH105520 Abies beshanzuensis isolate E21 clone 3
MH105518 Abies beshanzuensis isolate E21 clone 2
MH105500 Abies beshanzuensis isolate B2
MH105490 Abies beshanzuensis isolate A1 clone 5
MH105516 Abies beshanzuensis isolate D8 clone 4
MH105498 Abies beshanzuensis isolate A4 clone 1
MH105497 Abies beshanzuensis isolate A2 clone 1
MH177290 Abies chensiensis isolate a2
MH105499 Abies beshanzuensis isolate A3 clone 5
MH105494 Abies beshanzuensis isolate A1 clone 1
JF418974 Abies fargesii voucher He J. S. s.n. (PE)
JF418972 Abies fargesii voucher He J. S. s.n. (PE)
JF418971 Abies fanjingshanensis voucher Xiang Q.-P. s.n. (PE)
JF418973 Abies fanjingshanensis voucher Xiang Q.-P. s.n. (PE)
MH105494 Abies beshanzuensis isolate A1 clone 2
MH105503 Abies beshanzuensis isolate A4 clone 4
MH177279 Abies chensiensis isolate a1
EU196126 Abies kawakamii clone T1
MH105492 Abies beshanzuensis isolate A2 clone 3
MH105488 Abies beshanzuensis isolate A1 clone 3
MH177281 Abies fanjingshanensis isolate 1 clone 5
MH105496 Abies beshanzuensis isolate A3 clone 4
MH105495 Abies beshanzuensis isolate A3 clone 3
MH105511 Abies beshanzuensis isolate D1 clone 2
MH105512 Abies beshanzuensis isolate C2 clone 1
MH177282 Abies chensiensis isolate b1
MH177288 Abies fanjingshanensis isolate 1 clone 2
MH177285 Abies fanjingshanensis isolate 2
MH177284 Abies yuanbaoshanensis isolate 1
MH177286 Abies fanjingshanensis isolate 1 clone 4
MH177287 Abies fanjingshanensis isolate 1 clone 1
MH105502 Abies beshanzuensis isolate A4 clone 3
MH105513 Abies beshanzuensis isolate D8 clone 1
JF418975 Abies fargesii voucher He J. S. s.n. (PE)
JF418976 Abies fargesii voucher He J. S. s.n. (PE)
MH105507 Abies beshanzuensis isolate A4 clone 2
JF418977 Abies fargesii var. faxoniana (Rehder & E.H.Wilson) Cheng (PE)
MH105514 Abies beshanzuensis isolate D8 clone 2
MH105504 Abies beshanzuensis isolate A4 clone 5
MH177278 Abies ferreana isolate 2
MH105517 Abies beshanzuensis isolate E5
MH105519 Abies beshanzuensis isolate D8 clone 3
MH105515 Abies beshanzuensis isolate D8 clone 5
JF418981 Abies ziyuanensis voucher Xiang Q.-P. s.n. (PE)
MH177283 Abies chensiensis isolate b2
MH105521 Abies beshanzuensis isolate C2 clone 2
MH105489 Abies beshanzuensis isolate A1 clone 4
MH105501 Abies beshanzuensis isolate A4 clone 6
MH105518 Abies beshanzuensis isolate E21 clone 1
JF828774 Keteleeria pubescens voucher AHNU P1004 clone R810
JF828721 Keteleeria fortunei var. cyclolepis voucher AHNU P803 clone JW10
JF828776 Keteleeria fortunei clone 12 cyclolepis voucher AHNU P803 clone JW1
JN020110 Keteleeria davidiana var. calcarea clone H2
MF795421 Keteleeria fortunei
JN020111 Keteleeria davidiana clone T8
JN020112 Keteleeria davidiana clone T4
MH107922.1 Keteleeria davidiana voucher Cai130914
JF828774 Keteleeria pubescens voucher AHNU P1004 clone R844
MF348505 Tsuga caradensis isolate ACFH421
OQ078910.1 Tsuga canadensis voucher CCDB-18346-B09
MH705311.1 Tsuga chinensis voucher QZ41
AY692103 Larix speciosa
AY692102 Larix speciosa clone 8
AY692108 Larix speciosa clone 17
AY692105 Larix speciosa clone 33
AY692104 Larix speciosa clone 10
AY692106 Larix speciosa clone 25
AY523427 Larix speciosa clone 9
AY189854 Larix potaninii var. himalica clone 22
AY523428 Larix speciosa clone 19
AY692111 Larix speciosa clone 18
AY692115 Larix speciosa clone 18
AY189841 Larix potaninii var. potaninii clone 1
AY141149 Larix griffithiana
AY523416 Larix griffithiana clone 9
AY189836 Larix potaninii var. himalica clone 7
AY189838 Larix potaninii var. himalica clone 13
AY189842 Larix potaninii var. potaninii clone 2
AY189844 Larix potaninii var. potaninii clone 10
AY189839 Larix potaninii var. himalica clone 18
AY189836 Larix potaninii var. himalica clone 18
AY692101 Larix griffithiana clone 35
AY523421 Larix mastersiana clone 5
MF795491 Larix griffithii
AY189843 Larix potaninii var. potaninii clone 5
AY692100 Larix griffithiana clone 34
AY523426 Larix mastersiana clone 2
AY692105 Larix speciosa clone 23
AY189850 Larix potaninii var. australis clone 7
AY692107 Larix speciosa clone 13
AY692109 Larix speciosa clone 16
AY189853 Larix potaninii var. himalica clone 14
AY523414 Larix griffithiana clone 6
AY523415 Larix griffithiana clone 5
AY189847 Larix potaninii var. potaninii clone 15
AY189852 Larix potaninii var. australis clone 5
MH703285.1 Larix potaninii voucher QA72
AY189852 Larix potaninii var. australis clone 27
AY189830 Larix potaninii var. chinensis clone 1
AY189831 Larix potaninii var. chinensis clone 5
AY189832 Larix potaninii var. chinensis clone 9
AY189834 Larix potaninii var. chinensis clone 26
AY692118 Larix griffithiana clone 13
AY692102 Larix speciosa clone 7
AY523424 Larix speciosa clone 3
AY523425 Larix speciosa clone 4
AY692172 Larix speciosa clone 23
AY523408 Larix speciosa clone 6
AY523419 Larix mastersiana clone 1
AY523418 Larix griffithiana clone 2
AY523423 Larix mastersiana clone 25
AY523417 Larix griffithiana clone 1
AY692150 Larix speciosa clone 9
AY189851 Larix potaninii var. australis clone 17
AY189835 Larix potaninii var. himalica clone 5
AY189844 Larix potaninii var. potaninii clone 8
AY189829 Larix potaninii var. australis clone 20
AY692169 Larix speciosa clone 16
AY189827 Larix potaninii var. australis clone 2
AY692171 Larix speciosa clone 17
AY523422 Larix sibirica clone 1
AY523427 Larix griffithiana var. griffithii clone 12
AY523420 Larix griffithiana clone 3
GQ463498.1 Larix griffithii var. principis-rupprechtii voucher PS13154U703
MH703286.1 Larix potaninii var. australis voucher QA98
AY523432 Larix decidua clone 8
MG210978.1 Larix decidua voucher CCDB-16346-G08
AY523445 Larix gmelinii var. principis-rupprechtii clone 1
GQ805727.1 Larix gmelinii var. principis-rupprechtii voucher PS13154U701
AF041340 Larix sibirica
AF041344 Larix kaempferi
AY692177 Larix sibirica clone 2
AY523434 Larix decidua clone 14
AY523431 Larix decidua clone 2
AY523430 Larix decidua clone 12
AY692173 Larix sibirica clone 6
GQ805723.1 Larix gmelinii var. principis-rupprechtii voucher PS13154U704
AY523446 Larix kaempferi clone 24
AY523438 Larix gmelinii var. principis-rupprechtii clone 4
AY523444 Larix kaempferi clone 6
AY523440 Larix gmelinii var. principis-rupprechtii clone 12
AY523435 Larix decidua clone 18
AY523439 Larix gmelinii var. principis-rupprechtii clone 8
AY141351 Pseudotsuga wilsoniana
JF828663 Pseudotsuga sinensis var. gaussenii voucher AHNU P2090
AF041350 Pseudotsuga sinensis
AY692181 Larix gmelinii var. olgensis clone 19
AY692183 Larix sibirica clone 23
AY692179 Larix sibirica clone 21
AY692175 Larix sibirica clone 4
AY692172 Larix sibirica clone 2
AY523443 Larix sibirica clone 15
MF348854 Larix laricina isolate AC3HL21
AY523442 Larix sibirica clone 5
AY692180 Larix sibirica clone 19
AY141343 Larix decidua
AY523450 Larix sibirica clone 1
AY692174 Larix sibirica clone 24
AY692171 Larix sibirica clone 3
AY523447 Larix griffithiana clone 3
AY189843 Larix potaninii var. potaninii clone 3
AY189849 Larix laricina clone 1
AY523439 Larix laricina clone 7
AY523449 Larix occidentalis clone 4
AY141346 Larix laricina
AY141347 Larix occidentalis
AY141346 Larix lyallii
AY523452 Larix laricina clone 30
AY523460 Larix laricina clone 18
AY523448 Larix occidentalis clone 35
AY523461 Larix laricina clone 19
AY523455 Larix occidentalis clone 2
AY523457 Larix occidentalis clone 5
AY523454 Larix occidentalis clone 1
MG210815.1 Larix occidentalis voucher ERM1049
MG210816.1 Larix laricina voucher CCDB-24911-E01
AY692184 Larix laricina clone 27
EF660568 Podocarpus macrophyllus voucher HZ05070103
KM074139 Podocarpus neriifolius isolate YMM490914
AY583606 Podocarpus longifolioides

Table S1. Plant samples of Pinaceae used for analysis in this study.

Subfamily	Genus	No. of taxa	No. of sequences
Abietoideae	*Abies*	18	105
	Keteeleria	3	9
	Tsuga	2	4
Laricoideae	*Larix*	11	112
	Pseudotsuga	1	3
Pinoideae	*Picea*	7	24
	Pinus	29	111
Total	**7**	**71**	**368**

Table S2. List of taxa and records with accession numbers analyzed in this study.

Taxon	GeneBank accession no.
Abies amabilis voucher RBG Edinburgh UK198163a	EF057688
Abies amabilis voucher CCDB-23327-G03	MG216847.1
Abies beshanzuensis isolate A1 clone 1	MH165486
Abies beshanzuensis isolate A1 clone 2	MH165487
Abies beshanzuensis isolate A1 clone 3	MH165488
Abies beshanzuensis isolate A1 clone 4	MH165489
Abies beshanzuensis isolate A1 clone 5	MH165490
Abies beshanzuensis isolate A2 clone 1	MH165491
Abies beshanzuensis isolate A2 clone 2	MH165492
Abies beshanzuensis isolate A2 clone 3	MH165493
Abies beshanzuensis isolate A2 clone 4	MH165494
Abies beshanzuensis isolate A3 clone 1	MH165495
Abies beshanzuensis isolate A3 clone 2	MH165496
Abies beshanzuensis isolate A3 clone 3	MH165497
Abies beshanzuensis isolate A3 clone 4	MH165498
Abies beshanzuensis isolate A3 clone 5	MH165499
Abies beshanzuensis isolate A4 clone 1	MH165500
Abies beshanzuensis isolate A4 clone 2	MH165501
Abies beshanzuensis isolate A4 clone 3	MH165502
Abies beshanzuensis isolate A4 clone 4	MH165503
Abies beshanzuensis isolate B2	MH165505
Abies beshanzuensis isolate C2 clone 1	MH165506
Abies beshanzuensis isolate C2 clone 2	MH165507
Abies beshanzuensis isolate C4	MH165508
Abies beshanzuensis isolate C8	MH165509
Abies beshanzuensis isolate D1 clone 1	MH165510
Abies beshanzuensis isolate D1 clone 2	MH165511
Abies beshanzuensis isolate D4	MH165512
Abies beshanzuensis isolate D8 clone 1	MH165513
Abies beshanzuensis isolate D8 clone 2	MH165514
Abies beshanzuensis isolate D8 clone 3	MH165515
Abies beshanzuensis isolate D8 clone 4	MH165516
Abies beshanzuensis isolate E5	MH165517
Abies beshanzuensis isolate E21 clone 1	MH165518
Abies beshanzuensis isolate E21 clone 2	MH165519
Abies beshanzuensis isolate E21 clone 3	MH165520
Abies beshanzuensis isolate A4 clone 5	MH165504.1
Abies chensiensis voucher Q.P. Xiang s.n. (PE)	EF057702
Abies chensiensis isolate a3	MH17728
Abies chensiensis isolate a1	MH177279
Abies chensiensis isolate a2	MH177280

Abies chensiensis isolate b1	MH177282
Abies chensiensis isolate b2	MH177283
Abies chensiensis voucher Q316	MH703259.1
Abies concolor clone 7	DQ975359
Abies concolor clone 6	DQ975360
Abies concolor clone 8	DQ975361
Abies concolor clone 15	DQ975362
Abies concolor clone 45	DQ975363
Abies concolor clone 21	DQ975364
Abies concolor clone 29	DQ975365
Abies concolor voucher EBG 1938002A	EF057692
Abies delavayi clone 2	EU196133
Abies delavayi clone 4	EU196134
Abies delavayi clone 10	EU196135
Abies fabri clone 1	DQ975352
Abies fabri clone 2	DQ975353
Abies fabri voucher Guo Y.-Y. 28 (PE)	JF416968
Abies fabri voucher Guo Y.-Y. 28 (PE)	JF416969
Abies fanjingshanensis voucher Xiang Q.-P. s.n. (PE)	JF416970
Abies fanjingshanensis voucher Xiang Q.-P. s.n. (PE)	JF416971
Abies fanjingshanensis isolate 1 clone 1	MH177287
Abies fanjingshanensis isolate 1 clone 2	MH177288
Abies fanjingshanensis isolate 1 clone 3	MH177289
Abies fanjingshanensis isolate 1	MH177290
Abies fanjingshanensis isolate 1	MH177291
Abies fargesii voucher He J.-S. s.n. (PE)	JF416972
Abies fargesii voucher He J.-S. s.n. (PE)	JF416973
Abies fargesii voucher He J.-S. s.n. (PE)	JF416974
MH703258.1 Abies fargesii voucher Q315	MH703258.1
Abies firma isolate 1	MH177277.1
Abies firma isolate 2	MH177278.1
Abies firma voucher RBGE 19583927a	JF416975.1
Abies forrestii voucher DRB 61290 (PE) clone 1	EF057708
Abies forrestii clone 2	EU196131
Abies forrestii clone 3	EU196132
Abies forrestii voucher YLDP003A	MH117810.1
Abies forrestii voucher YLDP003B	MH117811.1
Abies forrestii voucher YLDP101A	MH117812.1
Abies forrestii voucher YLDP101B	MH117813.1
Abies forrestii voucher YLDP103A	MH117814.1
Abies kawakamii voucher C.J. Chen 9471 (PE) clone 1	EF063715
Abies kawakamii clone 11	EU196136
Abies koreana	AF283010
Abies koreana voucher Won M.-Y. 01	JF499944
Abies koreana f. chlorocarpa	AF283011

Abies koreana f. nigrocarpa	AF283013
Abies koreana f. rubrocarpa	AF283012
Abies lasiocarpa var. bifolia voucher CCDB-23954-B05	MG216802.1
Abies lasiocarpa voucher CCDB-18358-B12	MG216806.1
Abies pindrow clone 8	EU196128
Abies pindrow clone 14	EU196129
Abies pindrow clone 18	EU196130
Abies spectabilis voucher RBGE 19793335	JF416978
Abies spectabilis	MF785496
Abies squamata voucher RBGE 19890977	JF416980
Abies squamata voucher RBGE 19890977	JF416981
Abies squamata voucher RBGE 19890977	JF416982
Abies squamata voucher RBGE 19890977	JF416983
Abies squamata voucher RBGE 19890977	JF416979.1
Abies yuanbaoshanensis isolate 1	MH177284
Abies yuanbaoshanensis isolate 2	MH177285
Abies yuanbaoshanensis isolate 3	MH177286
Abies ziyuanensis voucher Xiang Q.-P. s.n. (PE)	JF416985.1
Abies ziyuanensis voucher Xiang Q.-P. s.n. (PE)	JF416984.1
Keteleeria davidiana clone T8	JN032111
Keteleeria davidiana var. calcarea clone H2	JN032110
Keteleeria davidiana voucher Ge130914	MH712627.1
Keteleeria fortunei clone Y12	JN032112
Keteleeria fortunei	MF785621
Keteleeria fortunei var. cyclolepis voucher AHNU:P883 clone JN19	JF829715
Keteleeria fortunei var. cyclolepis voucher AHNU:P883 clone JN1	JF829716.1
Keteleeria pubescens voucher AHNU:P1604 clone RM4	JF829713
Keteleeria pubescens voucher AHNU:P1604 clone RM10	JF829714
Larix decidua	AF041343
Larix decidua clone 2	AY523431
Larix decidua clone 8	AY523432
Larix decidua clone 12	AY523433
Larix decidua clone 14	AY523434
Larix decidua clone 18	AY523435
Larix decidua voucher CCDB-18346-E08	MG216878.1
Larix gmelinii var. gmelinii clone 12	AY523437.1
Larix gmelinii var. olgensis clone 19	AY523448.1
Larix gmelinii var. principis-rupprechtii clone 4	AY523438.1
Larix gmelinii var. principis-rupprechtii clone 8	AY523439.1
Larix gmelinii var. principis-rupprechtii clone 12	AY523440.1
Larix gmelinii var. principis-rupprechtii voucher PS1354MT01	GQ865722.1
Larix gmelinii var. principis-rupprechtii voucher	GQ463498.1

PS1354MT03	
Larix gmelinii var. principis-rupprechtii voucher PS1354MT04	GQ865723.1
Larix gmelinii voucher Q648	MH703336.1
Larix griffithiana	AF041349
Larix griffithiana clone 1	AY523413
Larix griffithiana clone 4	AY523414
Larix griffithiana clone 5	AY523415
Larix griffithiana clone 2	AY523416
Larix griffithiana clone 3	AY523417
Larix griffithiana clone 9	AY523418
Larix griffithiana clone 13	AY603159
Larix griffithiana clone 34	AY603160
Larix griffithiana clone 35	AY603161
Larix griffithii	MF785497
Larix speciosa clone 3	AY523424
Larix potaninii var. himalaica clone 22	AY188540
Larix kaempferi	AF041344
Larix kaempferi clone 1	AY523442
Larix kaempferi clone 6	AY523444
Larix kaempferi clone 15	AY523445
Larix kaempferi clone 24	AY523446
Larix kaempferi clone 31	AY523447
Larix laricina	AF041348
Larix laricina clone 1	AY188548
Larix laricina clone 2	AY188549
Larix laricina clone 6	AY523459
Larix laricina clone 18	AY523460
Larix laricina clone 19	AY523461
Larix laricina clone 30	AY523462
Larix laricina isolate AD3HL21	MF348954
Larix lyallii	AF041346
Larix lyallii voucher CCDB-24911-E01	MG216886.1
Larix mastersiana clone 1	AY523419
Larix mastersiana clone 2	AY523420
Larix mastersiana clone 5	AY523421
Larix mastersiana clone 10	AY523422
Larix mastersiana clone 25	AY523423
Larix occidentalis	AF041347
Larix occidentalis clone 1	AY523454
Larix occidentalis clone 2	AY523455
Larix occidentalis clone 4	AY523456
Larix occidentalis clone 5	AY523457
Larix occidentalis clone 35	AY523458
Larix occidentalis voucher ERM1049	MG216815.1

Larix potaninii var. australis clone 1	AY188526.1
Larix potaninii var. australis clone 2	AY188527.1
Larix potaninii var. australis clone 7	AY188550.1
Larix potaninii var. australis clone 17	AY188551.1
Larix potaninii var. australis clone 18	AY188528.1
Larix potaninii var. australis clone 20	AY188529.1
Larix potaninii var. chinensis clone 1	AY188530.1
Larix potaninii var. chinensis clone 5	AY188531.1
Larix potaninii var. chinensis clone 9	AY188532.1
Larix potaninii var. chinensis clone 26	AY188534.1
Larix potaninii var. chinensis clone 27	AY188552.1
Larix potaninii var. hiamlaica clone 14	AY188553.1
Larix potaninii var. himalaica clone 5	AY188535.1
Larix potaninii var. himalaica clone 7	AY188536.1
Larix potaninii var. himalaica clone 13	AY188538.1
Larix potaninii var. himalaica clone 18	AY188539.1
Larix potaninii var. potaninii clone 16	AY188547
Larix potaninii var. potaninii clone 1	AY188541.1
Larix potaninii var. potaninii clone 2	AY188542.1
Larix potaninii var. potaninii clone 3	AY188543.1
Larix potaninii var. potaninii clone 5	AY188544.1
Larix potaninii var. potaninii clone 8	AY188545.1
Larix potaninii var. potaninii clone 10	AY188546.1
Larix potaninii voucher Q472	MH703285.1
Larix sibirica	AF041345
Larix sibirica clone 1	AY523450
Larix sibirica clone 4	AY523451
Larix sibirica clone 9	AY523452
Larix sibirica clone 15	AY523453
Larix sibirica clone 2	AY603173
Larix sibirica clone 3	AY603174
Larix sibirica clone 24	AY603176
Larix sibirica clone 2	AY603177
Larix sibirica clone 4	AY603178
Larix sibirica clone 6	AY603179
Larix sibirica clone 18	AY603180
Larix sibirica clone 23	AY603181
Larix sibirica clone 4	AY603175.1
Larix speciosa clone 4	AY523425.1
Larix speciosa clone 6	AY523426.1
Larix speciosa clone 7	AY603162.1
Larix speciosa clone 8	AY603163.1
Larix speciosa clone 9	AY523427.1
Larix speciosa clone 9	AY603166.1
Larix speciosa clone 10	AY603164.1

Larix speciosa clone 13	AY603167.1
Larix speciosa clone 16	AY603168.1
Larix speciosa clone 17	AY603169.1
Larix speciosa clone 18	AY523428.1
Larix speciosa clone 18	AY603170.1
Larix speciosa clone 19	AY603171.1
Larix speciosa clone 20	AY523429.1
Larix speciosa clone 23	AY603172.1
Larix speciosa clone 23	AY603165.1
Larix speciosa clone 25	AY523430.1
Picea glauca isolate 64I5I2	AF136618
Picea glauca isolate 494I5I2	AF136619
Picea glauca isolate SAI5I2	AF136620
Picea glauca isolate VBI5I2	AF136621
Picea glauca isolate AD3HK69	MF349026
Picea likiangensis	MF785498
Picea likiangensis voucher YLDP021A	MH117817.1
Picea likiangensis voucher YLDP021B	MH117818.1
Picea mariana isolate 63I5I2	AF136614
Picea mariana isolate 4274I5I2	AF136615
Picea mariana isolate 4962I5I2	AF136616
Picea mariana isolate 5004I5I2	AF136617
Picea mariana voucher JEM 150	MG216912.1
Picea meyeri voucher PS1349MT03	GQ463495
Picea meyeri voucher PS1349MT02	GQ865721
Picea rubens isolate 61I5I2	AF136610
Picea rubens isolate 2032I5I2	AF136612
Picea rubens isolate 2505I5I2	AF136613
Picea sitchensis voucher CCDB-23395-F09	MG216825.1
Picea sitchensis voucher CCDB-23395-E09	MG216855.1
Picea wilsonii voucher PS1358MT01	GQ463500
Picea wilsonii voucher PS1358MT02	GQ463501
Picea wilsonii isolate AD3HJ29	MF349087
Picea wilsonii voucher Q673	MH703339.1
Pinus albicaulis	AF036983
Pinus albicaulis clone ITS2-08	AY430073.1
Pinus armandii	AF036980
Pinus armandii isolate AD3HK68	MF349015
Pinus armandii clone ITS2-07	AY430072.1
Pinus armandii voucher HSS18012801	MH444827.1
Pinus armandii voucher HSS18012802	MH444828.1
Pinus armandii voucher HSS18012803	MH444829.1
Pinus armandii voucher YLDP059B	MH117820.1
Pinus armandii voucher HSS18012804	MH444830.1
Pinus armandii voucher Q215	MH703240.1

Pinus armandii voucher YLDP059A	MH117819.1
Pinus cembra	AF036987
Pinus cembra clone 1	AF344002
Pinus cembroides	AF036997
Pinus cembroides clone 3	AF343983
Pinus cembroides clone 2	AF343994
Pinus cembroides subsp. lagunae clone 1	AF343985
Pinus cembroides subsp. orizabensis clone 1	AF343982
Pinus contorta	AF037014
Pinus contorta var. latifolia voucher CCDB-18346-D09	MG216817.1
Pinus contorta voucher ERM558	MG216868.1
Pinus contorta voucher ERM609	MG216850.1
Pinus culminicola	AF036995
Pinus culminicola clone 1	AF343988
Pinus echinata	AF037016
Pinus echinata	AF367378
Pinus flexilis clone 1	AF344001
Pinus flexilis clone ITS2-10	AY430075.1
Pinus flexilis voucher CCDB-23395-B04	MG216873.1
Pinus flexilis voucher CCDB-23395-D04	MG216819.1
Pinus hwangshanensis voucher AHNU:P458 clone 16-5	JF8296
Pinus hwangshanensis voucher AHNU:P458 clone 16-6	JF829697
Pinus hwangshanensis voucher AHNU:P453 clone 47-17	JF829698
Pinus hwangshanensis voucher AHNU:P557 clone 93-30	JF829699
Pinus hwangshanensis voucher AHNU:P567 clone 35-7	JF829700
Pinus hwangshanensis voucher AHNU:P567 clone 35-12	JF829701
Pinus krempfii	AF200521
Pinus krempfii	AF305061
Pinus luchuensis voucher AHNU:P2904 clone 1-11	JF829702
Pinus luchuensis voucher AHNU:P2904 clone 1-20	JF829703
Pinus massoniana	AF305063
Pinus massoniana voucher AHNU:P444 clone 126-K7	JF829704
Pinus massoniana voucher AHNU:P444 clone 126-K9	JF829705
Pinus massoniana voucher AHNU:P444 clone 126-K10	JF829706
Pinus massoniana var. massoniana voucher PS1356MT02	GQ463499.1
Pinus massoniana voucher MWS18012802	MH444832.1
Pinus maximartinezii	AF036994
Pinus maximartinezii clone 2	AF343995
Pinus monophylla clone 2	AF343984
Pinus monophylla clone 1	AF343986
Pinus monticola	AY619694
Pinus monticola voucher ERM764	MG216921.1
Pinus mugo isolate B40	MW816473
Pinus mugo isolate B41	MW816474
Pinus mugo isolate T6	MW816476

Pinus mugo isolate T1	MW816475
Pinus mugo voucher CCDB-24911-A01	MG216836.1
Pinus nelsonii	AF037001
Pinus nelsonii clone 1	AF343998
Pinus nelsonii clone 2	AF343999
Pinus pinceana	AF036998
Pinus pinceana clone 2	AF343996
Pinus quadrifolia clone 1	AF343990
Pinus quadrifolia clone 1	AF343991
Pinus remota clone 1	AF343980
Pinus remota clone 2	AF343989
Pinus remota clone 1	AF343992
Pinus rigida clone pcPR1_2_3	KC583358
Pinus rigida voucher MT00179249	MG216866.1
Pinus mugo subsp. rotundata isolate CB1	MW816477
Pinus mugo subsp. rotundata isolate CB20	MW816478
Pinus mugo subsp. rotundata isolate I1	MW816479
Pinus mugo subsp. rotundata isolate I21	MW816480
Pinus mugo subsp. rotundata isolate N4	MW816481
Pinus mugo subsp. rotundata isolate N5	MW816482
Pinus mugo subsp. rotundata isolate R1	MW816483
Pinus mugo subsp. rotundata isolate R12	MW816484
Pinus mugo subsp. rotundata isolate S18	MW816485
Pinus mugo subsp. rotundata isolate S20	MW816486
Pinus sylvestris	AF037003
Pinus sylvestris voucher 06_1000_0008	MW816487
Pinus sylvestris isolate P2	MW816488
Pinus sylvestris voucher NMW6748	KX167937.1
Pinus sylvestris voucher RBGE188	KX167550.1
Pinus sylvestris voucher RBGE186	KX167559.1
Pinus sylvestris voucher RBGE187	KX167560.1
Pinus sylvestris voucher RBGE189	KX167551.1
Pinus tabuliformis var. henryi voucher AHNU:P1753 clone 58-50	JF829691
Pinus tabuliformis var. tabuliformis voucher PS1361MT04	GQ865727
Pinus tabuliformis var. tabuliformis voucher PS1361MT01	GQ463502.1
Pinus tabuliformis var. tabuliformis voucher PS1361MT02	GQ865725.1
Pinus tabuliformis var. tabuliformis voucher PS1361MT03	GQ865726.1
Pinus tabuliformis var. tabuliformis voucher PS1361MT05	GQ865728.1
Pinus taeda	AF367379
Pinus taeda	MK895659.1
Pinus thunbergii	AF037025
Pinus thunbergii voucher PS1353MT01	GQ463496
Pinus thunbergii voucher PS1353MT02	GQ463497
Pinus thunbergii voucher HS18012801	MH444823.1

Pinus thunbergii voucher HS18012802	MH444824.1
Pinus thunbergii voucher HS18012803	MH444825.1
Pinus thunbergii voucher HS18012804	MH444826.1
Pinus mugo subsp. rotundata isolate B36	MW816489
Pinus mugo subsp. rotundata isolate B37	MW816490
Pinus mugo subsp. rotundata isolate W7	MW816491
Pinus mugo subsp. rotundata isolate W14	MW816492
Pinus virginiana	AF037015
Pinus virginiana isolate AD7LH42	MF348972
Pinus x rhaetica isolate C18	MW816494
Pinus x rhaetica isolate C23	MW816495
Pseudotsuga sinensis	AF041350
Pseudotsuga sinensis var. gaussenii voucher AHNU:P2006 clone HD14-5	JF829695.1
Pseudotsuga wilsoniana	AF041351
Tsuga canadensis isolate AD3HK27	MF349035
Tsuga canadensis voucher CCDB-18346-B09	MG216811.1
Tsuga chinensis clone 4	DQ975358
Tsuga chinensis voucher Q541	MH703311.1

Table S3. List of individuals representing *Pinus mugo* complex taxa sequenced in this study.

No.	Taxon name	Origin	Coordinates	Accession number
1	*Pinus mugo*	Tatra Mountains, Poland	49°14′30 N, 20°00′15 E	MW816475
2	*Pinus mugo*	Tatra Mountains, Poland	49°16′07 N, 20°02′41 E	MW816476
3	*Pinus uliginosa*	"Torfowisko pod Węglińcem" Nature Reserve, Poland	51°17′36 N, 15°13′37 E	MW816491
4	*Pinus uliginosa*	"Torfowisko pod Węglińcem" Nature Reserve, Poland	51°17′37 N, 15°13′34 E	MW816492
5	*Pinus* x *rhaetica*	"Bór nad Czerwonem" Nature Reserve, Poland	50° 15′42 N, 16° 8′ 31 E	MW816494
6	*Pinus* x *rhaetica*	"Bór nad Czerwonem" Nature Reserve, Poland	50° 15 42 N, 16° 8′ 31 E	MW816495
7	*Pinus uliginosa*	"Wielkie Torfowisko Batorowskie" Nature Reserve, Poland	50°27′30 N, 16°22′57 E	MW816489
8	*Pinus uliginosa*	"Wielkie Torfowisko Batorowskie" Nature Reserve, Poland	50°27′30 N, 16°22′57 E	MW816490
9	*Pinus mugo*	"Wielkie Torfowisko Batorowskie" Nature Reserve, Poland	50°49′51N, 16°20′30 E	MW816473
10	*Pinus mugo*	"Wielkie Torfowisko Batorowskie" Nature Reserve, Poland	50°50′29 N, 16°51′32 E	MW816474
11	*Pinus sylvestris*	Dendrological Garden, University of Life Sciences, Poland	52°25′32 N, 16°53′39 E	MW816487
12	*Pinus sylvestris*	Morasko, Poland	52°28′02 N, 16°55′30 E	MW816488
13	*Pinus rotundata*	Ibacher Moor, Germany	47°43′38 N, 8°03′01 E	MW816479
14	*Pinus rotundata*	Ibacher Moor, Germany	47°43′38 N, 8°03′01 E	MW816480
15	*Pinus rotundata*	Rotmeer, Germany	47°51′53 N, 8°06′00 E	MW816483
16	*Pinus rotundata*	Rotmeer, Germany	47°51′53 N, 8°06′00 E	MW816484
17	*Pinus rotundata*	Steerenmoos, Germany	47°48′21 N, 8°12′00 E	MW816485
18	*Pinus rotundata*	Steerenmoos, Germany	47°48′20 N, 8°12′00 E	MW816486
19	*Pinus rotundata*	Novohůrecká slať, Czech Republik	49°09′18 N,	MW816481

			13°19′46 E	
20	*Pinus rotundata*	Novohůrecká slať, Czech Republik	49°09′19 N, 13°19′45 E	MW816482
21	*Pinus rotundata*	Červené blato, Czech Republik	48°51′26 N, 14°48′08 E	MW816477
22	*Pinus rotundata*	Červené blato, Czech Republik	48°51′37 N, 14°48′20 E	MW816478

Table S4. Abundance of all intraspecific and all interspecific K2P pairwise distance in *Pinaceae*.

Distances	Intra-specific distance	Inter-specific distance
<= 0.0%	17,16	8,79
0.0% - 1.0%	46,04	14,74
1.0% - 2.0%	20,73	14,39
2.0% - 3.0%	3,01	9,71
3.0% - 4.0%	5,23	6,25
4.0% - 5.0%	1,44	2,75
5.0% - 6.0%	0	11,12
6.0% - 7.0%	0,22	5,59
7.0% - 8.0%	1,33	1,68
8.0% - 9.0%	0,55	4,86
9.0% - 10.0%	0,22	3,64
10.0% - 11.0%	0	2,62
11.0% - 12.0%	0	1,46
12.0% - 13.0%	0	1,74
13.0% - 14.0%	0	1,15
14.0% - 15.0%	0	0,25
15.0% - 16.0%	0	0,16
16.0% - 17.0%	0	0,15
17.0% - 18.0%	0	0,32
18.0% - 19.0%	0	0,6
19.0% - 20.0%	0	0,49
20.0% - 21.0%	0	0,42
21.0% - 22.0%	0	0,08
22.0% - 23.0%	0	0,04
23.0% - 24.0%	0	0,01
24.0% - 25.0%	0	0,02
25.0% - 26.0%	0	0
26.0% - 27.0%	0	0,01
27.0% - 28.0%	0	0,03
28.0% - 28.99%	0	0,09
28.99% - 30.0%	0	0,03
> 30.0%	4,01	3,37

Table S5. The results of species discrimination in seven genera of the *Pinaceae* family based on the BLAST1 method.

Genera	Correct	Ambiguous	Incorrect
Abies	43,81	51,43	4,76
Keteleeria	22,22	77,78	0,00
Tsuga	100,00	0,00	0,00
Larix	29,46	67,86	2,68
Pseudotsuga	0,00	100,00	0,00
Picea	8,33	66,67	25,00
Pinus	30,63	55,86	13,51

4.3. Publikacja 3

Wcześniejsze próby znalezienia markerów DNA specyficznych gatunkowo dla *Pinus mugo, Pinus uliginosa* i *Pinus uncinata* w oparciu o osiem chloroplastowych regionów barkodowych, wykazały brak zmienności zarówno między- jak i wewnątrzgatunkowej w analizowanym zakresie (Celiński i in. 2017a). Podobnie zresztą jak próba identyfikacji różnic na podstawie jądrowej sekwencji ITS2 (Sokołowska i in. 2022). Zwiększenie zakresu poszukiwań do prawie stu dwudziestu tysięcy nukleotydów w postaci kompletnych genomów chloroplastowych (Celiński i in. 2017b, Sokołowska i in. 2021) pozwoliło na wskazanie potencjalnych różnic w kilku regionach DNA u wyżej wymienionych taksonów. Konieczne jednak było poszerzenie badań o większą liczbę osobników z danego taksonu, w tym także taksonów pochodzących z różnych populacji.

W związku z tym, **w trzeciej części mojej rozprawy doktorskiej przeprowadziłam analizę relacji genetycznych i taksonomicznych w kompleksie *Pinus mugo*** na podstawie siedmiu zestawów danych wygenerowanych dzięki sekwencjonowaniu nowej generacji i metodzie przeczesywania genomu, w tym: A) 57 kompletnych genów chloroplastowych; B) 15 kompletnych genów mitochondrialnych; C) 18 chloroplastowych regionów międzygenowych (ang. *Intergenic Spacer*, IGS); D) kompletnego jądrowego cistronu nrDNA; E) sekwencji ITS (ang. *Internal Transcribed Spacer,* ITS); F) regionów barkodowych *matK* + *rbcL*; G) 5 regionów hiperzmiennych (tzw. hotspotów) (Sikora i Celiński 2024).

W tej pracy zwrócono szczególną uwagę na poszukiwanie zestawu danych, za pomocą którego będzie można zidentyfikować różnice międzygatunkowe, poprawić rozdzielczość filogenetyczną, a także opisać ogólną zmienność genetyczną w obrębie poszczególnych populacji i taksonów. Aby ocenić skuteczność wygenerowanych zestawów danych do dyskryminacji taksonów z kompleksu *Pinus mugo* wykorzystano trzy różne metody oparte na: 1) drzewie filogenetycznym, 2) odległościach genetycznych i 3) podobieństwie sekwencji z wykorzystaniem algorytmu ASAP (ang. *Assemble Species by Automatic Partitioning*). Zróżnicowanie genetyczne wśród populacji i taksonów z kompleksu *Pinus mugo* zostało ocenione na podstawie analizy haplotypów i analizy PCoA (ang. *Principal Coordinate Analysis*).

Zastosowanie metody przeczesywania genomu pozwoliło po raz pierwszy na uzyskanie i scharakteryzowanie kompletnej sekwencji ITS (ang. *Internal Transcribed Spacer*) i jądrowego cistronu nrDNA dla przedstawicieli kompleksu *Pinus mugo*. Jednakże pomimo znacznej długości (tj. 2881 pz), kompletny region ITS charakteryzuje się niską zmiennością

i podobnie jak region ITS2, nie pozwala na skuteczną dyskryminację tych blisko spokrewnionych taksonów.

Uzyskane wyniki delimitacji taksonów przy użyciu trzech różnych podejść znacząco się od siebie różniły. Najwyższy procent sukcesu w rozróżnianiu taksonów osiągnięto przy użyciu metody opartej na odległości i zestawu złożonego z regionów wysoce zmiennych (tzw. hotspotów), tj. *rps1*, *rps2*, *rps14*, *ycf3-psaA* i *trnE-clpP*. Z kolei, metoda ASAP (ang. *Assemble Species by Automatic Partitioning*) miała najniższy wskaźnik sukcesu, niezależnie od użytego zestawu danych.

Przeprowadzone analizy dostarczają również nowych danych na temat różnorodności genetycznej sosny błotnej z terenów torfowiskowych Czech i Niemiec. Ujawniony wzór zmienności genetycznej oparty na genach mitochondrialnym wyraźnie wskazuje na bardzo unikalny i lokalny charakter analizowanych populacji *Pinus rotundata*, jak i na fakt, że polskie populacje *Pinus uliginosa* są od nich odmienne, co może być związane ze stosunkowo długą izolacją przestrzenną jak i wskazywać na złożone tło historyczne poszczególnych populacji. Wykazano również, że *Pinus uncinata* stanowi wyraźnie odrębny klad w obrębie kompleksu *Pinus mugo* w przypadku większości zastosowanych metod i zestawów danych genetycznych.

Co więcej, uzyskane wyniki wskazują, że *Pinus* × *rhaetica* jest znacznie bardziej zbliżona genetycznie do *Pinus sylvestris* niż do taksonów należących do kompleksu *Pinus mugo*. Wyniki te są w pełni zgodne z wcześniej przeprowadzonymi badaniami przy użyciu profilowania białek nasion (Celińskii in. 2020) oraz analizy cytogenetycznej i cytometrii przepływowej (Celiński i in. 2019).

W świetle uzyskanych wyników można skonkludować, że *Pinus* × *rhaetica* nie należy do kompleksu *Pinus mugo* i nie powinna być utożsamiana ani z *Pinus uliginosa*, ani z *Pinus rotundata*. Co więcej, nazw tych nie należy używać jako synonimów.

PUBLIKACJA 3

Sikora, J.; Celiński, K. Exploring taxonomic and genetic relationships in the *Pinus mugo* complex using genome skimming data. International Journal of Molecular Sciences 2024, 25, 10178.

https://doi.org/10.3390/ijms251810178

International Journal of
Molecular Sciences

MDPI

Article

Exploring Taxonomic and Genetic Relationships in the *Pinus mugo* Complex Using Genome Skimming Data

Joanna Sikora and Konrad Celiński *

Department of Genetics, Institute of Experimental Biology, Faculty of Biology, School of Natural Sciences, Adam Mickiewicz University, Uniwersytetu Poznańskiego 6, 61-614 Poznań, Poland; joanna.sikora@amu.edu.pl
* Correspondence: konrad.celinski@amu.edu.pl

Abstract: Genome skimming is a novel approach that enables obtaining large-scale genomic information based on high-copy DNA fractions from shallow whole-genome sequencing. The simplicity of this method, low analysis costs, and large amounts of generated data have made it widely used in plant research, including species identification, especially in the case of protected or endangered taxa. This task is particularly difficult in the case of closely related taxa. The *Pinus mugo* complex includes several dozen closely related taxa occurring in the most important mountain ranges in Europe. The taxonomic rank, origin, or distribution of many of these taxa have been debated for years. In this study, we used genome skimming and multilocus DNA barcoding approaches to obtain different sequence data sets and also to determine their genetic diversity and suitability for distinguishing closely related taxa in the *Pinus mugo* complex. We generated seven different data sets, which were then analyzed using three discrimination methods, i.e., tree based, distance based, and assembling species by automatic partitioning. Genetic diversity among populations and taxa was also investigated using haplotype network analysis and principal coordinate analysis. The proposed data set based on divergence hotspots is even twenty-times more variable than the other analyzed sets and improves the phylogenetic resolution of the *Pinus mugo* complex. In light of the obtained results, *Pinus* × *rhaetica* does not belong to the *Pinus mugo* complex and should not be identified with either *Pinus uliginosa* or *Pinus rotundata*. It seems to represent a fixed hybrid or introgressant between *Pinus sylvestris* and *Pinus mugo*. In turn, *Pinus mugo* and *Pinus uncinata* apparently played an important role in the origins of *Pinus uliginosa* and *Pinus rotundata*.

Keywords: *Pinus mugo* complex; genome skimming; nrDNA; mitogenome; plastome; intergenic spacer; next-generation sequencing

Citation: Sikora, J.; Celiński, K. Exploring Taxonomic and Genetic Relationships in the *Pinus mugo* Complex Using Genome Skimming Data. *Int. J. Mol. Sci.* **2024**, *25*, 10178. https://doi.org/10.3390/ijms251810178

Academic Editor: Luigi De Masi

Received: 21 August 2024
Revised: 19 September 2024
Accepted: 20 September 2024
Published: 22 September 2024

1. Introduction

The rapid development of both high-throughput sequencing and advanced bioinformatics methods in recent years, combined with the decreasing costs of performing such analyses, has enabled the widespread use of these advanced methods not only for medical purposes but also in many fields of life sciences, including ecological studies [1–5]. With the use of these powerful tools, it seems that the search for markers to distinguish and identify species has become easier than ever. However, many works show that the use of even the most modern and advanced approaches does not always guarantee an effective solution to seemingly simple but fundamental problems in biological research, such as species identification [6–8]. This task remains a major challenge, especially in the case of closely related taxa (so-called complexes), regardless of the development of new and sophisticated research methods.

The *Pinus mugo* complex from the *Pinaceae* family is a large (including 16 species, 91 varieties, and 19 other forms) and taxonomically demanding group of closely related pines found in peat bogs and mountain areas in Central and Southern Europe, i.e., dwarf

pine (*Pinus mugo* Turra), peat bog pine (*Pinus uliginosa* Neumann ex Wimmer/*Pinus rotundata* Link), or mountain pine (*Pinus uncinata* Ramond ex DC) [9,10]. Taxa discrimination and identification are problematic in this complex, mainly due to the lack of clarity regarding the nomenclature of individual taxa (the presence of different synonyms in scientific literature, i.e., *Pinus rotundata* as *Pinus uliginosa*) [9,11]; postulated gene flow (both within this complex and with taxa outside it, i.e., *Pinus sylvestris* L. [12–18]; the formation of interspecific hybrids, characterized by the high morphological and anatomical variability of such individuals (i.e *Pinus* × *rhaetica* Brügger); or complex evolutionary history [19,20]. Although the pine complex is one of the best-studied complexes of closely related taxa, including morphological [21–24], anatomical [21,25], biochemical [24,26–29], phytochemical [30,31], as well as genetic and cytogenetic studies [32–35], the relationships between these taxa are the subject of further research, but their delimitation has remained a major challenge for scientists for many years. The clarification of these complicated relationships is also hampered by the lack of variability at the DNA level observed in selected barcoding regions of chloroplast [36] and nuclear DNA [37].

The latest research conducted on the *Pinus mugo* complex, which involved a comparative analysis of the complete chloroplast genome sequences of three representatives of the *Pinus mugo* complex [38], showed the occurrence of potential genetic differences between them. However, it is difficult to assess their species specificity due to the lack of broader population studies in this area. Undoubtedly, the extension of the analyzed sequences helped to identify these differences and showed that this is the right direction for the analysis of these closely related taxa in future studies.

Genome skimming is a novel approach that allows obtaining large-scale genomic information based on high-copy DNA fractions from shallow whole genome sequencing. This method involves the extraction, assembling, and analysis of genomic DNA fractions present in cells in multiple copies, i.e., nuclear ribosomal cistron (nrDNA), chloroplast DNA (cpDNA), and mitochondrial DNA (mtDNA) [39]. This enables obtaining partial or even complete organellar genomes (plastomes and mitogenomes) as well as large tandem repeats of ribosomal DNA (rDNA) in plants [40]. The indisputable advantage of the method is the generation of sequence sets without the necessity for primer design, PCR amplification, electrophoresis, cloning, or Sanger sequencing. It is notable that genome skimming relies on the analysis of numerous genetic regions, enabling the identification of hotspots and novel, species-specific genetic markers. Given its simplicity and relatively low analysis costs, this approach has been used in many plant studies, including plant systematics, plant identification, DNA barcoding, phylogeny, and even the discovery of new markers [41–43]. The method was used also to distinguish species in many taxonomically challenging plant groups, i.e., *Theobroma* [44], *Araucaria* [45], *Diospyros* [6], and *Panax* [46].

The European Mountain pines aggregate provides an ideal model to assess the usefulness of genome skimming and multilocus DNA barcoding approaches for improving phylogenetic resolution and extended barcodes for taxa discrimination. Therefore, in this study, by sampling multiple individuals of different members of the *Pinus mugo* complex and closely related taxa, i.e., *Pinus sylvestris* and *Pinus* × *rhaetica*, we useda genome skimming approach coupled with bioinformatic and statistic pipelines to: (1) obtain large genomic data sets representing the most abundant mitochondrial, chloroplast, and nuclear regions of the genome; (2) generate diverse sequence data sets and determine their genetic variability and usefulness in studies of closely related pine taxa; (3) determine the effectiveness of discriminating between analyzed taxa using different data sets and delimitation methods; (4) reveal the genetic relationships of *P.* × *rhaetica*, *P. rotundata*, and *P. ulignosa* to each other; and (5) screen the genetic diversity of taxa and populations of the *Pinus mugo* complex.

The results of our study not only provide new and extensive genomic resources for further studies on the *Pinus mugo* complex, but also allow us to answer the question about the effectiveness and validity of using the genome skimming approach to perform analyses of other complexes of closely related taxa.

2. Results

2.1. DNA Sequencing, Assembly, and Annotation

Ion Torrent sequencing generated between 5,846,195 and 10,076,095 reads per sample. In samples, fifty-seven complete chloroplast genes, eighteen plastid intergenic spacer regions, and fifteen complete mitochondrial coding genes were identified using the genome skimming approach and bioinformatics methods. Additionally, a complete nrDNA cistron was also obtained, and this is the first report of it in representatives of the *Pinus mugo* complex. The total length of the nrDNA sequence was 7848 bp. The lengths of the individual parts were 1812 bp, 2482 bp, 158 bp, 241 bp, and 3155 bp for 18S, ITS1, 5.8S, ITS2, and 26S, respectively. A set of five divergent regions, selected on the basis of their significantly higher-than-average level of genetic variability, was also analyzed. A list of analyzed sequences and accession numbers is provided in the Supplementary Materials (Tables S2, S7–S10).

2.2. Genetic Characterization of Diversity in Data Sets

In this study, seven different genetic data sets were analyzed, namely: plastid coding genes (A), mitochondrial coding genes (B), plastid intergenic regions (C), nrDNA cistron (D), ITS (E), *matK* + *rbcL* (F), and divergence hotspot regions (G). In each data set, all sequences were concatenated, and then we determined the alignment length, variable sites, and parsimony informative sites, as well as percentage of variable sites and parsimony informative sites. For an easier comparison, these parameters are presented in two separate tables. Table 1 includes the results for taxa from the *Pinus mugo* complex. Table S3 includes the results obtained for *Pinus* × *rhaetica* and *Pinus sylvestris*. Seven data sets differed in all the above-mentioned parameters. For the *Pinus mugo* complex, the longest data set obtained was set A (45,467 bp) and the shortest was set G (1021 bp). The number of variable sites varied between data sets from three for set E to 128 for set B. The highest number of informative parsimony sites was found for set B (68) and the lowest (0) for data set E. The highest percentage of variable sites was found for set G (1.959%) and it was almost twenty-times higher than that observed in set E (0.104%), which was the least variable in this parameter. The percentage of informative parsimony sites ranged from 0% for set E to 1.469% in set G. For *Pinus* × *rhaetica* and *Pinus sylvestris*, similar results were obtained to those observed for the *Pinus mugo* complex. The longest data set was set A (45,467 bp) and the shortest set G (1021 bp).

Table 1. Comparison of variability characteristics of seven data sets in the *Pinus mugo* complex.

Alignment	Data Set Code	Length (bp)	Variable Sites (% divergence)	Parsimony Informative Sites (%)
Plastid coding genes	A	45,467	75 (0.165%)	17 (0.037%)
Mitochondrial coding genes	B	8378	128 (1.528%)	68 (0.812%)
Plastid intergenic regions	C	7092	59 (0.832%)	33 (0.465%)
nrDNA cistron	D	7849	10 (0.127%)	2 (0.025%)
ITS	E	2881	3 (0.104%)	0 (0.000%)
matK + *rbcL*	F	2976	8 (0.269%)	1 (0.034%)
Divergence hotspot regions	G	1021	20 (1.959%)	15 (1.469%)

The number of variable sites varied between data sets from 0 for sets E and F to 33 for set B. There were no informative parsimony sites for *Pinus* × *rhaetica* and *Pinus sylvestris* in the tree data sets analyzed (sets D, E, and F) (Table S3). The level of genetic variation observed in *Pinus* × *rhaetica* and *Pinus sylvestris* was lower than that observed for the *Pinus mugo* complex. Overall, data set B is characterized by the largest number of variable sites in

both the *Pinus mugo* complex and *Pinus* × *rhaetica* + *Pinus sylvestris*. In percentage terms, the G data set is the most variable in both cases considered.

The order of the data sets from most to least variable in percentage terms for the *Pinus mugo* complex taxa is G > B > C > F > A > D > E.

2.3. Delimitation of Taxa

2.3.1. Tree-Based Method

In the phylogenetic analyses, a species was deemed successfully identified when all individuals of the same species formed a monophyletic group with a support value exceeding 50%. The results of the tree-based analyses performed with different input data sets were not fully consistent or unambiguous.

When considering individual data sets, it can be seen that data set A, although the longest in terms of the base pair, did not allow for the successful identification of the analyzed taxa, and the discrimination success using this data set was zero. The mitochondrial coding genes (data set B) showed higher discriminatory power than plastid coding genes (data set A) because two taxa could be distinguished, i.e., *Pinus sylvestris* and *Pinus uncinata*. The remaining taxa did not form highly supported monophyletic groups. The data set covering the plastid intergenic regions (data set C) allowed us to distinguish three taxa, while set D identified only one taxon (Table 2). Data sets E and F did not extract any taxon. The highest success of taxonomic discrimination was characterized by set G. Using divergence hotspots, three taxa can be distinguished.

Table 2. Summary of the rate of taxa discrimination using seven data sets based on tree-based, distance-based, and ASAP methods in combined *Pinus mugo* complex taxa, *Pinus* × *rhaetica*, and *Pinus sylvestris*. *—Considering all individuals from one taxon in the same operational taxonomic unit (OTU).

Alignment	Data Set Code	Tree-Based Method	Distance-Based Method	ASAP *
Plastid coding genes	A	0/6 (0.00%)	0/6 (0.0%)	0/6 (0.00%)
Mitochondrial coding genes	B	2/6 (33.33%)	2/6 (33.33%)	0/6 (0.00%)
Plastid intergenic regions	C	2/6 (33.33%)	2/6 (33.33%)	0/6 (0.00%)
nr DNA cistron	D	1/6 (16.67%)	1/6 (16.67%)	0/6 (0.00%)
ITS	E	0/6 (0.00%)	0/6 (0.00%)	0/6 (0.00%)
matK + *rbcL*	F	0/6 (0.00%)	0/6 (0.00%)	0/6 (0.00%)
Divergence hotspot regions	G	3/6 (50.00%)	3/6 (50.0%)	1/6 (16.67%)

Considering individual taxon in turn, *P. mugo*, *P. uncinata*, *P. uliginosa*, *P. rotundata* from the *Pinus mugo* complex, as well as *P.* × *rhaetica* and *P. sylvestris* cannot be distinguished using A, E, and F data sets (Figures S1, S4 and S5). All *Pinus uncinata* samples were considered monophyletic and distinguished from other taxa of the *Pinus mugo* complex using data sets B, C, D, and G with 63 to 95% bootstrap support (Figures 1, S2 and S3). Similarly, all *Pinus sylvestris* individuals formed a monophyletic clade in data sets B, C, and G with 56% to 98% bootstrap support. All *Pinus mugo* specimens were resolved as monophyletic only with data set C (34% bootstrap support), and *P.* × *rhaetica* with data set G (75% bootstrap support) (Figure 1). *Pinus uliginosa* and *Pinus rotundata* individuals were not resolved as monophyletic for either the ITS, the broader nrDNA, or any of the plastid and mitochondrial data sets. In turn, *P.* × *rhaetica* was placed on the maximum likelihood phylogenetic tree between *Pinus sylvestris* individuals and taxa from the *Pinus mugo* complex based on data set C, and in data set G in the same clade as *Pinus sylvestris* (Figure 1).

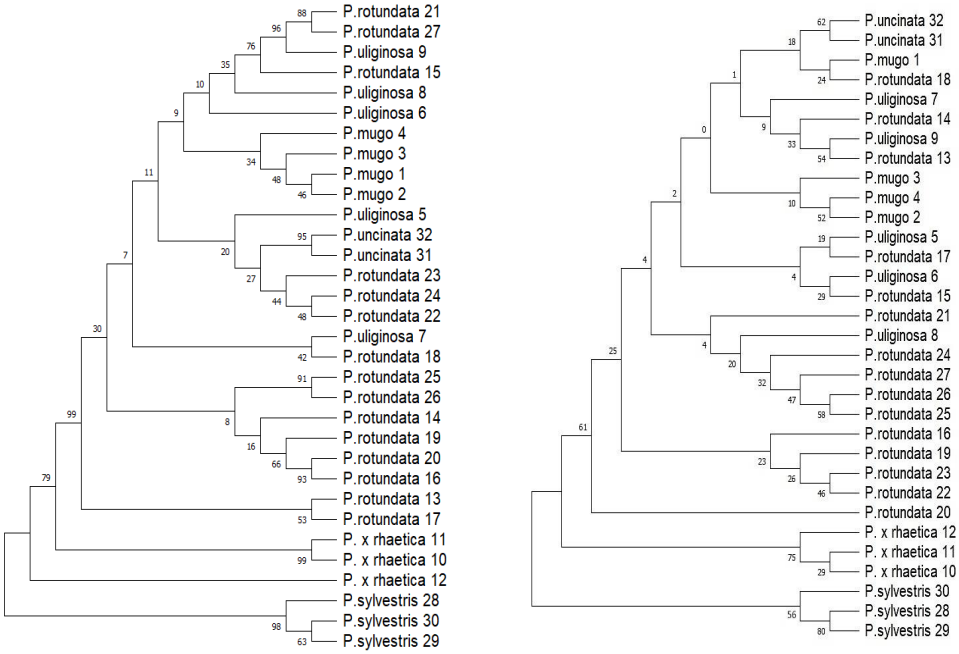


Figure 1. Maximum likelihood trees based on the plastid intergenic regions data set C (**left**) and divergence hotspot regions data set G (**right**).

2.3.2. Distance-Based Methods

In the distance analyses, a species was deemed successfully identified when its minimum interspecific distance exceeded its maximum intraspecific distance. The distance-based method exhibited a trend similar to that of the tree-based method.

Seven data sets (A–G) differed in their success in discriminating taxa. Sets A, E, and F did not differentiate taxa at all, while for sets D and B, C distinguished one (D) and two taxa (B,C), respectively. The highest discriminatory success (50%) was provided by data set G. It was possible to distinguish three taxa out of six analyzed. In data set G, the highest values were observed for both interspecific and intraspecific distances, 0.00295 and 0.01285, respectively (Table S4).

Considering individual taxa, for *Pinus mugo*, *Pinus rotundata*, and *Pinus uliginosa*, in each data set, their maximum value of intraspecific variability exceeded the minimum interspecific variability, making them impossible to delimit (Table S4). For *Pinus uncinata*, the minimum interspecific value was higher than the maximum intraspecific value in four of the seven data sets, i.e., B, C, D, and G, thus allowing the delimitation of this taxon. *Pinus* × *rhaetica* can be distinguished using three data sets, i.e., B, C, and G, while *Pinus sylvestris* can be distinguished only using data set G (Table S4).

2.3.3. Assembling Species by the Automatic Partitioning Method (ASAP)

In the case of the ASAP method, the obtained results differed significantly depending on the input data set used (Figure 2 and Figures S6–S10). For the plastid data (set A), ASAP indicates the presence of three or two partitions. In the case of two partitions, all *Pinus sylvestris* and *Pinus* × *rhaetica* individuals are in one partition and the rest of the individuals constitute the second one, corresponding as it were to the *Pinus mugo* complex. In the case of mitochondrial data, we observe a large fragmentation in the range of the partitions. There are 27 of them in total and they include mostly single individuals. One partition includes six individuals and contains representatives of *Pinus uliginosa*, *Pinus mugo*, and *Pinus* × *rhaetica*—although these are not all analyzed individuals from these taxa. The rest of the individuals representing these taxa occur in separate partitions. Set C (plastid intergenic regions) is basically consistent with set A. We observe two partitions: in one, 26 individuals of the *Pinus mugo* complex and, in the other, six individuals—all analyzed individuals of *Pinus sylvestris* and *Pinus* × *rhaetica*. The cistron data show four

partitions: one containing 29 individuals and three smaller partitions containing single individuals—each consisting of representatives of a different taxon. The ASAP score for the ITS data set is similar to the results obtained for cistron. Two partitions were distinguished: one has 31 individuals and the other only one. In terms of the barcode sequences (data set F), the lowest ASAP score was found for four partitions. The first partition contains 24 individuals from different taxa included in the *Pinus mugo* complex. The second largest partition contains six individuals, and these are all representatives of *Pinus sylvestris* and *Pinus* × *rhaetica*. Partitions three and four contain one individual each of *Pinus rotundata* (sample no. 27 from the Czech Republic) and *P. rotundata* (sample no. 21 from Germany). The divergence hotspots data indicate the presence of 23 partitions. Among them, three partitions contained three individuals each. The most homogeneous partition was the one composed exclusively of *Pinus* × *rhaetica* from "Bór na Czerwonem" (Poland) and exclusively of *Pinus rotundata* from the "Červené blato" population (Czech Republic). The third three-individual partition was already mixed and contained representatives of *P. uliginosa*, *P. mugo*, and *P. rotundata*.

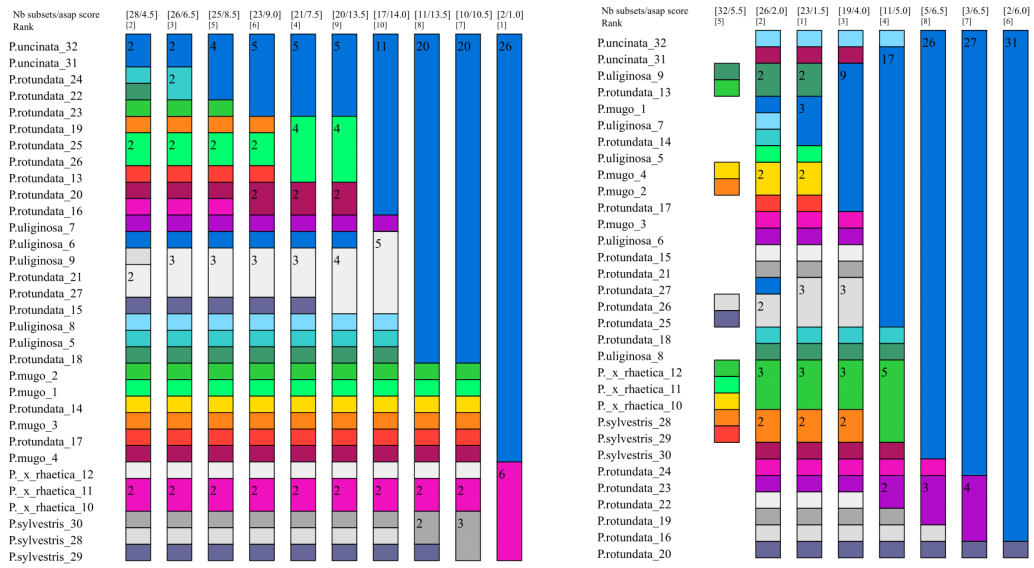


Figure 2. Species delimitation by ASAP analysis for two data sets: the plastid intergenic regions (**left**) and divergence hotspot regions (**right**). Graphical output showing the ten different delimitations: each column represents a partition, and the colors represent the operational taxonomic units (OTUs). Every field contains the number of individuals. Above the colorful bars, the coefficient asap-score (the lower value) and number of species (the upper value) recognized for whole data sets are presented.

In summary, the results of the taxon delimitation using three different approaches differ significantly from each other. The highest level of success in distinguishing taxa was achieved using the distance-based method and the G data set. The ASAP method had the lowest success rate, regardless of the input data used (Table 2).

2.4. Genetic Relationships of Taxa and Populations

Having a set of chloroplast data (set A) and mitochondrial data (set B), we determined the mutual relationships between the analyzed taxa using a median-joining haplotype network (Figure 3) constructed using PopArt [47]. Particular emphasis was placed on *Pinus* × *rhaetica* because the results obtained from taxon delimitation showed interesting trends.

This analysis revealed 26 chloroplast haplotypes (chlorotypes). Haplotype 9 is shared by two individuals of *Pinus mugo* and *Pinus rotundata*. Haplotype 18 is common to one individual of *Pinus uliginosa* and three individuals of *Pinus rotundata*. The highest genetic distance expressed in the number of nucleotide differences was observed between *Pinus* × *rhaetica* and *Pinus sylvestris* and the rest of the individuals belonging to the *Pinus*

mugo complex. The analysis of the relationship of the *Pinus mugo* complex, *Pinus* × *rhaetica*, and *Pinus sylvestris* based on complete mitochondrial genes (data set B) revealed the existence of 31 haplotypes (mitotypes). Two individuals of *Pinus* × *rhaetica* have the same haplotype (haplotype 9). Interestingly, all mitochondrial haplotypes of *Pinus rotundata*, *Pinus uliginosa*, and *Pinus uncinata* corresponded to the population from which the individuals were collected. The list of detected chloro- and mitotypes is provided in Tables S5 and S6.

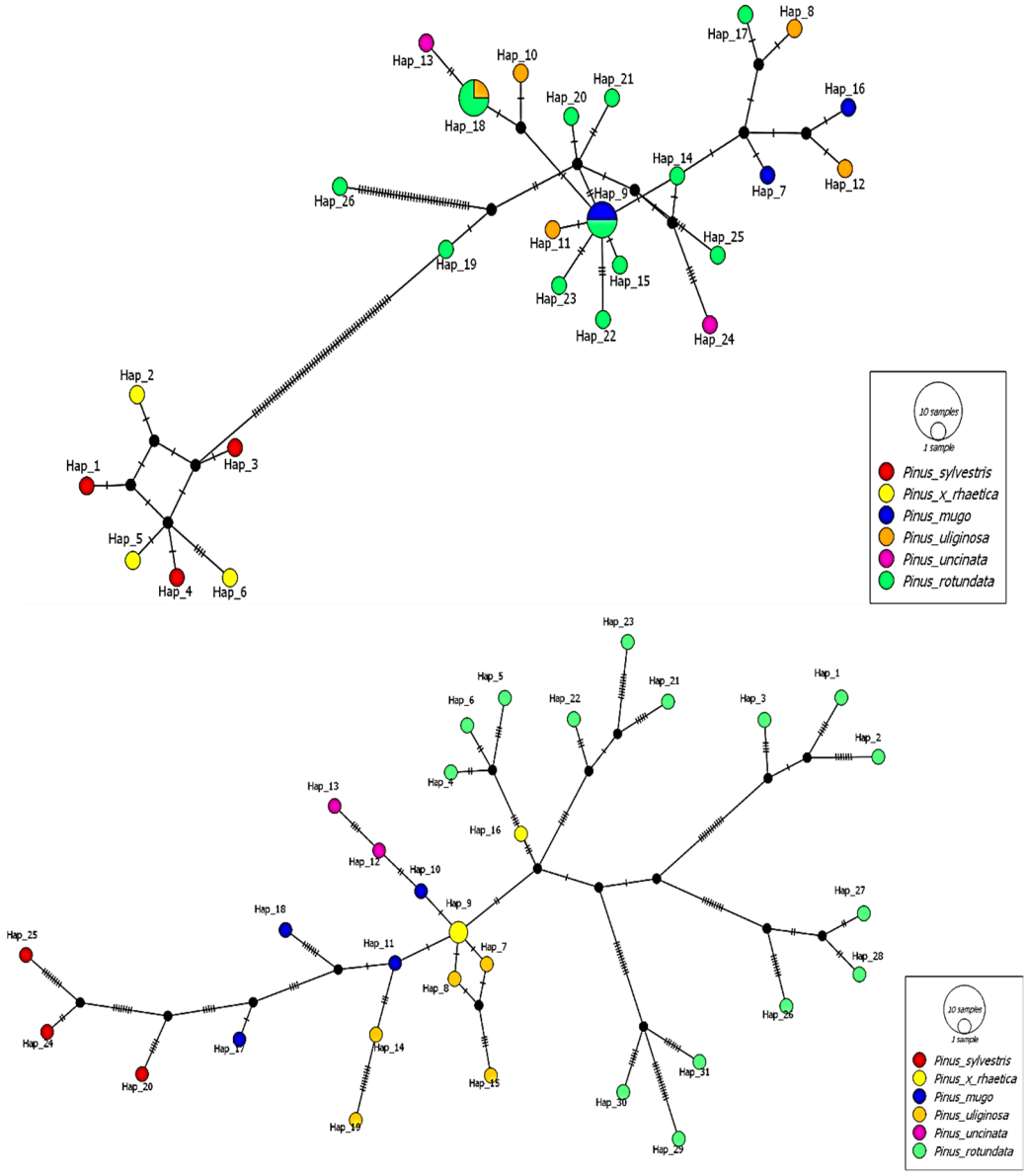


Figure 3. Median-joining haplotype network of the 32 individuals analyzed based on chloroplast protein coding genes (data set A) (**upper**) and based on mitochondrial protein coding genes (data set B) (**lower**).

Principal coordinate analysis (PCoA) performed on the divergence hotspots regions clearly distinguished taxa belonging to the *Pinus mugo* complex (larger group) from *Pinus sylvestris* and *Pinus* × *rhaetica* (smaller group) (Figure 4). Within each of these groups, two smaller subgroups can also be distinguished. Within this larger group, individuals of *Pinus uncinata* clearly stand out, while individuals of *Pinus mugo*, *Pinus rotundata*, and *Pinus uliginosa* form an extensive cloud. In a smaller group, differences between *Pinus sylvestris* and *Pinus* × *rhaetica* are visible, the latter being clearly closer to *Pinus sylvestris* than to taxa from the *Pinus mugo* complex.

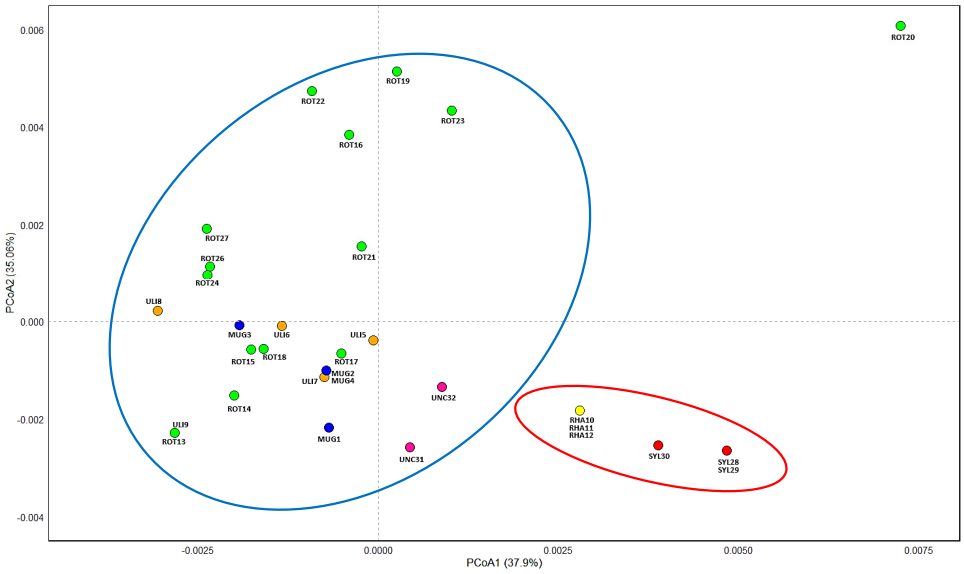


Figure 4. Results of PCoA analysis based on divergence hotspot regions. Abbreviations: MUG—*Pinus mugo*; UNC—*Pinus uncinata*; ULI—*Pinus uliginosa*; ROT—*Pinus rotundata*; RHA—*Pinus* × *rhaetica*; SYL—*Pinus sylvestris*.

The mean nucleotide divergence (Kimura 2-parameter) among the 13 samples ranged from 0.00095 to 0.00997 for the plastid intergenic region (data set C) (Figure 5). The largest genetic distance occurs between the *Pinus rotundata* from the "Rotmeer" Nature Reserve (Germany) and *Pinus sylvestris* from the Dendrological Garden, University of Life Sciences (US) (Poland). Conversely, the smallest genetic distance was observed between the Polish population of *Pinus uncinata* (US—Dendrological Garden, University of Life Sciences) and *Pinus rotundata* from the Czech Republic (NH—"Novohůrecká slat"). For divergence hotspot regions (data set G), the genetic distances range from 0.000981 to 0.008882 (Figure 5). The largest genetic distance occurs between the *P. rotundata* population from "Novohůrecká slat" (NH) (Czech Republic) and *Pinus sylvestris* from the Dendrological Garden, University of Life Sciences (US) (Poland). The smallest genetic distance is between the Polish populations of *Pinus sylvestris* (US—Dendrological Garden, University of Life Sciences and MO—Morasko, Poznań).

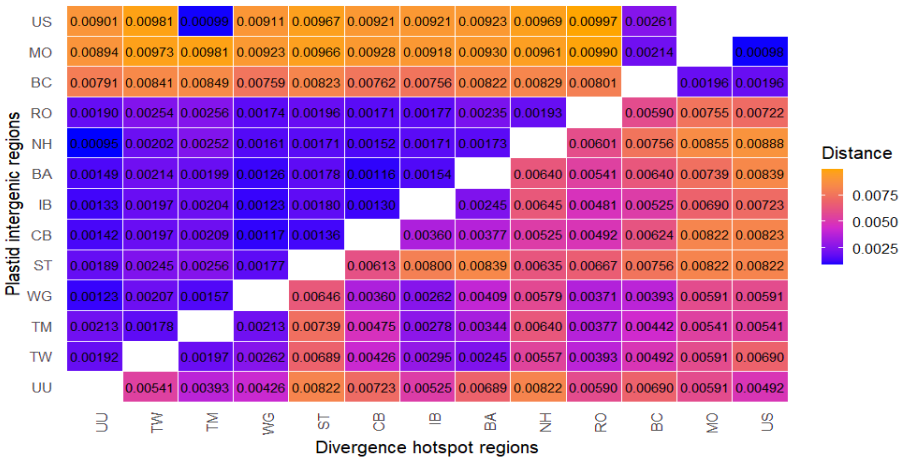


Figure 5. Heat map for pairwise genetic distances between populations. K2P genetic distance matrix on the plastid intergenic and divergence hotspot regions of 13 samples of the *Pinus mugo* complex, *Pinus* × *rhaetica*, and *Pinus sylvestris* taxa (the lower half of genetic distances belong to divergence hotspot regions and the upper half of distances to the plastid intergenic regions). Genetic distance is low to moderate: 0.0025 < x < 0.0050 (blue to purple); high: 0.0050 < x < 0.0075 (fandango to orange).

3. Discussion

The aims of this work were to generate a large amount of genomic data from multiple representatives of particular closely related taxa based on the genome skimming method and use them to characterize the genetic variability present in different data sets. We paid special attention to searching for a genetic data set in which one could try to identify interspecific differences, improve the phylogenetic resolution, and also describe the overall genetic variability present within individual populations and taxa. This task was a major challenge because previous attempts to find species-specific DNA markers for *Pinus mugo*, *Pinus uliginosa*, and *Pinus uncinata* based on eight selected chloroplast DNA barcoding regions, including traditional core barcodes (*matK* + *rbcL*), candidate (*trnH-psbA*), or even hypervariable regions (*ycf1* and *ycf2*) [36], showed that, in the case of these closely related taxa from the taxa of the *Pinus mugo* complex, these regions are not variable and there are no interspecific differences (at least in the analyzed area). Increasing the search area for differences from over five-thousand nucleotides (the work mentioned above) to almost one-hundred-and-twenty-thousand nucleotides, namely in the form of complete chloroplast genomes [38,48], allowed us to identify some potential species differences in a few regions. However, it is difficult to consider them as fully differentiating taxa because they were not tested on a larger number of individuals. Due to the complexity of the relationships in the *Pinus mugo* complex, there are few studies examining more than a few taxa in a multi-population system [19,49]. Unfortunately, without such a broader sample and knowledge of intraspecific variability, it is difficult to both search for and verify potential species-specific markers. Therefore, in this work, we focused on the analysis of many individuals from a given taxon in order to also determine the level of intraspecific variability and be able to compare it with the level of interspecific variability (and thus determine, among other things, the barcoding gap). We also wanted to understand the variability and relationships not only at the species level, but also at the population level.

Genome skimming is a shallow whole-genome sequencing technique that usually provides enough coverage for a complete assembly. In plants, this method often yields adequate coverage of high-copy regions, enabling the recovery of complete or partial organellar genomes and nuclear ribosomal DNA sequences. This method was used, among others, to recover the complete plastomes of *Theobroma* spp. [44], *Fargesia* [50], *Paris* [51], *Rhododendron* [52], *Panax* [46], *Areca* [53], *Piper*, as well as *Peperomia* [54] and *Salix* [55]. Numerous publications have highlighted the value of data obtained through genome skimming in relation to the valuable genomic resources archived in museums and herbaria [56,57]. Nevertheless, in complex plant groups like *Panax* [46] and *Araucaria* [45], the plastome and nrDNA data did not significantly enhance the discriminatory power. The recent diversification of several lineages of the *Panax bipinnatifidus* species complex may result in plant barcodes being shared with closely related species and thus not delimiting species boundaries [44].

Genome skimming proved to be extremely helpful in our studies, as it allowed us to obtain many very long regions representing the genomes of different organelles and to generate seven data covering: plastid coding genes; mitochondrial coding genes; plastid intergenic regions; nrDNA cistron; ITS; *matK* + *rbcL*; and divergence hotspot regions. The same method was recently used to generate sequence data also for *Cymbidium* [8] and *Rhododendron* [52]. In these studies, genetic diversity and species delimitation ability were analyzed using different combinations of DNA regions, including complete plastomes (superbarcoding approach), plastomes + nrDNA (ultrabarcoding approach), a mixture of barcoding regions in a narrower sense, i.e., nrDNA, ITS, *rbcL*, *matK*, *trnH-psbA* (core DNA barcode regions), and, in a broader sense: *matK* + *rbcL* + *trnH-psbA*, ITS + *matK* + *trnH-psbA*, and ITS + *matK* + *rbcL* + *trnH-psbA*. The level of observed genetic variation measured by the percentage of divergence was highly variable and differed significantly within individual taxa and regions. In *Rhododedron* species, the most variable region was the ITS region (17.05% divergence) before the combination of ITS + *matK* + *trnH-psbA* (15.88%) [52]. In turn, in *Cymbidium* species [8], the most variable region was the *matK* region (16.65%

divergence) before the combination of *rbcL* + *matK* + *trnH-psbA* (8.27% divergence). In *Panax* [46] and *Fargesia* [50] species, the ITS region was also the most variable (12.25% and 10.60% divergence values, respectively). It is worth noting that the length of the ITS region in all the species mentioned above is about 600 to 700 nt. These results confirm that the ITS is a very good candidate for an interspecific marker, at least in these plants.

The analysis of the ITS region in conifers has been, to date, significantly hampered by the size of the nuclear genome [58–60] and the length and complexity of the sequence itself [61,62]. Therefore, there are not many works using this region in the analyses of conifers [62–65]. However, the use of genome skimming in this study allowed us to determine and characterize for the first time the complete ITS and nrDNA sequence for representatives of the *Pinus mugo* complex.

Our results contradict previous literature reports of high ITS polymorphism in some species [46,50,52]. Both ITS and the whole nrDNA in the case of the analysis of taxa in the *Pinus mugo* complex were the regions with the lowest level of variability (0.104% and 0.127%, respectively) of all the data sets we analyzed. This is despite the fact that the ITS and nrDNA regions were significantly longer in the *Pinus mugo* complex representatives than in other taxa. The length of the ITS in the *Pinus mugo* complex was 2881 nucleotides, and the length of the nrDNA was 7849 nucleotides. Previous studies of ITS2, which is ten-times shorter than ITS1 [37], showed that ITS2 is quite good for identifying taxa in the *Pinaceae* family, but only for some genera. In the case of the *Pinus* genus and the *Pinus mugo* complex, this was not particularly the case. This region was simply too short and showed too little variability in this group of plants. It was therefore suggested that perhaps ITS1, as a longer fragment, would be better in this respect. Our results clearly indicate that neither ITS1 nor ITS2 is sufficiently variable to be used for the analysis and differentiation of closely related taxa in the *Pinus mugo* complex. Despite our observations, ITS2 still remains a very good, or even ideal, barcode for many other plant groups than the *Pinus* genus, as it has many advantages [50,66,67].

Compared to ITS, the nrDNA sequence had a slightly higher discrimination rate, probably due to the larger number of parsimony information sites. However, neither ITS nor nrDNA was able to distinguish all taxa. It therefore seems reasonable to continue searching for other nuclear regions. Finding them would certainly be very helpful for identifying hybrid individuals. Perhaps a good direction is the analysis of single copy genes. Genome skimming can be difficult to use for this purpose because only well-represented fragments are sequenced in this approach and there is not much pure nuclear DNA in the sample. Probably with very high genome coverage there would be a chance to capture single copy genes. Many studies indicate that single gene copies are an interesting tool [68,69].

Chloroplast core DNA coding regions (*matK* + *rbcL*) analyzed in this work as set F, similarly to ITS and nrDNA, were characterized by a low level of variability (0.269% divergence). Using a combination of these regions did not allow us to distinguish closely related taxa from the *Pinus mugo* complex, and this was an expected result after our previous analyses conducted in 2017 [36]. The reasons why these taxa cannot be distinguished using these DNA markers may be many, ranging from incomplete lineage sorting (ILS) to gene exchange (hybridization) or recent reticulate divergence/evolution [19]. In the *Pinaceae* family, chloroplast DNA is assumed to be inherited from the paternal line, although recent studies shed a slightly different light on this [70], which may also have influenced our results.

Our results show that intergenic plastid and hotspot divergence regions (*rps1*, *rps2*, *rps14*, *ycf3-psaA*, and *trnE-clpP*) discriminate between 33.3% and 50.0% of the analyzed taxa using the tree-based and distance-based methods. Hotspot divergence regions, despite having the shortest alignment length, had the highest evolutionary rate of all data sets. This was the only data set in the ASAP method that was able to discriminate any species. In contrast, mitochondrial coding genes showed greater discrimination power than plastid coding genes, up to tenfold. Not surprisingly, the phylogenetic trees obtained from these

data differ significantly. According to Wang [71], these discrepancies suggest a series of mtDNA capture events during past range shifts of pine species, indicating that vertical and horizontal inheritance have played a role in the evolution of mtDNA in *Pinus*.

Taxonomists have been using molecular methods for many years to differentiate species more effectively. DNA barcoding in this regard is still a valuable method. However, over time, instead of one or two core barcode regions, researchers have begun to analyze several or a dozen of them (multilocus DNA barcoding) to improve the efficiency of delimitation. Sometimes, a complete chloroplast genome sequence is also very helpful, proposed as a so-called superbarcode [72]. Recently, very satisfactory results have been obtained by using ultrabarcoding—a promising approach that solves many challenges associated with currently common DNA barcoding methods. Ultrabarcoding involves expanding from traditional DNA barcodes to entire plastomes and nuclear ribosomal DNA (nrDNA) sequences [44]. This approach generates a significant amount of data and, as a result, can provide information necessary to investigate variations below the species level. In the case of the *Pinus mugo* complex, ultrabarcoding is unlikely to be useful because nrDNA has poor discriminatory power. Although, of course, there are studies on other plant species where this approach is effective [44,51].

The correct identification and differentiation of species are of great importance not only for purely cognitive or scientific reasons, but also for practical reasons. Problems with the unambiguous identification of protected or endangered taxa translate directly into problems with the preparation, undertaking, and implementation of appropriate protective measures. Closely related taxa, which are characterized by very similar morphologies and a lack of clear species determinants, are usually grouped into larger taxonomic units called complexes. They also pose a great challenge for researchers due to the complex or unknown origin of individual taxa and the presence of many different synonymous names in the literature. The *Pinus mugo* complex is one of many such complexes, but several others are known and this is in the genus *Pinus* itself, e.g., [10,73,74]. Additionally, in the case of sympatric populations of closely related taxa, hybridization and introgression processes occur. The resulting hybrids and introgressants representing an intermediate phenotype between the habit of the parent taxa make it even more difficult to correctly identify individuals from individual taxa. Hence, it is so important to develop appropriate methods and markers that would facilitate the reliable and unambiguous identification of individuals, including hybrids, regardless of the morphological determinants that depend on environmental conditions or require a specific developmental phase.

The *Pinus mugo* complex includes closely related taxa that occur in the most important mountain ranges in Europe, such as the Pyrenees, the Alps, and the Carpathians [9]. These taxa fulfill very important ecological functions and are the basis of many mountain ecosystems. At the same time, discussions have been ongoing for many years on the taxonomic status, origin, and final number of taxa included in this complex. Some researchers distinguish one main species, *Pinus mugo*, with two subspecies, *Pinus mugo* ssb. *mugo* and *Pinus mugo* ssb. *uncinata*, which occur in the mountains of Central and Eastern Europe and Western Europe, respectively [9]. Others, however, distinguish three main species in this complex, i.e., *Pinus mugo*, *Pinus uncinata*, and *Pinus uliginosa*. The latter taxon—*Pinus uliginosa*—is one of the most enigmatic and controversial taxa from the *Pinus mugo* complex [48]. This pine was first described at a site in Wielki Torfowisko Batorowskie in Poland in 1837 [75]. In Poland, this species represents one of the four native species of pines (apart from Scots pine (*Pinus sylvestris* L.), dwarf mountain pine (*Pinus mugo*), and Swiss pine (*Pinus cembra*)) and occurs only at four sites, mainly peat bogs or wetlands. It is a protected species [76] and is threatened with extinction. The most serious threat to the continuity of this species is primarily the poor health condition of the trees, poor natural regeneration, and the risk of contamination of the gene pool of swamp pines due to the uncontrolled gene exchange with Scots pine and dwarf pine that occurs at sympatric sites [12,14].

Pinus uliginosa, *mugo*, and *uncinata* are characterized by high morphological variability on the one hand and genetic invariability on the other hand, which does not allow for their effective and unambiguous distinction [36]. Another problem is the identification of individuals of the protected peat bog pine and its postulated hybrids with dwarf mountain pine. Effective protection of the peat bog pine in Poland is additionally complicated by its legal status, because in the Regulation of the Minister of the Environment of October 9, 2014, the peat bog pine is listed under the name *Pinus* × *rhaetica* [76]. It is not clear whether the name *Pinus* × *rhaetica*, which refers to peat bog pine, is identical to the name *Pinus uliginosa*. The relationship between *Pinus uliginosa* and *Pinus* × *rhaetica* remains unclear. It is also not known whether *Pinus uliginosa* is a separate species or, as some researchers postulate, a hybrid of Scots pine (*Pinus sylvestris* L.) with dwarf mountain pine (*Pinus mugo* Turra (as *Pinus* × *rhaetica*)) or a hybrid of mountain pine (*Pinus uncinata* Rammond) with dwarf mountain pine (*Pinus mugo* Turra (*Pinus uncinata* × *Pinus mugo*)) [9]. The results of our research rather confirm the latter hypothesis. Moreover, the results of the research on Polish populations of the peat bog pine—*Pinus uliginosa* or *Pinus* × *rhaetica*—are difficult to directly relate to the results of the research from abroad, especially from the Czech Republic and Germany, because, in the scientific literature, peat bog pine is listed under the name *Pinus rotundata* [11,77]. Therefore, one of the goals of this work was to conduct a phylogenetic inference to confirm or exclude the hypothesis that *Pinus uliginosa* from Poland is the genetic equivalent of *Pinus rotundata* from the Czech Republic and Germany, and both names are synonyms referring to the same taxon.

There are many more questions regarding *Pinus uliginosa*/*Pinus rotundata*, e.g., did it arise independently as a hybrid at different times and in different places (peat bogs in Poland, Germany, and the Czech Republic), giving rise to a taxon with a similar habit and genetic background, due to the similar parental composition? Perhaps it arose in one place and then colonized further regions. Or, maybe, the currently observed populations are relicts and represent the remains of an ancestral species with a wider range. To answer these questions, it is necessary to conduct further research with geographically broader sampling.

The data obtained in this study indicate that *Pinus* × *rhaetica* is not identical to *Pinus uliginosa* or any other analyzed taxon belonging to the *Pinus mugo* complex. These results are in full accordance with the previously conducted research on *Pinus mugo* complex taxa discrimination using seed storage protein profiling [29]. Celiński et al. (2020) analyzed the patterns of seed total proteins and showed that each of the closely related taxon from this complex has a different and unique protein pattern. Moreover, based on the results obtained, *Pinus* × *rhaetica* is much closer to *Pinus sylvestris* than to *Pinus mugo* and *Pinus uliginosa* from the *Pinus mugo* complex. The distinctiveness of *Pinus* × *rhaetica* has also been demonstrated previously in cytological studies using C-banding methods and flow cytometric analysis [33]. This study showed that three closely related pines from the *Pinus mugo* complex (*Pinus mugo*, *Pinus uliginosa*, and *Pinus* × *rhaetica*) differ in some karyotypical features and DNA content. Celiński et al. (2019) therefore postulated not to use the names *Pinus* × *rhaetica* and *Pinus uliginosa* interchangeably as synonyms because these taxa differ from each other.

Our work also presents new data on the genetic diversity of bog pines from the Czech Republic and Germany. Knowledge regarding this subject is still relatively scarce because, in many studies on the *Pinus mugo* complex, *Pinus rotundata* is omitted, and researchers focus mainly on the *Pinus mugo*–*Pinus uncinata* system.

The revealed pattern of genetic variability based on the mitochondrial genome clearly indicates the very unique, local, and isolated character of the analyzed populations from the *Pinus mugo* complex. Attention is paid to both the peripheral location of the Czech and German populations of *Pinus rotundata* and the fact that the Polish populations of *Pinus uliginosa* show a different character. Such a result may indicate the relatively long spatial isolation and the complex historical background of individual populations, especially taking into account the phylogeographic structure of *Pinus mugo* or *Pinus uncinata* as putative parent taxa of *Pinus rotundata* and *Pinus ulignosa* [20].

4. Materials and Methods

4.1. Plant Material and DNA Extraction

In order to best represent the *Pinus mugo* complex, four taxa from the *Pinus mugo* complex were analyzed: *P. mugo, P. uncinata, P. uliginosa*, and *P. rotundata*. Each of these taxa was represented by several individuals from one to five populations. Three *Pinus sylvestris* and three *Pinus* × *rhaetica* individuals were also included in this study as the taxa most closely related to the *Pinus mugo* complex. A total of 32 pine-needle samples was sequenced. Details of individual specimens are presented in Table S1 and the habitats of selected taxa from the *Pinus mugo* complex are presented in Figure 6. Vouchers of the analyzed individuals were deposited in the Natural History Collections, Faculty of Biology, Adam Mickiewicz University, Poznań, Poland (Table S1).

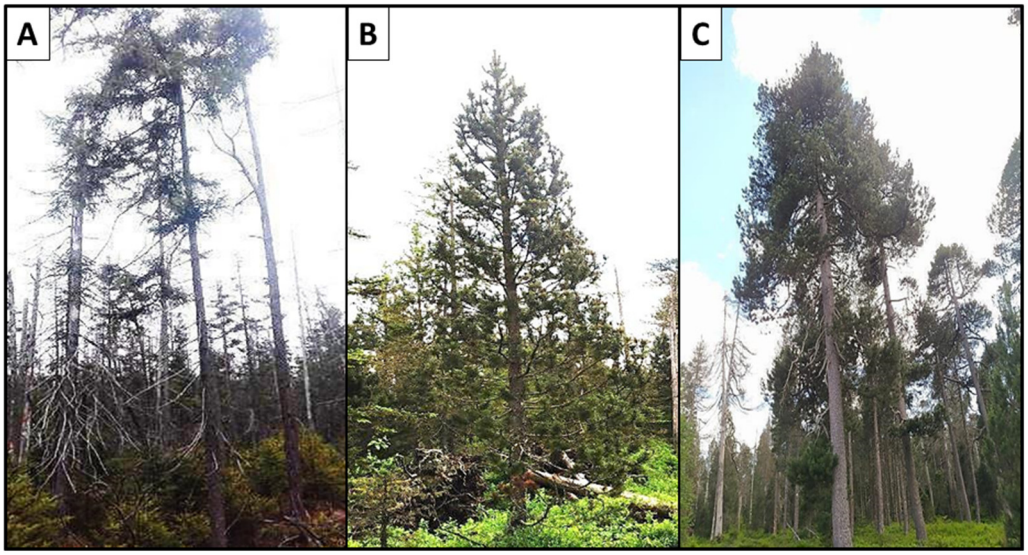


Figure 6. The habitats of selected taxa from the *Pinus mugo* complex. (**A**) *Pinus uliginosa*—"Wielkie Torfowisko Batorowskie" Nature Reserve (BA), Poland; (**B**) *Pinus rotundata*—"Novohůrecká slat" (NH), the Czech Republic; (**C**) *Pinus rotundata*—"Kirchspielwald-Ibacher Moos" Nature and Forest Reserve (IB), Germany.

Genomic DNA extraction was performed from 100 mg of fresh plant tissue (needles) using the Genomic Mini AX Plant Spin Kit (A&A Biotechnology, Gdynia, Poland). Sample DNA quality was assessed using the Agilent High-Sensitivity D1000 ScreenTape System (Agilent Technologies, Inc., Waldbronn, Germany). Until analysis, the isolated DNA was stored in a freezer at −20 °C.

4.2. DNA Sequencing, Assembly, and Annotation

Ion Torrent barcoded libraries with approximately 300 bp insert fragments were constructed using the Ion Xpress™ Plus Fragment Library Kit and Ion Xpress™ Barcode Adapters Kit (ThermoFisher Scientific, Waltham, MA, USA). Libraries were sequenced using the Ion GeneStudio™ S5 System (Thermo Fisher Scientific, Waltham, MA, USA). Adapters and low-quality reads were filtered using BBDuk Adapter/Quality Trimming V. 38.84 by Brian Bushnell implemented in Geneious Prime 2020.2.5 [78].

A genome skimming method was used to generate nucleotide sequences from the chloroplast, mitochondrial, and nuclear genomes for *Pinus mugo* complex taxa, as well as *Pinus sylvestris* and *Pinus* × *rhaetica* individuals. For this purpose, reads were assembled de novo using Geneious Assembler on default settings. Resulting contigs were mapped to the reference chloroplast genome of *Pinus sylvestris* (NC_035069.1), *Pinus taeda* mitochondrion (NC039746), and *Pinus sylvestris* nrDNA cistron (MT735327) using Bowtie2 v.2.4.5 [79]. Then, the complete protein coding genes and intergenic spacer regions were extracted from the resulting incomplete chloroplast genomes using Geneious Prime 2020.2.5 [78]. The

internal transcribed spacer sequences (ITS1, 5.8S, and ITS2) were extracted from nrDNA cistron in Geneious Prime 2020.2.5 [78].

4.3. Data Analysis

4.3.1. Genetic Characterization of Diversity in Data Sets

We constructed seven DNA data sets by concatenating different aligned regions, including: (A) the concatenated plastid coding genes; (B) the concatenated mitochondrial coding genes; (C) the plastid intergenic regions; (D) the complete nuclear ribosomal DNA (nrDNA); (E) the internal transcribed spacer (ITS) consisting of ITS1-5.8S-ITS2 sequences; (F) core barcode regions *matK* + *rbcL*; and (G) divergence hotspot regions (*rps1*, *rps2*, *rps14*, *ycf3-psaA*, and *trnE-clpP*). The alignment of each region was conducted using MAFFT v7.490 [80,81]. MEGA11 [82] was used to calculate the alignment length, number of variables, and parsimony informative sites, as well as the percentage of variable sites and percentage of parsimony informative sites among representatives of the *Pinus mugo* complex, *Pinus* × *rhaetica*, and *Pinus sylvestris*. The diversity thresholds, used to select variable regions, were calculated by the sum of the average and double the standard deviation [83].

4.3.2. Delimitation of Taxa

To evaluate the effectiveness of taxa discrimination, we performed analyses based on the created seven data sets using three different delimitation methods. The tree-based method was performed using the maximum likelihood method with the option of 1000 bootstrap replicates conducted in MEGA 11 [82]. To estimate the best nucleotide substitution model for each data set, the "Find Best Fit Substitution Model" option available in MEGA11 was used [84]. Phylogenetic trees were created for the 26 analyzed individuals belonging to the *Pinus mugo* complex, 3 individuals of *Pinus* × *rhaetica*, and 3 individuals of *Pinus sylvestris*. The distance-based method was performed using the pairwise distance calculated in MEGA 11 [82] using the Kimura 2-parameter (K2P) model. The taxa with the larger minimum interspecific distance than the maximum intraspecific distance were consider as successfully identified. We also used the ASAP (Assemble Species by Automatic Partitioning) web server available at https://bioinfo.mnhn.fr/abi/public/asap. The ASAP analysis was conducted using the Kimura K80 substitution model (ts/tv = 2.0). ASAP employs automatic partitioning algorithms to delineate species through hierarchical clustering based on sequence similarity [85]. A lower ASAP score indicates a better partition of the data set.

4.3.3. Genetic Diversity of Taxa and Populations

To assess the taxon genealogy evolutionary lineages of the analyzed individuals, haplotype networks were constructed based on the chloroplast and mitochondrial data sets using the median-joining approach available in PopArt [47]. Genetic relationships between samples at the taxon level were visualized using principal coordinate analysis (PCoA) based on K2P distances and divergence hotspot data in the ggplot2 R package. The Kimura two-parameter (K2P) model [86] with 1000 bootstrap replicates was used to quantify divergence between populations of the *Pinus mugo* complex, as implemented in MEGA11 [82]. Average K2P distances were calculated from pairwise comparisons of all sequences within and between populations.

5. Conclusions

Based on the obtained results, we can conclude that genome skimming is very useful for the analysis of closely related taxa, as in the case of the *Pinus mugo* complex, because it allows the generation of a large amount of sequence data. It allows the acquisition of many (sometimes difficult in traditional sequencing) regions of mitochondrial, chloroplast, and nuclear DNA. The ITS1 region for representatives of the *Pinus mugo* complex, fully sequenced for the first time, although very long, is not too variable and, similarly to the ITS2 region, does not allow for the effective delimitation of these closely related taxa. The

most promising data set for further discrimination analysis of the *Pinus mugo* complex is the divergence hotspot data set, which is nearly twenty-times more variable than traditional core chloroplast or nuclear DNA barcodes. *Pinus* × *rhaetica* should not be considered a synonym for *Pinus uliginosa* and *Pinus rotundata* because it differs significantly genetically from these two taxa and most likely represents a fixed hybrid or introgressant between *Pinus sylvestris* and *Pinus mugo*, as some previous studies have postulated. *Pinus uliginosa* and *Pinus rotundata*, despite the different scientific names, origins, and population histories, seem to have a similar genetic background, which is perhaps shaped by environmental pressures and the adaptation to the specific habitat they occupy. *Pinus uncinata* represents a clearly distinct clade within the *Pinus mugo* complex in most of the applied genetic data sets and analytical methods. The analysis of the pattern of genetic differentiation at the population level allows for a clear distinction of populations and taxa of the *Pinus mugo* complex from *Pinus sylvestris* and *Pinus* × *rhaetica*.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms251810178/s1.

Author Contributions: Conceptualization, J.S. and K.C.; methodology, J.S. and K.C.; software, J.S.; validation, J.S. and K.C.; formal analysis, J.S.; investigation, J.S.; resources, J.S.; data curation, J.S.; writing—original draft preparation, J.S. and K.C.; writing—review and editing, J.S. and K.C.; visualization, J.S.; supervision, K.C.; project administration, J.S. All authors have read and agreed to the published version of the manuscript.

Funding: Research paper financed from the budget for science in 2018–2020, as a research project under the "Diamond Grant" program No. DI2017003147.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article/Supplementary Materials. Further inquiries can be directed to the corresponding author/s.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Wang, J.; McLenachan, P.A.; Biggs, P.J.; Winder, L.H.; Schoenfeld, B.I.K.; Narayan, V.V.; Phiri, B.J.; Lockhart, P.J. Environmental bio-monitoring with high-throughput sequencing. *Briefings Bioinform.* **2013**, *14*, 575–588. [CrossRef] [PubMed]
2. Shafer, A.B.A.; Northrup, J.M.; Wikelski, M.; Wittemyer, G.; Wolf, J.B.W. Forecasting Ecological Genomics: High-Tech Animal Instrumentation Meets High-Throughput Sequencing. *PLoS Biol.* **2016**, *14*, e1002350. [CrossRef] [PubMed]
3. Titcomb, G.C.; Jerde, C.L.; Young, H.S. High-Throughput Sequencing for Understanding the Ecology of Emerging Infectious Diseases at the Wildlife-Human Interface. *Front. Ecol. Evol.* **2019**, *7*, 126. [CrossRef]
4. Ríos-Castro, R.; Romero, A.; Aranguren, R.; Pallavicini, A.; Banchi, E.; Novoa, B.; Figueras, A. High-Throughput Sequencing of Environmental DNA as a Tool for Monitoring Eukaryotic Communities and Potential Pathogens in a Coastal Upwelling Ecosystem. *Front. Veter-Sci.* **2021**, *8*, 765606. [CrossRef]
5. Wang, S.; Schneider, D.; Hartke, T.R.; Ballauff, J.; Moura, C.C.d.M.; Schulz, G.; Li, Z.; Polle, A.; Daniel, R.; Gailing, O.; et al. Optimising high-throughput sequencing data analysis, from gene database selection to the analysis of compositional data: A case study on tropical soil nematodes. *Front. Ecol. Evol.* **2024**, *12*, 1168288. [CrossRef]
6. Turner, B.; Paun, O.; Munzinger, J.; Chase, M.W.; Samuel, R. Sequencing of whole plastid genomes and nuclear ribosomal DNA of *Diospyros* species (Ebenaceae) endemic to New Caledonia: Many species, little divergence. *Ann. Bot.* **2016**, *117*, 1175–1185. [CrossRef]
7. Li, Y.; Sylvester, S.P.; Li, M.; Zhang, C.; Li, X.; Duan, Y.; Wang, X. The Complete Plastid Genome of *Magnolia zenii* and Genetic Comparison to Magnoliaceae species. *Molecules* **2019**, *24*, 261. [CrossRef] [PubMed]
8. Zhang, L.; Huang, Y.; Huang, J.; Ya, J.; Zhe, M.; Zeng, C.; Zhang, Z.; Zhang, S.; Li, D.; Li, H.; et al. DNA barcoding of *Cymbidium* by genome skimming: Call for next-generation nuclear barcodes. *Mol. Ecol. Resour.* **2022**, *23*, 424–439. [CrossRef]
9. Hamerník, J.; Musil, I. The *Pinus mugo* complex—Its structuring and general overview of the used nomenclature. *J. For. Sci.* **2007**, *53*, 253–266. [CrossRef]

10. Christensen, K.L. Taxonomic revision of the *Pinus mugo* complex and *P. rhaetica* (*P. mugo* sylvestris) (*Pinaceae*). *Nord. J. Bot.* **1987**, *7*, 383–408. [CrossRef]
11. Bastl, M.; Burian, M.; Kučera, J.; Prach, K.; Rektoris, L.; Štech, M. Central European Pine Bogs Change along an Altitudinal Gradient. *Preslia* **2008**, *80*, 349–363.
12. Wachowiak, W.; Celiński, K.; Prus-Głowacki, W. Evidence of natural reciprocal hybridisation between *Pinus uliginosa* and *P. sylvestris* in the sympatric population of the species. *Flora-Morphol. Distrib. Funct. Ecol. Plants* **2005**, *200*, 563–568. [CrossRef]
13. Wachowiak, W.; Prus-Głowacki, W. Hybridisation processes in sympatric populations of pines *Pinus sylvestris* L., *P. mugo* Turra and *P. uliginosa* Neumann. *Plant Syst. Evol.* **2007**, *271*, 29–40. [CrossRef]
14. Lewandowski, A.; Wiśniewska, M. Short Note: Crossability Between Pinus uliginosa and its Putative Parental Species *Pinus sylvestris* and *Pinus mugo*. *Silvae Genet.* **2006**, *55*, 52–54. [CrossRef]
15. Danusevičius, D.; Marozas, V.; Brazaitis, G.; Petrokas, R.; Christensen, K.I. Spontaneous Hybridization between *Pinus mugo* and *Pinus sylvestris* at the Lithuanian Seaside: A Morphological Survey. *Sci. World J.* **2012**, *2012*, 172407. [CrossRef]
16. Kormutak, A.; Demankova, B.; Gömöry, D. Spontaneous Hybridization between *Pinus sylvestris* L. and *P. mugo* Turra in Slovakia. *Silvae Genet.* **2008**, *57*, 76–82. [CrossRef]
17. Wachowiak, W.; Żukowska, W.B.; Wójkiewicz, B.; Cavers, S.; Litkowiec, M. Hybridization in contact zone between temperate European pine species. *Tree Genet. Genomes* **2016**, *12*, 48. [CrossRef]
18. Sobierajska, K.; Wachowiak, W.; Zaborowska, J.; Łabiszak, B.; Wójkiewicz, B.; Sękiewicz, M.; Jasińska, A.K.; Sękiewicz, K.; Boratyńska, K.; Marcysiak, K.; et al. Genetic Consequences of Hybridization in Relict Isolated Trees *Pinus sylvestris* and the *Pinus mugo* Complex. *Forests* **2020**, *11*, 1086. [CrossRef]
19. Łabiszak, B.; Wachowiak, W. Molecular Signatures of Reticulate Evolution within the Complex of European Pine Taxa. *Forests* **2021**, *12*, 489. [CrossRef]
20. Zaborowska, J.; Łabiszak, B.; Wachowiak, W. Population history of European mountain pines *Pinus mugo* and *Pinus uncinata* revealed by mitochondrial DNA markers. *J. Syst. Evol.* **2019**, *58*, 474–486. [CrossRef]
21. Staszkiewicz, J.; Tyszkiewicz, M. Variability of natural hybrids of *Pinus sylvestris* L. × *P. mugo* Turra (= *P.* × *rotundata* Link) in south-western Poland and in selected localities in Bohemia and Moravia. *Fragm. Florist. Et Geobot. Pol.* **1972**, *18*, 173–191.
22. Boratyńska, K.; Bobowicz, M.A. *Pinus uncinata* Ramond Taxonomy Based on Needle Characters. *Plant Syst. Evol.* **2001**, *227*, 183–194. [CrossRef]
23. Boratyńska, K.; Boratyński, A.; Lewandowski, A. Morphology of *Pinus uliginosa* (*Pinaceae*) needles from populations exposed to and isolated from the direct influence of *Pinus sylvestris*. *Bot. J. Linn. Soc.* **2003**, *142*, 83–91. [CrossRef]
24. 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. [CrossRef]
25. Boratyńska, K.; Boratyński, A. Taxonomic differences among closely related pines *Pinus sylvestris*, *P. mugo*, *P. uncinata*, *P. rotundata* and *P. uliginosa* as revealed in needle sclerenchyma cells. *Flora* **2007**, *202*, 555–569. [CrossRef]
26. Prus-Głowacki, W.; Bajus, E.; Ratyńska, H. Taxonomic position of *Pinus uliginosa* Neumann as related to other taxa of *Pinus mugo* complex. *Acta Soc. Bot. Pol.* **1998**, *67*, 269–274. [CrossRef]
27. Lewandowski, A.; Boratyński, 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]
28. Siedlewska, A.; Prus-Głowacki, W. Genetic structure and taxonomic position of *Pinus uliginosa* Neumann population from Wielkie Torfowisko Batorowskie in Stołowe Mts. (locus classicus). *Acta Soc. Bot. Pol.* **1995**, *64*, 51–58. [CrossRef]
29. Celiński, K.; Sokołowska, J.; Zemleduch-Barylska, A.; Kuna, R.; Kijak, H.; Staszak, A.M.; Wojnicka-Półtorak, A.; Chudzińska, E. Seed Total Protein Profiling in Discrimination of Closely Related Pines: Evidence from the *Pinus mugo* Complex. *Plants* **2020**, *9*, 872. [CrossRef]
30. Bonikowski, R.; Celiński, K.; Wojnicka-Półtorak, A.; Maliński, T. Composition of Essential Oils Isolated from the Needles of *Pinus uncinata* and *P. uliginosa* Grown in Poland. *Nat. Prod. Commun.* **2015**, *10*, 371–373. [CrossRef]
31. Celiński, K.; Bonikowski, R.; Wojnicka-Półtorak, A.; Chudzińska, E.; Maliński, T. Volatiles as Chemosystematic Markers for Distinguishing Closely Related Species within the *Pinus mugo* Complex. *Chem. Biodivers.* **2015**, *12*, 1208–1213. [CrossRef] [PubMed]
32. Bogunić, F.; Siljak-Yakovlev, S.; Muratović, E.; Pustahija, F.; Medjedović, S. Molecular cytogenetics and flow cytometry reveal conserved genome organization in *Pinus mugo* and *P. uncinata*. *Ann. For. Sci.* **2011**, *68*, 179–187. [CrossRef]
33. Celiński, K.; Chudzińska, E.; Gmur, A.; Piosik, Ł.; Wojnicka-Półtorak, A. Cytological characterization of three closely related pines—*Pinus mugo*, *P. uliginosa* and P. × rhaetica from the *Pinus mugo* complex (*Pinaceae*). *Biologia* **2019**, *74*, 751–756. [CrossRef]
34. Celiński, K.; Zbránková, V.; Wojnicka-Półtorak, A.; Chudzińska, E. Biogeography and evolutionary factors determine genetic differentiation of *Pinus mugo* (Turra) in the Tatra Mountains (Central Europe). *J. Mt. Sci.* **2015**, *12*, 549–557. [CrossRef]
35. Celiński, K.; Pawlaczyk, E.M.; Wojnicka-Półtorak, A.; Chudzińska, E.; Prus-Głowacki, W. Cross-species amplification and characterization of microsatellite loci in *Pinus mugo* Turra. *Biologia* **2013**, *68*, 621–626. [CrossRef]
36. Celiński, K.; Kijak, H.; Wojnicka-Półtorak, A.; Buczkowska-Chmielewska, K.; Sokołowska, J.; Chudzińska, E. Effectiveness of the DNA barcoding approach for closely related conifers discrimination: A case study of the *Pinus mugo* complex. *Comptes Rendus Biol.* **2017**, *340*, 339–348. [CrossRef]

37. Sokołowska, J.; Fuchs, H.; Celiński, K. Assessment of ITS2 Region Relevance for Taxa Discrimination and Phylogenetic Inference among *Pinaceae*. *Plants* **2022**, *11*, 1078. [CrossRef]
38. Sokołowska, J.; Fuchs, H.; Celiński, K. New Insight into Taxonomy of European Mountain Pines, *Pinus mugo* Complex, Based on Complete Chloroplast Genomes Sequencing. *Plants* **2021**, *10*, 1331. [CrossRef]
39. Straub, S.C.K.; Parks, M.; Weitemier, K.; Fishbein, M.; Cronn, R.C.; Liston, A. Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. *Am. J. Bot.* **2012**, *99*, 349–364. [CrossRef]
40. Dodsworth, S. Genome skimming for next-generation biodiversity analysis. *Trends Plant Sci.* **2015**, *20*, 525–527. [CrossRef]
41. Weitemier, K.; Straub, S.C.K.; Cronn, R.C.; Fishbein, M.; Schmickl, R.; McDonnell, A.; Liston, A. Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics. *Appl. Plant Sci.* **2014**, *2*, 1400042. [CrossRef] [PubMed]
42. Nevill, P.G.; Zhong, X.; Tonti-Filippini, J.; Byrne, M.; Hislop, M.; Thiele, K.; van Leeuwen, S.; Boykin, L.M.; Small, I. Large scale genome skimming from herbarium material for accurate plant identification and phylogenomics. *Plant Methods* **2020**, *16*, 1. [CrossRef]
43. Meena, R.K.; Negi, N.; Uniyal, N.; Bhandari, M.S.; Sharma, R.; Ginwal, H.S. Genome skimming-based STMS marker discovery and its validation in temperate hill bamboo *Drepanostachyum falcatum*. *J. Genet.* **2021**, *100*, 28. [CrossRef] [PubMed]
44. Kane, N.; Sveinsson, S.; Dempewolf, H.; Yang, J.Y.; Zhang, D.; Engels, J.M.M.; Cronk, Q. Ultra-barcoding in cacao (*Theobroma* spp.; Malvaceae) using whole chloroplast genomes and nuclear ribosomal DNA. *Am. J. Bot.* **2012**, *99*, 320–329. [CrossRef]
45. Ruhsam, M.; Rai, H.S.; Mathews, S.; Ross, T.G.; Graham, S.W.; Raubeson, L.A.; Mei, W.; Thomas, P.I.; Gardner, M.F.; Ennos, R.A.; et al. Does complete plastid genome sequencing improve species discrimination and phylogenetic resolution in *Araucaria*? *Mol. Ecol. Resour.* **2015**, *15*, 1067–1078. [CrossRef]
46. Ji, Y.; Liu, C.; Yang, Z.; Yang, L.; He, Z.; Wang, H.; Yang, J.; Yi, T. Testing and using complete plastomes and ribosomal DNA sequences as the next generation DNA barcodes in *Panax* (Araliaceae). *Mol. Ecol. Resour.* **2019**, *19*, 1333–1345. [CrossRef] [PubMed]
47. Leigh, J.W.; Bryant, D. Popart: Full-feature software for haplotype network construction. *Methods Ecol. Evol.* **2015**, *6*, 1110–1116. [CrossRef]
48. Celiński, K.; Kijak, H.; Barylski, J.; Grabsztunowicz, M.; Wojnicka-Półtorak, A.; Chudzińska, E. Characterization of the complete chloroplast genome of *Pinus uliginosa* (Neumann) from the *Pinus mugo* complex. *Conserv. Genet. Resour.* **2016**, *9*, 209–212. [CrossRef]
49. Heuertz, M.; Teufel, J.; González-Martínez, S.C.; Soto, A.; Fady, B.; Alía, 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]
50. Lv, S.-Y.; Ye, X.-Y.; Li, Z.-H.; Ma, P.-F.; Li, D.-Z. Testing complete plastomes and nuclear ribosomal DNA sequences for species identification in a taxonomically difficult bamboo genus Fargesia. *Plant Divers.* **2023**, *45*, 147–155. [CrossRef]
51. Ji, Y.; Liu, C.; Yang, J.; Jin, L.; Yang, Z.; Yang, J.-B. Ultra-Barcoding Discovers a Cryptic Species in Paris yunnanensis (Melanthiaceae), a Medicinally Important Plant. *Front. Plant Sci.* **2020**, *11*, 411. [CrossRef] [PubMed]
52. Fu, C.; Mo, Z.; Yang, J.; Cai, J.; Ye, L.; Zou, J.; Qin, H.; Zheng, W.; Hollingsworth, P.M.; Li, D.; et al. Testing genome skimming for species discrimination in the large and taxonomically difficult genus *Rhododendron*. *Mol. Ecol. Resour.* **2021**, *22*, 404–414. [CrossRef] [PubMed]
53. Raimondeau, P.; Manzi, S.; Brucato, N.; Kinipi, C.; Leavesley, M.; Ricaut, F.-X.; Besnard, G. Genome skims analysis of betel palms (*Areca* spp., Arecaceae) and development of a profiling method to assess their plastome diversity. *Gene* **2021**, *800*, 145845. [CrossRef]
54. Simmonds, S.E.; Smith, J.F.; Davidson, C.; Buerki, S. Phylogenetics and comparative plastome genomics of two of the largest genera of angiosperms, *Piper* and *Peperomia* (Piperaceae). *Mol. Phylogenetics Evol.* **2021**, *163*, 107229. [CrossRef]
55. Wagner, N.D.; Volf, M.; Hörandl, E. Highly Diverse Shrub Willows (*Salix* L.) Share Highly Similar Plastomes. *Front. Plant Sci.* **2021**, *12*, 662715. [CrossRef]
56. Bakker, F.T. Herbarium genomics: Skimming and plastomics from archival specimens. *Webbia* **2017**, *72*, 35–45. [CrossRef]
57. Zeng, C.-X.; Hollingsworth, P.M.; Yang, J.; He, Z.-S.; Zhang, Z.-R.; Li, D.-Z.; Yang, J.-B. Genome skimming herbarium specimens for DNA barcoding and phylogenomics. *Plant Methods* **2018**, *14*, 43. [CrossRef]
58. Morse, A.M.; Peterson, D.G.; Islam-Faridi, M.N.; Smith, K.E.; Magbanua, Z.; Garcia, S.A.; Kubisiak, T.L.; Amerson, H.V.; Carlson, J.E.; Nelson, C.D.; et al. Evolution of Genome Size and Complexity in Pinus. *PLoS ONE* **2009**, *4*, e4332. [CrossRef] [PubMed]
59. Zonneveld, B.J.M. Conifer genome sizes of 172 species, covering 64 of 67 genera, range from 8 to 72 picogram. *Nord. J. Bot.* **2012**, *30*, 490–502. [CrossRef]
60. De La Torre, A.R.; Birol, I.; Bousquet, J.; Ingvarsson, P.K.; Jansson, S.; Jones, S.J.M.; Keeling, C.I.; MacKay, J.; Nilsson, O.; Ritland, K.; et al. Insights into Conifer Giga-Genomes. *Plant Physiol.* **2014**, *166*, 1724–1732. [CrossRef]
61. Álvarez, I.; Wendel, J. Ribosomal ITS sequences and plant phylogenetic inference. *Mol. Phylogenetics Evol.* **2003**, *29*, 417–434. [CrossRef]
62. Liston, A.; Robinson, W.A.; Oliphant, J.M.; Alvarez-Buylla, E.R. Length Variation in the Nuclear Ribosomal DNA Internal Transcribed Spacer Region of Non-Flowering Seed Plants. *Syst. Bot.* **1996**, *21*, 109–120. [CrossRef]
63. Marrocco, R.; Gelati, M.T.; Maggini, F. Nucleotide sequence of the internal transcribed spacers and 5.8s region of ribosomal DNA in *Pinus pinea* L. *DNA Seq.* **1996**, *6*, 175–177. [CrossRef]

64. Gernandt, D.S.; Liston, A. Internal transcribed spacer region evolution in *Larix* and *Pseudotsuga* (*Pinaceae*). *Am. J. Bot.* **1999**, *86*, 711–723. [CrossRef]
65. Maggini, F.; Frediani, M.; Gelati, M.T. Nucleotide Sequence of the Internal Transcribed Spacers of Ribosomal DNA in *Picea Abies* Karst. *DNA Seq.* **2000**, *11*, 87–89. [CrossRef]
66. Liu, Z.; Zeng, X.; Yang, D.; Chu, G.; Yuan, Z.; Chen, S. Applying DNA barcodes for identification of plant species in the family Araliaceae. *Gene* **2012**, *499*, 76–80. [CrossRef]
67. Gao, T.; Sun, Z.; Yao, H.; Song, J.; Zhu, Y.; Ma, X.; Chen, S. Identification of Fabaceae plants using the DNA barcode matK. *Planta Medica* **2011**, *77*, 92–94. [CrossRef]
68. Liu, B.; Ma, Z.; Ren, C.; Hodel, R.G.J.; Sun, M.; Liu, X.; Liu, G.; Hong, D.; Zimmer, E.A.; Wen, J. Capturing single-copy nuclear genes, organellar genomes, and nuclear ribosomal DNA from deep genome skimming data for plant phylogenetics: A case study in Vitaceae. *J. Syst. Evol.* **2021**, *59*, 1124–1138. [CrossRef]
69. Su, N.; Liu, B.-B.; Wang, J.-R.; Tong, R.-C.; Ren, C.; Chang, Z.-Y.; Zhao, L.; Potter, D.; Wen, J. On the Species Delimitation of the *Maddenia* Group of *Prunus* (Rosaceae): Evidence From Plastome and Nuclear Sequences and Morphology. *Front. Plant Sci.* **2021**, *12*, 743643. [CrossRef]
70. Kormutak, A.; Galgoci, M.; Sukenikova, D.; Bolecek, P.; Libantova, J.; Gőmőry, D. Maternal inheritance of chloroplast DNA in *Pinus mugo* Turra: A case study of *Pinus mugo* × *Pinus sylvestris* crossing. *Plant Syst. Evol.* **2017**, *304*, 71–76. [CrossRef]
71. Wang, B.; Wang, X.-R. Mitochondrial DNA capture and divergence in Pinus provide new insights into the evolution of the genus. *Mol. Phylogenetics Evol.* **2014**, *80*, 20–30. [CrossRef]
72. Kane, N.C.; Cronk, Q. Botany without borders: Barcoding in focus. *Mol. Ecol.* **2008**, *17*, 5175–5176. [CrossRef]
73. Bucci, G.; Anzidei, M.; Madaghiele, A.; Vendramin, G.G. Detection of haplotypic variation and natural hybridization in *halepensis*-complex pine species using chloroplast simple sequence repeat (SSR) markers. *Mol. Ecol.* **1998**, *7*, 1633–1643. [CrossRef]
74. Businský, R.; Frantík, T.; Vít, P. Morphological evaluation of the Pinus kesiya complex (*Pinaceae*). *Plant Syst. Evol.* **2013**, *300*, 273–285. [CrossRef]
75. Neumann, C. Über Eine Auf Den Seefeldern Bei Reinerz u. Einigen Ähnlichen Gebirgsmooren Der Königl. Oberförsterei Karlsberg in Der Graftschaft Glatz Vorkommende Noch Unbeschrieben Form Der Gattung Pinus. *Jahresber Schlesische Ges. Für Vaterländische Kult.* **1837**, *11*, 52–57.
76. Rozporządzenie Ministra Środowiska Z Dnia 9 Października 2014 R. W Sprawie Ochrony Gatunkowej Roślin. W: Dz.U. 2014 Poz. 1409.; Poland. Internetowy System Aktów Prawnych. Available online: https://isap.sejm.gov.pl/isap.nsf (accessed on 20 July 2024).
77. Bastl, M.; Štechová, T.; Prach, K. Effect of Disturbance on the Vegetation of Peat Bogs with Pinus Rotundata in the Třeboň Basin, Czech Republic. *Preslia* **2009**, *81*, 105–117.
78. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649. [CrossRef]
79. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [CrossRef]
80. Kazutaka, K.; Misakwa, K.; Kei-ichi, K.; Miyata, T. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **2002**, *30*, 3059–3066.
81. Katoh, K.; Standley, D.M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef]
82. Tamura, K.; Stecher, G.; Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.* **2021**, *38*, 3022–3027. [CrossRef] [PubMed]
83. Bi, Y.; Zhang, M.-F.; Xue, J.; Dong, R.; Du, Y.-P.; Zhang, X.-H. Chloroplast genomic resources for phylogeny and DNA barcoding: A case study on Fritillaria. *Sci. Rep.* **2018**, *8*, 1184. [CrossRef] [PubMed]
84. Nei, M.; Kumar, S. *Molecular Evolution and Phylogenetics*; Oxford University Press (OUP): Oxford, UK, 2000; pp. 3–295.
85. Puillandre, N.; Brouillet, S.; Achaz, G. ASAP: Assemble species by automatic partitioning. *Mol. Ecol. Resour.* **2020**, *21*, 609–620. [CrossRef]
86. Kimura, M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **1980**, *16*, 111–120. [CrossRef]

Table S1. Plant materials used in this study.

Taxon	Location	Latitude (N), Longitude (E)	Population code	Sample size (individual code)	Country	Voucher
Pinus mugo	Tatra National Park	49°14′30 N, 20°00′15 E	TM	2 (1-2)	Poland	POZG-V-0150537 POZG-V-0150538
Pinus mugo	Tatra National Park	49°16′06 N, 20°02′40E	TW	2 (3-4)	Poland	POZG-V-0150539 POZG-V-0150540
Pinus uliginosa	"Torfowisko pod Węglińcem" Nature Reserve	51°17′36 N, 15°13′37 E	WG	3 (5-7)	Poland	POZG-V-0150541 POZG-V-0150542 POZG-V-0150543
Pinus uliginosa	"Wielkie Torfowisko Batorowskie" Nature Reserve	50°27′30 N, 16°22′57 E	BA	2 (8-9)	Poland	POZG-V-0150544 POZG-V-0150545
Pinus × rhaetica	"Bór nad Czerwonem" Nature Reserve	49° 27′30N 20° 02′25 E	BC	3 (10-12)	Poland	POZG-V-0150546 POZG-V-0150547 POZG-V-0150548
Pinus rotundata	"Kirchspielwald-Ibacher Moos" Nature and Forest Reserve	47°43′38 N, 8°03′01 E	IB	3 (13-15)	Germany	POZG-V-0150549 POZG-V-0150550 POZG-V-0150551
Pinus rotundata	"Rotmeer" Nature Reserve	47°51′53 N, 8°06′00 E	RO	3 (16-18)	Germany	POZG-V-0150552 POZG-V-0150553 POZG-V-0150554
Pinus rotundata	"Steerenmoos", Fauna-flora-habitat area: "Valley of Schwarza, Mettma, Schlücht and the Steina"	47°48′21 N, 8°12′00 E	ST	3 (19-21)	Germany	POZG-V-0150555 POZG-V-0150556 POZG-V-0150557
Pinus rotundata	"Novohůrecká slať", Šumava National Park	49°09′18 N, 13°19′46 E	NH	3 (22-24)	Czech Republic	POZG-V-0150558 POZG-V-0150559 POZG-V-0150560
Pinus rotundata	"Červené blato" National Nature Reserve	48°51′26 N, 14°48′08 E	CB	3 (25-27)	Czech Republic	POZG-V-0150561 POZG-V-0150562 POZG-V-0150563
Pinus sylvestris	Dendrological Garden, University of Life Sciences	52°25′32 N, 16°53′39 E	US	1 (28)	Poland	POZG-V-0150564
Pinus sylvestris	Morasko, Poznań	52°28′02 N, 16°55′30 E	MO	2 (29-30)	Poland	POZG-V-0150565 POZG-V-0150566
Pinus uncinata	Dendrological Garden, University of Life Sciences, Poznań	52°25′32 N, 16°53′39 E	UU	2 (31-32)	Poland	POZG-V-0150567 POZG-V-0150568

Table S2. List of sequence names used in datasets A, B, C, G.

No.	Plastid coding genes (data set A)	Mitochondrial coding genes (data set B)	Intergenic regions of the plastid (data set C)	Divergence hotspot regions (data set G)
1	*accD*	*ATP4*	*accD-psaI*	*rps1*
2	*atpA*	*ATP6*	*rps2-atpI*	*rps2*
3	*atpB*	*ATP8*	*clpP-rps12*	*rps14*
4	*atpF*	*ATP9*	*petB-petD*	*ycf3-psaA*
5	*atpH*	*ccmB*	*psaM-trnS*	*trnE-clpP*
6	*atpI*	*ccmC*	*psbJ-petA*	
7	*ccsA*	*cox3*	*psbL-psbJ*	
8	*cemA*	*NAD3*	*trnS-psbZ*	
9	*chlB*	*NAD4L*	*atpB-rbcL*	
10	*chlL*	*NAD6*	*rpoB-trnC*	
11	*chlN*	*rpl5*	*rrn4.5-rrn5*	
12	*matK*	*rps1*	*trnC-petN*	
13	*petA*	*rps2*	*trnD-psbM*	
14	*petB*	*rps14*	*trnE-clpP*	
15	*petD*	*rps19*	*trnL-trnF*	
16	*petG*		*trnM-ndhC*	
17	*petN*		*trnS-psaM*	
18	*psaB*		*ycf3-psaA*	
19	*psaC*			
20	*psbB*			
21	*psbC*			
22	*psbD*			
23	*psbE*			
24	*psbF*			
25	*psbH*			
26	*psbI*			
27	*psbJ*			
28	*psbK*			
29	*psbL*			
30	*psbM*			
31	*psbN*			
32	*psbT*			
33	*psbZ*			
34	*rbcL*			
35	*rpl2*			
36	*rpl14*			
37	*rpl16*			
38	*rpl20*			
39	*rpl22*			
40	*rpl23*			
41	*rpl33*			
42	*rpl36*			
43	*rpoA*			
44	*rpoB*			
45	*rpoC1*			
46	*rpoC2*			
47	*rps2*			
48	*rps3*			
49	*rps4*			

50	*rps7*
51	*rps11*
52	*rps14*
53	*rps15*
54	*rps18*
55	*rps19*
56	*ycf3*
57	*ycf4*

Table S3. Comparison of characteristics of seven data sets in *Pinus sylvestris* and *Pinus* × *rhaetica.*

Alignment	Data set code	Length (bp)	Variable sites (% divergence)	Parsimony informative sites (%)
Plastid coding genes	A	45,467	11 (0.024%)	2 (0.004%)
Mitochondrial coding genes	B	8,378	33 (0.394%)	15 (0.179%)
Plastid intergenic regions	C	7,092	22 (0.310%)	19 (0.268%)
nr DNA cistron	D	7,849	5 (0.064%)	0 (0.000%)
ITS	E	2,881	0 (0.000%)	0 (0.000%)
matK + *rbcL*	F	2,976	0 (0.000%)	0 (0.000%)
Divergence hotspot regions	G	1,021	3 (0.294%)	2 (0.196%)

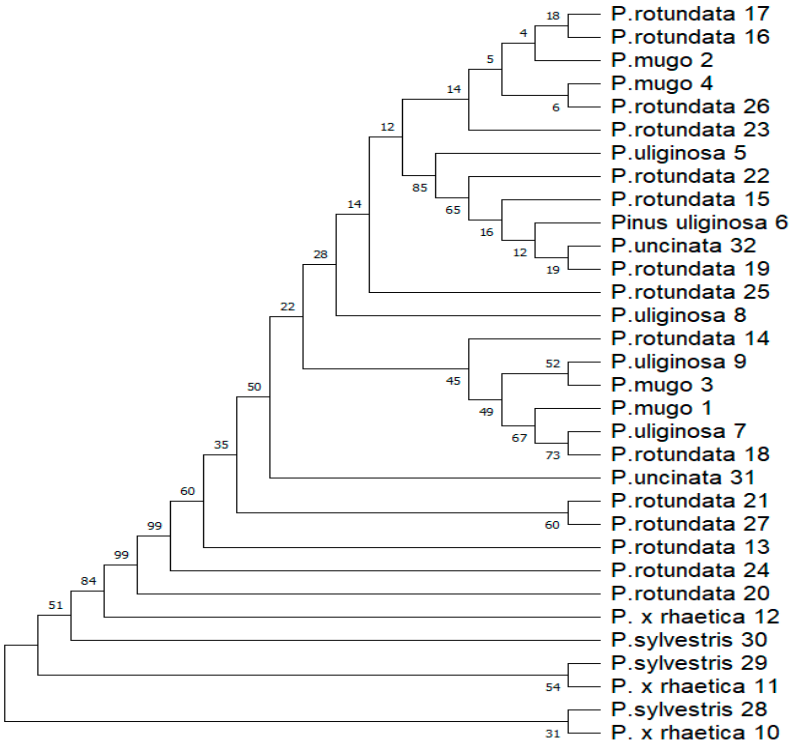


Figure S1. Maximum likelihood tree based on the plastid coding genes data set A.

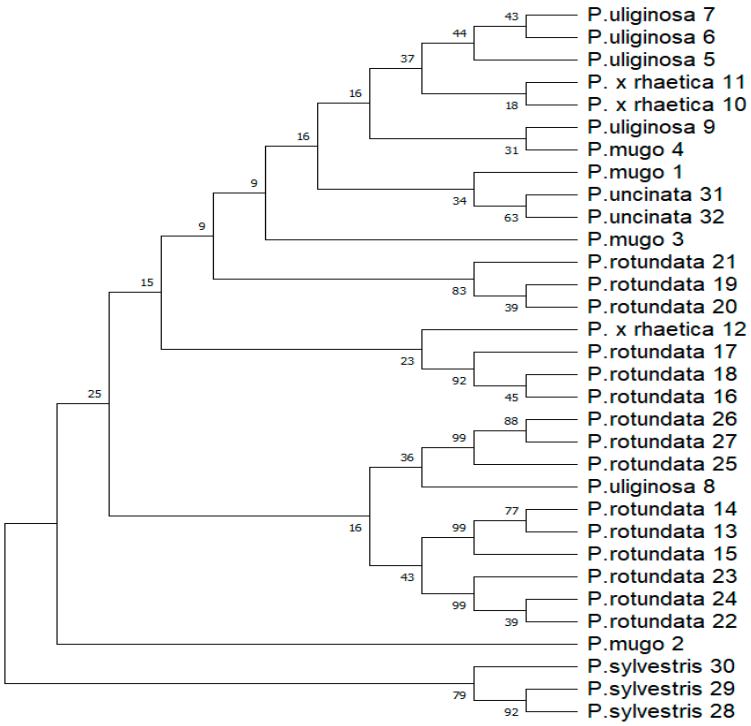


Figure S2. Maximum likelihood tree based on the mitochondrial coding genes data set B.

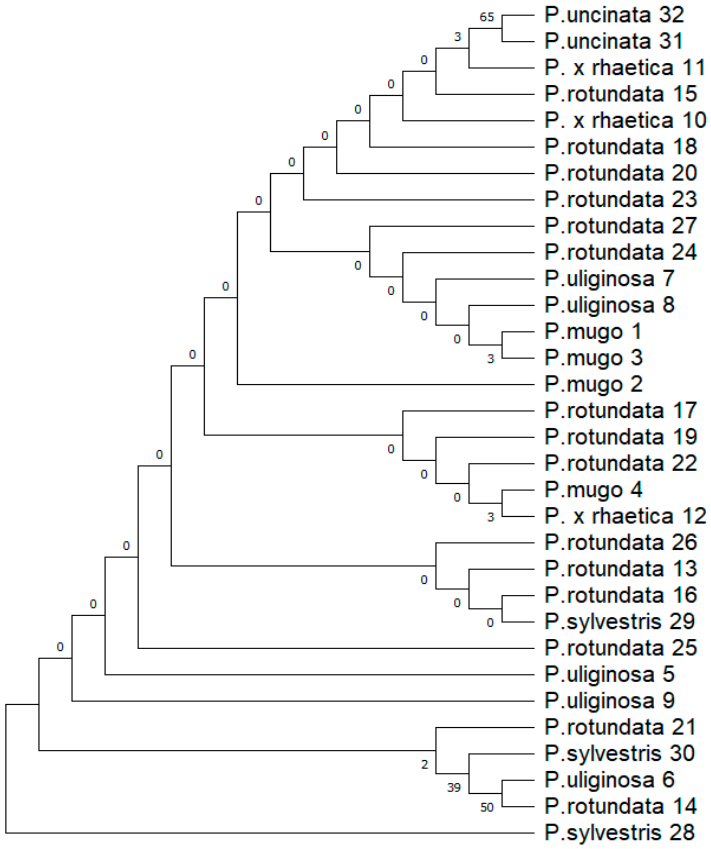


Figure S3. Maximum likelihood tree based on the nrDNA cistron data set D.

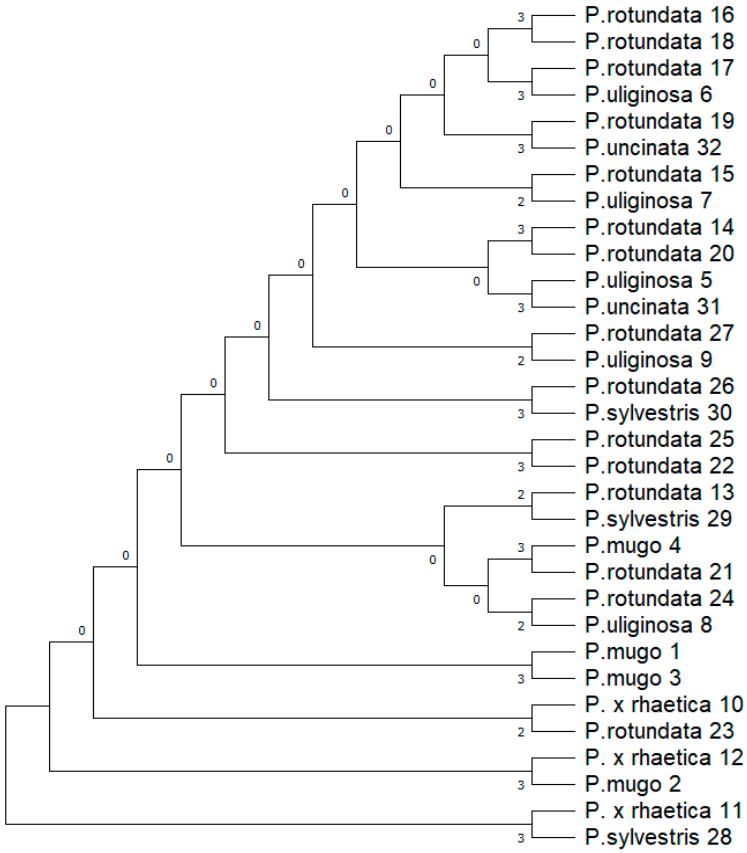


Figure S4. Maximum likelihood tree based on the ITS data set E.

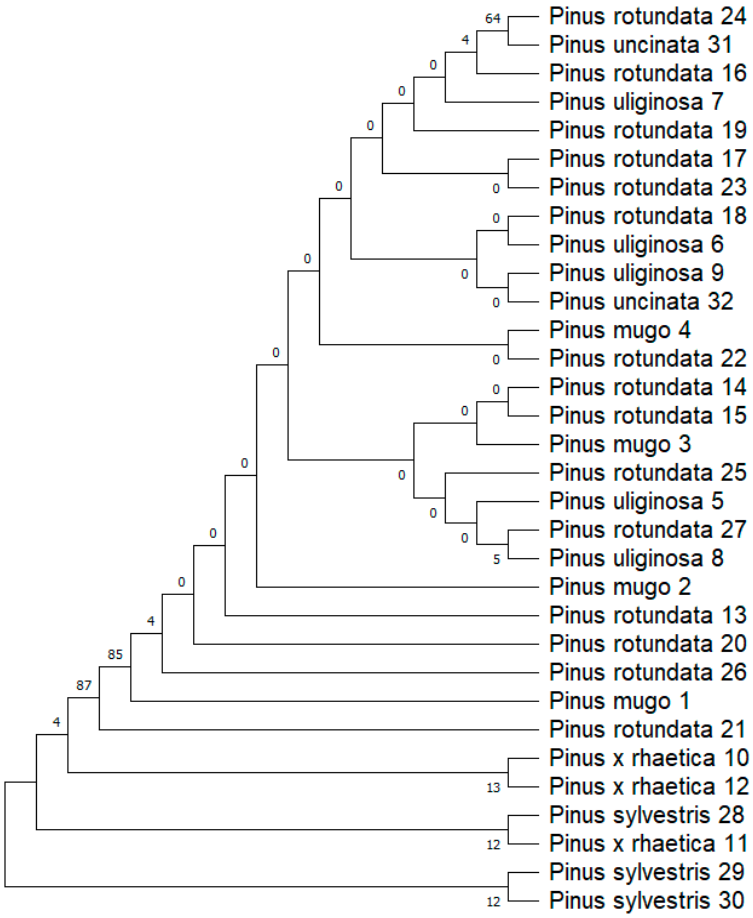


Figure S5. Maximum likelihood tree based on the *matK* + *rbcL* data set F.

Table S4. Minimum interspecific distance and maximum intraspecific distance for each taxa in the different data sets.

Taxon	Data set													
	A		B		C		D		E		F		G	
	min_inter	max_intra	min_inter	max_intra	min_inter	max_intra	min_inter	max_intra	min_inter	max_intra	min_inter	max_intra	min_inter	max_intra
Pinus mugo	0.00000	0.00009	0.00012	0.00143	0.00099	0.00227	0.00000	0.00013	0.00000	0.00000	0.00000	0.00000	0.00098	0.00294
Pinus uliginosa	0.00000	0.00015	0.00012	0.00227	0.00028	0.00170	0.00000	0.00013	0.00000	0.00000	0.00000	0.00000	0.00000	0.00591
Pinus × rhaetica	0.00007	0.00015	0.00071	0.00060	0.00199	0.00099	0.00000	0.00013	0.00000	0.00000	0.00000	0.00000	0.00196	0.00000
Pinus rotundata	0.00000	0.00106	0.00096	0.00491	0.00028	0.00355	0.00000	0.00064	0.00000	0.00000	0.00000	0.00236	0.00000	0.01285
Pinus sylvestris	0.00004	0.00009	0.00131	0.00215	0.00085	0.00114	0.00000	0.00051	0.00000	0.00000	0.00000	0.00000	0.00196	0.00195
Pinus uncinata	0.00004	0.00022	0.00060	0.00048	0.00029	0.00028	0.00040	0.00035	0.00000	0.00104	0.00000	0.00034	0.00295	0.00196

Note: the numbers colored by red represent the minimum interspecific distance larger than maximum intraspecific distance.

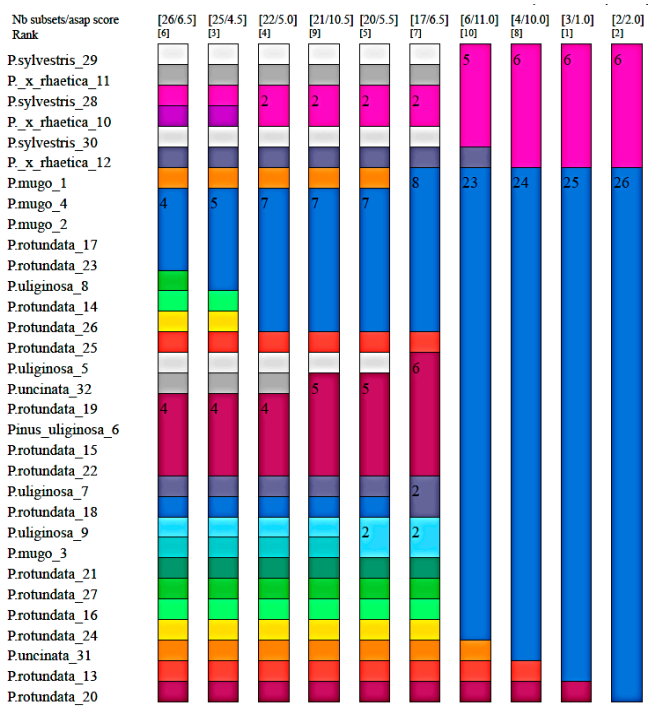


Figure S6. Species delimitation by ASAP analysis based on plastid coding genes. Graphical output showing the ten different delimitations; each column represents a partition, and the colors represent the OTUs. Every field contains the number of individuals. Above the colorful bars, the coefficient asap-score (the lower value) and number of species (the upper value) recognized for whole dataset are presented.

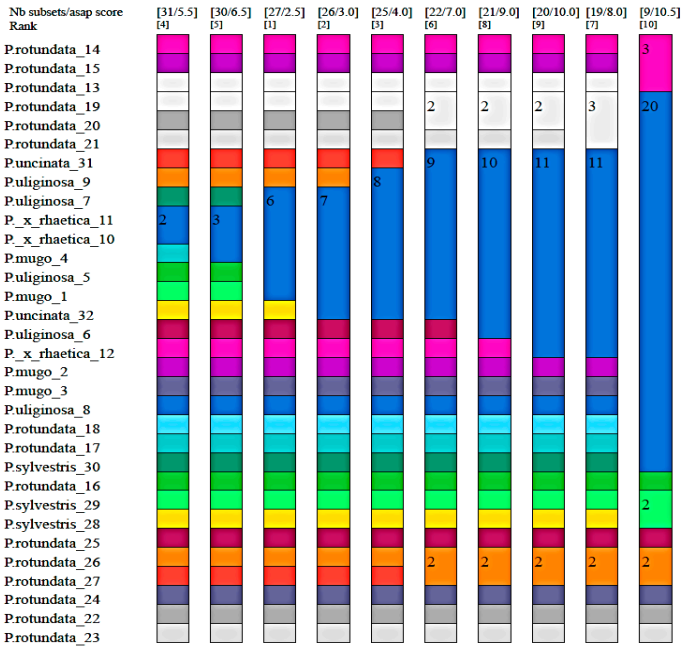


Figure S7. Species delimitation by ASAP analysis based on mitochondrial coding genes. Graphical output showing the ten different delimitations; each column represents a partition, and the colors represent the OTUs. Every field contains the number of individuals. Above the colorful bars, the coefficient asap-score (the lower value) and number of species (the upper value) recognized for whole dataset are presented.

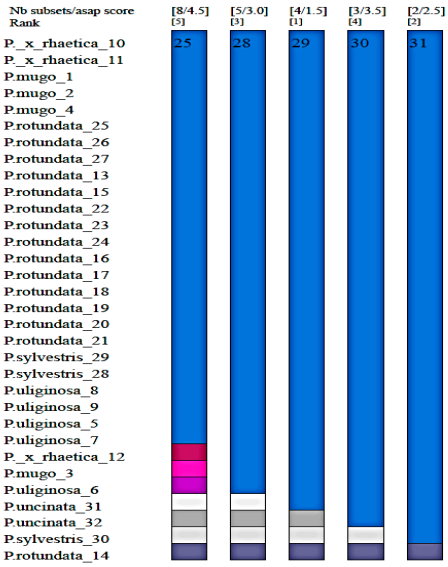


Figure S8. Species delimitation by ASAP analysis based on nrDNA cistron. Graphical output showing the ten different delimitations; each column represents a partition, and the colors represent the OTUs. Every field contains the number of individuals. Above the colorful bars, the coefficient asap-score (the lower value) and number of species (the upper value) recognized for whole dataset are presented.

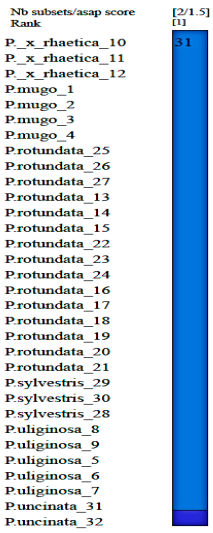


Figure S9. Species delimitation by ASAP analysis based on the internal transcribed spacer (ITS). Graphical output showing the ten different delimitations; each column represents a partition, and the colors represent the OTUs. Every field contains the number of individuals. Above the colorful bars, the coefficient asap-score (the lower value) and number of species (the upper value) recognized for whole dataset are presented.

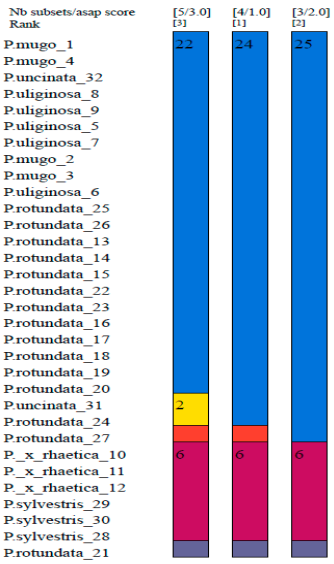


Figure S10. Species delimitation by ASAP analysis based on *matK* + *rbcL*. Graphical output showing the ten different delimitations; each column represents a partition, and the colors represent the OTUs. Every field contains the number of individuals. Above the colorful bars, the coefficient asap-score (the lower value) and number of species (the upper value) recognized for whole dataset are presented.

Table S5. Various haplotypes of *Pinus mugo* complex*, P. sylvestris and P.* × *rhaetica* taxa identified in the current study based on plastid coding genes.

Haplotype	**No. of sequences**	**Taxa**
Hap_1	1	*P. sylvestris* 29
Hap_2	1	*P.* × *rhaetica* 11
Hap_3	1	*P. sylvestris* 30
Hap_4	1	*P. sylvestris* 28
Hap_5	1	*P.* × *rhaetica* 10
Hap_6	1	*P.* × *rhaetica* 12
Hap_7	1	*P. mugo* 1
Hap_8	1	*P. uliginosa* 7
Hap_9	4	*P. mugo* 4, *P. mugo* 2, *P. rotundata* 17, *P. rotundata* 23
Hap_10	1	*P. uliginosa* 5
Hap_11	1	*P. uliginosa* 8
Hap_12	1	*P. uliginosa* 9
Hap_13	1	*P. uncinata* 32
Hap_14	1	*P. rotundata* 14
Hap_15	1	*P. rotundata* 26
Hap_16	1	*P. mugo* 3
Hap_17	1	*P. rotundata* 18
Hap_18	4	*P. rotundata* 19, *P. uliginosa* 6, *P. rotundata* 15, *P. rotundata* 22
Hap_19	1	*P. rotundata* 24
Hap_20	1	*P. rotundata* 21
Hap_21	1	*P. rotundata* 27
Hap_22	1	*P. rotundata* 16
Hap_23	1	*P. rotundata* 25
Hap_24	1	*P. uncinata* 31
Hap_25	1	*P. rotundata* 13
Hap_26	1	*P. rotundata* 20

Table S6. Various haplotypes of *Pinus mugo* complex, *P. sylvestris and P. × rhaetica* taxa identified in the current study based on mitochondrial coding genes.

Haplotype	No. of sequences	Taxa
Hap_1	1	*P. rotundata* 14
Hap_2	1	*P. rotundata* 13
Hap_3	1	*P. rotundata* 15
Hap_4	1	*P. rotundata* 19
Hap_5	1	*P. rotundata* 21
Hap_6	1	*P. rotundata* 20
Hap_7	1	*P. uliginosa* 7
Hap_8	1	*P. uliginosa* 5
Hap_9	2	*P. × rhaetica* 11, *P. × rhaetica* 10
Hap_10	1	*P. mugo* 1
Hap_11	1	*P. mugo* 4
Hap_12	1	*P. uncinata* 32
Hap_13	1	*P. uncinata* 31
Hap_14	1	*P. uliginosa* 9
Hap_15	1	*P. uliginosa* 6
Hap_16	1	*P. × rhaetica* 12
Hap_17	1	*P. mugo* 2
Hap_18	1	*P. mugo* 3
Hap_19	1	*P. uliginosa* 8
Hap_20	1	*P. sylvestris* 30
Hap_21	1	*P. rotundata* 18
Hap_22	1	*P. rotundata* 17
Hap_23	1	*P. rotundata* 16
Hap_24	1	*P. sylvestris* 28
Hap_25	1	*P. sylvestris* 29
Hap_26	1	*P. rotundata* 25
Hap_27	1	*P. rotundata* 26
Hap_28	1	*P. rotundata* 27
Hap_29	1	*P. rotundata* 24
Hap_30	1	*P. rotundata* 22

Hap_31	1	*P. rotundata* 23

Table S7. List of accession numbers of complete nrDNA cistron sequences used in this study.

No.	Taxa name	Accession number
1	*P.× rhaetica* 10	PQ094679
2	*P. × rhaetica* 11	PQ094680
3	*P. × rhaetica* 12	PQ094699
4	*P. mugo* 1	PQ094688
5	*P. mugo* 2	PQ094692
6	*P. mugo* 3	PQ094691
7	*P. mugo* 4	PQ094687
8	*P. rotundata* 25	PQ094668
9	*P. rotundata* 26	PQ094694
10	*P. rotundata* 27	PQ094667
11	*P. rotundata* 13	PQ094676
12	*P. rotundata* 14	PQ094697
13	*P. rotundata* 15	PQ094675
14	*P. rotundata* 22	PQ094688
15	*P. rotundata* 23	PQ094689
16	*P. rotundata* 24	PQ094695
17	*P. rotundata* 16	PQ094674
18	*P. rotundata* 17	PQ094690
19	*P. rotundata* 18	PQ094673
20	*P. rotundata* 19	PQ094696
21	*P. rotundata* 20	PQ094672
22	*P. rotundata* 21	PQ094671
23	*P. sylvestris* 29	PQ094677
24	*P. sylvestris* 30	PQ094689
25	*P. sylvestris* 28	PQ094678
26	*P. uliginosa* 8	PQ094682
27	*P. uliginosa* 9	PQ094681
28	*P. uliginosa* 5	PQ094686
29	*P. uliginosa* 6	PQ094693
30	*P. uliginosa* 7	PQ094685
31	*P. uncinata* 31	PQ094698
32	*P. uncinata* 32	PQ094700

Table S8. List of accession numbers of complete mitochondrial genes used in this study.

No.	Gene / Taxa	*ATP4*	*ATP6*	*ATP8*	*ATP9*	*ccmB*	*ccmC*	*COX3*	*NAD3*	*NAD4L*	*NAD6*	*rpl5*	*rps1*	*rps2*	*rps14*	*rps19*
1	*P. rotundata* 13	PQ255545	PQ255579	PQ255613	PQ255647	PQ342009	PQ342043	PQ342077	PQ342111	PQ342145	PQ342179	PQ342213	PQ342247	PQ342281	PQ342315	PQ342349
2	*P. rotundata* 14	PQ255546	PQ255580	PQ255614	PQ255648	PQ342010	PQ342044	PQ342078	PQ342112	PQ342146	PQ342180	PQ342214	PQ342248	PQ342282	PQ342316	PQ342350
3	*P. rotundata* 15	PQ255547	PQ255581	PQ255615	PQ255649	PQ342011	PQ342045	PQ342079	PQ342113	PQ342147	PQ342181	PQ342215	PQ342249	PQ342283	PQ342317	PQ342351
4	*P. rotundata* 19	PQ255548	PQ255582	PQ255616	PQ255650	PQ342012	PQ342046	PQ342080	PQ342114	PQ342148	PQ342182	PQ342216	PQ342250	PQ342284	PQ342318	PQ342352
5	*P. rotundata* 20	PQ255549	PQ255583	PQ255617	PQ255651	PQ342013	PQ342047	PQ342081	PQ342115	PQ342149	PQ342183	PQ342217	PQ342251	PQ342285	PQ342319	PQ342353
6	*P. rotundata* 21	PQ255550	PQ255584	PQ255618	PQ255652	PQ342014	PQ342048	PQ342082	PQ342116	PQ342150	PQ342184	PQ342218	PQ342252	PQ342286	PQ342320	PQ342354
7	*P. rotundata* 16	PQ255551	PQ255585	PQ255619	PQ255653	PQ342015	PQ342049	PQ342083	PQ342117	PQ342151	PQ342185	PQ342219	PQ342253	PQ342287	PQ342321	PQ342355
8	*P. rotundata* 17	PQ255552	PQ255586	PQ255620	PQ255654	PQ342016	PQ342050	PQ342084	PQ342118	PQ342152	PQ342186	PQ342220	PQ342254	PQ342288	PQ342322	PQ342356
9	*P. rotundata* 18	PQ255553	PQ255587	PQ255621	PQ255655	PQ342017	PQ342051	PQ342085	PQ342119	PQ342153	PQ342187	PQ342221	PQ342255	PQ342289	PQ342323	PQ342357
10	*P. rotundata* 25	PQ255554	PQ255588	PQ255622	PQ255656	PQ342018	PQ342052	PQ342086	PQ342120	PQ342154	PQ342188	PQ342222	PQ342256	PQ342290	PQ342324	PQ342358
11	*P. rotundata* 26	PQ255555	PQ255589	PQ255623	PQ255657	PQ342019	PQ342053	PQ342087	PQ342121	PQ342155	PQ342189	PQ342223	PQ342257	PQ342291	PQ342325	PQ342359
12	*P. rotundata* 27	PQ255556	PQ255590	PQ255624	PQ255658	PQ342020	PQ342054	PQ342088	PQ342122	PQ342156	PQ342190	PQ342224	PQ342258	PQ342292	PQ342326	PQ342360
13	*P. rotundata* 22	PQ255557	PQ255591	PQ255625	PQ255659	PQ342021	PQ342055	PQ342089	PQ342123	PQ342157	PQ342191	PQ342225	PQ342259	PQ342293	PQ342327	PQ342361
14	*P. rotundata* 23	PQ255558	PQ255592	PQ255626	PQ255660	PQ342022	PQ342056	PQ342090	PQ342124	PQ342158	PQ342192	PQ342226	PQ342260	PQ342294	PQ342328	PQ342362
15	*P. rotundata* 24	PQ255559	PQ255593	PQ255627	PQ255661	PQ342023	PQ342057	PQ342091	PQ342125	PQ342159	PQ342193	PQ342227	PQ342261	PQ342295	PQ342329	PQ342363
16	*P. uliginosa* 9	PQ255560	PQ255594	PQ255628	PQ255662	PQ342024	PQ342058	PQ342092	PQ342126	PQ342160	PQ342194	PQ342228	PQ342262	PQ342296	PQ342330	PQ342364
17	*P. uliginosa* 7	PQ255561	PQ255595	PQ255629	PQ255663	PQ342025	PQ342059	PQ342093	PQ342127	PQ342161	PQ342195	PQ342229	PQ342263	PQ342297	PQ342331	PQ342365
18	*P. uliginosa* 6	PQ255562	PQ255596	PQ255630	PQ255664	PQ342026	PQ342060	PQ342094	PQ342128	PQ342162	PQ342196	PQ342230	PQ342264	PQ342298	PQ342332	PQ342366
19	*P. uliginosa* 5	PQ255563	PQ255597	PQ255631	PQ255665	PQ342027	PQ342061	PQ342095	PQ342129	PQ342163	PQ342197	PQ342231	PQ342265	PQ342299	PQ342333	PQ342367
20	*P. uliginosa* 8	PQ255564	PQ255598	PQ255632	PQ255666	PQ342028	PQ342062	PQ342096	PQ342130	PQ342164	PQ342198	PQ342232	PQ342266	PQ342300	PQ342334	PQ342368
21	*P.* × *rhaetica* 10	PQ255565	PQ255599	PQ255633	PQ255667	PQ342029	PQ342063	PQ342097	PQ342131	PQ342165	PQ342199	PQ342233	PQ342267	PQ342301	PQ342335	PQ342369
22	*P.* × *rhaetica* 11	PQ255566	PQ255600	PQ255634	PQ255668	PQ342030	PQ342064	PQ342098	PQ342132	PQ342166	PQ342200	PQ342234	PQ342268	PQ342302	PQ342336	PQ342370
23	*P.* × *rhaetica* 12	PQ255567	PQ255601	PQ255635	PQ255669	PQ342031	PQ342065	PQ342099	PQ342133	PQ342167	PQ342201	PQ342235	PQ342269	PQ342303	PQ342337	PQ342371
24	*P. mugo* 1	PQ255568	PQ255602	PQ255636	PQ255670	PQ342032	PQ342066	PQ342100	PQ342134	PQ342168	PQ342202	PQ342236	PQ342270	PQ342304	PQ342338	PQ342372
25	*P. mugo* 2	PQ255569	PQ255603	PQ255637	PQ255671	PQ342033	PQ342067	PQ342101	PQ342135	PQ342169	PQ342203	PQ342237	PQ342271	PQ342305	PQ342339	PQ342373
26	*P. mugo* 3	PQ255570	PQ255604	PQ255638	PQ255672	PQ342034	PQ342068	PQ342102	PQ342136	PQ342170	PQ342204	PQ342238	PQ342272	PQ342306	PQ342340	PQ342374
27	*P. mugo* 4	PQ255571	PQ255605	PQ255639	PQ255673	PQ342035	PQ342069	PQ342105	PQ342139	PQ342171	PQ342205	PQ342239	PQ342273	PQ342307	PQ342343	PQ342375
28	*P. uncinata* 31	PQ255574	PQ255608	PQ255642	PQ255676	PQ342038	PQ342072	PQ342106	PQ342140	PQ342174	PQ342208	PQ342242	PQ342276	PQ342310	PQ342344	PQ342378
29	*P. uncinata* 32	PQ255575	PQ255609	PQ255643	PQ255677	PQ342039	PQ342073	PQ342107	PQ342141	PQ342175	PQ342209	PQ342243	PQ342277	PQ342311	PQ342345	PQ342379
30	*P. sylvestris* 30	PQ255576	PQ255610	PQ255644	PQ255678	PQ342040	PQ342074	PQ342108	PQ342142	PQ342176	PQ342210	PQ342244	PQ342278	PQ342312	PQ342346	PQ342380
31	*P. sylvestris* 28	PQ255577	PQ255611	PQ255645	PQ255679	PQ342041	PQ342075	PQ342109	PQ342143	PQ342177	PQ342211	PQ342245	PQ342279	PQ342313	PQ342347	PQ342381
32	*P. sylvestris* 29	PQ255578	PQ255612	PQ255646	PQ255680	PQ342042	PQ342076	PQ342110	PQ342144	PQ342178	PQ342212	PQ342246	PQ342280	PQ342314	PQ342348	PQ342382

Table S9. List of accession numbers of complete chloroplast genes used in this study.

No.	Taxa / Gene	*P.× rhaetica* 10	*P.× rhaetica* 11	*P.× rhaetica* 12	*P. mugo* 1	*P. mugo* 2	*P. mugo* 3	*P. mugo* 4	*P. uliginosa* 5	*P. uliginosa* 6	*P .uliginosa* 7	*P. uliginosa* 8	*P. uliginosa* 9	*P. uncinata* 31	*P. uncinata* 32	*P. sylvestris* 29	*P. sylvestris* 30
1	*accD*	PQ342383	PQ342384	PQ342385	PQ165116	PQ342386	PQ342387	MZ333466	PQ165118	PQ342388	PQ165119	PQ165117	MZ333465	PQ342389	MZ333464	PQ342390	PQ342391
2	*atpA*	PQ342708	PQ342709	PQ342710	PQ165116	PQ342711	PQ342712	MZ333466	PQ165118	PQ342713	PQ165119	PQ165117	MZ333465	PQ342714	MZ333464	PQ342715	PQ342716
3	*atpB*	PQ342408	PQ342409	PQ342410	PQ165116	PQ342411	PQ342412	MZ333466	PQ165118	PQ342413	PQ165119	PQ165117	MZ333465	PQ342414	MZ333464	PQ342415	PQ342416
4	*atpF*	PQ342733	PQ342734	PQ342735	PQ165116	PQ342736	PQ342737	MZ333466	PQ165118	PQ342738	PQ165119	PQ165117	MZ333465	PQ342739	MZ333464	PQ342740	PQ342741
5	*atpH*	PQ342758	PQ342759	PQ342760	PQ165116	PQ342761	PQ342762	MZ333466	PQ165118	PQ342763	PQ165119	PQ165117	MZ333465	PQ342764	MZ333464	PQ342765	PQ342766
6	*atpI*	PQ342558	PQ342559	PQ342560	PQ165116	PQ342561	PQ342562	MZ333466	PQ165118	PQ342563	PQ165119	PQ165117	MZ333465	PQ342564	MZ333464	PQ342565	PQ34256
7	*ccsA*	PQ342783	PQ342784	PQ342785	PQ165116	PQ342786	PQ342787	MZ333466	PQ165118	PQ342788	PQ165119	PQ165117	MZ333465	PQ342789	MZ333464	PQ342790	PQ342791
8	*cemA*	PQ342808	PQ342809	PQ342810	PQ165116	PQ342811	PQ342812	MZ333466	PQ165118	PQ342813	PQ165119	PQ165117	MZ333465	PQ342814	MZ333464	PQ342815	PQ342816
9	*chlB*	PQ342833	PQ342834	PQ342835	PQ165116	PQ342836	PQ342837	MZ333466	PQ165118	PQ342838	PQ165119	PQ165117	MZ333465	PQ342839	MZ333464	PQ342840	PQ342841
10	*chlL*	PQ342858	PQ342859	PQ342860	PQ165116	PQ342861	PQ342862	MZ333466	PQ165118	PQ342863	PQ165119	PQ165117	MZ333465	PQ342864	MZ333464	PQ342865	PQ342866
11	*chlN*	PQ342883	PQ342884	PQ342885	PQ165116	PQ342886	PQ342887	MZ333466	PQ165118	PQ342888	PQ165119	PQ165117	MZ333465	PQ342889	MZ333464	PQ342890	PQ342891
12	*matK*	PQ342908	PQ342909	PQ342910	PQ165116	PQ342911	PQ342912	MZ333466	PQ165118	PQ342913	PQ165119	PQ165117	MZ333465	PQ342914	MZ333464	PQ342915	PQ342916
13	*petA*	PQ342483	PQ342484	PQ342485	PQ165116	PQ342486	PQ342487	MZ333466	PQ165118	PQ342488	PQ165119	PQ165117	MZ333465	PQ342489	MZ333464	PQ342490	PQ342491
14	*petB*	PQ342433	PQ342434	PQ342435	PQ165116	PQ342436	PQ342437	MZ333466	PQ165118	PQ342438	PQ165119	PQ165117	MZ333465	PQ342439	MZ333464	PQ342440	PQ342441
15	*petD*	PQ342433	PQ342434	PQ342435	PQ165116	PQ342436	PQ342437	MZ333466	PQ165118	PQ342438	PQ165119	PQ165117	MZ333465	PQ342439	MZ333464	PQ342440	PQ342441
16	*petG*	PQ342933	PQ342934	PQ342935	PQ165116	PQ342936	PQ342937	MZ333466	PQ165118	PQ342938	PQ165119	PQ165117	MZ333465	PQ342939	MZ333464	PQ342940	PQ342941
17	*petN*	PQ342533	PQ342534	PQ342535	PQ165116	PQ342536	PQ342537	MZ333466	PQ165118	PQ342538	PQ165119	PQ165117	MZ333465	PQ342539	MZ333464	PQ342539	PQ342539
18	*psaB*	PQ342958	PQ342959	PQ342960	PQ165116	PQ342961	PQ342962	MZ333466	PQ165118	PQ342963	PQ165119	PQ165117	MZ333465	PQ342964	MZ333464	PQ342965	PQ342966
19	*psaC*	PQ342983	PQ342984	PQ342985	PQ165116	PQ342986	PQ342987	MZ333466	PQ165118	PQ342988	PQ165119	PQ165117	MZ333465	PQ342989	MZ333464	PQ342990	PQ342991
20	*psbB*	PQ343008	PQ343009	PQ343010	PQ165116	PQ343011	PQ343012	MZ333466	PQ165118	PQ343013	PQ165119	PQ165117	MZ333465	PQ343014	MZ333464	PQ343015	PQ343016
21	*psbC*	PQ343033	PQ343034	PQ343035	PQ165116	PQ343036	PQ343037	MZ333466	PQ165118	PQ343038	PQ165119	PQ165117	MZ333465	PQ343039	MZ333464	PQ343040	PQ343041
22	*psbD*	PQ343058	PQ343059	PQ343060	PQ165116	PQ343061	PQ343062	MZ333466	PQ165118	PQ343063	PQ165119	PQ165117	MZ333465	PQ343064	MZ333464	PQ343065	PQ343066
23	*psbE*	PQ343083	PQ343084	PQ343085	PQ165116	PQ343086	PQ343087	MZ333466	PQ165118	PQ343088	PQ165119	PQ165117	MZ333465	PQ343089	MZ333464	PQ343090	PQ343091
24	*psbF*	PQ343108	PQ343109	PQ343110	PQ165116	PQ343111	PQ343112	MZ333466	PQ165118	PQ343113	PQ165119	PQ165117	MZ333465	PQ343114	MZ333464	PQ343115	PQ343116
25	*psbH*	PQ343133	PQ343134	PQ343135	PQ165116	PQ343136	PQ343137	MZ333466	PQ165118	PQ343138	PQ165119	PQ165117	MZ333465	PQ343139	MZ333464	PQ343140	PQ343141
26	*psbI*	PQ343158	PQ343159	PQ343160	PQ165116	PQ343161	PQ343162	MZ333466	PQ165118	PQ343163	PQ165119	PQ165117	MZ333465	PQ343164	MZ333464	PQ343165	PQ343166
27	*psbJ*	PQ342483	PQ342484	PQ342485	PQ165116	PQ342486	PQ342487	MZ333466	PQ165118	PQ342488	PQ165119	PQ165117	MZ333465	PQ342489	MZ333464	PQ342490	PQ342491
28	*psbK*	PQ343183	PQ343184	PQ343185	PQ165116	PQ343186	PQ343187	MZ333466	PQ165118	PQ343188	PQ165119	PQ165117	MZ333465	PQ343189	MZ333464	PQ343190	PQ343191
29	*psbL*	PQ342483	PQ342484	PQ342485	PQ165116	PQ342486	PQ342487	MZ333466	PQ165118	PQ342488	PQ165119	PQ165117	MZ333465	PQ342489	MZ333464	PQ342490	PQ342491
30	*psbM*	PQ342608	PQ342609	PQ342610	PQ165116	PQ342611	PQ342612	MZ333466	PQ165118	PQ342613	PQ165119	PQ165117	MZ333465	PQ342614	MZ333464	PQ342615	PQ342616
31	*psbN*	PQ343208	PQ343209	PQ343210	PQ165116	PQ343211	PQ343212	MZ333466	PQ165118	PQ343213	PQ165119	PQ165117	MZ333465	PQ343214	MZ333464	PQ343215	PQ343216
32	*psbT*	PQ343233	PQ343234	PQ343235	PQ165116	PQ343236	PQ343237	MZ333466	PQ165118	PQ343238	PQ165119	PQ165117	MZ333465	PQ343239	MZ333464	PQ343240	PQ343241

33	*psbZ*	PQ342508	PQ342509	PQ342510	PQ165116	PQ342511	PQ342512	MZ333466	PQ165118	PQ342513	PQ165119	PQ165117	MZ333465	PQ342514	MZ333464	PQ342515	PQ342516
34	*rbcL*	PQ342408	PQ342409	PQ342410	PQ165116	PQ342411	PQ342412	MZ333466	PQ165118	PQ342413	PQ165119	PQ165117	MZ333465	PQ342414	MZ333464	PQ342415	PQ342416
35	*rpl2*	PQ343258	PQ343259	PQ343260	PQ165116	PQ343261	PQ343262	MZ333466	PQ165118	PQ343263	PQ165119	PQ165117	MZ333465	PQ343264	MZ333464	PQ343265	PQ343266
36	*rpl14*	PQ343283	PQ343284	PQ343285	PQ165116	PQ343286	PQ343287	MZ333466	PQ165118	PQ343288	PQ165119	PQ165117	MZ333465	PQ343289	MZ333464	PQ343290	PQ343291
37	*rpl16*	PQ343308	PQ343309	PQ343310	PQ165116	PQ343311	PQ343312	MZ333466	PQ165118	PQ343313	PQ165119	PQ165117	MZ333465	PQ343314	MZ333464	PQ343315	PQ343316
38	*rpl20*	PQ343333	PQ343334	PQ343335	PQ165116	PQ343336	PQ343337	MZ333466	PQ165118	PQ343338	PQ165119	PQ165117	MZ333465	PQ343339	MZ333464	PQ343340	PQ343341
39	*rpl22*	PQ343358	PQ343359	PQ343360	PQ165116	PQ343361	PQ343362	MZ333466	PQ165118	PQ343363	PQ165119	PQ165117	MZ333465	PQ343364	MZ333464	PQ343365	PQ343366
40	*rpl23*	PQ343383	PQ343384	PQ343385	PQ165116	PQ343386	PQ343387	MZ333466	PQ165118	PQ343388	PQ165119	PQ165117	MZ333465	PQ343389	MZ333464	PQ343390	PQ343391
41	*rpl33*	PQ343408	PQ343409	PQ343410	PQ165116	PQ343411	PQ343412	MZ333466	PQ165118	PQ343413	PQ165119	PQ165117	MZ333465	PQ343414	MZ333464	PQ343415	PQ343416
42	*rpl36*	PQ343433	PQ343434	PQ343435	PQ165116	PQ343436	PQ343437	MZ333466	PQ165118	PQ343438	PQ165119	PQ165117	MZ333465	PQ343439	MZ333464	PQ343440	PQ343441
43	*rpoA*	PQ343458	PQ343459	PQ343460	PQ165116	PQ343461	PQ343462	MZ333466	PQ165118	PQ343463	PQ165119	PQ165117	MZ333465	PQ343464	MZ333464	PQ343465	PQ343466
44	*rpoB*	PQ342533	PQ342534	PQ342535	PQ165116	PQ342536	PQ342537	MZ333466	PQ165118	PQ342538	PQ165119	PQ165117	MZ333465	PQ342539	MZ333464	PQ342539	PQ342539
45	*rpoC1*	PQ343483	PQ343484	PQ343485	PQ165116	PQ343486	PQ343487	MZ333466	PQ165118	PQ343488	PQ165119	PQ165117	MZ333465	PQ343489	MZ333464	PQ343490	PQ343491
46	*rpoC2*	PQ343508	PQ343509	PQ343510	PQ165116	PQ343511	PQ343512	MZ333466	PQ165118	PQ343513	PQ165119	PQ165117	MZ333465	PQ343514	MZ333464	PQ343515	PQ343516
47	*rps2*	PQ342558	PQ342559	PQ342560	PQ165116	PQ342561	PQ342562	MZ333466	PQ165118	PQ342563	PQ165119	PQ165117	MZ333465	PQ342564	MZ333464	PQ342565	PQ34256
48	*rps3*	PQ343533	PQ343534	PQ343535	PQ165116	PQ343536	PQ343537	MZ333466	PQ165118	PQ343538	PQ165119	PQ165117	MZ333465	PQ343539	MZ333464	PQ343540	PQ343541
49	*rps4*	PQ343558	PQ343559	PQ343560	PQ165116	PQ343561	PQ343562	MZ333466	PQ165118	PQ343563	PQ165119	PQ165117	MZ333465	PQ343564	MZ333464	PQ343565	PQ343566
50	*rps7*	PQ343583	PQ343584	PQ343585	PQ165116	PQ343586	PQ343587	MZ333466	PQ165118	PQ343588	PQ165119	PQ165117	MZ333465	PQ343589	MZ333464	PQ343590	PQ343591
51	*rps11*	PQ343608	PQ343609	PQ343610	PQ165116	PQ343611	PQ343612	MZ333466	PQ165118	PQ343613	PQ165119	PQ165117	MZ333465	PQ343614	MZ333464	PQ343615	PQ343616
52	*rps14*	PQ343633	PQ343634	PQ343635	PQ165116	PQ343636	PQ343637	MZ333466	PQ165118	PQ343638	PQ165119	PQ165117	MZ333465	PQ343639	MZ333464	PQ343640	PQ343641
53	*rps15*	PQ343658	PQ343659	PQ343660	PQ165116	PQ343661	PQ343662	MZ333466	PQ165118	PQ343663	PQ165119	PQ165117	MZ333465	PQ343664	MZ333464	PQ343665	PQ343666
54	*rps18*	PQ343683	PQ343684	PQ343685	PQ165116	PQ343686	PQ343687	MZ333466	PQ165118	PQ343688	PQ165119	PQ165117	MZ333465	PQ343689	MZ333464	PQ343690	PQ343691
55	*rps19*	PQ343708	PQ343709	PQ343710	PQ165116	PQ343711	PQ343712	MZ333466	PQ165118	PQ343713	PQ165119	PQ165117	MZ333465	PQ343714	MZ333464	PQ343715	PQ343716
56	*ycf3*	PQ342658	PQ342659	PQ342660	PQ165116	PQ342661	PQ342662	MZ333466	PQ165118	PQ342663	PQ165119	PQ165117	MZ333465	PQ342664	MZ333464	PQ342665	PQ342666
57	*ycf4*	PQ343733	PQ343734	PQ343735	PQ165116	PQ343736	PQ343737	MZ333466	PQ165118	PQ343738	PQ165119	PQ165117	MZ333465	PQ343739	MZ333464	PQ343740	PQ343741

No.	**Taxa** / **Gene**	*P. sylvestris* 28	*P. rotundata* 25	*P. rotundata* 26	*P. rotundata* 27	*P. rotundata* 13	*P. rotundata* 14	*P. rotundata* 15	*P. rotundata* 22	*P. rotundata* 23	*P. rotundata* 24	*P. rotundata* 16	*P. rotundata* 17	*P. rotundata* 18	*P. rotundata* 19	*P. rotundata* 20	*P. rotundata* 21
1	*accD*	PQ342392	PQ342393	PQ342394	PQ342395	PQ342396	PQ342397	PQ342398	PQ342399	PQ342400	PQ342401	PQ342402	PQ342403	PQ342404	PQ342405	PQ342406	PQ342407
2	*atpA*	PQ342717	PQ342718	PQ342719	PQ342720	PQ342721	PQ342722	PQ342723	PQ342724	PQ342725	PQ342726	PQ342727	PQ342728	PQ342729	PQ342730	PQ342731	PQ3427232
3	*atpB*	PQ342417	PQ342418	PQ342419	PQ342420	PQ342421	PQ342422	PQ342423	PQ342424	PQ342425	PQ342426	PQ342427	PQ342428	PQ342429	PQ342430	PQ342431	PQ342432
4	*atpF*	PQ342742	PQ342743	PQ342744	PQ342745	PQ342746	PQ342747	PQ342748	PQ342749	PQ342750	PQ342751	PQ342752	PQ342753	PQ342754	PQ342755	PQ342756	PQ342757
5	*atpH*	PQ342767	PQ342768	PQ342769	PQ342770	PQ342771	PQ342772	PQ342773	PQ342774	PQ342775	PQ342776	PQ342777	PQ342778	PQ342779	PQ342780	PQ342781	PQ342782
6	*atpI*	PQ342567	PQ342568	PQ342569	PQ342570	PQ342571	PQ342572	PQ342573	PQ342574	PQ342575	PQ342576	PQ342577	PQ342578	PQ342579	PQ342580	PQ342581	PQ342582
7	*ccsA*	PQ342792	PQ342793	PQ342794	PQ342795	PQ342796	PQ342797	PQ342798	PQ342799	PQ342800	PQ342801	PQ342802	PQ342803	PQ342804	PQ342805	PQ342806	PQ342807

8	*cemA*	PQ342817	PQ342818	PQ342819	PQ342820	PQ342821	PQ342822	PQ342823	PQ342824	PQ342825	PQ342826	PQ342827	PQ342828	PQ342829	PQ342830	PQ342831	PQ342832
9	*chlB*	PQ342842	PQ342843	PQ342844	PQ342845	PQ342846	PQ342847	PQ342848	PQ342849	PQ342850	PQ342851	PQ342852	PQ342853	PQ342854	PQ342855	PQ342856	PQ342857
10	*chlL*	PQ342867	PQ342868	PQ342869	PQ342870	PQ342871	PQ342872	PQ342873	PQ342874	PQ342875	PQ342876	PQ342877	PQ342878	PQ342879	PQ342880	PQ342881	PQ342882
11	*chlN*	PQ342892	PQ342893	PQ342894	PQ342895	PQ342896	PQ342897	PQ342898	PQ342899	PQ342900	PQ342901	PQ342902	PQ342903	PQ342904	PQ342905	PQ342906	PQ342907
12	*matK*	PQ342917	PQ342918	PQ342919	PQ342920	PQ342921	PQ342922	PQ342923	PQ342924	PQ342925	PQ342926	PQ342927	PQ342928	PQ342929	PQ342930	PQ342931	PQ342932
13	*petA*	PQ342492	PQ342493	PQ342494	PQ342495	PQ342496	PQ342497	PQ342498	PQ342499	PQ342500	PQ342501	PQ342502	PQ342503	PQ342504	PQ342505	PQ342506	PQ342507
14	*petB*	PQ342442	PQ342443	PQ342444	PQ342445	PQ342446	PQ342447	PQ342448	PQ342449	PQ342450	PQ342451	PQ342452	PQ342453	PQ342454	PQ342455	PQ342456	PQ342457
15	*petD*	PQ342442	PQ342443	PQ342444	PQ342445	PQ342446	PQ342447	PQ342448	PQ342449	PQ342450	PQ342451	PQ342452	PQ342453	PQ342454	PQ342455	PQ342456	PQ342457
16	*petG*	PQ342942	PQ342943	PQ342944	PQ342945	PQ342946	PQ342947	PQ342948	PQ342949	PQ342950	PQ342951	PQ342952	PQ342953	PQ342954	PQ342955	PQ342956	PQ342957
17	*petN*	PQ342542	PQ342543	PQ342544	PQ342545	PQ342546	PQ342547	PQ342548	PQ342549	PQ342550	PQ342551	PQ342552	PQ342553	PQ342554	PQ342555	PQ342556	PQ342557
18	*psaB*	PQ342967	PQ342968	PQ342969	PQ342970	PQ342971	PQ342972	PQ342973	PQ342974	PQ342975	PQ342976	PQ342977	PQ342978	PQ342979	PQ342980	PQ342981	PQ342982
19	*psaC*	PQ342992	PQ342993	PQ342994	PQ342995	PQ342996	PQ342997	PQ342998	PQ342999	PQ343000	PQ343001	PQ343002	PQ343003	PQ343004	PQ343005	PQ343006	PQ343007
20	*psbB*	PQ343017	PQ343018	PQ343019	PQ343020	PQ343021	PQ343022	PQ343023	PQ343024	PQ343025	PQ343026	PQ343027	PQ343028	PQ343029	PQ343030	PQ343031	PQ343032
21	*psbC*	PQ343042	PQ343043	PQ343044	PQ343045	PQ343046	PQ343047	PQ343048	PQ343049	PQ343050	PQ343051	PQ343052	PQ343053	PQ343054	PQ343055	PQ343056	PQ343057
22	*psbD*	PQ343067	PQ343068	PQ343069	PQ343070	PQ343071	PQ343072	PQ343073	PQ343074	PQ343075	PQ343076	PQ343077	PQ343078	PQ343079	PQ343080	PQ343081	PQ343082
23	*psbE*	PQ343092	PQ343093	PQ343094	PQ343095	PQ343096	PQ343097	PQ343098	PQ343099	PQ343100	PQ343101	PQ343102	PQ343103	PQ343104	PQ343105	PQ343106	PQ343107
24	*psbF*	PQ343117	PQ343118	PQ343119	PQ343120	PQ343121	PQ343122	PQ343123	PQ343124	PQ343125	PQ343126	PQ343127	PQ343128	PQ343129	PQ343130	PQ343131	PQ343132
25	*psbH*	PQ343142	PQ343144	PQ343145	PQ343146	PQ343147	PQ343148	PQ343149	PQ343150	PQ343151	PQ343151	PQ343152	PQ343153	PQ343154	PQ343155	PQ343156	PQ343157
26	*psbI*	PQ343167	PQ343168	PQ343169	PQ343170	PQ343171	PQ343172	PQ343173	PQ343174	PQ343175	PQ343176	PQ343177	PQ343178	PQ343179	PQ343180	PQ343181	PQ343182
27	*psbJ*	PQ342492	PQ342493	PQ342494	PQ342495	PQ342496	PQ342497	PQ342498	PQ342499	PQ342500	PQ342501	PQ342502	PQ342503	PQ342504	PQ342505	PQ342506	PQ342507
28	*psbK*	PQ343192	PQ343193	PQ343194	PQ343195	PQ343196	PQ343197	PQ343198	PQ343199	PQ343200	PQ343201	PQ343202	PQ343203	PQ343204	PQ343205	PQ343206	PQ343207
29	*psbL*	PQ342492	PQ342493	PQ342494	PQ342495	PQ342496	PQ342497	PQ342498	PQ342499	PQ342500	PQ342501	PQ342502	PQ342503	PQ342504	PQ342505	PQ342506	PQ342507
30	*psbM*	PQ342617	PQ342618	PQ342619	PQ342620	PQ342621	PQ342622	PQ342623	PQ342624	PQ342625	PQ342626	PQ342627	PQ342628	PQ342629	PQ342630	PQ342631	PQ342632
31	*psbN*	PQ343217	PQ343218	PQ343219	PQ343220	PQ343221	PQ343222	PQ343223	PQ343224	PQ343225	PQ343226	PQ343227	PQ343228	PQ343229	PQ343230	PQ343231	PQ3432232
32	*psbT*	PQ343242	PQ343243	PQ343244	PQ343245	PQ343246	PQ343247	PQ343248	PQ343249	PQ343250	PQ343251	PQ343252	PQ343253	PQ343254	PQ343255	PQ343256	PQ343257
33	*psbZ*	PQ342517	PQ342518	PQ342519	PQ342520	PQ342521	PQ342522	PQ342523	PQ342524	PQ342525	PQ342526	PQ342527	PQ342528	PQ342529	PQ342530	PQ342531	PQ342532
34	*rbcL*	PQ342417	PQ342418	PQ342419	PQ342420	PQ342421	PQ342422	PQ342423	PQ342424	PQ342425	PQ342426	PQ342427	PQ342428	PQ342429	PQ342430	PQ342431	PQ342432
35	*rpl2*	PQ343267	PQ343268	PQ343269	PQ343270	PQ343271	PQ343272	PQ343273	PQ343274	PQ343275	PQ343276	PQ343277	PQ343278	PQ343279	PQ343280	PQ343281	PQ343282
36	*rpl14*	PQ343292	PQ343293	PQ343294	PQ343295	PQ343296	PQ343297	PQ343298	PQ343299	PQ343300	PQ343301	PQ343302	PQ343303	PQ343304	PQ343305	PQ343306	PQ343307
37	*rpl16*	PQ343317	PQ343318	PQ343319	PQ343320	PQ343321	PQ343322	PQ343323	PQ343324	PQ343325	PQ343326	PQ343327	PQ343328	PQ343329	PQ343330	PQ343331	PQ343332
38	*rpl20*	PQ343342	PQ343343	PQ343344	PQ343345	PQ343346	PQ343347	PQ343348	PQ343349	PQ343350	PQ343351	PQ343352	PQ343353	PQ343354	PQ343355	PQ343356	PQ343357
39	*rpl22*	PQ343367	PQ343368	PQ343369	PQ343370	PQ343371	PQ343372	PQ343373	PQ343374	PQ343375	PQ343376	PQ343377	PQ343378	PQ343379	PQ343380	PQ343381	PQ34332
40	*rpl23*	PQ343392	PQ343393	PQ343394	PQ343395	PQ343396	PQ343397	PQ343398	PQ343399	PQ343400	PQ343401	PQ343402	PQ343403	PQ343404	PQ343405	PQ343406	PQ343407
41	*rpl33*	PQ343417	PQ343418	PQ343419	PQ343420	PQ343421	PQ343422	PQ343423	PQ343424	PQ343425	PQ343426	PQ343427	PQ343428	PQ343429	PQ343430	PQ343431	PQ343432
42	*rpl36*	PQ343442	PQ343443	PQ343444	PQ343445	PQ343446	PQ343447	PQ343448	PQ343449	PQ343450	PQ343451	PQ343452	PQ343453	PQ343454	PQ343455	PQ343456	PQ343457
43	*rpoA*	PQ343467	PQ343468	PQ343469	PQ343470	PQ343471	PQ343472	PQ343473	PQ343474	PQ343475	PQ343476	PQ343477	PQ343478	PQ343479	PQ343480	PQ343481	PQ343482
44	*rpoB*	PQ342542	PQ342543	PQ342544	PQ342545	PQ342546	PQ342547	PQ342548	PQ342549	PQ342550	PQ342551	PQ342552	PQ342553	PQ342554	PQ342555	PQ342556	PQ342557
45	*rpoC1*	PQ343492	PQ343493	PQ343494	PQ343495	PQ343496	PQ343497	PQ3434938	PQ343499	PQ343500	PQ343501	PQ343502	PQ343503	PQ343504	PQ343505	PQ343506	PQ343507
46	*rpoC2*	PQ343517	PQ343518	PQ343519	PQ343520	PQ343521	PQ343522	PQ343523	PQ343524	PQ343525	PQ343526	PQ343527	PQ343528	PQ343529	PQ343530	PQ343531	PQ343532
47	*rps2*	PQ342567	PQ342568	PQ342569	PQ342570	PQ342571	PQ342572	PQ342573	PQ342574	PQ342575	PQ342576	PQ342577	PQ342578	PQ342579	PQ342580	PQ342581	PQ342582

48	*rps3*	PQ343542	PQ343543	PQ343544	PQ343545	PQ343546	PQ343547	PQ343548	PQ343549	PQ343550	PQ343551	PQ343552	PQ343553	PQ343554	PQ343555	PQ343556	PQ343557
49	*rps4*	PQ343567	PQ343568	PQ343569	PQ343570	PQ343571	PQ343572	PQ343573	PQ343574	PQ343575	PQ343576	PQ343577	PQ343578	PQ343579	PQ343580	PQ343581	PQ343582
50	*rps7*	PQ343592	PQ343593	PQ343594	PQ343595	PQ343596	PQ343597	PQ343598	PQ343599	PQ343600	PQ343601	PQ343602	PQ343603	PQ343604	PQ343605	PQ343606	PQ343607
51	*rps11*	PQ343617	PQ343618	PQ343619	PQ343620	PQ343621	PQ343622	PQ343623	PQ343624	PQ343625	PQ343626	PQ343627	PQ343628	PQ343629	PQ343630	PQ343631	PQ343632
52	*rps14*	PQ343642	PQ343643	PQ343644	PQ343645	PQ343646	PQ343647	PQ343648	PQ343649	PQ343650	PQ343651	PQ343652	PQ343653	PQ343654	PQ343655	PQ343656	PQ343657
53	*rps15*	PQ343667	PQ343668	PQ343669	PQ343670	PQ343671	PQ343672	PQ343673	PQ343674	PQ343675	PQ343676	PQ343677	PQ343678	PQ343679	PQ343680	PQ343681	PQ343682
54	*rps18*	PQ343692	PQ343693	PQ343694	PQ343695	PQ343696	PQ343697	PQ343698	PQ343699	PQ343700	PQ343701	PQ343702	PQ343703	PQ343704	PQ343705	PQ343706	PQ343707
55	*rps19*	PQ343717	PQ343718	PQ343719	PQ343720	PQ343721	PQ343722	PQ343723	PQ343724	PQ343725	PQ343726	PQ343727	PQ343728	PQ343729	PQ343730	PQ343731	PQ343732
56	*ycf3*	PQ342667	PQ342668	PQ342669	PQ342670	PQ342671	PQ342672	PQ342673	PQ342674	PQ342675	PQ342676	PQ342677	PQ342678	PQ342679	PQ342680	PQ342681	PQ342682
57	*ycf4*	PQ343742	PQ343743	PQ343744	PQ343745	PQ343746	PQ343747	PQ343748	PQ343749	PQ343750	PQ343751	PQ343752	PQ343753	PQ343754	PQ343755	PQ343756	PQ343757

Table S10. List of accession numbers of plastid intergenic regions used in this study.

No.	Taxa / Intergenic spacer	*P.× rhaetica* 10	*P.× rhaetica* 11	*P.× rhaetica* 12	*P. mugo* 1	*P. mugo* 2	*P. mugo* 3	*P. mugo* 4	*P. uliginosa* 5	*P. uliginosa* 6	*P. uliginosa* 7	*P. uliginosa* 8	*P. uliginosa* 9	*P. uncinata* 31	*P. uncinata* 32	*P. sylvestris* 29	*P. sylvestris* 30
1	*accD-psaI*	PQ342383	PQ342384	PQ342385	PQ165116	PQ342386	PQ342387	MZ333466	PQ165118	PQ342388	PQ165119	PQ165117	MZ333465	PQ342389	MZ333464	PQ342390	PQ342391
2	*rps2-atpI*	PQ342558	PQ342559	PQ342560	PQ165116	PQ342561	PQ342562	MZ333466	PQ165118	PQ342563	PQ165119	PQ165117	MZ333465	PQ342564	MZ333464	PQ342565	PQ34256
3	*clpP-rps12*	PQ343758	PQ343759	PQ343760	PQ165116	PQ343761	PQ343762	MZ333466	PQ165118	PQ343763	PQ165119	PQ165117	MZ333465	PQ343764	MZ333464	PQ343765	PQ343766
4	*petB-petD*	PQ342433	PQ342434	PQ342435	PQ165116	PQ342436	PQ342437	MZ333466	PQ165118	PQ342438	PQ165119	PQ165117	MZ333465	PQ342439	MZ333464	PQ342440	PQ342441
5	*psaM-trnS*	PQ342458	PQ342459	PQ342460	PQ165116	PQ342461	PQ342462	MZ333466	PQ165118	PQ342463	PQ165119	PQ165117	MZ333465	PQ342464	MZ333464	PQ342465	PQ342466
6	*psbJ-petA*	PQ342483	PQ342484	PQ342485	PQ165116	PQ342486	PQ342487	MZ333466	PQ165118	PQ342488	PQ165119	PQ165117	MZ333465	PQ342489	MZ333464	PQ342490	PQ342491
7	*psbL-psbJ*	PQ342483	PQ342484	PQ342485	PQ165116	PQ342486	PQ342487	MZ333466	PQ165118	PQ342488	PQ165119	PQ165117	MZ333465	PQ342489	MZ333464	PQ342490	PQ342491
8	*trnS-psbZ*	PQ342508	PQ342509	PQ342510	PQ165116	PQ342511	PQ342512	MZ333466	PQ165118	PQ342513	PQ165119	PQ165117	MZ333465	PQ342514	MZ333464	PQ342515	PQ342516
9	*atpB-rbcL*	PQ342408	PQ342409	PQ342410	PQ165116	PQ342411	PQ342412	MZ333466	PQ165118	PQ342413	PQ165119	PQ165117	MZ333465	PQ342414	MZ333464	PQ342415	PQ342416
10	*rpoB-trnC*	PQ342533	PQ342534	PQ342535	PQ165116	PQ342536	PQ342537	MZ333466	PQ165118	PQ342538	PQ165119	PQ165117	MZ333465	PQ342539	MZ333464	PQ342539	PQ342539
11	*rrn4.5-rrn5*	PQ342583	PQ342584	PQ342585	PQ165116	PQ342586	PQ342587	MZ333466	PQ165118	PQ342588	PQ165119	PQ165117	MZ333465	PQ342589	MZ333464	PQ342590	PQ342591
12	*trnC-petN*	PQ342533	PQ342534	PQ342535	PQ165116	PQ342536	PQ342537	MZ333466	PQ165118	PQ342538	PQ165119	PQ165117	MZ333465	PQ342539	MZ333464	PQ342539	PQ342539
13	*trnD-psbM*	PQ342608	PQ342609	PQ342610	PQ165116	PQ342611	PQ342612	MZ333466	PQ165118	PQ342613	PQ165119	PQ165117	MZ333465	PQ342614	MZ333464	PQ342615	PQ342616
14	*trnE-clpP*	PQ343758	PQ343759	PQ343760	PQ165116	PQ343761	PQ343762	MZ333466	PQ165118	PQ343763	PQ165119	PQ165117	MZ333465	PQ343764	MZ333464	PQ343765	PQ343766
15	*trnL-trnF*	PQ343783	PQ343784	PQ343785	PQ165116	PQ343786	PQ343787	MZ333466	PQ165118	PQ343788	PQ165119	PQ165117	MZ333465	PQ343789	MZ333464	PQ343790	PQ343791
16	*trnM-ndhC*	PQ342633	PQ342634	PQ342635	PQ165116	PQ342636	PQ342637	MZ333466	PQ165118	PQ342638	PQ165119	PQ165117	MZ333465	PQ342639	MZ333464	PQ342640	PQ342641
17	*trnS-psaM*	PQ342683	PQ342684	PQ342685	PQ165116	PQ342686	PQ342687	MZ333466	PQ165118	PQ342688	PQ165119	PQ165117	MZ333465	PQ342689	MZ333464	PQ342690	PQ342691
18	*ycf3-psaA*	PQ342658	PQ342659	PQ342660	PQ165116	PQ342661	PQ342662	MZ333466	PQ165118	PQ342663	PQ165119	PQ165117	MZ333465	PQ342664	MZ333464	PQ342665	PQ342666

No.	Taxa / Intergenic spacer	*P. sylvestris* 28	*P. rotundata* 25	*P. rotundata* 26	*P. rotundata* 27	*P. rotundata* 13	*P. rotundata* 14	*P. rotundata* 15	*P. rotundata* 22	*P. rotundata* 23	*P. rotundata* 24	*P. rotundata* 16	*P. rotundata* 17	*P. rotundata* 18	*P. rotundata* 19	*P. rotundata* 20	*P. rotundata* 21
1	*accD-psaI*	PQ342392	PQ342393	PQ342394	PQ342395	PQ342396	PQ342397	PQ342398	PQ342399	PQ342400	PQ342401	PQ342402	PQ342403	PQ342404	PQ342405	PQ342406	PQ342407
2	*rps2-atpI*	PQ342567	PQ342568	PQ342569	PQ342570	PQ342571	PQ342572	PQ342573	PQ342574	PQ342575	PQ342576	PQ342577	PQ342578	PQ342579	PQ342580	PQ342581	PQ342582
3	*clpP-rps12*	PQ343767	PQ343768	PQ343769	PQ343770	PQ343771	PQ343772	PQ343773	PQ343774	PQ343775	PQ343776	PQ343777	PQ343778	PQ343779	PQ343780	PQ343781	PQ343782

4	*petB-petD*	PQ342442	PQ342443	PQ342444	PQ342445	PQ342446	PQ342447	PQ342448	PQ342449	PQ342450	PQ342451	PQ342452	PQ342453	PQ342454	PQ342455	PQ342456	PQ342457
5	*psaM-trnS*	PQ342467	PQ342468	PQ342469	PQ342470	PQ342471	PQ342472	PQ342473	PQ342474	PQ342475	PQ342476	PQ342477	PQ342478	PQ342479	PQ342480	PQ342481	PQ342482
6	*psbJ-petA*	PQ342492	PQ342493	PQ342494	PQ342495	PQ342496	PQ342497	PQ342498	PQ342499	PQ342500	PQ342501	PQ342502	PQ342503	PQ342504	PQ342505	PQ342506	PQ342507
7	*psbL-psbJ*	PQ342492	PQ342493	PQ342494	PQ342495	PQ342496	PQ342497	PQ342498	PQ342499	PQ342500	PQ342501	PQ342502	PQ342503	PQ342504	PQ342505	PQ342506	PQ342507
8	*trnS-psbZ*	PQ342517	PQ342518	PQ342519	PQ342520	PQ342521	PQ342522	PQ342523	PQ342524	PQ342525	PQ342526	PQ342527	PQ342528	PQ342529	PQ342530	PQ342531	PQ342532
9	*atpB-rbcL*	PQ342417	PQ342418	PQ342419	PQ342420	PQ342421	PQ342422	PQ342423	PQ342424	PQ342425	PQ342426	PQ342427	PQ342428	PQ342429	PQ342430	PQ342431	PQ342432
10	*rpoB-trnC*	PQ342542	PQ342543	PQ342544	PQ342545	PQ342546	PQ342547	PQ342548	PQ342549	PQ342550	PQ342551	PQ342552	PQ342553	PQ342554	PQ342555	PQ342556	PQ342557
11	*rrn4.5-rrn5*	PQ342592	PQ342593	PQ342594	PQ342595	PQ342596	PQ342597	PQ342598	PQ342599	PQ342600	PQ342601	PQ342602	PQ342603	PQ342604	PQ342605	PQ342606	PQ342607
12	*trnC-petN*	PQ342542	PQ342543	PQ342544	PQ342545	PQ342546	PQ342547	PQ342548	PQ342549	PQ342550	PQ342551	PQ342552	PQ342553	PQ342554	PQ342555	PQ342556	PQ342557
13	*trnD-psbM*	PQ342617	PQ342618	PQ342619	PQ342620	PQ342621	PQ342622	PQ342623	PQ342624	PQ342625	PQ342626	PQ342627	PQ342628	PQ342629	PQ342630	PQ342631	PQ342632
14	*trnE-clpP*	PQ343767	PQ343768	PQ343769	PQ343770	PQ343771	PQ343772	PQ343773	PQ343774	PQ343775	PQ343776	PQ343777	PQ343778	PQ343779	PQ343780	PQ343781	PQ343782
15	*trnL-trnF*	PQ343792	PQ343793	PQ343794	PQ343795	PQ343796	PQ343797	PQ343798	PQ343799	PQ343800	PQ343801	PQ343802	PQ343803	PQ343804	PQ343805	PQ343806	PQ343807
16	*trnM-ndhC*	PQ342642	PQ342643	PQ342644	PQ342645	PQ342646	PQ342647	PQ342648	PQ342649	PQ342650	PQ342651	PQ342652	PQ342653	PQ342654	PQ342655	PQ342656	PQ342657
17	*trnS-psaM*	PQ342692	PQ342693	PQ342694	PQ342695	PQ342696	PQ342697	PQ342698	PQ342699	PQ342700	PQ342701	PQ342702	PQ342703	PQ342704	PQ342705	PQ342706	PQ342707
18	*ycf3-psaA*	PQ342667	PQ342668	PQ342669	PQ342670	PQ342671	PQ342672	PQ342673	PQ342674	PQ342675	PQ342676	PQ342677	PQ342678	PQ342679	PQ342680	PQ342681	PQ342682

5. PODSUMOWANIE

W niniejszej rozprawie doktorskiej po raz pierwszy dokonano szczegółowej analizy porównawczej kompletnych sekwencji genomów chloroplastowych trzech głównych przedstawicieli kompleksu *Pinus mugo,* tj. *Pinus mugo, P. rotundata* i *P. uncinata*. Wyniki ukazały wysoką konserwatywność genomów chloroplastowych pod względem długości, struktury i liczby genów. Analiza porównawcza kompletnych plastomów ujawniła występowanie kilku wysoce zmiennych rejonów (tzw. hotspotów) oraz *loci* mikrosatelitarnych, potencjalnie przydatnych do identyfikacji poszczególnych taksonów z kompleksu *Pinus mugo*. Wnioskowanie filogenetyczne potwierdziło bliskie relacje genetyczne pomiędzy analizowanymi taksonami i wykazało, że sosny z kompleksu *Pinus mugo* tworzą jedną odrębną grupę w obrębie rodzaju *Pinus* z silnym wsparciem.

Wyodrębnione sekwencje jądrowego regionu ITS2 (ang. *Internal Transcribed Spacer* 2) posłużyły do oceny jego przydatności do dyskryminacji taksonów drzew iglastych i wnioskowania filogenetycznego u *Pinaceae,* w tym także dla kompleksu *Pinus mugo*. Pomimo iż, ITS2 posiada cechy idealnego regionu barkodowego DNA, to jego przydatność do rozróżniania taksonów wśród rodziny *Pinaceae* jest mocno ograniczona. Wykazano, że wartości podstawowych parametrów opisujących zmienność genetyczną były najniższe na poziomie blisko spokrewnionych taksonów, tj. kompleksu *Pinus mugo*, u którego nie zaobserwowano żadnej zmienności w obrębie analizowanej sekwencji, a najwyższe na poziomie rodziny. Skuteczna identyfikacja taksonów w rodzinie *Pinaceae* z wykorzystaniem regionu ITS2 jest możliwa wyłącznie dla tych taksonów, które są filogenetycznie odległe i reprezentują raczej odmienne rodzaje niż gatunki.

Na podstawie wyników uzyskanych przez analizę siedmiu wygenerowanych zestawów sekwencji nukleotydowych pochodzących z trzech genomów organellowych, wywnioskowano, że nazwa *Pinus* × *rhaetica* nie powinna być używana jako synonim *Pinus uliginosa* i *Pinus rotundata*, ponieważ różni się od nich znacznie pod względem genetycznym i najprawdopodobniej jest mieszańcem lub introgresantem, pomiędzy *Pinus sylvestris* i *Pinus mugo,* jak postulowały niektóre wcześniejsze badania. Co więcej, *Pinus uliginosa* i *Pinus rotundata*, pomimo różnych nazw, pochodzenia i historii populacji, posiadają zbliżone tło genetyczne, które jest prawdopodobnie kształtowane przez presję środowiskową i adaptację do zajmowanego przez nie siedliska. *Pinus uncinata* stanowi wyraźnie odrębny klad w obrębie kompleksu *Pinus mugo* w większości zastosowanych metod i zestawów danych genetycznych. Analiza zróżnicowania genetycznego na poziomie populacji pozwala na

wyraźne rozróżnienie populacji i taksonów z kompleksu *Pinus mugo* od *Pinus sylvestris* i *Pinus* × *rhaetica*.

Podsumowując, zastosowanie sekwencjonowania nowej generacji (ang. *next-generation sequencing*, NGS) i metody przeczesywania genomu (ang. *genome skimming*) umożliwiło uzyskanie bardzo dużej ilości danych sekwencyjnych, które pozwoliły na wieloregionową i wielogenomową charakterystykę zmienności w obrębie kompleksu *Pinus mugo*. Wyniki uzyskane przy pomocy tej metody zwiększają wiedzę na temat różnorodności genetycznej, jak również filogenezy kompleksu *Pinus mugo* i zapewniają cenną genomiczną podstawę dla przyszłych badań nad historią ewolucji i ochroną tej wysoce polimorficznej grupy, a także rodziny *Pinaceae*. Wskazują również, które regiony mogą być przydatne w poszukiwaniu diagnostycznych markerów DNA dla przedstawicieli kompleksu *Pinus mugo* i wyznaczają punkt odniesienia w ochronie zasobów genetycznych jego zagrożonych taksonów.

6. BIBLIOGRAFIA

1. Rozporządzenie Ministra Środowiska z dnia 9 października 2014 r. w sprawie ochrony gatunkowej roślin. **2014.** Dz. U., Poz. 1409.

2. Agarwal, M.; Shrivastava, N.; Padh, H. Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Rep*. **2008**, 27, 617–631. https://doi.org/10.1007/s00299-008-0507-z

3. Álvarez, I.; Wendel, J. F. Ribosomal ITS sequences and plant phylogenetic inference. *Mol. Phylogenet. Evol.* **2003**, *29* (3), 417–434. https://doi.org/10.1016/S1055-7903(03)00208-2.

4. Bastl, M.; Burian, M.; Kučera, J.; Prach, K.; Rektoris, L.; Štech, M. Central European pine bogs change along an altitudinal gradient. *Preslia* **2008**, *80* (4), 349–363.

5. Bobowicz, A. M.; Danielelewicz, W.; Pieczyńska, B.; Wojnicka-Półtorak, A.; Prus-Głowacki, W. Isoenzymatic variability in progeny of *Pinus mugo* Turra × *Pinus sylvestris* L. hybrids from Bór nad Czerwonem, in experimental culture. *Acta Soc. Bot. Pol.* **2000**, *69* (2), 137–144.

6. Bogunić, F.; Siljak-Yakovlev, S.; Muratović, E.; Pustahija, F.; Medjedović, S. Molecular cytogenetics and flow cytometry reveal conserved genome organization in *Pinus mugo* and *P. uncinata*. *Annals of Forest Science*. **2011**, *68*, 179–187. https://doi.org/10.1007/s13595-011-0019-9.

7. Bonikowski, R.; Celiński, K.; Wojnicka-Półtorak, A.; Maliński, T. Composition of essential oils isolated from the needles of *Pinus uncinata* and *P. uliginosa* grown in Poland. *Nat. Prod. Commun.* **2015**, *10* (2), 371–373. https://doi.org/10.1177/1934578x1501000243.

8. Boratyńska, K.; Bobowicz, M. A. *Pinus uncinata* Ramond Taxonomy based on needle characters. *Plant Syst. Evol.* **2001**, *227*, 183–194. https://doi.org/10.1007/s006060170047.

9. Boratyńska, K. Relacje Taksonomiczne między sosnami z kompleksu *Pinus mugo* (Pinaceae) na podstawie cech igieł. *Fragm. Florist. Geobot. Pol.* **2004**, *11* (2), 235–255.

10. Boratyńska, K.; Boratyński, A. Taxonomic differences among closely related *Pines Pinus sylvestris, P. mugo, P. uncinata, P. rotundata* and *P. uliginosa* as revealed in needle sclerenchyma cells. *Flora Morphol. Distrib. Funct. Ecol. Plants* **2007**, *202* (7), 555–569. https://doi.org/10.1016/j.flora.2006.11.004.

11. Boratyńska, K.; Boratyński, A.; Lewandowski, A. Morphology of *Pinus uliginosa* (Pinaceae) needles from populations exposed to and isolated from the direct influence of *Pinus sylvestris*. *Bot. J. Linn. Soc.* **2003**, *124*, 83–91. https://doi.org/10.1046/j.1095-8339.2003.00156.x.

12. Boratyńska, K.; Jasińska, A. K.; Boratyński, A. Taxonomic and geographic differentiation of *Pinus mugo* complex on the needle characteristics. *Syst. Biodivers.* **2015**, *13* (6), 581–595. https://doi.org/10.1080/14772000.2015.1058300.

13. Businský, R. Taxonomická studie agregátu *Pinus mugo* a jeho hybridních populací. W: Studium Domácích a Introdukovaných Druhů Rodu *Pinus*. Sborník Referátů. *Acta Průhoniciana* **1999**, *68*, 123–144.

14. Businský, R.; Kirschner, J. Nomenclatural notes on the *Pinus mugo* complex in central europe. *Phyt. - Ann. Rei Bot.* **2006**, *46* (1), 129–139.

15. Celiński, K.; Sokołowska, J.; Zemleduch-Barylska, A.; Kuna, R.; Kijak, H.; Staszak, A.M.; Wojnicka-Półtorak, A.; Chudzińska, E. Seed total protein profiling in discrimination of closely related pines: evidence from the *Pinus mugo* complex. Plants **2020**, *9*(7), 872. https://doi.org/10.3390/plants9070872.

16. Celiński, K.; Chudzińska, E.; Gmur, A.; Piosik, Ł.; Wojnicka-Półtorak, A. Cytological characterization of three closely related pines - *Pinus mugo, P. uliginosa* and *P. × rhaetica* from the *Pinus mugo* complex (Pinaceae). *Biologia (Bratisl).* **2019**, *74* (7), 751–756. https://doi.org/10.2478/s11756-019-00201-6.

17. Celiński, K.; Kijak, H.; Wojnicka-Półtorak, A.; Buczkowska-Chmielewska, K.; Sokołowska, J.; Chudzińska, E. Effectiveness of the DNA barcoding approach for closely related conifers discrimination: a case study of the *Pinus mugo* complex. *Comptes Rendus Biol.* **2017a**, *340* (6–7), 339–348. https://doi.org/10.1016/j.crvi.2017.06.002.

18. Celiński, K.; Kijak, H.; Barylski, J.; Grabsztunowicz, M.; Wojnicka-Półtorak, A.; Chudzińska, E. Characterization of the complete chloroplast genome of *Pinus uliginosa* (Neumann) from the *Pinus mugo* complex. *Conserv. Genet. Resour.* **2017b**, *9* (2), 209–212. https://doi.org/10.1007/s12686-016-0652-6.

19. Celiński, K.; Bonikowski, R.; Wojnicka-Półtorak, A.; Chudzińska, E., Maliński, T. Volatiles as chemosystematic markers for distinguishing closely related species within the *Pinus mugo* complex. *Chem Biodivers.* **2015**, *12*(8), 1208–1213. https://doi.org/10.1002/cbdv.201400253.

20. China Plant BOL Group; Li, D.Z.; Gao, L.M.; Li, H.T.; Wang, H.; Ge, X.J.; Liu, J.Q.; Chen, Z.D.; Zhou, S.L.; Chen, S.L.; Yang, J.B.; Fu, C.X.; Zeng, C.X.; Yan, H.F., Zhu, Y.J.; Sun, Y.S.; Chen, S.Y.; Zhao, L.; Wang, K.; Yang, T.; Duan, G.W. Comparative analysis of a large dataset indicates that Internal Transcribed Spacer (ITS) should be incorporated into the core barcode for seed plants. P*roc. Natl. Acad. Sci. USA.* **2011**, *108*(49), 19641–19646. https://doi.org/10.1073/pnas.1104551108.

21. Christ, H. Ubersicht Der Europaischen Ahietineen (*Pinus* L.). *Verh. Naturf. Ges. Basel* **1863**, *3*, 1–19.

22. Christensen, K. I. A Morphometric Study of the *Pinus mugo* turra complex and its natural hybridization with *P. sylvestris* L. (Pinaceae). *Feddes Repert.* **1987a**, *98*, 623–635.

23. Christensen, K. I. Taxonomic Revision of the *Pinus mugo* complex and *P. × rhaetica* (*P. mugo × sylvestris*) (Pinaceae). *Nord. J. Bot.* **1987b**, *7*, 383–408.

24. Dabral, A.; Shamoon, A.; Meena, R. K.; Kant, R.; Pandey, S.; Ginwal, H. S.; Bhandari, M. S. Genome skimming-based simple sequence repeat (SSR) marker discovery and characterization in *Grevillea robusta*. *Physiol. Mol. Biol. Plants* **2021**, *27* (7), 1623–1638. https://doi.org/10.1007/s12298-021-01035-w.

25. de Boer, H.; Rydmark, M. O.; Verstraete, B.; Gravendeel, B. Molecular Identification of plants: from sequence to species. *Advanced Books* **2022,** pp 1–396. https://doi.org/10.3897/ab.e98875.

26. de Queiroz, K. Species concepts and species delimitation. *Syst. Biol.* **2007**, *56* (6), 879–886. https://doi.org/10.1080/10635150701701083.

27. Dodsworth, S. Genome skimming for next-generation biodiversity analysis. *Trends Plant Sci.* **2015**, *20* (9), 525–527. https://doi.org/10.1016/j.tplants.2015.06.012.

28. Dodsworth, S.; Pokorny, L.; Johnson, M. G.; Kim, J. T.; Maurin, O.; Wickett, N. J.; Forest, F.; Baker, W. J. Hyb-Seq for flowering plant systematics. *Trends Plant Sci.* **2019**, *24* (10), 887–891. https://doi.org/10.1016/j.tplants.2019.07.011.

29. Domin, K. Plantarum Čechoslovakiae Enumeratio. *Preslia* **1936**, 13-15.

30. Fonseca, L. H. M.; Lohmann, L. G. Exploring the potential of nuclear and mitochondrial sequencing data generated through genome-skimming for plant phylogenetics: a case study from a clade of neotropical lianas. *J. Syst. Evol.* **2020**, *58* (1), 18–32. https://doi.org/10.1111/jse.12533.

31. Franco, J. M. A. P. do A. *Pinus* L. W: Flora iberica. Plantas vasculares de la Península Ibérica e Islas Baleares.Vol. I. *Lycopodiaceae-Papaveraceae*; Castroviejo Bolibar, Santiago ; Laínz Gallo, Manuel ; López González, Ginés Alejandro ; Montserrat Recoder, Pedro ; Muñoz Garmendia, Félix ; Paiva, Jorge Américo Rodrigues ; Villar Pérez, L., Ed.; Real Jardín Botánico, C.S.I.C.: Madrid, **1986**; pp 168–174.

32. Gernandt, D. S.; Geada López, G.; Ortiz García, S.; Liston, A. Phylogeny and classification of *Pinus*. *Taxon* **2005**, *54* (1), 29–42. https://doi.org/10.2307/25065300.

33. Gołąb Z. Sosna błotna (*Pinus uliginosa* Neumann) na Wielkim Torfowisku Batorowskim w Górach Stołowych. *Szczeliniec* **1999**, *3*, 41–48.

34. Goppert, H. R. Bemerkungen uber die formen der *Pinus montana* Mill. *Bot. Zeitung Berlin* **1864**, *22*, 41–43.

35. Grube, M.; Kroken, S. Molecular Approaches and the concept of species and species complexes in lichenized fungi. *Mycol. Res.* **2000**, *104* (11), 1284–1294. https://doi.org/10.1017/S0953756200003476.

36. Hamerník, J.; Musil, I. The *Pinus mugo* complex - Its structuring and general overview of the used nomenclature. *Journal of Forest Science*. **2007**, 53(6), 253-266. https://doi.org/10.17221/2020-jfs.

37. Harrison, R. G.; Larson, E. L. Hybridization, introgression, and the nature of species boundaries. *J. Hered.* **2014**, *105* (S1), 795–809. https://doi.org/10.1093/jhered/esu033.

38. He, J.; Lyu, R.; Luo, Y.; Lin, L.; Yao, M.; Xiao, J.; Xie, L.; Wen, J.; Pei, L.; Yan, S. An updated phylogenetic and biogeographic analysis based on genome skimming data reveals convergent evolution of shrubby habit in *Clematis* in the pliocene and pleistocene. *Mol. Phylogenet. Evol.* **2021**, *164* (December 2020), 107259. https://doi.org/10.1016/j.ympev.2021.107259.

39. Heath, T. A.; Hedtke, S. M.; Hillis, D. M. Taxon sampling and the accuracy of phylogenetic analyses. *J. Syst. Evol.* **2008**, *46* (3), 239–257. https://doi.org/10.3724/SP.J.1002.2008.08016.

40. Hebert, P. D. N.; Cywinska, A.; Ball, S. L.; DeWaard, J. R. Biological identifications through DNA barcodes. *Proc. R. Soc. B Biol. Sci.* **2003**, *270*, 313–322. https://doi.org/10.1098/rspb.2002.2218.

41. Hernández-León, S.; Gernandt, D. S.; Pérez de la Rosa, J. A.; Jardón-Barbolla, L. Phylogenetic relationships and species delimitation in *Pinus* section *Trifoliae* inferrred from plastid DNA. *PLoS One* **2013**, *8* (7), e705. https://doi.org/10.1371/journal.pone.0070501.

42. Heuertz, M.; Teufel, J.; González-Martínez, S. C.; Soto, A.; Fady, B.; Alía, R.; Vendramin, G. G. Geography determines genetic relationships between species of mountain pine (*Pinus mugo* complex) in Western Europe. *J. Biogeogr.* **2010**, *37* (3), 541–556. https://doi.org/10.1111/j.1365-2699.2009.02223.x.

43. Hudson, R. R.; Coyne, J. A. Mathematical Consequences of the genealogical species concept. *Evolution (N. Y).* **2002**, *56* (8), 1557–1565. https://doi.org/10.1111/j.0014-3820.2002.tb01467.x.

44. Jalas, J.; Suominen, J. Atlas Florae Europaeae: Distribution of vascular plants in europe. Gymnospermae, Volume 2. The Commitee for Mapping the Flora of Europe and Soc. Biol. Fenn: Vanamo, Helsinki, **1973**; p 40.

45. Jarman, S. N.; Elliott, N. G. DNA Evidence for Morphological and cryptic cenozoic speciations in the *Anaspididae*, "living fossils" from the triassic. *J. Evol. Biol.* **2000**, *13* (4), 624–633. https://doi.org/10.1046/j.1420-9101.2000.00207.x.

46. Jasičová, M. *Pinus* L. In *Flóra Slovenska 2*; Futák, J., Ed.; Vydavatel'stvo Slovenskej akadémie vied, **1966**; pp 278 – 294.

47. Ji, Y.; Yang, J.; Landis, J. B.; Wang, S.; Jin, L.; Xie, P.; Liu, H.; Yang, J. B.; Yi, T. S. Genome Skimming contributes to clarifying species limits in *Paris* section *Axiparis* (*Melanthiaceae*). *Front. Plant Sci.* **2022**, *13* (4), 1–13. https://doi.org/10.3389/fpls.2022.832034.

48. Kang, H.-I.; Lee, H.O.; Lee, I.H.; Kim, I.S.; Lee, S.-W.; Yang, T.J.; Shim, D. Complete chloroplast genome of *Pinus densiflora* Siebold & Zucc. and comparative analysis with five pine trees. *Forests* **2019**, *10* (7), 600. https://doi.org/10.3390/f10070600.

49. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28* (12), 1647–1649. https://doi.org/10.1093/bioinformatics/bts199.

50. Lewandowski A.; Boratyński 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. https://doi.org/10.1007/BF01086377.

51. Lewandowski, A.; Wiśniewska, M. Short note: Crossability between *Pinus uliginosa* and its putative parental species *Pinus sylvestris* and *Pinus mugo*. *Silvae Genet.* **2006**, *55* (2), 52–54. https://doi.org/10.1515/sg-2006-0008.

52. Li, X.; Yang, Y.; Henry, R. J.; Rossetto, M.; Wang, Y.; Chen, S. Plant DNA barcoding : from gene to genome. *Biol. Rev. Camb. Philos. Soc.* **2014**, *90* (1), 157–166. https://doi.org/10.1111/brv.12104.

53. Liu, B. Bin; Hong, D. Y.; Zhou, S. L.; Xu, C.; Dong, W. P.; Johnson, G.; Wen, J. Phylogenomic analyses of the photinia complex support the recognition of a new genus *Phippsiomeles* and the resurrection of a redefined *Stranvaesia* in *Maleae* (Rosaceae). *J. Syst. Evol.* **2019**, *57* (6), 678–694. https://doi.org/10.1111/jse.12542.

54. Liu, L.; Low, S. L.; Sakaguchi, S.; Feng, Y.; Ge, B.; Konowalik, K.; Li, P. Development of nuclear and chloroplast polymorphic microsatellites for *Crossostephium chinense* (Asteraceae). *Mol. Biol. Rep.* **2021**, *48* (9), 6259–6267. https://doi.org/10.1007/s11033-021-06590-9.

55. Łabiszak, B.; Zaborowska, J.; Wachowiak, W. Patterns of mtDNA variation reveal complex evolutionary history of relict and endangered peat bog pine (*Pinus uliginosa*). *AoB Plants* **2019**, 11(2), plz015. https://doi.org/10.1093/aobpla/plz015.

56. Malé, P. J. G.; Bardon, L.; Besnard, G.; Coissac, E.; Delsuc, F.; Engel, J.; Lhuillier, E.; Scotti-Saintagne, C.; Tinaut, A.; Chave, J. Genome skimming by shotgun sequencing helps resolve the phylogeny of a pantropical tree family. *Mol. Ecol. Resour.* **2014**, *14* (5), 966–975. https://doi.org/10.1111/1755-0998.12246.

57. Marcysiak, K.; Boratyński, A. Contribution to the taxonomy of *Pinus uncinata* (Pinaceae) based on cone characters. *Plant Syst. Evol.* **2007**, *264* (1–2), 57–73. https://doi.org/10.1007/s00606-006-0501-2.

58. Mascarello, M.; Amalfi, M.; Asselman, P.; Smets, E.; Hardy, O. J.; Beeckman, H.; Janssens, S. B. Genome skimming reveals novel plastid markers for the molecular identification of illegally logged african timber species. *PLoS One* **2021**, *16* (6 June 2021), 1–18. https://doi.org/10.1371/journal.pone.0251655.

59. 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* (3), 391–406. https://doi.org/10.14214/sf.476.

60. Neumann, C. Über eine auf den Seefeldern bei Reinerz u. einigen ähnlichen Gebirgsmooren der königl. Oberförsterei Karlsberg in der Graftschaft Glatz vorkommende noch unbeschrieben Form der Gattung *Pinus*. Jahresber. *Schlesische Gesellschaft für Vaterländische Kult.* **1837**, *11*, 52–57.

61. Nevill, P. G.; Zhong, X.; Tonti-Filippini, J.; Byrne, M.; Hislop, M.; Thiele, K.; Van Leeuwen, S.; Boykin, L. M.; Small, I. Large Scale genome skimming from herbarium material for accurate plant identification and phylogenomics. *Plant Methods* **2020**, *16* (1), 1–8. https://doi.org/10.1186/s13007-019-0534-5.

62. Olsson, S.; Grivet, D.; Cid Vian, J. Species-diagnostic markers in the genus *Pinus*: Evaluation of the chloroplast regions *matK* and *ycf1*. *For. Syst.* **2018**, *27* (3). https://doi.org/10.5424/fs/2018273-13688.

63. Parks, M.; Cronn, R.; Liston, A. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biol.* **2009**, *7* (84). https://doi.org/10.1186/1741-7007-7-84.

64. Prus-Głowacki, W.; Bujas, E.; Ratyńska, H. Taxonomic position of *Pinus uliginosa* Neumann as related to other taxa of *Pinus mugo* complex. *Acta Soc. Bot. Pol.* **1998**, *67* (3–4), 269–274. https://doi.org/10.5586/asbp.1998.035.

65. Prus-Głowacki, W.; Szweykowski, J. Studies on antigenic differences in needle proteins of *Pinus sylvestris* L., *P. mugo* Turra, *P. uliginosa* Neumann and *P. nigra* Arnold. *Acta Soc. Bot. Pol.* **1979**, *48*, 138–217. https://doi.org/10.5586/asbp.1979.019.

66. Prus-Głowacki, W.; Szweykowski, J. Studies on isoenzyme variability in populations of *Pinus sylvestris* L., *Pinus mugo* Turra, *Pinus uliginosa* Neumann and individuals from a hybrid swarm population. *Bull. la Soc. des Amis des Aciences des Lettres Poznań* **1983**, *Ser. D*, *22*, 107–122.

67. Prus-Głowacki, W.; Szweykowski, J.; Nowak, R. Serotaxonomical Investigation of the european pine species. *Silvae Genet.* **1985**, *34* (4), 162–170.

68. Ragupathy, S.; Newmaster, S. G.; Murugesan, M.; Balasubramaniam, V. DNA barcoding discriminates a new cryptic grass species revealed in an ethnobotany study by the hill tribes of the Western Ghats in Southern India. *Mol. Ecol. Resour.* **2009**, *9* (SUPPL. 1), 164–171. https://doi.org/10.1111/j.1755-0998.2009.02641.x.

69. Schlechtendal, D. F. L. De *Pinastris germaniae* et helvetiae observationes. *Linnaea* **1857**, *29*, 357–384.

70. Shaffer, H. B., and Thomson, R. C. Delimiting species in recent radiations. *Syst. Biol.* **2007**, *56*, 896–906. doi:10.1080/10635150701772563.

71. Shaw, G. R. The Genus *Pinus*. *J. Arnold Arbor.* **1914**, *5*, 1–96.

72. Simmonds, S. E.; Smith, J. F.; Davidson, C.; Buerki, S. Phylogenetics and comparative plastome genomics of two of the largest genera of angiosperms, *Piper* and *Peperomia* (Piperaceae). *Mol. Phylogenet. Evol.* **2021**, *163* (June), 107229. https://doi.org/10.1016/j.ympev.2021.107229.

73. Skalická, A.; Skalický, V. *Pinus* L. W: *Květena České Socialistické Republiky 1*; Hejny, S., Slavik, B., Eds.; Akademia: Praha, **1988**; pp 289–308.

74. Skalický, V. *Pinus sylvestris, P. mugo, P. rotundata, P. × pseudopumilio, P. × digenea*. W: *Květena České republiky 1*; Hejný, S., Slavík, B., Eds.; Academia: Praha, 1988; pp 291–298, 308.

75. Staszkiewicz, J. *Pinus* × *rhaetica* Brūgger sosna drzewokosa. W: *Polska Czerwona Księga Roślin : paprotniki i rośliny kwiatowe*; Kaźmierczakowa, R., Zarzycki, K., Eds.; Instytut Ochrony Przyrody Polskiej Akademii Nauk: Kraków, **1993**; pp 38–39.

76. Straub, S. C. K.; Parks, M.; Weitemier, K.; Fishbein, M.; Cronn, R. C.; Liston, A. Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. *Am. J. Bot.* **2012**, *99* (2), 349–364. https://doi.org/10.3732/ajb.1100335.

77. Wachowiak, W.; Prus-Głowacki, W. Hybridisation processes in sympatric populations of pines *Pinus sylvestris* L., *P. mugo* Turra and *P. uliginosa* Neumann. *Plant Syst. Evol.* **2008**, *271* (1–2), 29–40. https://doi.org/10.1007/s00606-007-0609-z.

78. Wachowiak, W.; PalmÉ, 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* (8), 1729–1743. https://doi.org/10.1111/j.1365-294X.2011.05037.x.

79. Wachowiak, W.; Boratyńska, 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. https://doi.org/10.1111/boj.12049.

80. Wachowiak, W.; Trivedi, U.; Perry, A.; Cavers, S. Comparative transcriptomics of a complex of four european pine species. *BMC Genomics* **2015**, *16* (1), 1–9. https://doi.org/10.1186/s12864-015-1401-z.

81. Wang, Y. B.; Liu, B. Bin; Nie, Z. L.; Chen, H. F.; Chen, F. J.; Figlar, R. B.; Wen, J. Major clades and a revised classification of *Magnolia* and Magnoliaceae based on whole plastid genome sequences via genome skimming. *J. Syst. Evol.* **2020**, *58* (5), 673–695. https://doi.org/10.1111/jse.12588.

82. Whittall, J. B.; Syring, J.; Parks, M.; Buenrostro, J.; Dick, C.; Liston, A.; Cronn, R. Finding a (pine) needle in a haystack: chloroplast genome sequence divergence in rare and widespread pines. *Mol. Ecol.* **2010**, *19* (SUPPL. 1), 100–114. https://doi.org/10.1111/j.1365-294X.2009.04474.x.

83. Xia, Q.; Zhang, H.; Lv, D.; El-Kassaby, Y. A.; Li, W. Insights into phylogenetic relationships in *Pinus* inferred from a comparative analysis of complete chloroplast genomes. *BMC Genomics* **2023**, *24*, 346. https://doi.org/10.1186/s12864-023-09439-6.

84. Zeng, C. X.; Hollingsworth, P. M.; Yang, J.; He, Z. S.; Zhang, Z. R.; Li, D. Z.; Yang, J. B. Genome skimming herbarium specimens for DNA barcoding and phylogenomics. *Plant Methods* **2018**, *14* (1), 1–14. https://doi.org/10.1186/s13007-018-0300-0.

UNIWERSYTET IM. ADAMA MICKIEWICZA W POZNANIU
Szkoła Nauk Przyrodniczych

Poznań, 01.08.2024r.

Mgr Joanna Sikora
Zakład Genetyki
Instytut Biologii Eksperymentalnej
Wydział Biologii, UAM

Oświadczenie określające wkład doktoranta w powstanie publikacji

Niniejszym oświadczam, że jestem współautorką jak również pierwszą autorką publikacji:

Sokołowska, J.; Fuchs, H.; Celiński, K. New Insight into Taxonomy of European Mountain Pines, Pinus mugo Complex, Based on Complete Chloroplast Genomes Sequencing. Plants 2021, 10, 1331. https://doi.org/10.3390/plants10071331.

Mój udział polegał na: zaplanowaniu badań, zebraniu materiału badawczego i wykonaniu większości eksperymentów laboratoryjnych i przeprowadzeniu analizy danych, interpretacji wyników, współudziale w złożeniu kompletnych genomów chloroplastowych trzech przedstawicieli kompleksu *Pinus mugo,* przesłaniu kompletnych genomów chloroplastowych do GenBank, zredagowaniu manuskryptu, jego korekcie i odpowiedzi na recenzje.

Potwierdzenie promotora:

Konrad Celiński

Joanna Sikora

ul. H. Wieniawskiego 1, 61-712 Poznań
tel. +48 61 829 41 61
snp@amu.edu.pl
snp.home.amu.edu.pl

UNIWERSYTET IM. ADAMA MICKIEWICZA W POZNANIU
Szkoła Nauk Przyrodniczych

Poznań, 01.08.2024r.

Mgr Joanna Sikora
Zakład Genetyki
Instytut Biologii Eksperymentalnej
Wydział Biologii, UAM

Oświadczenie określające wkład doktoranta w powstanie publikacji

Niniejszym oświadczam, że jestem współautorką jak również pierwszą autorką publikacji:

Sokołowska, J.: Fuchs, H.; Celiński, K. Assessment of ITS2 Region Relevance for Taxa Discrimination and Phylogenetic Inference among Pinaceae. Plants 2022, 11, 1078. https://doi.org/10.3390/plants11081078.

Mój udział w powstanie publikacji polegał na zaplanowaniu badań, zebraniu materiału badawczego i wykonaniu eksperymentów laboratoryjnych, wykonaniu analizy danych, interpretacji wyników, współudziału w złożeniu kompletnych sekwencji ITS2 przedstawicieli kompleksu *Pinus mugo,* przesłaniu kompletnych sekwencji ITS2 do GenBank, zredagowaniu manuskryptu, jego korekcie i odpowiedzi na recenzje.

Potwierdzenie promotora:

Konrad Celiński

Joanna Sikora

ul. H. Wieniawskiego 1, 61-712 Poznań
tel. +48 61 829 41 61
snp@amu.edu.pl
snp.home.amu.edu.pl

UNIWERSYTET IM. ADAMA MICKIEWICZA W POZNANIU
Szkoła Nauk Przyrodniczych

Poznań, 01.08.2024r.

Mgr Joanna Sikora
Zakład Genetyki
Instytut Biologii Eksperymentalnej
Wydział Biologii, UAM

Oświadczenie określające wkład doktoranta w powstanie publikacji

Niniejszym oświadczam, że jestem współautorką jak również pierwszą autorką publikacji:

Sikora, J.; Celiński, K. Exploring Taxonomic and Genetic Relationships in the Pinus mugo Complex Using Genome Skimming Data. Int. J. Mol. Sci. 2024, 25, 10178. https://doi.org/10.3390/ijms251810178.

Mój udział w powstaniu publikacji polegał na: zaplanowaniu badań, zebraniu materiału badawczego i wykonaniu większości eksperymentów laboratoryjnych i przeprowadzeniu analizy danych, interpretacji wyników, złożeniu kompletnych sekwencji pochodzących z genomu chloroplastowego, mitochondrialnego i jądrowego przedstawicieli kompleksu *Pinus mugo,* przesłaniu kompletnych sekwencji do GenBank, zredagowaniu manuskryptu, jego korekcie i odpowiedzi na recenzje.

Potwierdzenie promotora:

Konrad Celiński

Joanna Sikora

UNIWERSYTET IM. ADAMA MICKIEWICZA W POZNANIU
Wydział Biologii
Instytut Biologii Eksperymentalnej, Zakład Genetyki

Poznań, 01.08.2024r.

Prof. UAM dr hab. Konrad Celiński
Zakład Genetyki
Instytut Biologii Eksperymentalnej
Wydział Biologii, UAM

Oświadczenie określające wkład współautorów w powstanie artykułu

Niniejszym oświadczam, że jestem współautorem i autorem korespondencyjnym artykułu, będącego częścią cyklu publikacji, wchodzących w skład rozprawy doktorskiej mgr Joanny Sikory:

Sokołowska, J.; Fuchs, H.; **Celiński, K.** New Insight into Taxonomy of European Mountain Pines, Pinus mugo Complex, Based on Complete Chloroplast Genomes Sequencing. Plants 2021, 10, 1331. https://doi.org/10.3390/plants10071331.

Mgr Joanna Sikora przygotowując powyższą publikację, zaplanowała badania i opracowała metodologię, zebrała materiał badawczy, wykonała większość eksperymentów laboratoryjnych, przeprowadziła analizy danych i interpretację wyników, współdziałała w złożeniu kompletnych genomów chloroplastowych trzech przedstawicieli kompleksu *Pinus mugo,* przesłała kompletne genomy chloroplastowe do GenBank, zredagowała manuskrypt, współuczestniczyła przy jego korekcie i odpowiedzi na recenzje.

Mój wkład w powstanie artykułu polegał na koordynacji pracy doktorantki, w tym współudziale w zebraniu materiału badawczego, zaplanowaniu badań, uczestnictwie w analizie wyników, a także współudziale w zredagowaniu treści artykułu i jego korekcie.

Konrad Celiński

ul. Uniwersytetu Poznańskiego 6, 61-614 Poznań
tel. +48 61 829 55 53
biologia@amu.edu.pl
biologia.amu.edu.pl

UNIWERSYTET IM. ADAMA MICKIEWICZA W POZNANIU
Wydział Biologii
Instytut Biologii Eksperymentalnej, Zakład Genetyki

Poznań, 01.08.2024r.

Prof. UAM dr hab. Konrad Celiński
Zakład Genetyki
Instytut Biologii Eksperymentalnej
Wydział Biologii, UAM

Oświadczenie określające wkład współautorów w powstanie artykułu

Niniejszym oświadczam, że jestem współautorem i autorem korespondencyjnym artykułu, będącego częścią cyklu publikacji, wchodzących w skład rozprawy doktorskiej mgr Joanny Sikory:

Sokołowska, J.; Fuchs, H.; **Celiński, K.** Assessment of ITS2 Region Relevance for Taxa Discrimination and Phylogenetic Inference among Pinaceae. Plants 2022, 11, 1078. https://doi.org/10.3390/plants11081078.

Mgr Joanna Sikora przygotowując powyższą publikację, zaplanowała metodologię, zebrała materiał badawczy i przeprowadziła eksperymenty laboratoryjne, przeprowadziła analizy danych i dokonała interpretacji wyników, brała udział w złożeniu kompletnych sekwencji ITS2 przedstawicieli kompleksu *Pinus mugo,* przesłała kompletne sekwencje ITS2 do GenBank, zredagowała manuskrypt, współdziałała przy jego korekcie i odpowiedzi na recenzje.

Mój wkład w powstanie artykułu polegał na koordynacji pracy doktorantki, w tym udziale w zaplanowaniu badań i zebraniu materiału badawczego, uczestnictwie w analizie wyników, a także współudziale w zredagowaniu artykułu i jego korekcie.

Konrad Celiński

UNIWERSYTET IM. ADAMA MICKIEWICZA W POZNANIU
Wydział Biologii
Instytut Biologii Eksperymentalnej, Zakład Genetyki

Poznań, 01.08.2024r.

Prof. UAM dr hab. Konrad Celiński
Zakład Genetyki
Instytut Biologii Eksperymentalnej
Wydział Biologii, UAM

Oświadczenie określające wkład współautorów w powstanie artykułu

Niniejszym oświadczam, że jestem współautorem i autorem korespondencyjnym artykułu, będącego częścią cyklu publikacji, wchodzących w skład rozprawy doktorskiej mgr Joanny Sikory:

Sikora, J.; **Celiński, K.** Exploring Taxonomic and Genetic Relationships in the Pinus mugo Complex Using Genome Skimming Data. Int. J. Mol. Sci. 2024, 25, 10178. https://doi.org/10.3390/ijms251810178.

Mgr Joanna Sikora przygotowując powyższą publikację, zaplanowała badania, zebrała materiał badawczy, wykonała większość eksperymentów laboratoryjnych, przeprowadziła analizy danych, zinterpretowała wyniki, złożyła kompletne sekwencje pochodzące z genomu chloroplastowego, mitochondrialnego i jądrowego przedstawicieli kompleksu *Pinus mugo*, przesłała kompletne sekwencje do GenBank, zredagowała manuskrypt, brała udział w jego korekcie i odpowiedzi na recenzje.

Mój wkład w powstanie artykułu polegał na koordynacji pracy doktorantki, w tym udziale w zaplanowaniu badań i zebraniu materiału badawczego, uczestnictwie w analizie wyników, a także współudziale w zredagowaniu artykułu i jego korekcie.

Konrad Celiński

Kórnik, 01.08.2024 r.

dr Hanna Fuchs
Zakład Biologii Rozwoju
Instytut Dendrologii Polskiej Akademii Nauk
ul. Parkowa 5; 62-035 Kórnik

Oświadczenie określające wkład współautorów w powstanie artykułu

Niniejszym oświadczam, że jestem współautorem artykułu, będącego częścią cyklu publikacji, wchodzących w skład rozprawy doktorskiej mgr Joanny Sikory:

Sokołowska, J.; **Fuchs, H.**; Celiński, K. New Insight into Taxonomy of European Mountain Pines, Pinus mugo Complex, Based on Complete Chloroplast Genomes Sequencing. Plants 2021, 10, 1331. https://doi.org/10.3390/plants10071331.

Mgr Joanna Sikora przygotowując powyższą publikację, zaplanowała badania i opracowała metodologię, zebrała materiał badawczy, wykonała większość eksperymentów laboratoryjnych, przeprowadziła analizy danych i interpretację wyników, współdziałała w złożeniu kompletnych genomów chloroplastowych trzech przedstawicieli kompleksu *Pinus mugo,* przesłała kompletne genomy chloroplastowe do GenBank, zredagowała manuskrypt, współdziałała przy jego korekcie i odpowiedzi na recenzje.

Mój wkład w powstanie artykułu polegał na współudziale w złożeniu trzech kompletnych genomów chloroplastowych przedstawicieli kompleksu *Pinus mugo*, tj. *Pinus mugo subsp. mugo*, *Pinus mugo subsp. rotundata, Pinus mugo subsp. uncinata.*

Hanna Fuchs

Kórnik, 01.08.2024 r.

dr Hanna Fuchs
Zakład Biologii Rozwoju
Instytut Dendrologii Polskiej Akademii Nauk
ul. Parkowa 5; 62-035 Kórnik

Oświadczenie określające wkład współautorów w powstanie artykułu

Niniejszym oświadczam, że jestem współautorem artykułu, będącego częścią cyklu publikacji, wchodzących w skład rozprawy doktorskiej mgr Joanny Sikory:

Sokołowska, J.; **Fuchs, H.**; Celiński, K. Assessment of ITS2 Region Relevance for Taxa Discrimination and Phylogenetic Inference among Pinaceae. Plants 2022, 11, 1078. https://doi.org/10.3390/plants11081078.

Mgr Joanna Sikora przygotowując powyższą publikację, zaplanowała metodologię, zebrała materiał badawczy i przeprowadziła eksperymenty laboratoryjne, przeprowadziła analizy danych i dokonała interpretacji wyników, brała udział w złożeniu kompletnych sekwencji ITS2 przedstawicieli kompleksu *Pinus mugo,* przesłała kompletne sekwencje ITS2 do GenBank, zredagowała manuskrypt, współdziałała przy jego korekcie i odpowiedzi na recenzje.

Mój wkład w powstanie artykułu polegał na współudziale w złożeniu kompletnych sekwencji ITS2 (ang. internal transcribed spacer 2) przedstawicieli kompleksu *Pinus mugo* i udziale w przeprowadzeniu analizy danych.

Hanna Fuchs