Institute of Dendrology, Polish Academy of Sciences

mgr Weronika Barbara Żukowska

Processes shaping genetic variation of European populations of dwarf mountain pine (*Pinus mugo* Turra)

PhD Thesis
in the Field of Biological Sciences
Discipline: Biology

Kórnik, 2017

http://rcin.org.pl

The title of the thesis in Polish:
Procesy kształtujące zmienność genetyczną europejskich populacji kosodrzewiny (*Pinus mugo* Turra)

The research was carried out in the Laboratory of Molecular Biology of the Institute of Dendrology, Polish Academy of Sciences in Kórnik.

The main supervisor: dr hab. Witold Wachowiak

The study was funded through a grant from the Polish National Science Centre (Grant No. DEC-2012/05/E/NZ9/03476) “Association genetics of nucleotide polymorphisms related to adaptive variation and population history in the European range of pine species *Pinus sylvestris* and *Pinus mugo*”.

http://rcin.org.pl

Acknowledgements

I would like to express my sincere gratitude to my supervisor dr hab. Witold Wachowiak
for his guidance, inspiring ideas and insightful discussion.
You showed me the ways to approach my research problems
that I would never have thought about.

With great appreciation I shall acknowledge prof. Andrzej Lewandowski,
dr Monika Litkowiec, mgr inż. Błażej Wójkiewicz and Maria Ratajczak
for their constant support and encouragement.

I owe a special thanks to my parents.
You had faith in me when nobody else had.

And finally, to my beloved husband, Grzegorz Przybylski,
words cannot express how grateful I am that you have been there for me.

http://rcin.org.pl

Table of contents

List of publications included in the doctoral dissertation……………………………………… 6

Abstract……………………………………………………………………………………7

Streszczenie…………………………………………………………………………………9

1. Introduction………………………………………………………………………………11

2. The aim of the study……………………………………………………………………14

3. Major outcomes…………………………………………………………………………15

3.1 Cross-amplification and multiplexing of *cp*SSRs and *n*SSRs. The level of genetic diversity of *P. mugo*…………………………………………………………………… 15

3.2 Interpopulation differentiation and phylogeographic structure. The role of genetic drift, mutations, isolation by distance and gene flow in shaping the genetic variation of *P. mugo*……………………………………………………………………………………17

3.3 The role of demographic processes, phenotypic plasticity and local adaptation in shaping the variation of *P. mugo* as inferred from the comparison between genetic and morphological data…………………………………………………………………………18

3.4 Hybridising taxa as useful objects for genetic studies of adaptive variation and speciation. How hybridisation affects the gene pool of *P. mugo*?…………………………19

4. Conclusions………………………………………………………………………………22

5. References…………………………………………………………………………………23

Statements……………………………………………………………………………………25

Appendices……………………………………………………………………………………35

Publication 1*:* Żukowska WB, Wójkiewicz B, Litkowiec M, Wachowiak W (2017). Cross-amplification and multiplexing of *cp*SSRs and *n*SSRs in two closely related pine species (*Pinus sylvestris* L. and *P. mugo* Turra). *Dendrobiology* 77: 59-64………………………36

Publication 2: Żukowska WB, Wachowiak W (2017). Nuclear microsatellite markers reveal the low genetic structure of *Pinus mugo* Turra (dwarf mountain pine) populations in Europe. *Plant Systematics and Evolution*. doi: 10.1007/s00606-017-1395-x…………… 43

Publication 3: Żukowska WB, Boratyńska K, Wachowiak W (2017). Comparison of range-wide chloroplast microsatellite and needle trait variation patterns in *Pinus mugo* Turra (dwarf mountain pine). *iForest – Biogeosciences and Forestry*. doi: 10.3832/ifor1860-009. ……………………………………………………………………………………………60

Publication 4: Żukowska WB, Wachowiak W (2016). Utility of closely related taxa for genetic studies of adaptive variation and speciation: current state and perspectives in plants with focus on forest tree species. *Journal of Systematics and Evolution* 54: 17-28………76

http://rcin.org.pl

Publication 5: Wachowiak W, Cavers S, Żukowska WB (2015). Interspecific gene flow and ecological selection in a pine (*Pinus* sp.) contact zone. *Plant Systematics and Evolution* 301: 1643-1652……………………………….…….….………..………………89

Publication 6: Wachowiak W, Żukowska WB, Wójkiewicz B, Cavers S, Litkowiec M (2016). Hybridization in contact zone between temperate European pine species. *Tree Genetics & Genomes* 12: 48……………………………….……………………………103

http://rcin.org.pl

List of publications included in the doctoral dissertation

1. **Żukowska WB**[*], Wójkiewicz B[*], Litkowiec M, Wachowiak W (2017). Cross-amplification and multiplexing of *cp*SSRs and *n*SSRs in two closely related pine species (*Pinus sylvestris* L. and *P. mugo* Turra). *Dendrobiology* 77: 59-64.

2. **Żukowska WB,** Wachowiak W (2017). Nuclear microsatellite markers reveal the low genetic structure of *Pinus mugo* Turra (dwarf mountain pine) populations in Europe. *Plant Systematics and Evolution*. doi: 10.1007/s00606-017-1395-x.

3. **Żukowska WB**, Boratyńska K, Wachowiak W (2017). Comparison of range-wide chloroplast microsatellite and needle trait variation patterns in *Pinus mugo* Turra (dwarf mountain pine). *iForest – Biogeosciences and Forestry*. doi: 10.3832/ifor1860-009.

4. **Żukowska WB**, Wachowiak W (2016). Utility of closely related taxa for genetic studies of adaptive variation and speciation: current state and perspectives in plants with focus on forest tree species. *Journal of Systematics and Evolution* 54: 17-28.

5. Wachowiak W, Cavers S, **Żukowska WB** (2015). Interspecific gene flow and ecological selection in a pine (*Pinus* sp.) contact zone. *Plant Systematics and Evolution* 301: 1643-1652.

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

[*] WB Żukowska and B Wójkiewicz contributed equally to this work

http://rcin.org.pl

Abstract

The genetic structure of a population is affected by complex demographic (i.e. associated with population history) and evolutionary processes. Mutations, recombination, genetic drift, hybridisation and natural selection all leave signatures that can be detected at the genomic level. Tree species, which are long-lived and highly outcrossing, harbour the vast amount of genetic variation, but their populations are differentiated only to a small extent due to the high level of gene flow. The geographic distribution of genetic variation of forest trees was predominantly shaped during the postglacial recolonisation of Europe, but it is hard to assess how various processes influenced this pattern.

The main aim of my doctoral dissertation was to assess how different demographic and evolutionary processes contribute to the present genetic variation of *Pinus mugo* Turra (dwarf mountain pine) in Europe. The analyses involved 21 populations of *P. mugo* (555 individuals) inhabiting the Alps, Sudetes, Carpathians, Apennines and the Balkans as well as two hybrid zones occupied by *P. mugo* and its relatives. The genetic variation was screened with 13 nuclear and 11 chloroplast microsatellite markers. An additional analysis was made with the available data on the morphological variation of *P. mugo* (needle traits) that were compared to the genetic data.

The results showed that the populations of *P. mugo* have the high level of genetic variation, but they are only little differentiated. The level and distribution of genetic variation indicate that the range of *P. mugo* during the cold periods of the Pleistocene was much larger. The present populations of the species have probably diverged only recently mainly due to genetic drift. The pattern of differentiation at chloroplast microsatellites can be explained by the effect of isolation by distance and the existence of some barriers to gene flow. As shown by the analysis of nuclear microsatellites, the Sudeten populations of *P. mugo* have been most likely established via the founder effect by the Alpine migrants. No apparent barriers to gene flow between the Sudetes and Alps support this hypothesis. The comparison between the genetic and morphological data revealed that the eastern stands of *P. mugo* differ morphologically from the other populations, but they do not possess a distinct gene pool. Therefore, the morphological variation of *P. mugo* may (to some degree) be shaped by natural selection. Finally, ongoing hybridisation can be observed in the sympatric populations of *P. mugo* and its relatives, but the possible historical hybridisation processes have not affected the present gene pool of the allopatric populations of this species.

http://rcin.org.pl

The outcomes of my research give a valuable insight into how various demographic and evolutionary processes interact to shape the genetic variation of *P. mugo*. Studies of neutral genetic variation are necessary to advance research focusing on the genetic background of the variation of adaptive traits. Knowledge about the adaptive potential and response of tree species to environmental changes will be also useful to develop sustainable breeding programs and management strategies for the conservation of genetic resources of forest tree species.

The results of my PhD dissertation constitute an independent and separate part of the collective work (6 publications) with my individual contribution in developing the concept, carrying out the experimental part and the development and interpretation of the results.

http://rcin.org.pl

Streszczenie

Złożone procesy demograficzne (tj. związane z historią populacji) i ewolucyjne mają wpływ na strukturę genetyczną populacji. Mutacje, rekombinacja, dryf genetyczny, proces hybrydyzacji i naturalna selekcja pozostawiają wzorce, które mogą zostać wykryte na poziomie genomu. Poziom zmienności genetycznej gatunków drzewiastych, jako długożyjących i charakteryzujących się wysoką częstością zapłodnień krzyżowych, jest bardzo wysoki, jednak ich zróżnicowanie międzypopulacyjne jest niewielkie z powodu intensywnego przepływu genów. Geograficzne rozmieszczenie zmienności genetycznej drzew leśnych zostało w głównej mierze ukształtowane w okresie polodowcowej rekolonizacji Europy, jednak trudno jest ocenić jak różne procesy wpłynęły na ten wzór.

Głównym celem mojej rozprawy doktorskiej była ocena jak różne procesy demograficzne i ewolucyjne przyczyniają się do obecnej zmienności genetycznej *Pinus mugo* Turra (kosodrzewiny) w Europie. Analizy obejmowały 21 populacji *P. mugo* (555 osobników) z terenów Alp, Sudetów, Karpat, Apeninów i Bałkanów, a także dwie strefy hybrydyzacyjne zajmowane przez *P. mugo* i taksony spokrewnione z tym gatunkiem. Zmienność genetyczna została przeanalizowana przy pomocy 13 jądrowych i 11 chloroplastowych markerów mikrosatelitarnych. Dodatkowa analiza porównawcza została przeprowadzona przy wykorzystaniu dostępnych danych dotyczących zmienności morfologicznej *P. mugo* (cechy igieł) oraz danych genetycznych.

Wyniki pokazały, że poziom zmienności genetycznej populacji *P. mugo* jest wysoki, jednak zróżnicowanie międzypopulacyjne jest niskie. Poziom i rozmieszczenie zmienności genetycznej wskazują, że zasięg *P. mugo* w okresie glacjałów w czasie plejstocenu był znacznie większy. Obecne populacje gatunku prawdopodobnie uległy niedawnemu zróżnicowaniu głównie w wyniku działania dryfu genetycznego. Izolacja na dystans oraz istnienie pewnych barier dla przepływu genów tłumaczą wzorzec zróżnicowania mikrosatelit chloroplastowych. Zgodnie z analizą mikrosatelit jądrowych populacje sudeckie zostały najprawdopodobniej zapoczątkowane przez migrantów alpejskich w wyniku tzw. efektu założyciela. Brak barier dla przepływu genów między Sudetami i Alpami popiera tę hipotezę. Porównanie danych genetycznych i morfologicznych wykazało, że wschodnie populacje *P. mugo* różnią się pod względem morfologicznym od innych populacji, jednak nie posiadają odmiennej puli genowej. Zatem zmienność morfologiczna *P. mugo* może (do pewnego stopnia) być kształtowana przez działanie selekcji naturalnej. W sympatrycznych populacjach *P. mugo* i taksonów

http://rcin.org.pl

spokrewnionych z tym gatunkiem można zaobserwować trwający proces hybrydyzacji, jednak możliwe historyczne procesy hybrydyzacji nie miały wpływu na obecną pulę genową allopatrycznych populacji tego gatunku.

Rezultaty moich badań dostarczają cennej wiedzy na temat współdziałania różnych procesów demograficznych i ewolucyjnych w kształtowaniu zmienności genetycznej *P. mugo*. Badania nad neutralną zmiennością genetyczną są niezbędne dla dalszych dociekań koncentrujących się wokół zagadnień genetycznych podstaw zmienności cech adaptacyjnych. Wiedza na temat potencjału adaptacyjnego i odpowiedzi gatunków drzewiastych na zmiany środowiska będzie także użyteczna dla rozwoju zrównoważonych programów hodowli i strategii zarządzania w celu zachowania zasobów genowych gatunków drzew leśnych.

Wyniki mojej rozprawy doktorskiej stanowią samodzielną i wyodrębnioną część pracy zbiorowej (6 publikacji), wykazującej mój indywidualny wkład przy opracowywaniu koncepcji, wykonywaniu części eksperymentalnej oraz opracowaniu i interpretacji wyników.

http://rcin.org.pl

1. Introduction

Genetic variation found among individuals and populations means that there are differences in the sequence of their DNA. The main sources of genetic variation are mutations and recombination that together drive the evolution of species. But the gene pool of a population is influenced by numerous processes, such as genetic drift, hybridisation and natural selection (Nosil and Feder 2013; Fig. 1). Such processes leave specific genetic signatures. In general, demographic processes (i.e. processes that are associated with population history) affect the whole genome in the same way, whereas natural selection that acts on genomic regions that are of adaptive importance leaves a more localised pattern (Nielsen 2005).

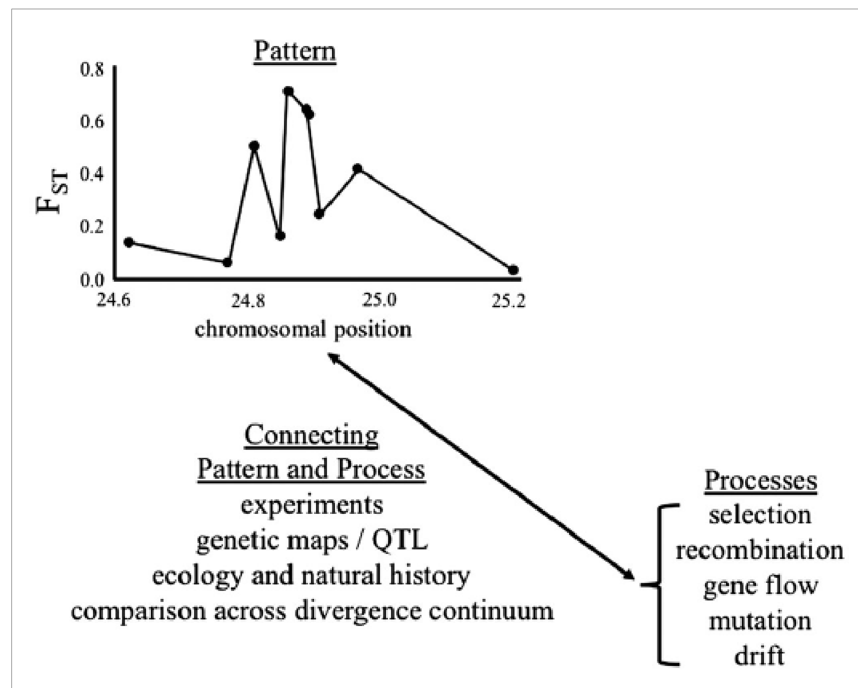


Fig. 1. A given genomic pattern can be generated via different combinations of demographic and evolutionary processes (Nosil and Feder 2013). F_{ST} – fixation index; QTL – quantitative trait locus.

The present geographic distribution of genetic variation of forest trees was largely shaped during the postglacial recolonisation of Europe (Huntley 1990), but it remains to be resolved how various demographic and evolutionary processes contributed to this pattern. As life form and breeding system have a significant influence on genetic diversity, populations of forest tree species, which are long-lived and highly outcrossing, are usually characterised by high genetic diversity and low interpopulation differentiation (Hamrick and Godt 1996).

http://rcin.org.pl

The main object of my research was *Pinus mugo* Turra (known as dwarf mountain pine, mountain pine, mugo pine or Swiss mountain pine). It is a pine species belonging to the Pinaceae family. *P. mugo* forms shrubs up to a few metres in height that grow in the subalpine belt of mountain ranges in Central and Southern Europe (Richardson 1998; Fig. 2). It plays a very important role in these areas, preventing avalanches and soil erosion. *P. mugo* can reproduce both generatively and vegetatively (Prus-Głowacki et al. 2005). The species is frost-resistant and can adapt to various environmental conditions. Also, it is often planted as an ornamental shrub.

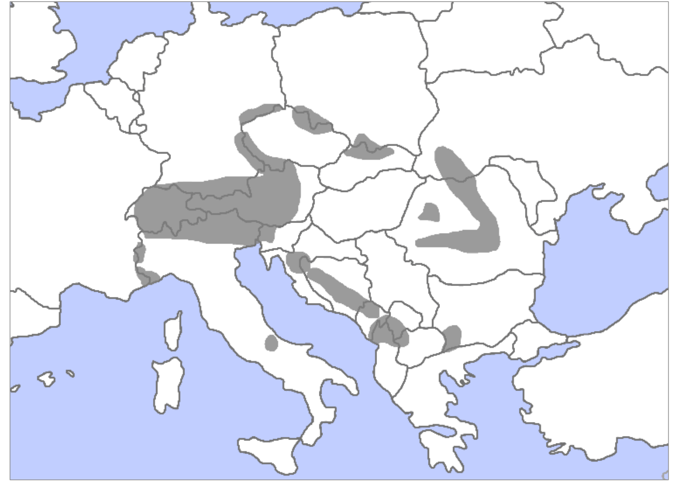


Fig. 2. Geographic range of *Pinus mugo* Turra.

To date, studies on *P. mugo* have mainly concentrated on the variability of its needles and cones (e.g. Boratyńska et al. 2004; Boratyńska et al. 2015). The results indicate that some morphological traits may be adaptively important. What is more, the eastern populations of *P. mugo* are different from the other stands with regard to needle trait variation. In turn, the research concentrating on the genetic variation of the species has been generally limited to some parts of its geographic range (e.g. Slavov and Zhelev 2004; Dzialuk et al. 2012). So far, the relationship between the genotype and phenotype as well as genes that are involved in the evolution and adaptation of *P. mugo* have been poorly understood.

From a taxonomical point of view, *P. mugo* belongs to the *P. mugo* complex – a group of closely related pine species that have a various geographic distribution, morphology and ecology. The complex include i.a. *Pinus uncinata* Ramond (Domin), that is a single-stemmed tree and can be as high as 20 m. It grows in the Pyrenees and co-occurs with *P. mugo* in the Alps, where intermediate form of both species can be found (Monteleone et al. 2006). On the other hand, populations of *Pinus uliginosa* Neumann, which also belongs to the *P. mugo* complex, are isolated and inhabit Central European peat-bogs. Species that form the *P. mugo* complex are interfertile. They can also hybridise with *Pinus sylvestris* (Scots pine) (e.g. Wachowiak et al. 2005; Wachowiak and Prus-Głowacki 2008), which is one of the major forest-forming species in Europe and Asia. The above-mentioned pine species have probably diverged relatively recently due to their adaptation to various habitats. Thus, they constitute a good study system

http://rcin.org.pl

to search for the genetic background of adaptive variation and speciation (Wachowiak et al. 2011).

Despite the fact that species from the *P. mugo* complex and *P. sylvestris* have currently mostly an allopatric distribution, it is likely that the genetic variation of their present populations has been to some degree affected by historical interspecific hybridisation processes. The postglacial recolonisation of Europe could have created temporary contact zones in which the pines were able to interbreed. Hybridisation increases the overall genetic variation of the species and may also lead to some changes in the morphology and adaptive potential of hybrid individuals. The past hybridisation events are the most probable cause of the present high morphological variation found within the *P. mugo* complex. At present, contact zones of the above-mentioned pine species exist in Southern and Central Europe (Christensen 1987), giving a unique opportunity to study how hybridisation contributes to both genetic and morphological variation and how it affects the adaptive potential of the species. Such zones are also very useful to study the initial stages of speciation.

A detailed description of the genetic variation of allopatric and sympatric populations of *P. mugo* will give us an insight into how it is shaped by various demographic and evolutionary processes. Such studies are the first step to further research on the genetic architecture of adaptive variation. The level and distribution of neutral genetic variation need to be elucidated, because genomic patterns generated by neutral processes may mimic the ones which arise due to natural selection. An additional comparison between genetic and morphological data will make it possible to better understand the relationship between the geographic patterns of genetic and morphological variation of *P. mugo* and possibly to identify which morphological traits are of adaptive importance. Finally, genetic analyses in the pine hybrid zones and reference allopatric populations will be useful to make inferences about the influence of historical and present hybridisation processes among *P. mugo* and its relatives on their genetic and morphological variability and adaptive potential.

http://rcin.org.pl

2. The aim of the study

The main objective of my doctoral dissertation was to identify which demographic and evolutionary processes contribute to the present genetic variation of *P. mugo* across its geographic range (with the exception of anthropological factors). I also aimed to describe genetic variation and differentiation at the within and among population level to make some inferences about the postglacial history of the species. To this end, nuclear and chloroplast microsatellite markers (= SSRs – simple sequence repeats) were used. To my best knowledge, nuclear SSRs (*n*SSRs) have not been used to study the genetic variation of *P. mugo*. I also made use of the previously published data on the morphological variation of the species (needle morphology and anatomy) and compared them to the genetic data. By making such comparison, I was able to assess whether the geographic patterns of neutral genetic and morphological variation overlap, suggesting that morphological variation is neutral, or if they do not match, indicating the possible adaptive importance of some needle traits. The final step of my research included the analysis of the genetic composition of two hybrid zones occupied by *P. mugo* and its relatives as well as the reference populations of each species. The purpose of this analysis was to assess the influence of hybridisation on the gene pool of *P. mugo* and other pines, both in the past and at present.

I focused on the following research hypotheses:

(1) Populations of *P. mugo*, as a highly outcrossing and wind-pollinated tree species, have the high level of genetic diversity.
(2) Populations of *P. mugo* occupying different mountain chains display high genetic differentiation at neutral genetic markers resulting from their putative origin from different glacial refugia, barriers to gene flow and a predominant role of genetic drift.
(3) Geographic patterns of neutral genetic and morphological variation do not overlap, because a specific phenotype is a result of the interplay between neutral processes and adaptation to the local environment.
(4) The gene pool of the present allopatric populations of *P. mugo* and its relatives has been affected by historical hybridisation processes.

http://rcin.org.pl

3. Major outcomes

The results of my doctoral dissertation concentrate mainly on the neutral genetic diversity and differentiation of the present populations of *P. mugo*. I divided the results into 4 sections. The first and second one focus on the level and distribution of the genetic variation of *P. mugo*, which were studied using *n*SSRs and chloroplast SSRs (*cp*SSRs). The third part concentrates on the comparison between genetic (*cp*SSRs) and morphological (needle morphology and anatomy) data scored for the same individuals of *P. mugo*. The last section focuses on the genetic structure of two hybrid zones formed by *P. mugo* and its close relatives.

3.1 Cross-amplification and multiplexing of *cp*SSRs and *n*SSRs. The level of genetic diversity of *P. mugo*

Publication 1: Cross-amplification and multiplexing of *cp*SSRs and *n*SSRs in two closely related pine species (*Pinus sylvestris* L. and *P. mugo* Turra). *Dendrobiology* 77: 59-64.

Publication 2: Nuclear microsatellite markers reveal the low genetic structure of *Pinus mugo* Turra (dwarf mountain pine) populations in Europe. *Plant Systematics and Evolution*. doi: 10.1007/s00606-017-1395-x.

Publication 3: Comparison of range-wide chloroplast microsatellite and needle trait variation patterns in *Pinus mugo* Turra (dwarf mountain pine). *iForest – Biogeosciences and Forestry*. doi: 10.3832/ifor1860-009.

SSRs, commonly known as microsatellites or, less often, short tandem repeats (STRs) are the class of repetitive DNA in which motifs of 2-5 bp are repeated. They can be found in both eukaryotic and prokaryotic genomes. They are very useful neutral genetic markers due to their ubiquity, hypervariability and a co-dominant mode of inheritance (Ellegren 2004).

Conifers are a specific group of angiosperms, because their chloroplast (*cp*) genome is inherited paternally via pollen – a phenomenon not found in other taxa of higher plants – whereas their mitochondrial (*mt*) genome is inherited maternally via seeds. Therefore, *cp*DNA in conifers is useful to study the level, direction and barriers to gene flow, but *mt*DNA is better to identify the origin of a certain population or to track the postglacial recolonisation routes of the species. Unfortunately, the resolution of *mt*DNA markers initially developed for *P. sylvestris* is too low for fine-scale phylogeographic studies in *P. mugo* (Wachowiak et al. 2013).

The initial stage of my research involved the cross-amplification and development of multiplex polymerase chain reactions (PCRs) for *cp*SSRs and *n*SSRs, as described in *Publication 1*. To

http://rcin.org.pl

this end, 14 chloroplast and 22 nuclear SSRs initially developed for *Pinus thunbergii*, *P. sylvestris* and *Pinus taeda* were pre-tested on 4 populations of *P. mugo*. The transfer rate was higher for *cp*SSRs (86%; 12 loci) than for *n*SSRs (59%; 13 loci), probably because the mutation rate of *cp*DNA is lower than that of *n*DNA (Willyard et al. 2007). The markers were successfully amplified in 5 multiplex PCRs. The allelic variation of the analysed SSRs varied from only one allele for the chloroplast locus PCP102652 to 28 alleles for the nuclear locus SPAG 7.14, and it was generally lower for nuclear markers developed by Sebastiani et al. (2012) (the "psyl" series). The mean observed and expected heterozygosity of *n*SSRs was equal to 0.44 and 0.46, respectively. A few nuclear markers had a significant percentage of null alleles. They thus require proper corrections methods if applied in population genetic studies. Chloroplast locus PCP30277 turned out to be a useful species-specific marker that distinguishes
P. mugo from *P. sylvestris*. The study proves that cross-amplification is a good first choice alternative to the *de novo* development of microsatellite markers for species with poor genomic resources such as *P. mugo*.

The developed multiplex PCRs were used in two further studies of 21 populations of *P. mugo* (555 individuals) that cover the native range and putative refugia of the species (*Publication 2* and *3*).

Publication 2 describes the genetic variation of *P. mugo* analysed with the use of *n*SSRs. As expected in the hypothesis (1), the neutral genetic variation of *P. mugo* turned out to be high. 133 alleles were detected for 13 loci, with a mean number of 4.98 alleles per locus for each population. 15 alleles were private to particular *P. mugo* stands. 95% of the genetic variation was observed within the populations. The average level of observed and unbiased expected heterozygosity was 0.40 and 0.44, respectively. The results showed that inbreeding did not have a significant influence on the genetic diversity of *P. mugo*, so deviations from Hardy-Weinberg equilibrium at some loci resulted from the presence of null alleles.

The genetic diversity of one peripheral population form Italy (A5) was apparently lower as compared to other populations. It should be pointed, however, that no signs of recent bottlenecks were detected for A5 or any other stand of *P. mugo*. Still, A5 may be more prone to environmental changes. The smaller number of alleles found in this population indicate a lower chance for adaptation from standing genetic variation, which is considered to be faster than from new mutations (Barrett and Schluter 2008).

http://rcin.org.pl

The high genetic diversity of *P. mugo* was also confirmed by the analysis of 11 *cp*SSRs (locus PCP102652 was monomorphic) (*Publication 3*). The mean number of alleles varied between 3 to 14 (average 6.63), but the mean effective number of alleles was lower (average 2.49) as the frequency of many alleles was low. As *cp*DNA does not recombine or have a heterozygous nature, alleles found in a single individual were combined into haplotypes. Overall, 311 haplotypes were scored and most of them (201; 65%) were population-private. Hence, the haplotypic diversity within each population was high (average 96%). The highest number of shared haplotypes was found among the Sudeten populations. Again, the genetic diversity of the A5 population was lower as compared to the other stands of *P. mugo*.

3.2. Interpopulation differentiation and phylogeographic structure. The role of genetic drift, mutations, isolation by distance and gene flow in shaping the genetic variation of *P. mugo*

Publication 2: Nuclear microsatellite markers reveal the low genetic structure of *Pinus mugo* Turra (dwarf mountain pine) populations in Europe. *Plant Systematics and Evolution*. doi: 10.1007/s00606-017-1395-x.

Publication 3: Comparison of range-wide chloroplast microsatellite and needle trait variation patterns in *Pinus mugo* Turra (dwarf mountain pine). *iForest – Biogeosciences and Forestry*. doi: 10.3832/ifor1860-009.

The genetic analysis of the interpopulation differentiation of *P. mugo*, using both *cp*SSRs and *n*SSRs, pointed to the low among population genetic differentiation. The value of F_{ST} was equal to 5.2% for *n*SSRs (corrected for the presence of null alleles) and 6.4% for *cp*SSRs. Contrary to what was expected in the hypothesis (2), the differentiation among the mountain regions was almost none (1% for *n*SSRs and 3% for *cp*SSRs).

The results did not support the existence of phylogeographic structure, both for *cp*SSRs and *n*SSRs. This suggests that mutations have not played a significant role in shaping the genetic structure of the populations of *P. mugo* analysed in the study. Any new mutations are probably quickly spread by gene flow, and genetic drift has a predominant role in shaping the genetic variation of the species.

The results of the Bayesian clustering and Mantel test depended on the type of markers used. With respect to *cp*SSRs, which are transferred solely via pollen, the genetic variation tended to have a geographic pattern. The observed structure could be explained by isolation by distance

http://rcin.org.pl

(IBD). All Balkan, most Carpathian and two Alpine populations belonged to the same genetic cluster. Some populations from the Alps clustered together with the Sudeten stands, suggesting that the Sudeten populations could have been established by the Alpine migrants. On the other hand, one population from the Apennines and the south-westernmost stand form Italy grouped together, but they were clearly separated from the other populations of *P. mugo*, indicating the presence of barriers to gene flow, their different origin, or, as already suggested by Boratyńska and Boratyński (2007) and Boratyńska et al. (2015), hybridisation with *P. uncinata* (*Publication 3*).

The assignment of populations using the Bayesian clustering methods based on the analysis of *n*SSRs resulted in only two groups of populations. The first group comprised only the Sudeten populations, whereas the remaining 17 populations form the Alps, Carpathians, Apennines and the Balkans grouped into the second genetic cluster. This result along with the lack of the IBD signal support the isolation by colonisation (IBC) scenario (Orsini et al. 2013), confirming the outcomes of the analysis of *cp*SSRs – the Sudeten populations of *P. mugo* have most likely been established by the Alpine migrants (*Publication 2*).

Taking into account the results of both analyses (*cp*SSRs and *n*SSRs), it appears that the present populations of *P. mugo* originate from a larger glacial distribution of the species. Palynological data indicate that *P. mugo* indeed covered a broader area during the cold periods of the Pleistocene (e.g. Farcas et al. 1999). The isolation of particular stands in the most elevated mountains likely started only 8000-9000 years ago (Boratyńska et al. 2004). Such a relatively short period of time was not long enough to cause any significant genetic differentiation among the present populations of *P. mugo*.

3.3 The role of demographic processes, phenotypic plasticity and local adaptation in shaping the variation of *P. mug*o as inferred from the comparison between genetic and morphological data

Publication 3: Comparison of range-wide chloroplast microsatellite and needle trait variation patterns in *Pinus mugo* Turra (dwarf mountain pine). *iForest – Biogeosciences and Forestry*. doi: 10.3832/ifor1860-009.

Publication 3 compares the geographic patterns of neutral genetic and morphological variation. 21 populations of *P. mugo* and 11 *cp*SSRs were used in the genetic analysis, as described in section 3.1. From these, for 18 stands the morphological data were retrieved and reanalysed from the previous study by Boratyńska et al. (2015). The populations covered the same

http://rcin.org.pl

individuals of *P. mugo* that were analysed with the use of *cp*SSRs. In total, 22 anatomical and morphological needle traits were assessed. The research aimed to test whether the distribution of morphological variation is shaped by: (1) demographic factors, (2) environmental factors or (3) a combination of demographic and environmental factors.

The geographic distribution of morphological variation was different than in the case of neutral genetic variation, as expected by the hypothesis (3). The genetic variation had a rather south-north pattern, whereas the morphological variation tended to display a west-east trend. The eastern populations of *P. mugo* formed a separate group in the morphological analysis. Nevertheless, this result was not obtained in the genetic assay.

As the patterns of neutral genetic and morphological variation overlapped only to some extent, it is possible that some needle traits may be adaptively important in the eastern part of the range of *P. mugo*. The additional correlation analysis showed that a few needle features correlate with longitude and climate variables. Needles of eastern populations of *P. mugo* are narrower and thinner and therefore have thinner epidermal cells and fewer resin canals. It is hard to unambiguously conclude if these traits are simply plastic or remain under genetic control, because the information about their heritability is limited. Thickness of epidermal cells, however, was shown to have a hereditary nature (Fedorkov 2002). The smaller area of the needles of *P. mugo* growing in the east might be beneficial in xeric conditions, reducing the water loss.

3.4 Hybridising taxa as useful objects for genetic studies of adaptive variation and speciation. How hybridisation affects the gene pool of *P. mugo*?

Publication 4: Utility of closely related taxa for genetic studies of adaptive variation and speciation: current state and perspectives in plants with focus on forest tree species. *Journal of Systematics and Evolution* 54: 17-28.

Publication 5: Interspecific gene flow and ecological selection in a pine (*Pinus* sp.) contact zone. *Plant Systematics and Evolution* 301: 1643-1652.

Publication 6: Hybridization in contact zone between temperate European pine species. *Tree Genetics & Genomes* 12: 48.

Publication 4 discusses how comparative genomics of closely related taxa can advance evolutionary genetic studies. As many closely related species show high ecological and phenotypic differentiation but low background neutral genetic variation, they are ideal for this

http://rcin.org.pl

type of research. The publication gives examples and summarises studies that used closely related hybridising plant taxa to decipher the genetic architecture of adaptive traits and reproductive isolation. Such studies proved e.g. that most adaptive traits are polygenic or that "genomic islands of divergence" are mostly small and spread throughout a genome.

A particular attention is given to forest trees and the features which make *P. mugo* and its relatives a promising study system to search for genetic signatures of adaptive variation and speciation.

The final part of my research consisted of the analysis of the genetic composition of two hybrid zones occupied by the members of the *P. mugo* complex and *P. sylvestris*.

The first contact area is located in the Bór na Czerwonem reserve (*Publication 5*). In total, 60 individuals from this hybrid zone and 134 specimens from the reference allopatric populations of *P. mugo* and *P. sylvestris* were analysed. Trees from the sympatric population were classified as either pure species (PM – *P. mugo*; PS – *P. sylvestris*) or hybrids (HB) based on their phenotype, prior to the genetic analysis. Each group was represented by 20 individuals. A species-specific DNA marker and nucleotide polymorphisms in a set of 8 nuclear genes clearly distinguished the pure species. Evidence of selection was found at several genes. According to the Bayesian clustering, the PM and HB groups and the reference populations of *P. mugo* were assigned to the first cluster, whereas the PS group clustered together with the allopatric populations of *P. sylvestris*. Furthermore, 10 putative hybrid individuals were identified: 5 of them were in the PM group, 4 in the HB group and one in the PS group. Therefore, some hybrids were cryptic, i.e. they had a phenotype characteristic of one of the pure species. All hybrids carried *cp*DNA of *P. mugo*. The untypical morphology of some individuals found in the Bór na Czerwonem reserve stems probably from environmental variation.

The second contact zone comprised *P. mugo*, *P. uliginosa*, *P. sylvestris* and their intermediate forms (96 individuals) growing at the Zieleniec reserve (*Publication 6*) and 287 samples from the reference populations. Nucleotide polymorphisms were analysed in a set of 26 nuclear genes, and a species-specific *cp*DNA marker was used to identify which species (*P. sylvestris* vs. the *P. mugo* complex) was the pollen donor. 5 groups of individuals were distinguished based on their morphology and the pollen donor: PM – *P. mugo*, PS – *P. sylvestris*, PU – *P. uliginosa*, HB – group of *P. uliginosa*-like trees of untypical morphology and HPS – *P. sylvestris*-like trees carrying *cp*DNA diagnostic for the *P. mugo* complex.

http://rcin.org.pl

The Bayesian clustering revealed that the reference populations of the particular species formed separate genetic clusters except for *P. uliginosa*. Two populations of this species grouped with *P. mugo* and one population clustered with *P. uncinata*. Most individuals from the contact zone (PM, PU and HB) grouped with the reference populations of *P. mugo*. There was evidence of genetic admixture in all three groups (PM, PU and HB) resulting most likely from hybridisation with *P. sylvestris* and *P. uncinata*. The HB group should be considered as hybrids between *P. mugo* and *P. uliginosa*. The second genetic cluster included PS. The third genetic cluster was formed by the HPS group – these individuals are apparent hybrids with *P. mugo*/*P. uliginosa* and *P. uncinata* acting as pollen donors and *P. sylvestris* as a mother tree. These results suggest that hybrids succeed in peat-bogs which are unfavourable for any of the parental species, indicating the possible role of natural selection in maintaining intermediate phenotypes.

Contrary to the hypothesis (4), no genetic admixture was found in the reference allopatric populations of *P. mugo* in both studies (*Publication 5* and *6*). Hence, any hybridisation processes that might have occurred during the postglacial recolonisation of Europe when the ranges of *P. mugo* and *P. sylvestris* overlapped did not affect the present genetic variation of the allopatric populations of *P. mugo*.

http://rcin.org.pl

4. Conclusions

The results presented in my doctoral dissertation provide important information about the genetic variation of *P. mugo* and how it is shaped by different demographic and evolutionary processes. The protocols for the multiplex PCRs for *cp*SSRs and *n*SSRs can be used in other population and conservation genetic studies of *P. mugo* and its close relatives and for the tracking of plant material and gene flow. The most important findings include:

(1) Populations of *P. mugo* have the high level of genetic diversity, but they are little differentiated. The neutral genetic differentiation among the particular mountain ranges is even lower that the among population differentiation. These results indicate a relatively recent fragmentation of a historically larger glacial range of the species.

(2) The outcomes of my research point to the predominant role of genetic drift in shaping the genetic divergence of *P. mugo*. Any new mutations are most likely quickly spread by gene flow. Nevertheless, there are some barriers to gene flow, and the pattern of differentiation found at *cp*SSRs can be explained by IBD.

(3) The Sudeten populations of *P. mugo* have most likely been established by the Alpine migrants via the founder effect.

(4) The comparison between the geographic patterns of neutral genetic and morphological variation of *P. mugo* showed that they do not entirely overlap. The eastern populations of *P. mugo* differ morphologically from the other stands, but they do not form a separate group with respect to the neutral genetic variation. This finding suggests the possible role of natural selection in shaping the morphological variation of *P. mugo*.

(5) Hybrid zones occupied by *P. mugo* and its relatives at the Bór na Czerwonem and Zieleniec reserves constitute active examples of ongoing hybridisation. Nevertheless, the gene pool of the allopatric populations of *P. mugo* was not affected by possible historical processes of hybridisation when the ranges of different taxa overlapped.

The presented findings greatly complement our knowledge about the genetic variation of *P. mugo* assessed with different types of markers, some of which (*n*SSRs) used for the first time in this species. In addition to this, the dissertation presents the unique comparative analysis of genetic and morphological data obtained for the same plant material as well as the description of the genetic composition of two hybrid zones. Such studies constitute the first step in further research on the adaptive variation of *P. mugo* and its relatives, which are of utmost importance concerning ongoing and predicted environmental changes.

http://rcin.org.pl

5. References

Barrett RDH, Schluter D (2008). Adaptation from standing genetic variation. *Trends in Ecology & Evolution* 23: 38-44.

Boratyńska K, Boratyński A (2007). Taxonomic differences among closely related pines *Pinus sylvestris*, *P. mugo*, *P. uncinata*, *P. rotundata* and *P. uliginosa* as revealed in needle sclerenchyma cells. *Flora – Morphology, Distribution, Functional Ecology of Plants* 202: 555-569.

Boratyńska K, Jasińska AK, Boratyński A (2015). Taxonomic and geographic differentiation of *Pinus mugo* complex on the needle characteristics. *Systematics and Biodiversity* 13: 581-595.

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

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

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

Ellegren H (2004). Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics* 5: 435-445.

Farcas S, De Beaulieu J-L, Reille M, Coldea G, Diaconeasa B, Goeury C, Goslar T, Jull T (1999). First 14C datings of Late Glacial and Holocene pollen sequences from Romanian Carpathes. *Comptes Rendus de l'Académie des Sciences – Series III – Sciences de la Vie* 322: 799-807.

Fedorkov AL (2002). Variation in the anatomical characters of needles and their resistance to technogenic and climatic stress in Scotch pine. *Russian Journal of Ecology* 33: 65-67.

Hamrick JL, Godt MJW (1996). Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society B: Biological Sciences* 351: 1291-1298.

Huntley B (1990). European vegetation history: Palaeovegatation maps from pollen data – 13 000 yr BP to present. *Journal of Quaternary Science* 5: 103-122.

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

Nielsen R (2005). Molecular signatures of natural selection. *Annual Review of Genetics* 39: 197-218.

Nosil P, Feder JL (2013). Genome evolution and speciation: toward quantitative descriptions of pattern and process. *Evolution* 67: 2461-2467.

Orsini L, Vanoverbeke J, Swillen I, Mergeay J, De Meester L (2013). Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Molecular Ecology* 22: 5983-5999.

http://rcin.org.pl

Prus-Głowacki W, Bączkiewicz A, Wysocka D (2005). Clonal structure of small isolated populations of *Pinus mugo* Turra from peatbogs in the Tatra Mts. *Acta Biologica Cracoviensia Series Botanica* 47: 53-59.

Richardson DM (1998). Ecology and biogeography of *Pinus*. Cambridge University Press, Cambridge, UK.

Sebastiani F, Pinzauti F, Kujala ST, González-Martínez SC, Vendramin GG (2012). Novel polymorphic nuclear microsatellite markers for *Pinus sylvestris* L. *Conservation Genetics Resources* 4: 231-234.

Slavov GT, Zhelev P (2004). Allozyme variation, differentiation, and inbreeding in populations of *Pinus mugo* in Bulgaria. *Canadian Journal of Forest Research* 34: 2611-2617.

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

Wachowiak W, Lewandowski A, Prus-Głowacki W (2005). Reciprocal controlled crosses between *Pinus sylvestris* and *P. mugo* verified by a species-specific *cp*DNA marker. *Journal of Applied Genetics* 46: 41-43.

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

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

Willyard A, Syring J, Gernandt DS, Liston A, Cronn R (2007). Fossil calibration of molecular divergence infers a moderate mutation rate and recent radiations for *Pinus*. *Molecular Biology and Evolution* 24: 90-101.

http://rcin.org.pl

Statements

http://rcin.org.pl

Oświadczenie kierującego pracą

Oświadczam, że niniejsza praca została przygotowana pod moim kierunkiem i stwierdzam, że spełnia ona warunki do przedstawienia jej w postępowaniu o nadanie stopnia doktora nauk biologicznych w dyscyplinie biologia.

22.02.2017

Data

Podpis kierującego pracą

Oświadczenie autora pracy

Świadoma odpowiedzialności karnej z tytułu naruszenia przepisów ustawy z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych (Dz. U. z 2016 r., poz. 666) i konsekwencji dyscyplinarnych określonych w ustawie Prawo o szkolnictwie wyższym (Dz. U. z 2016 r., poz. 1842), a także odpowiedzialności cywilnoprawnej, oświadczam, że niniejsza rozprawa doktorska została napisana przeze mnie samodzielnie.

Jednocześnie oświadczam, że rozprawa ta nie zawiera danych oraz informacji pozyskanych w sposób nielegalny i nie była wcześniej przedmiotem procedur związanych z uzyskaniem stopnia naukowego doktora w innej jednostce.

Ponadto, potwierdzam zgodność niniejszej wersji rozprawy doktorskiej z załączoną wersją elektroniczną.

22.02.2017

Data

Podpis autora pracy

http://rcin.org.pl

Kórnik, 22.02.2017

mgr Weronika Barbara Żukowska
Instytut Dendrologii Polskiej Akademii Nauk
Pracownia Biologii Molekularnej

OŚWIADCZENIE

Oświadczam, że w pracy:

Publikacja 1
Żukowska WB, Wójkiewicz B, Litkowiec M, Wachowiak W (2017). Cross-amplification and multiplexing of cpSSRs and nSSRs in two closely related pine species (Pinus sylvestris L. and P. mugo Turra). Dendrobiology 77: 59-64.

ŻWB, WB i WW opracowali koncepcję pracy; ŻWB, WB i LM opracowali reakcje multipleks PCR; ŻWB i WB przeprowadzili analizy laboratoryjne i statystyczne oraz napisali manuskrypt; WW i LM dokonali krytycznej oceny pracy pod kątem zawartości; ŻWB i WB dokonali poprawy pracy zgodnie z uwagami recenzentów; wszyscy współautorzy przeczytali i zaakceptowali ostateczną wersję pracy.

Publikacja 2
Żukowska WB, Wachowiak W (2017). Nuclear microsatellite markers reveal the low genetic structure of Pinus mugo Turra (dwarf mountain pine) populations in Europe. Plant Systematics and Evolution. doi: 10.1007/s00606-017-1395-x.

ŻWB opracowała koncepcję pracy, przygotowała materiał do analiz, przeprowadziła analizy laboratoryjne i statystyczne oraz napisała manuskrypt; WW dokonał krytycznej oceny pracy pod kątem zawartości; ŻWB i WW dokonali poprawy pracy zgodnie z uwagami recenzentów; wszyscy współautorzy przeczytali i zaakceptowali ostateczną wersję pracy.

Publikacja 3
Żukowska WB, Boratyńska K, Wachowiak W (2017). Comparison of range-wide chloroplast microsatellite and needle trait variation patterns in Pinus mugo Turra (dwarf mountain pine). iForest – Biogeosciences and Forestry. doi: 10.3832/ifor1860-009.

ŻWB opracowała koncepcję pracy, przygotowała materiał do analiz genetycznych oraz przeprowadziła analizy laboratoryjne i statystyczne danych genetycznych; BK przygotowała materiał do analiz morfologicznych oraz przeprowadziła pomiary cech igieł; ŻWB i BK przeprowadziły analizy statystyczne danych morfologicznych; ŻWB pozyskała dane klimatyczne, przeprowadziła analizy porównawcze danych genetycznych, morfologicznych i klimatycznych oraz napisała manuskrypt; WW dokonał krytycznej oceny pracy pod kątem zawartości; ŻWB i WW dokonali poprawy pracy zgodnie z uwagami recenzentów; wszyscy współautorzy przeczytali i zaakceptowali ostateczną wersję pracy.

http://rcin.org.pl

Publikacja 4
Żukowska WB, Wachowiak W (2016). Utility of closely related taxa for genetic studies of adaptive variation and speciation: current state and perspectives in plants with focus on forest tree species. Journal of Systematics and Evolution 54: 17-28.

ŻWB i WW opracowali koncepcję pracy; ŻWB dokonała przeglądu literatury i napisała manuskrypt; WW dokonał krytycznej oceny pracy pod kątem zawartości; ŻWB i WW dokonali poprawy pracy zgodnie z uwagami recenzentów; wszyscy współautorzy przeczytali i zaakceptowali ostateczną wersję pracy.

Publikacja 5
Wachowiak W, Cavers S, Żukowska WB (2015). Interspecific gene flow and ecological selection in a pine (Pinus sp.) contact zone. Plant Systematics and Evolution 301: 1643-1652.

WW wygenerował dane; WW i ŻWB przeprowadzili analizy statystyczne; WW napisał manuskrypt; ŻWB asystowała w redagowaniu manuskryptu; CS i ŻWB dokonali krytycznej oceny i poprawy pracy pod kątem zawartości; wszyscy współautorzy przeczytali i zaakceptowali ostateczną wersję pracy.

Publikacja 6
Wachowiak W, Żukowska WB, Wójkiewicz B, Cavers S, Litkowiec M (2016). Hybridization in contact zone between temperate European pine species. Tree Genetics & Genomes 12: 48.

WW, ŻWB i WB wygenerowali dane oraz przeprowadzili analizy statystyczne; WW napisał manuskrypt; ŻWB i WB asystowali w redagowaniu manuskryptu; wszyscy współautorzy dokonali krytycznej oceny i poprawy pracy pod kątem zawartości oraz przeczytali i zaakceptowali ostateczną wersję pracy.

Podpis

http://rcin.org.pl

Kórnik, 22.02.2017

dr hab. Witold Wachowiak
Instytut Dendrologii Polskiej Akademii Nauk
Pracownia Biologii Molekularnej

OŚWIADCZENIE

Oświadczam, że w pracy:

Publikacja 1
Żukowska WB, Wójkiewicz B, Litkowiec M, Wachowiak W (2017). Cross-amplification and multiplexing of cpSSRs and nSSRs in two closely related pine species (Pinus sylvestris L. and P. mugo Turra). Dendrobiology 77: 59-64.

ŻWB, WB i WW opracowali koncepcję pracy; ŻWB, WB i LM opracowali reakcje multipleks PCR; ŻWB i WB przeprowadzili analizy laboratoryjne i statystyczne oraz napisali manuskrypt; WW i LM dokonali krytycznej oceny pracy pod kątem zawartości; ŻWB i WB dokonali poprawy pracy zgodnie z uwagami recenzentów; wszyscy współautorzy przeczytali i zaakceptowali ostateczną wersję pracy.

Publikacja 2
Żukowska WB, Wachowiak W (2017). Nuclear microsatellite markers reveal the low genetic structure of Pinus mugo Turra (dwarf mountain pine) populations in Europe. Plant Systematics and Evolution. doi: 10.1007/s00606-017-1395-x.

ŻWB opracowała koncepcję pracy, przygotowała materiał do analiz, przeprowadziła analizy laboratoryjne i statystyczne oraz napisała manuskrypt; WW dokonał krytycznej oceny pracy pod kątem zawartości; ŻWB i WW dokonali poprawy pracy zgodnie z uwagami recenzentów; wszyscy współautorzy przeczytali i zaakceptowali ostateczną wersję pracy.

Publikacja 3
Żukowska WB, Boratyńska K, Wachowiak W (2017). Comparison of range-wide chloroplast microsatellite and needle trait variation patterns in Pinus mugo Turra (dwarf mountain pine). iForest – Biogeosciences and Forestry. doi: 10.3832/ifor1860-009.

ŻWB opracowała koncepcję pracy, przygotowała materiał do analiz genetycznych oraz przeprowadziła analizy laboratoryjne i statystyczne danych genetycznych; BK przygotowała materiał do analiz morfologicznych oraz przeprowadziła pomiary cech igieł; ŻWB i BK przeprowadziły analizy statystyczne danych morfologicznych; ŻWB pozyskała dane klimatyczne, przeprowadziła analizy porównawcze danych genetycznych, morfologicznych i klimatycznych oraz napisała manuskrypt; WW dokonał krytycznej oceny pracy pod kątem zawartości; ŻWB i WW dokonali poprawy pracy zgodnie z uwagami recenzentów; wszyscy współautorzy przeczytali i zaakceptowali ostateczną wersję pracy.

http://rcin.org.pl

Publikacja 4

Żukowska WB, Wachowiak W (2016). Utility of closely related taxa for genetic studies of adaptive variation and speciation: current state and perspectives in plants with focus on forest tree species. Journal of Systematics and Evolution 54: 17-28.

ŻWB i WW opracowali koncepcję pracy; ŻWB dokonała przeglądu literatury i napisała manuskrypt; WW dokonał krytycznej oceny pracy pod kątem zawartości; ŻWB i WW dokonali poprawy pracy zgodnie z uwagami recenzentów; wszyscy współautorzy przeczytali i zaakceptowali ostateczną wersję pracy.

Publikacja 5

Wachowiak W, Cavers S, Żukowska WB (2015). Interspecific gene flow and ecological selection in a pine (Pinus sp.) contact zone. Plant Systematics and Evolution 301: 1643-1652.

WW wygenerował dane; WW i ŻWB przeprowadzili analizy statystyczne; WW napisał manuskrypt; ŻWB asystowała w redagowaniu manuskryptu; CS i ŻWB dokonali krytycznej oceny i poprawy pracy pod kątem zawartości; wszyscy współautorzy przeczytali i zaakceptowali ostateczną wersję pracy.

Publikacja 6

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

WW, ŻWB i WB wygenerowali dane oraz przeprowadzili analizy statystyczne; WW napisał manuskrypt; ŻWB i WB asystowali w redagowaniu manuskryptu; wszyscy współautorzy dokonali krytycznej oceny i poprawy pracy pod kątem zawartości oraz przeczytali i zaakceptowali ostateczną wersję pracy.

Podpis

http://rcin.org.pl

Kórnik, 22.02.2017

dr Monika Litkowiec
Instytut Dendrologii Polskiej Akademii Nauk
Pracownia Biologii Molekularnej

OŚWIADCZENIE

Oświadczam, że w pracy:

Publikacja 1
Żukowska WB, Wójkiewicz B, Litkowiec M, Wachowiak W (2017). Cross-amplification and multiplexing of cpSSRs and nSSRs in two closely related pine species (Pinus sylvestris L. and P. mugo Turra). Dendrobiology 77: 59-64.

ŻWB, WB i WW opracowali koncepcję pracy; ŻWB, WB i LM opracowali reakcje multipleks PCR; ŻWB i WB przeprowadzili analizy laboratoryjne i statystyczne oraz napisali manuskrypt; WW i LM dokonali krytycznej oceny pracy pod kątem zawartości; ŻWB i WB dokonali poprawy pracy zgodnie z uwagami recenzentów; wszyscy współautorzy przeczytali i zaakceptowali ostateczną wersję pracy.

Publikacja 6
Wachowiak W, Żukowska WB, Wójkiewicz B, Cavers S, Litkowiec M (2016). Hybridization in contact zone between temperate European pine species. Tree Genetics & Genomes 12: 48.

WW, ŻWB i WB wygenerowali dane oraz przeprowadzili analizy statystyczne; WW napisał manuskrypt; ŻWB i WB asystowali w redagowaniu manuskryptu; wszyscy współautorzy dokonali krytycznej oceny i poprawy pracy pod kątem zawartości oraz przeczytali i zaakceptowali ostateczną wersję pracy.

Podpis

http://rcin.org.pl

Kórnik, 22.02.2017

mgr Błażej Wójkiewicz
Instytut Dendrologii Polskiej Akademii Nauk
Pracownia Biologii Molekularnej

OŚWIADCZENIE

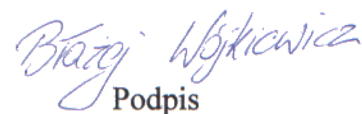
Oświadczam, że w pracy:

Publikacja 1
Żukowska WB, Wójkiewicz B, Litkowiec M, Wachowiak W (2017). Cross-amplification and multiplexing of cpSSRs and nSSRs in two closely related pine species (Pinus sylvestris L. and P. mugo Turra). Dendrobiology 77: 59-64.

ŻWB, WB i WW opracowali koncepcję pracy; ŻWB, WB i LM opracowali reakcje multipleks PCR; ŻWB i WB przeprowadzili analizy laboratoryjne i statystyczne oraz napisali manuskrypt; WW i LM dokonali krytycznej oceny pracy pod kątem zawartości; ŻWB i WB dokonali poprawy pracy zgodnie z uwagami recenzentów; wszyscy współautorzy przeczytali i zaakceptowali ostateczną wersję pracy.

Publikacja 6
Wachowiak W, Żukowska WB, Wójkiewicz B, Cavers S, Litkowiec M (2016). Hybridization in contact zone between temperate European pine species. Tree Genetics & Genomes 12: 48.

WW, ŻWB i WB wygenerowali dane oraz przeprowadzili analizy statystyczne; WW napisał manuskrypt; ŻWB i WB asystowali w redagowaniu manuskryptu; wszyscy współautorzy dokonali krytycznej oceny i poprawy pracy pod kątem zawartości oraz przeczytali i zaakceptowali ostateczną wersję pracy.



Podpis

http://rcin.org.pl

Kórnik, 22.02.2017

prof. Krystyna Boratyńska
Instytut Dendrologii Polskiej Akademii Nauk
Pracownia Biologii Molekularnej

OŚWIADCZENIE

Oświadczam, że w pracy:

Publikacja 3
Żukowska WB, Boratyńska K, Wachowiak W (2017). Comparison of range-wide chloroplast microsatellite and needle trait variation patterns in Pinus mugo Turra (dwarf mountain pine). iForest – Biogeosciences and Forestry. doi: 10.3832/ifor1860-009.

ŻWB opracowała koncepcję pracy, przygotowała materiał do analiz genetycznych oraz przeprowadziła analizy laboratoryjne i statystyczne danych genetycznych; BK przygotowała materiał do analiz morfologicznych oraz przeprowadziła pomiary cech igieł; ŻWB i BK przeprowadziły analizy statystyczne danych morfologicznych; ŻWB pozyskała dane klimatyczne, przeprowadziła analizy porównawcze danych genetycznych, morfologicznych i klimatycznych oraz napisała manuskrypt; WW dokonał krytycznej oceny pracy pod kątem zawartości; ŻWB i WW dokonali poprawy pracy zgodnie z uwagami recenzentów; wszyscy współautorzy przeczytali i zaakceptowali ostateczną wersję pracy.

Podpis

http://rcin.org.pl

Cc: Weronika B. Żukowska
Institute of Dendrology PAS
Parkowa 5
62-035 Kórnik
Poland

Edinburgh, 14.06.2016

To Whom It May Concern:

This is to certify that as a part of our collaborative research we were co-authors of two publications listed below.

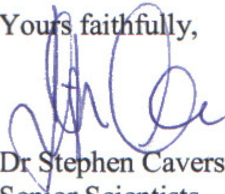
Paper 1:
Wachowiak W., Cavers S., Żukowska WB. (2015). Interspecific gene flow and ecological selection in a pine (*Pinus* sp.) contact zone. ***Plant Systematics and Evolution*** 301: 1643-1652

Authors' contributions: WW wrote the manuscript; WW generated data; WW, ŻWB analysed data; ŻWB assisted in drafting the manuscript. SC, ŻWB critically reviewed and revised the manuscript for content; all authors read and approved the final manuscript.
Author`s percentage share (%): 60:10:30

Paper 2:
Wachowiak W., Żukowska WB., Wójkiewicz B., Cavers S., Litkowiec M. (2016) Hybridization in contact zone between temperate European pine species. ***Tree Genetics & Genomes*** 12: 48 DOI: 10.1007/s11295-016-1007-x

Authors' contributions: WW wrote the manuscript, WW, ŻWB, WB generated data and analysed data; ŻWB, WB assisted with drafting the manuscript. All authors critically reviewed and revised the manuscript for content; all authors read and approved the final manuscript.
Author`s percentage share (%): 55:20:15:5:5

Yours faithfully,



Dr Stephen Cavers
Senior Scientists
Centre for Ecology & Hydrology, Edinburgh, UK
scav@ceh.ac.uk

http://rcin.org.pl

Appendices

http://rcin.org.pl

Publication 1

Żukowska WB, Wójkiewicz B, Litkowiec M, Wachowiak W

Cross-amplification and multiplexing of *cp*SSRs and *n*SSRs in two closely related pine species (*Pinus sylvestris* L. and *P. mugo* Turra)

Dendrobiology 77: 59-64

Impact factor: 0.643

MNiSW points: 20

http://rcin.org.pl

DENDROBIOLOGY

2017, vol. 77, 59–64

http://dx.doi.org/10.12657/

Weronika B. Żukowska[*]*, Błażej Wójkiewicz*, Monika Litkowiec, Witold Wachowiak*

Cross-amplification and multiplexing of cpSSRs and nSSRs in two closely related pine species (*Pinus sylvestris* L. and *P. mugo* Turra)

Received: 14 April 2016; Accepted: 16 November 2016

Abstract: Background: Simple sequence repeats (SSRs) are widespread molecular markers commonly used in population genetic studies. Nowadays, next-generation sequencing (NGS) methods allow identifying thousands of SSRs in one sequencing run, which greatly facilitates isolation and development of new SSRs. However, their usefulness as molecular markers still must be tested empirically on a number of populations to select SSRs with best parameters for future population genetic research. An alternative approach, cheaper and faster than isolation and characterization of new SSRs, involves cross-amplification of SSRs in closely related species.

Aims: Our goal was to develop multiplex PCR protocols that will be useful in population genetic studies of Scots pine (*Pinus sylvestris* L.) and dwarf mountain pine (*P. mugo* Turra), and possibly other pine species.

Methods: We tested 14 chloroplast (cpSSRs) and 22 nuclear (nSSRs) microsatellite markers originally designed for Japanese black pine (*P. thunbergii* Parl.), *P. sylvestris* and loblolly pine (*P. taeda* L.) in four populations of *P. sylvestris* and *P. mugo* across different locations in Europe. We designed six multiplex PCRs, which were subsequently screened for their ability to provide repeatable and high quality amplification products using capillary electrophoresis.

Results: The transfer rate in our study was similar in both pine species, and it was very high for cpSSRs (93% and 86% for *P. sylvestris* and *P. mugo*, respectively) and moderate for nSSRs (59% for both species). We managed to design five well-performing multiplex reactions out of six initially tested. Most of the tested loci were polymorphic. Moreover, the allelic patterns detected at some cpSSRs were species-specific.

Conclusions: We provide a set of five multiplexes which can be used in genetic studies of both *P. sylvestris* and *P. mugo*. Chloroplast marker PCP30277 is a good candidate for a cheap species diagnostic marker suitable for tracking interspecific gene flow between hybridizing species of *P. sylvestris* and *P. mugo*.

Keywords: chloroplast microsatellites, dwarf mountain pine, hybridization, nuclear microsatellites, Scots pine

Addresses: W. B. Żukowska, Institute of Dendrology, Polish Academy of Sciences, Parkowa 5, 62-035 Kórnik, Poland, e-mail: wzukowska@man.poznan.pl

B. Wójkiewicz, Institute of Dendrology, Polish Academy of Sciences, Parkowa 5, 62-035 Kórnik, Poland, email: bwojkiew@man.poznan.pl

M. Litkowiec, Institute of Dendrology, Polish Academy of Sciences, Parkowa 5, 62-035 Kórnik, Poland, email: mlit@man.poznan.pl

W. Wachowiak, Institute of Environmental Biology, Faculty of Biology, Adam Mickiewicz University, Umultowska 89, 61-614 Poznań, Poland; Institute of Dendrology, Polish Academy of Sciences, Parkowa 5, 62-035 Kórnik, Poland, e-mail: witwac@amu.edu.pl

* – W.B. Żukowska and B. Wójkiewicz contributed equally to this work.

http://rcin.org.pl

Introduction

Microsatellites (=simple sequence repeats; SSRs or short tandem repeats; STRs) are the class of repetitive DNA sequences present in both eukaryotic and prokaryotic genomes. With respect to population genetics of forest tree species, microsatellites have proved to be useful neutral molecular markers in studies focusing on genetic diversity (e.g. Chybicki et al., 2011; Litkowiec et al., 2015; Wójkiewicz & Wachowiak, 2016), mating systems (e.g. Lian et al., 2001) and gene mapping (e.g. Echt et al., 2011) due to their high level of allelic variation and co-dominant mode of inheritance. The popularity of SSRs in genetic research of trees is also connected with the fact that they can be genotyped in one multiplex polymerase chain reaction (PCR). This technique allows amplification of two or more DNA fragments simultaneously. The possibility of multiplexing combined with capillary electrophoresis, which is based on a laser-induced fluorescence DNA technology, results in a cost-effective tool for genotyping large quantities of independent samples.

Till the next-generation sequencing (NGS) era, the development of novel microsatellite markers for forest tree species was difficult, costly and time-consuming. Currently, it is possible to identify thousands of microsatellite regions during one sequencing run of a genome or transcriptome. As a result, the isolation of new SSRs is no longer a real challenge, practically for any organism, including trees. Regardless of this, the usefulness of novel SSRs for population genetic studies still must be tested to verify which of them 1) provide repeatable, polymorphic and high quality amplification products, 2) are the most informative and 3) are transferable, which gives opportunity to perform genetic analyses at interspecific level.

The objects of our study were two very closely related pine species: Scots pine (*Pinus sylvestris* L.) and dwarf mountain pine (*P. mugo* Turra). At present these species have mostly allopatric distribution. *P. sylvestris* is the most widespread conifer in Europe and Asia, whereas *P. mugo* is typical to the mountain regions of Europe. We aimed at developing of efficient multiplex protocols for the amplification of chloroplast and nuclear SSRs (cpSSRs and nSSRs, respectively) in *P. sylvestris* and *P. mugo*, which we had pre-selected from a collection of 36 SSRs originally designed for *P. thunbergii* Parl., *P. sylvestris* and *P. taeda* L. (Table 1). The results of the cross-species amplification of cpSSRs and nSSRs are discussed in the light of their utility for future genetic research.

Methods

Four populations of *P. sylvestris* (128 individuals) and four populations of *P. mugo* (105 individuals) across different locations in Europe were analysed in this study (Table 2). The collected samples were stored in -20°C until DNA extraction. Genomic DNA was extracted from 50-100 mg of needle tissue, following the CTAB protocol as described by Dumolin et al. (1995). RNase A was added to the final incubation step. The DNA concentration was measured with BioPhotometer (Eppendorf AG, Germany) and adjusted to 15 ng/μl.

We selected 14 chloroplast and 22 nuclear microsatellite markers available in the published literature (Table 1). CpSSRs were initially developed for *P. thunbergii*, whereas nSSRs for *P. sylvestris* and *P. taeda*. The markers were combined into six multiplex PCRs and screened for their ability to provide repeatable and high quality polymorphic amplification products of expected size. The loci were finally amplified in five multiplex PCRs in Applied Biosystems Veriti and 2720 thermal cyclers (Life Technologies, USA). The PCRs were carried out in a total volume of 10 μl, using the Qiagen Multiplex PCR kit (Qiagen, Germany). Each reaction contained about 45 ng of template DNA, 1x Qiagen Multiplex PCR Master Mix, 0.5x Q-Solution and 0.05-0.1 μM each of forward and reverse primers. All primers were tested individually prior to the performance of multiplex reactions. We used equimolar concentration of primers in the initial amplification procedures, which were subsequently adjusted to obtain an even intensity of the fluorescence signal. Amplification conditions were optimised across all multiplexes for both pine species. Details of final PCR parameters are described in Table 1. The fluorescently labelled PCR products were separated on a capillary sequencer, the Applied Biosystems 3130 Genetic Analyzer (Life Technologies, USA). The GeneScan 500 LIZ Size Standard (Life Technologies, USA) was used as an internal size standard. The raw data were scored with the GeneMapper Software ver 4.0 (Life Technologies, USA), checked manually and converted into discrete allele sizes with the use of the AlleloBin software (Prasanth et al., 2006).

Two parameters were calculated for each species for cpSSRs: the number of alleles (A_N) and unbiased diversity (A_{uh}) using GenAlEx ver 6.5 (Peakall & Smouse, 2006). A_{uh} was computed as mean across all populations for each species. With regard to nSSRs, we used the multiple sample score test (U test for heterozygote deficit, Raymond and Rousset 1995), implemented in GENEPOP ver 4.3 (Rousset, 2008), to assess the significance of departures from Hardy-Weinberg equilibrium (HWE) for each locus, separately for each species. The frequency of null alleles (NAF) was estimated using FreeNA (Chapuis & Estoup, 2007) separately for each population and each species. A_N, effective number of alleles (A_E), observed and expected heterozygosity (H_O and H_E, re-

http://rcin.org.pl

Table 1. A list of multiplexes and thermocycling conditions for *P. sylvestris* and *P. mugo*. Multiplex 4 (nSSR) is omitted as the loci (psyl17 (Sebastiani et al., 2012), ptTX3116 (Elsik & Williams, 2001), SPAC11.6, SPAC 11.8, SPAC 12.5 (Soranzo et al., 1998)) failed to amplify in both *P. sylvestris* and *P. mugo*. Each reaction consisted of the following steps: I – initial denaturation, II – denaturation, III – annealing, IV – elongation, V – final elongation

Multiplex	Loci	Step	*P. sylvestris*	*P. mugo*
1 (cpSSR)	Pt15169, Pt26081, Pt30204, Pt36480, Pt45002, Pt71936 (Vendramin et al., 1996)	I	95°C, 15 min.	95°C, 15 min.
		II	94°C, 15 sec.	94°C, 30 sec.
		III	58°C, 90 sec.	58°C, 45 sec.
		IV	72°C, 90 sec.; go to II × 27	72°C, 90 sec.; go to II × 30
		V	72°C, 10 min.	72°C, 10 min.
2 (cpSSR)	PCP1289, PCP26106, PCP30277, PCP36567, PCP41131, PCP45071, PCP87314, PCP102652 (Provan et al., 1998)	I	95°C, 15 min.	95°C, 15 min.
		II	94°C, 15 sec.	94°C, 30 sec.
		III	60°C, 90 sec.	60°C, 45 sec.
		IV	72°C, 90 sec.; go to II × 27	72°C, 90 sec.; go to II × 30
		V	72°C, 10 min.	72°C, 10 min.
3 (nSSR)	psyl2, psyl16, psyl18, psyl19, psyl25, psyl36, psyl42, psyl44, psyl57 (Sebastiani et al., 2012)	I	95°C, 15 min.	95°C, 15 min.
		II	94°C, 30 sec.	94°C, 30 sec.
		III	57°C, 90 sec.	55°C, 90 sec.
		IV	72°C, 90 sec.; go to II × 37	72°C, 90 sec.; go to II × 37
		V	72°C, 10 min.	72°C, 15 min.
5 (nSSR)	ptTX2146 (Elsik et al., 2000), ptTX3107 (Elsik & Williams, 2001), SPAG 7.14 (Soranzo et al., 1998)	I	95°C, 15 min.	95°C, 15 min.
		II	94°C, 30 sec.	94°C, 30 sec.
		III	55°C, 90 sec.	56°C, 90 sec.
		IV	72°C, 90 sec.; go to II × 29	72°C, 90 sec.; go to II × 34
		V	72°C, 10 min.	72°C, 15 min.
6 (nSSR)	ptTX3025, ptTX3032 (Elsik et al., 2000), ptTX4001, ptTX4011 (Zhou et al., 2002), SPAC 11.4 (Soranzo et al., 1998)	I	95°C, 15 min.	95°C, 15 min.
		II-1	94°C, 30 sec.	94°C, 30 sec.
		III-1	60°C Δ↓1°C/cycle, 40 sec.	65°C Δ↓1°C/cycle, 40 sec.
		IV-1	72°C, 90 sec.; go to II-1 × 9	72°C, 60 sec.; go to II-1 × 9
		II-2	94°C, 30 sec.	94°C, 30 sec.
		III-2	50°C, 40 sec.	55°C, 60 sec.
		IV-2	72°C, 90 sec.; go to II-2 × 35	72°C, 60 sec.; go to II-2 × 31
		V	72°C, 10 min.	72°C, 7 min.

spectively) were calculated in GenAlEx ver 6.5 across all populations separately for each species.

Results & Discussion

The transfer rates were very similar in both *P. sylvestris* and *P. mugo*. We managed to transfer 13 (93%) and 12 (86%) out of 14 initially tested chloroplast microsatellites to *P. sylvestris* and *P. mugo*, respectively. Locus Pt36480 was successfully transferred only to *P. sylvestris*. Similar high values of transfer rates for cpSSRs were noted previously by Dzialuk and Burczyk (2004), who proposed a multiplex PCR that consisted of six loci for population studies in *P. sylvestris*. With regard to nuclear microsatellites, the transfer rates were moderate (59%) for both pines. Similarly to our results, moderately low (26%) transfer rates were demonstrated by Celiński et al. (2013), who tested the transferability of 19 nSSRs from *P. sylvestris* and *P. taeda* to *P. mugo*. In our study, 13 out of 22 nSSRs were amplified successfully in both species, but some loci that failed to amplify or gave poor results in *P. sylvestris* turned out to be useful for *P. mugo* and vice-versa (ptTX3107 and SPAC 11.4 only for *P. sylvestris*, whereas psyl16 and ptTX4001 only for *P. mugo*). Our results clearly show that the amplification of cpSSRs was more successful than nSSRs, which is most likely associated with the fact that the mutation rate of chloroplast DNA is lower than of nuclear DNA (Willyard et al., 2007). As a result, the high sequence conservation among chloroplast genomes of conifers allows successful amplification of cpSSRs designed for *P. thunbergii* in closely (as in our study) or more distantly related conifer species.

Allelic variation of the analysed loci was high with mean 7.12 and 6.32 alleles per locus for *P. sylvestris* and *P. mugo*, respectively. Nearly all successfully am-

http://rcin.org.pl

Table 2. Descriptive statistics of the studied cpSSR and nSSR markers in *P. sylvestris* (S) and *P. mugo* (M)*. A_N – number of alleles; A_{uh} – unbiased diversity (mean for all populations); A_E – effective number of alleles; H_O – observed heterozygosity; H_E – expected heterozygosity; NAF – null allele frequency (range for all populations). Test for heterozygote deficit: ns – not significant; * – $p < 0.05$; ** – $p < 0.01$; *** – $p < 0.001$

Locus	Size range [bp]	A_N	A_{uh}	A_E	H_O	H_E	NAF
	S/M	S/M	S/M	S/M	S/M	S/M	S/M
Pt15169	124–130/121–126	7/5	0.75/0.56	–/–	–/–	–/–	–/–
Pt26081	110–112/109–112	3/4	0.26/0.48	–/–	–/–	–/–	–/–
Pt30204	140–148/143–149	9/7	0.79/0.79	–/–	–/–	–/–	–/–
Pt36480	143–145/–	3/–	0.18/–	–/–	–/–	–/–	–/–
Pt71936	148–154/145–149	7/5	0.64/0.62	–/–	–/–	–/–	–/–
PCP1289	108–111/107–108	4/2	0.35/0.17	–/–	–/–	–/–	–/–
PCP26106	146–148/145–148	3/4	0.28/0.50	–/–	–/–	–/–	–/–
PCP30277	134–140/115–120	7/6	0.77/0.75	–/–	–/–	–/–	–/–
PCP36567	110–112/110–112	3/3	0.12/0.47	–/–	–/–	–/–	–/–
PCP41131	139–143/140–159	5/10	0.16/0.69	–/–	–/–	–/–	–/–
PCP45071	153–156/146–151	4/6	0.45/0.55	–/–	–/–	–/–	–/–
PCP87314	112–114/112–116	3/5	0.32/0.68	–/–	–/–	–/–	–/–
PCP102652	114–116/114	3/1	0.03/0.00	–/–	–/–	–/–	–/–
psyl2	207–213/198–210	3/5	–/–	1.29/1.54	0.21/0.32	0.22/0.34	0.00–0.13^{ns}/ 0.00–0.13^{ns}
psyl16	–/201–213	–/6	–/–	–/3.03	–/0.64	–/0.67	–/0.00–0.06^{ns}
psyl18	292–307/292–304	6/5	–/–	1.28/1.18	0.16/0.12	0.21/0.15	0.00–0.08*/ 0.00–0.12**
psyl25	216–219/213–219	2/3	–/–	1.02/1.57	0.02/0.38	0.02/0.36	0.00^{ns}/ 0.00–0.03^{ns}
psyl36	250–262/250–262	5/5	–/–	1.27/1.12	0.22/0.07	0.21/0.10	0.00^{ns}/ 0.00–0.13**
psyl42	167–179/169–177	7/4	–/–	3.25/2.10	0.69/0.51	0.69/0.50	0.00–0.03^{ns}/ 0.00–0.05^{ns}
psyl44	169–178/169–175	4/2	–/–	1.19/1.29	0.15/0.26	0.16/0.22	0.00–0.06^{ns}/ 0.00^{ns}
psyl57	190–208/190–205	7/6	–/–	2.35/2.63	0.62/0.62	0.57/0.61	0.00–0.02^{ns}/ 0.00–0.09^{ns}
ptTX2146	180–252/153–264	17/17	–/–	3.86/3.24	0.74/0.64	0.74/0.63	0.00–0.04^{ns}/ 0.00–0.01^{ns}
ptTX3107	153–183/–	8/–	–/–	4.39/–	0.44/–	0.77/–	0.15–0.26***/–
SPAG 7.14	177–257/185–265	30/28	–/–	14.29/11.56	0.77/0.80	0.93/0.91	0.00–0.14***/ 0.00–0.15***
ptTX3025	266–299/266–275	7/4	–/–	1.90/1.27	0.43/0.19	0.47/0.21	0.00–0.12*/ 0.00–0.11*
ptTX4001	–/205–221	–/6	–/–	–/2.66	–/0.53	–/0.58	–/0.00–0.05^{ns}
ptTX4011	256–280/262–284	10/9	–/–	3.10/3.27	0.62/0.60	0.68/0.67	0.00–0.15**/ 0.00–0.18**
SPAC 11.4	130–166/–	18/–	–/–	7.10/–	0.88/–	0.85/–	0.00–0.02^{ns}/–
Mean		**7.12/6.32**	**0.39/0.52**	**3.56/2.80**	**0.46/0.44**	**0.50/0.46**	**0.04/0.03**

*Populations analysed in the study (long./lat.):

S: Joutsa, Finland (25°45'0"/64°41'24"); Tatras, Poland (20°21'36"/49°25'12"); Divčibare Mts, Serbia (44°6'0"/19°59'24"); St. Miguel d'Engolasters, Andorra (42°40'12"/0°46'12").

M: Sudetes, Poland (15°47'50"/50°44'40"); Carnic Alps, Italy (13°15'35"/46°32'45"); Carpathians, Romania (24°32'19"/45°36'30"); Dinaric Alps, Bosnia and Herzegovina (18°13'8"/43°45'0").

plified cpSSRs were polymorphic, exhibiting between two to ten alleles. Only PCP102652 was monomorphic in *P. mugo* (114 bp), whereas almost all individuals of *P. sylvestris* (99%) carried the 115 bp variant. In the case of nSSRs, A_N was lower for markers developed by Sebastiani et al. (2012) (the 'psyl' series; A_N between two and seven) than for other nSSRs (from four for ptTX3025 in *P. mugo* up to 30 for SPAG 7.14 in *P. sylvestris*). The mean value of unbiased diversity (mean A_{uh}) parameter, calculated for cpSSRs, did not differ statistically between the studied pines (mean A_{uh} = 0.39 and mean A_{uh} = 0.52 for *P. sylvestris* and *P. mugo*, respectively; Student's t-test: $p = 0.20$). As for nSSRs, the difference between the mean effective number of alleles (A_E) was also not significant (3.56 for *P. sylvestris* vs. 2.80 for *P. mugo*; U Mann-Whit-

http://rcin.org.pl

ney test: p = 0.63). Significant heterozygote deficit was observed for six loci (psyl18, psyl36, ptTX3107, SPAG 7.14, ptTX3025, and ptTX4011). The frequency of null alleles (NAF) differed across loci and, to a lesser extent, between species (NAF = 0.00-0.26). The mean observed and expected heterozygosity (H_O and H_E, respectively) were similar in both species (mean H_O = 0.46, mean H_E = 0.50 and mean H_O = 0.44, mean H_E = 0.46 for *P. sylvestris* and *P. mugo*, respectively; Student's t-test: p = 0.84 for H_O and p = 0.69 for H_E). For most loci H_O was only slightly lower than H_E. Some loci, however, displayed H_O greater than H_E. Microsatellites with higher number of repeats generally displayed higher heterozygosity values (Table 2). Based on our results, we recommend to omit some nSSR loci with the frequency of null alleles exceeding 5%, including psyl18, ptTX3107, SPAG 7.14, and ptTX4011. Alternatively, a proper correction methods should be applied as, according to the simulation study by Chapuis and Estoup (2007), the levels of classical parameters used to describe population differentiation are overestimated in the presence of null alleles.

Loci that exhibit species-specific allelic patterns are ideal for studies of interspecific gene flow and identification of hybrid zones. In the present work, the most pronounced differences were apparent for 2 cpSSRs: PCP45071 and PCP30277. Alleles scored for these loci did not overlap when the two species were taken into account. Only 2 bp difference was observed for PCP45071 and it does not seem to be a species-specific polymorphism as compared to other studies (Wójkiewicz & Wachowiak, 2016). The difference for PCP30277 was at least 14 bp (Table 2), and this locus can be useful as a diagnostic marker to track interspecific gene flow in the species' contact zones. Regarding interspecific differences for nSSRs, we observed opposing tendencies for psyl2 and SPAG 7.14. Higher variants in *P. sylvestris* as compared to *P. mugo* were identified for psyl2, whereas lower sizes were typical for SPAG 7.14. Variants scored for *P. mugo* represented a subset of those identified in *P. sylvestris* for four loci: psyl42, psyl44, psyl57, and ptTX3025. For these markers, longer alleles, preferred in *P. sylvestris*, were absent in *P. mugo*. The same A_N was observed for ptTX2146 for both *P. sylvestris* and *P. mugo*, but some individuals of *P. mugo* had alleles shorter and others longer than *P. sylvestris*. As oppose to cpSSRs, there was no locus which had non-overlapping alleles when compared in both pine species (Table 2).

Conclusions

We provide five well-performing multiplexes consisting of sets of chloroplast and nuclear microsatellites that can be applied in population and conservation genetic studies of both *P. sylvestris* and *P. mugo*, and possibly of other pine species, e.g. from the *P. mugo* complex. The markers seem particularly useful for the assessment of the background neutral genetic variation that is necessary to further look for genetic signatures of natural selection in candidate genomic regions. Due to their high genetic variability, they could also be applied in the identification and tracking of plant material. Furthermore, the marker that exhibits species-specific allelic patterns (PCP30277) seems ideal for studies of interspecific gene flow in the species' contact zones. Such studies accompanied by analyses of sequence variation at candidate genomic regions will help to address questions related to the role of hybridization in evolution of *P. sylvestris* and *P. mugo* (Wachowiak et al., 2015, 2016). Our study clearly confirms that cross-amplification seems to be a good first choice alternative to the *de novo* development of microsatellite markers, especially for species with poor genomic resources. The possibility of genotyping using multiplex PCRs makes their application additionally time and cost-effective.

Acknowledgments

This work was financially supported by the Polish National Science Centre (Grant No. DEC-2012/05/E/NZ9/03476), and the Institute of Dendrology of the Polish Academy of Sciences provided additional funding.

References

Celiński K, Pawlaczyk EM, Wojnicka-Półtorak A, Chudzińska E & Prus-Głowacki W (2013) Cross-species amplification and characterization of microsatellite loci in *Pinus mugo* Turra. Biologia 68: 621-626. doi:10.2478/s11756-013-0189-z.

Chapuis MP & Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. Molecular Biology and Evolution 24: 621–631. doi:10.1093/molbev/msl191.

Chybicki IJ, Oleksa A & Burczyk J (2011) Increased inbreeding and strong kinship structure in *Taxus baccata* estimated from both AFLP and SSR data. Heredity 107: 589–600. doi:10.1038/hdy.2011.51.

Dumolin S, Demesure B & Petit RJ (1995) Inheritance of chloroplast and mitochondrial genomes in pedunculate oak investigated with an efficient PCR method. Theoretical and Applied Genetics 91: 1253–1256. doi:10.1007/BF00220937.

Dzialuk A & Burczyk J (2004) PCR-multiplex of six chloroplast microsatellites for population studies and genetic typing in *Pinus sylvestris*. Silvae Genetica 53: 246–248.

http://rcin.org.pl

Echt CS, Saha S, Krutovsky KV, Wimalanathan K, Erpelding JE, Liang C & Nelson CD (2011) An annotated genetic map of loblolly pine based on microsatellite and cDNA markers. BMC Genetics 12: 17. doi:10.1186/1471-2156-12-17.

Elsik CG, Minihan VT, Hall SE, Scarpa AM & Williams CG (2000) Low-copy microsatellite markers for *Pinus taeda* L. Genome 43: 550–555. doi:10.1139/g00-002.

Elsik CG & Williams CG (2001) Low-copy micro-satellite recovery from a conifer genome. Theoretical and Applied Genetics 103: 1189–1195. doi:10.1007/s001220100725.

Lian C, Miwa M & Hogetsu T (2001) Outcrossing and paternity analysis of *Pinus densiflora* (Japanese red pine) by microsatellite polymorphism. Heredity 87: 88–98. doi:10.1046/j.1365-2540.2001.00913.x.

Litkowiec M, Plitta-Michalak BP, Lewandowski A & Iszkuło G (2015) Homogenous genetic structure in populations of *Taxus baccata* with varied proportions of male and female individuals. Silva Fennica 49. doi:10.14214/sf.1236.

Peakall R & Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6: 288–295. doi:10.1111/j.1471-8286.2005.01155.x.

Prasanth VP, Chandra S, Hoisington DA & Jayashree B (2006) AlleloBin – A program for allel binning of microsatellite markers based on the algorithm of Idury and Cardon (1997). ICRISAT. International Crops Research Institute for the Semi-Arid Tropics. doi:10.13140/2.1.4397.2165.

Provan J, Soranzo N, Wilson NJ, McNicol JW, For-rest GI, Cottrell J & Powell W (1998) Gene-pool variation in Caledonian and European Scots pine (*Pinus sylvestris* L.) revealed by chloroplast simple-sequence repeats. Proceedings of the Royal Society B: Biological Sciences 265: 1697–1705. doi:10.1098/rspb.1998.0491.

Raymond M & Rousset F (1995) An exact test for population differentiation. Evolution 49: 1280– 1283. doi:10.2307/2410454.

Rousset F (2008) Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Molecular Ecology Resources 8: 103–106. doi:10.1111/j.1471-8286.2007.01931.x.

Sebastiani F, Pinzauti F, Kujala ST, González-Martínez SC & Vendramin GG (2012) Novel polymorphic nuclear microsatellite markers for *Pinus sylvestris* L. Conservation Genetics Resources 4: 231–234. doi:10.1007/s12686-011-9513-5.

Soranzo N, Provan J & Powell W (1998) Characterization of microsatellite loci in *Pinus sylvestris* L. Molecular Ecology 7: 1260–1261.

Vendramin GG, Lelli L, Rossi P & Morgante M (1996) A set of primers for the amplification of 20 chloroplast microsatellites in Pinaceae. Molecular Ecology 5: 595–598. doi:10.1111/j.1365-294X.1996.tb00353.x.

Wachowiak W, Cavers S & Żukowska WB (2015) Interspecific gene flow and ecological selection in a pine (*Pinus* sp.) contact zone. Plant Systematics and Evolution 301: 1643–1652. doi: 10.1007/ s00606-014-1182-x.

Wachowiak W, Żukowska WB, Wójkiewicz B, Cavers S & Litkowiec M (2016) Hybridization in contact zone between temperate European pine species. Tree Genetics & Genomes 12: 48. doi: 10.1007/s11295-016-1007-x.

Willyard A, Syring J, Gernandt DS, Liston A & Cronn R (2007) Fossil calibration of molecular divergence infers a moderate mutation rate and recent radiations for *Pinus*. Molecular Biology and Evolution 24: 90–101. doi:10.1093/molbev/msl131.

Wójkiewicz B & Wachowiak W (2016) Substructuring of Scots pine in Europe based on polymorphism at chloroplast microsatellite loci. Flora – Morphology, Distribution, Functional Ecology of Plants 220: 142–149. doi:10.1016/j. flora.2016.03.005.

Zhou Y, Bui T, Auckland LD & Williams CG (2002) Undermethylated DNA as a source of microsatellites from a conifer genome. Genome 45: 91–99. doi:10.1139/g01-119.

http://rcin.org.pl

Publication 2

Żukowska WB, Wachowiak W

Nuclear microsatellite markers reveal the low genetic structure of *Pinus mugo* Turra (dwarf mountain pine) populations in Europe

Plant Systematics and Evolution

doi: 10.1007/s00606-017-1395-x

Impact factor: 1.361

MNiSW points: 20

http://rcin.org.pl

Plant Syst Evol
DOI 10.1007/s00606-017-1395-x

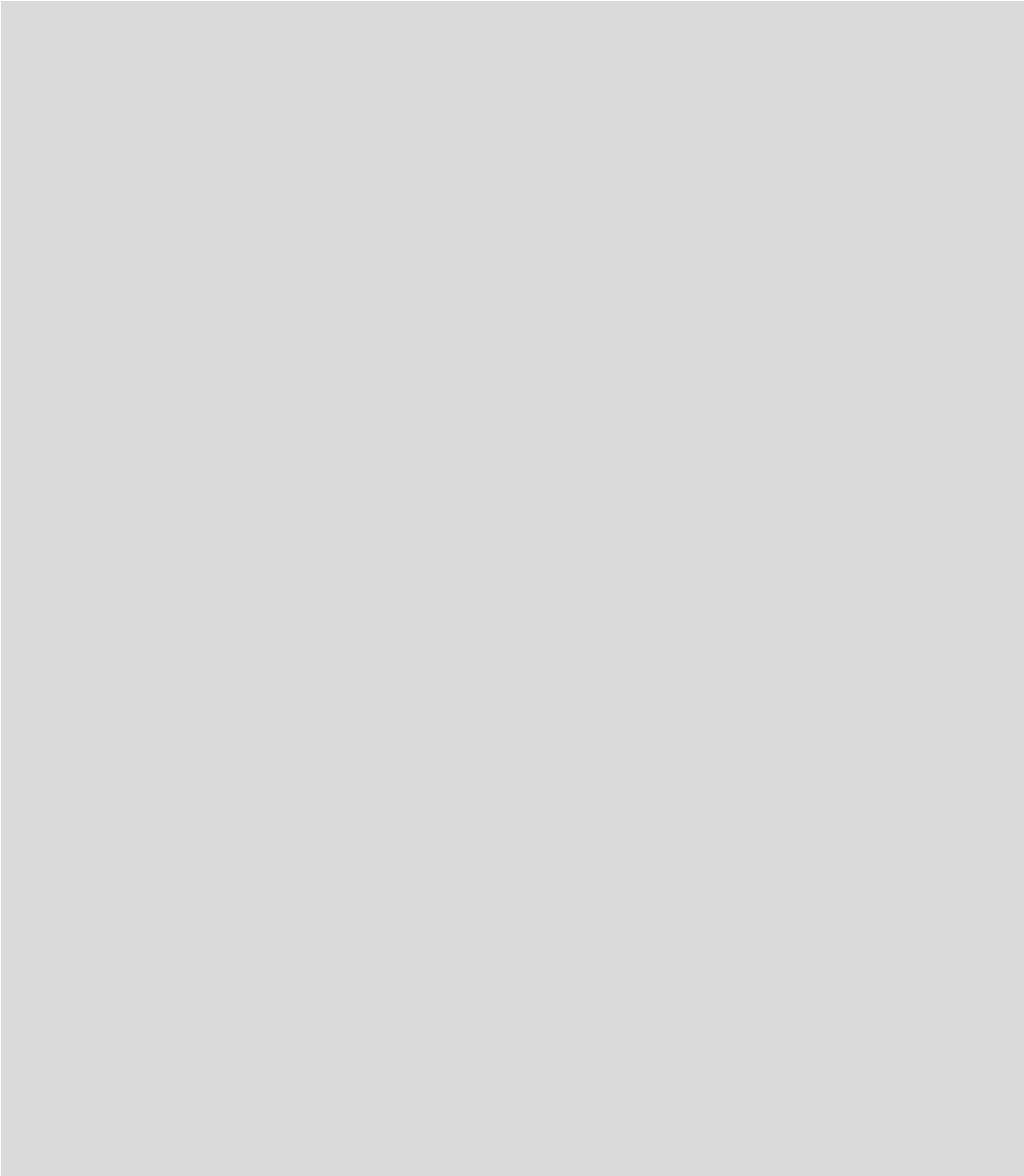
CrossMark

ORIGINAL ARTICLE

Nuclear microsatellite markers reveal the low genetic structure of *Pinus mugo* Turra (dwarf mountain pine) populations in Europe

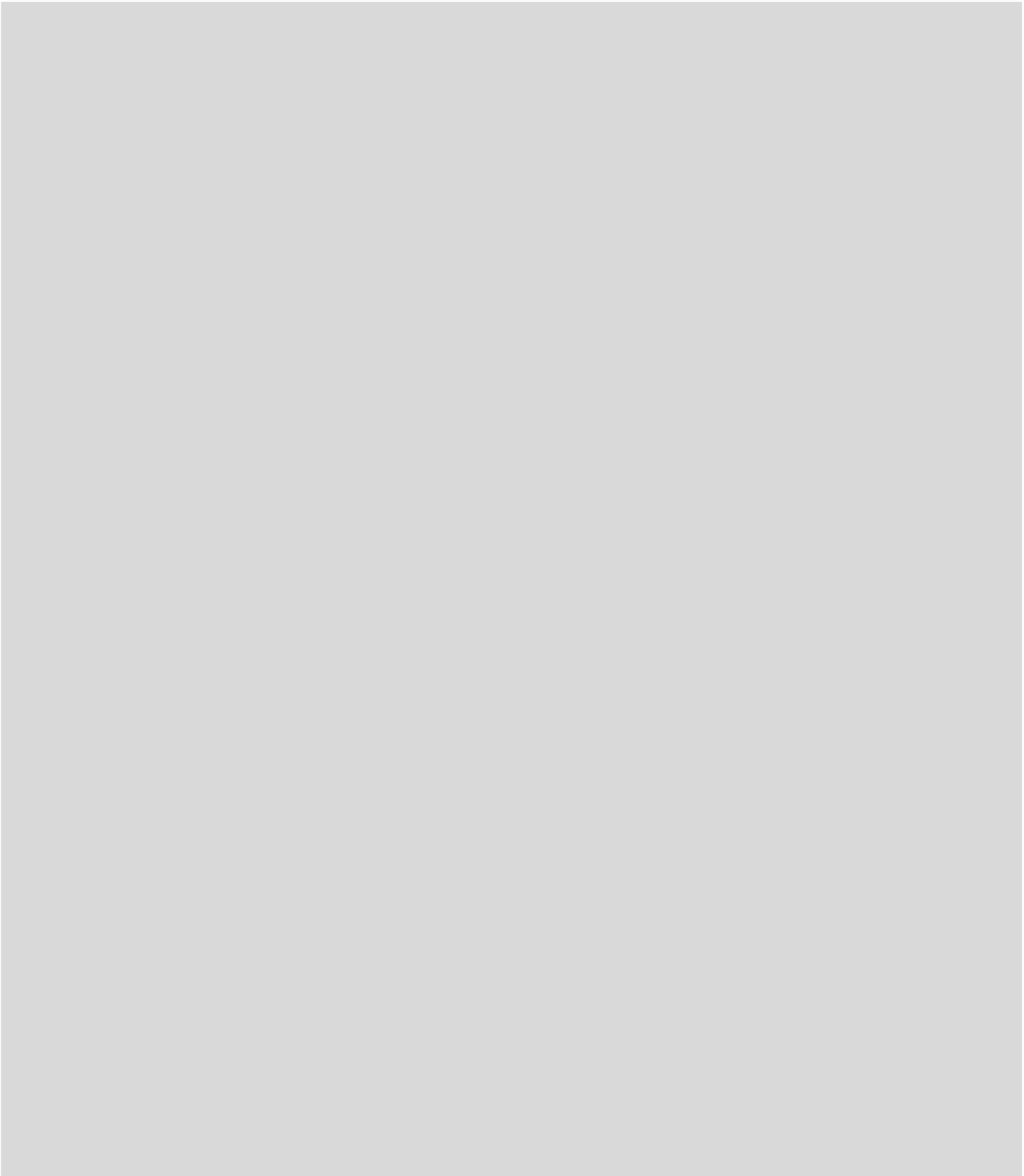
Weronika B. Żukowska[1] · Witold Wachowiak[1,2]

http://rcin.org.pl



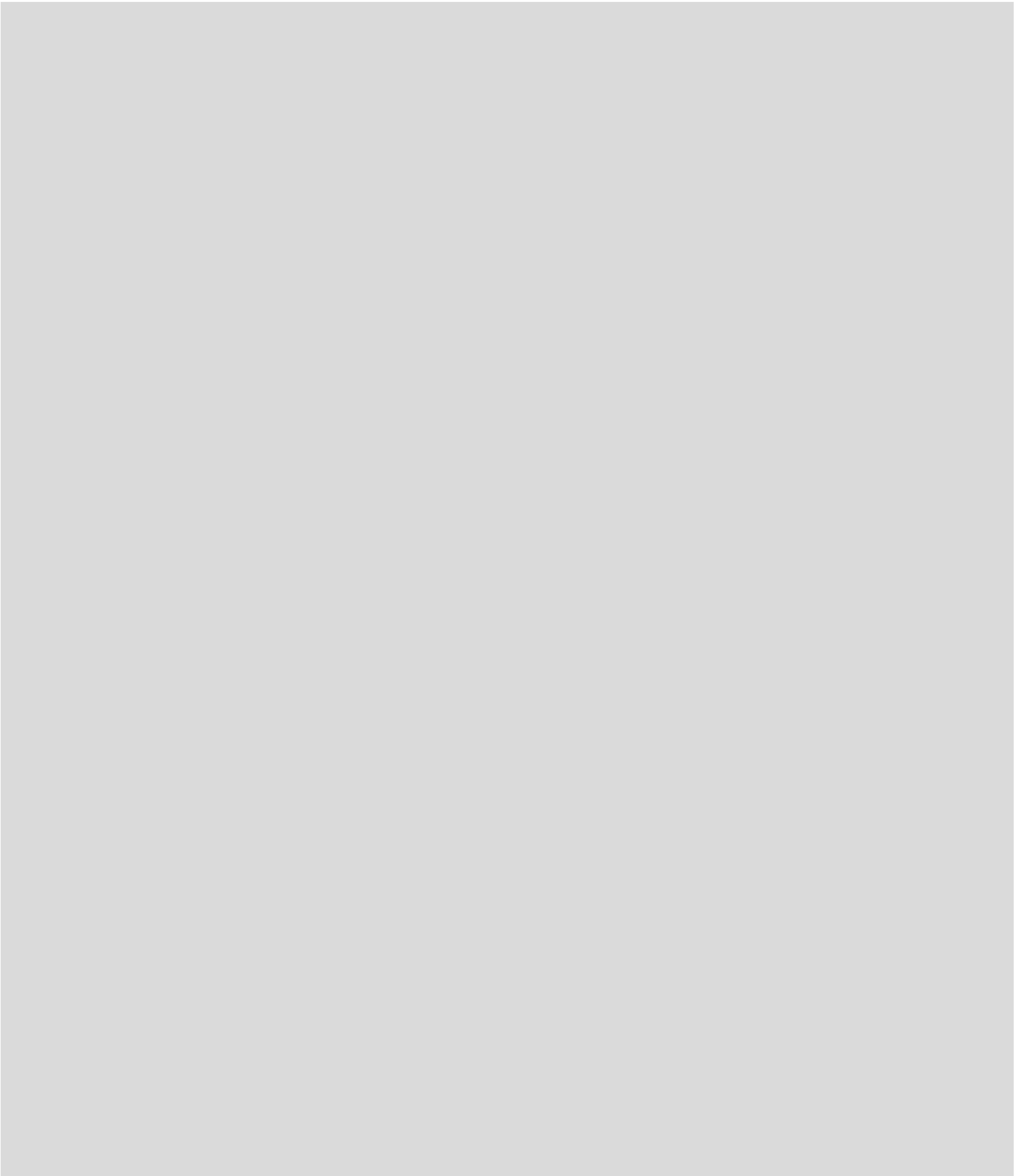
http://rcin.org.pl

http://rcin.org.pl

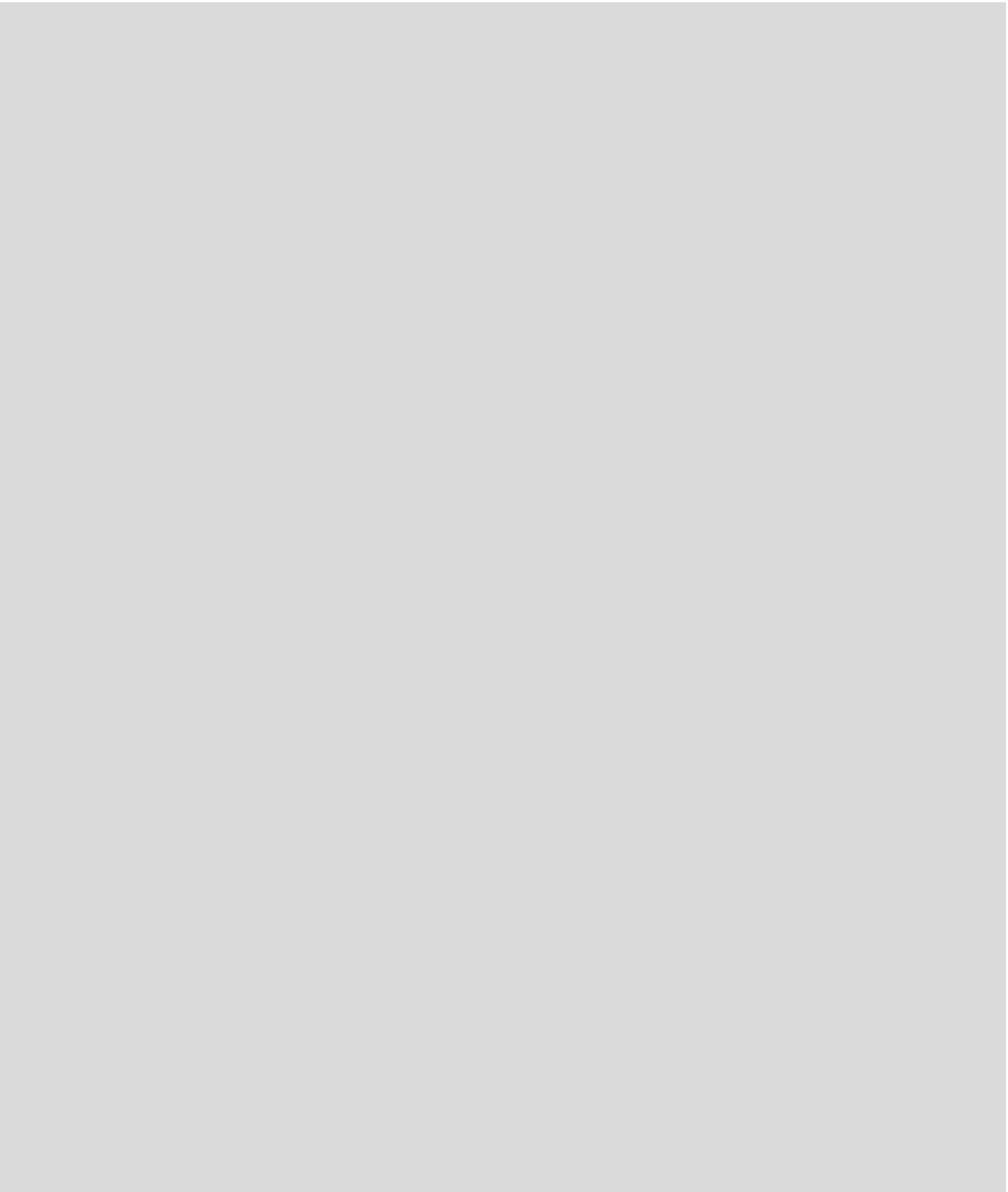


http://rcin.org.pl

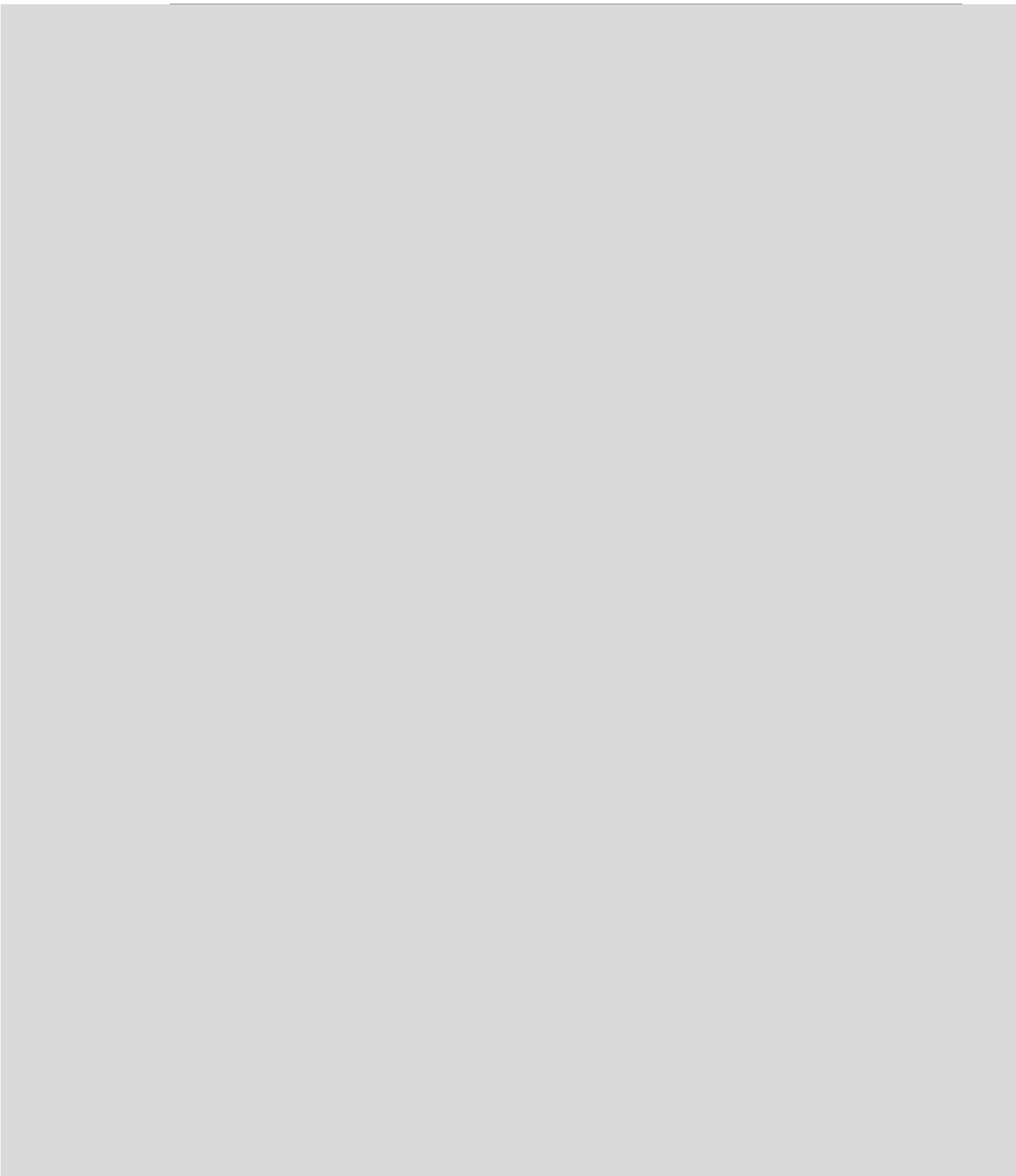
http://rcin.org.pl



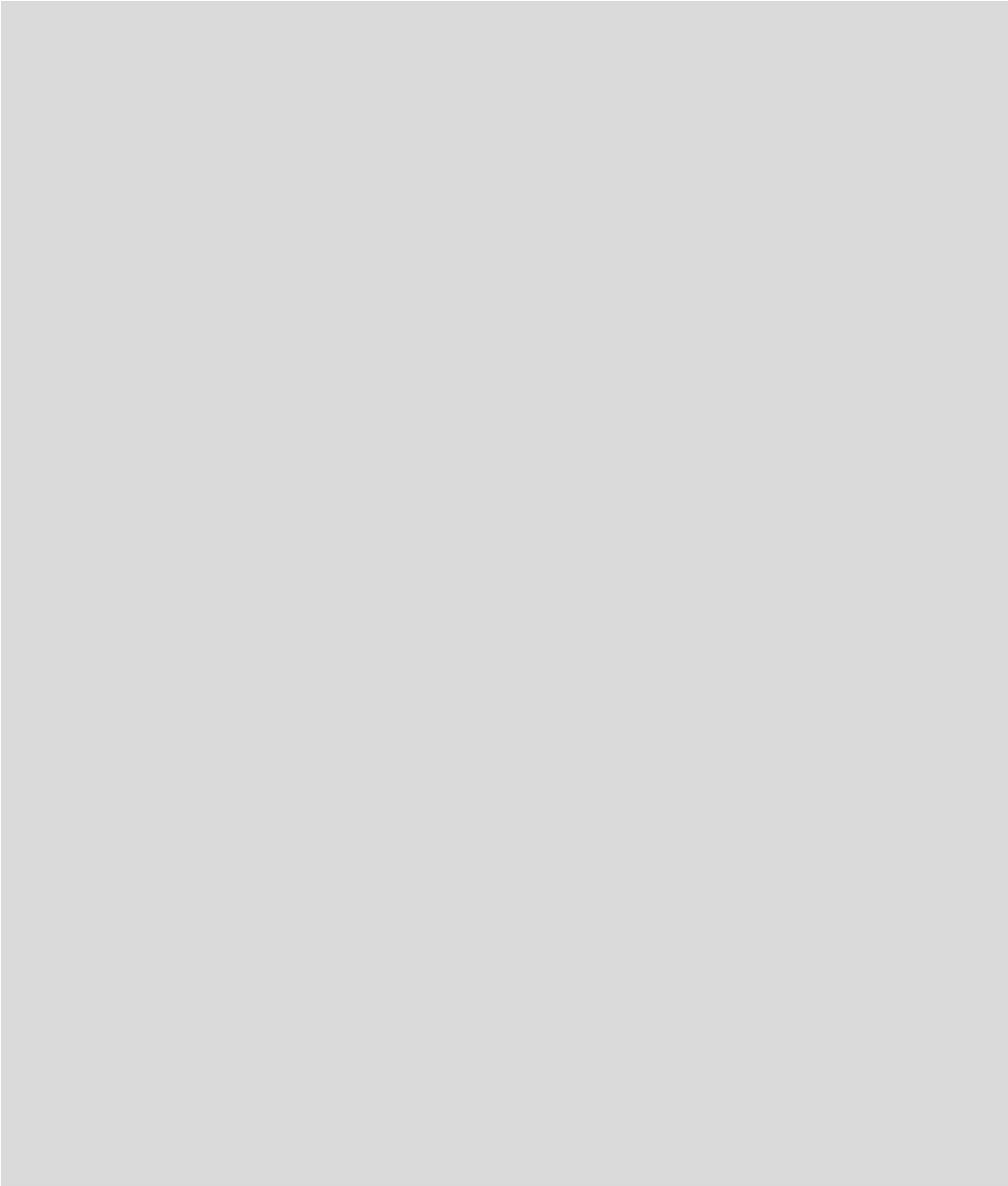
http://rcin.org.pl



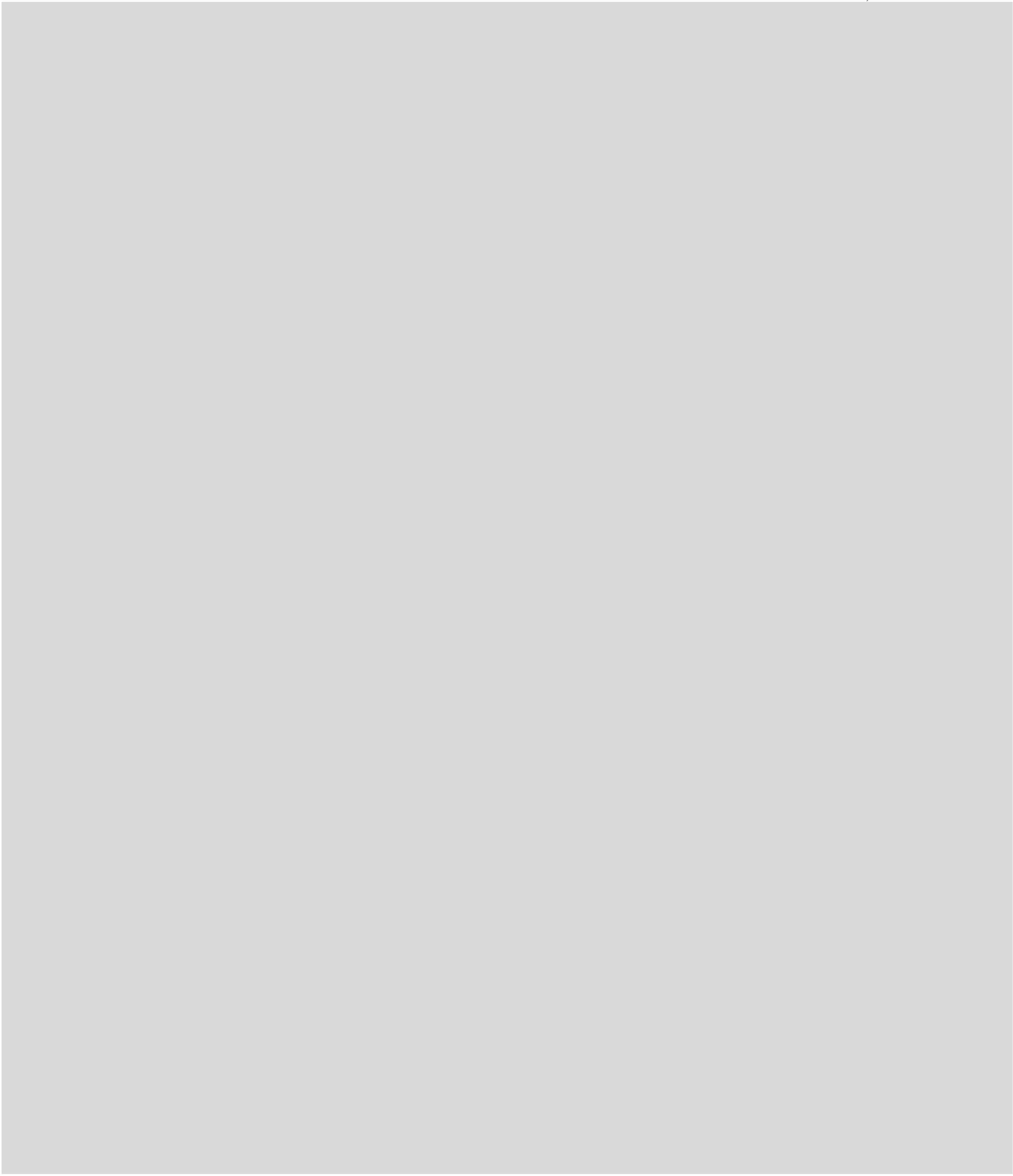
http://rcin.org.pl



http://rcin.org.pl

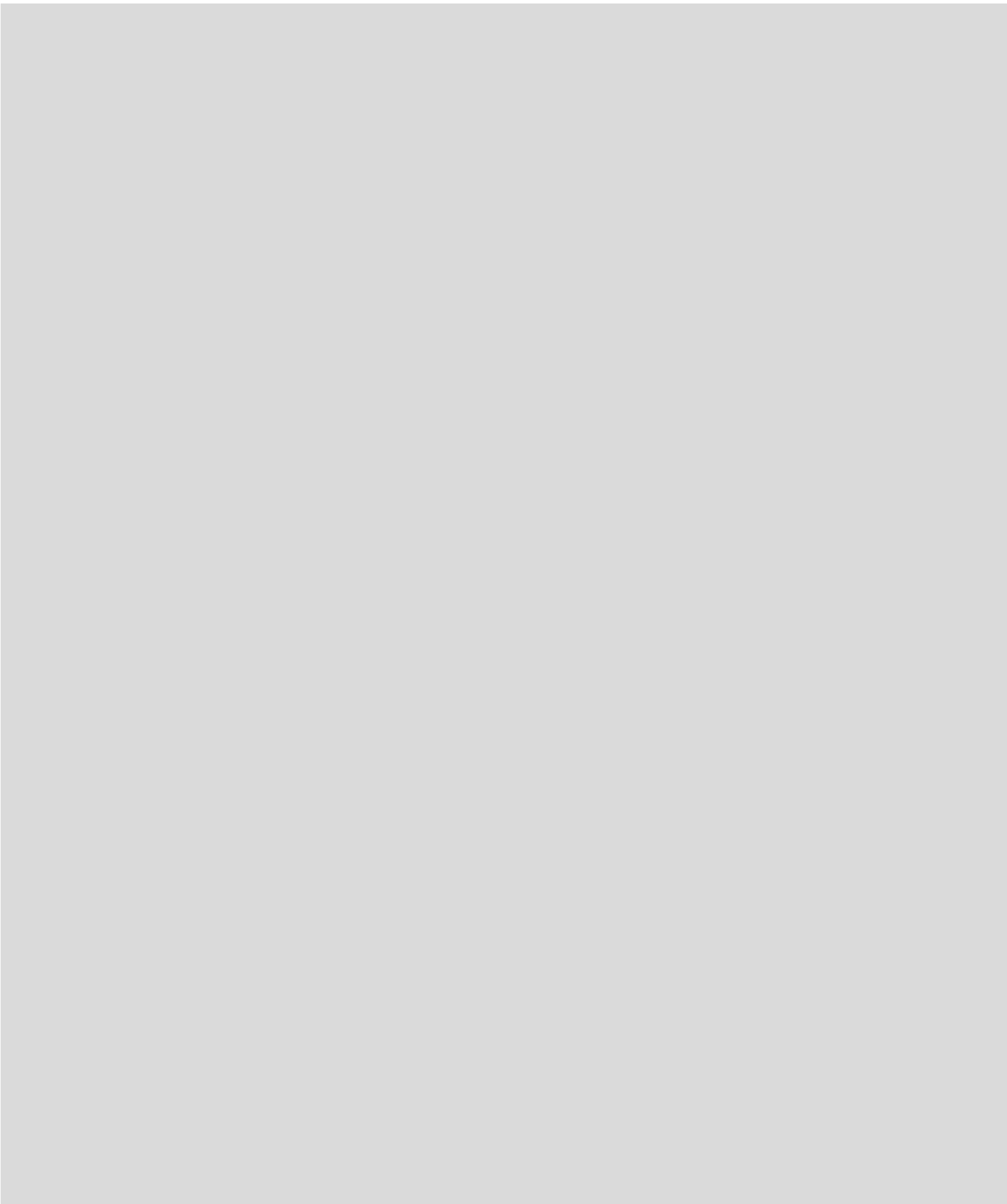


http://rcin.org.pl

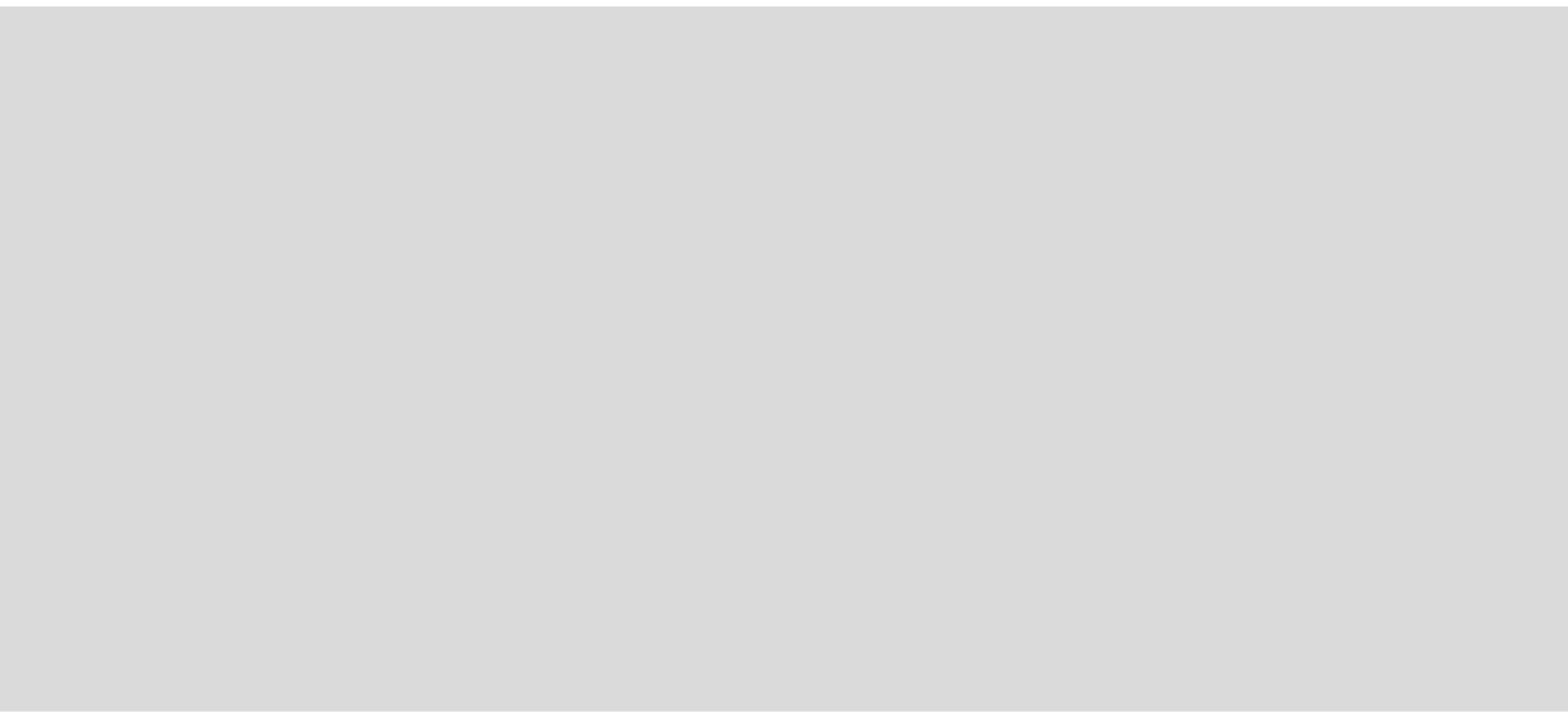


http://rcin.org.pl

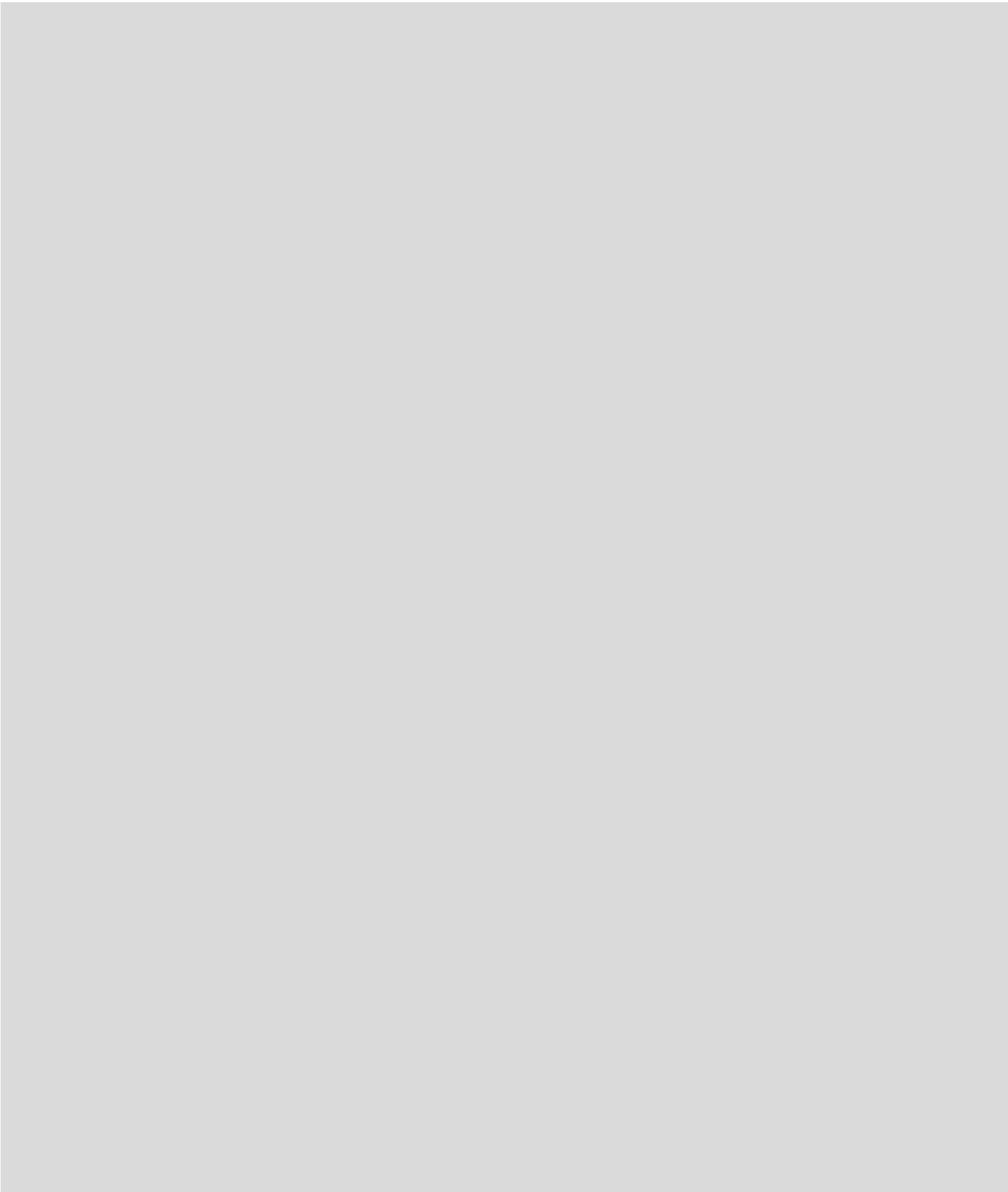
http://rcin.org.pl



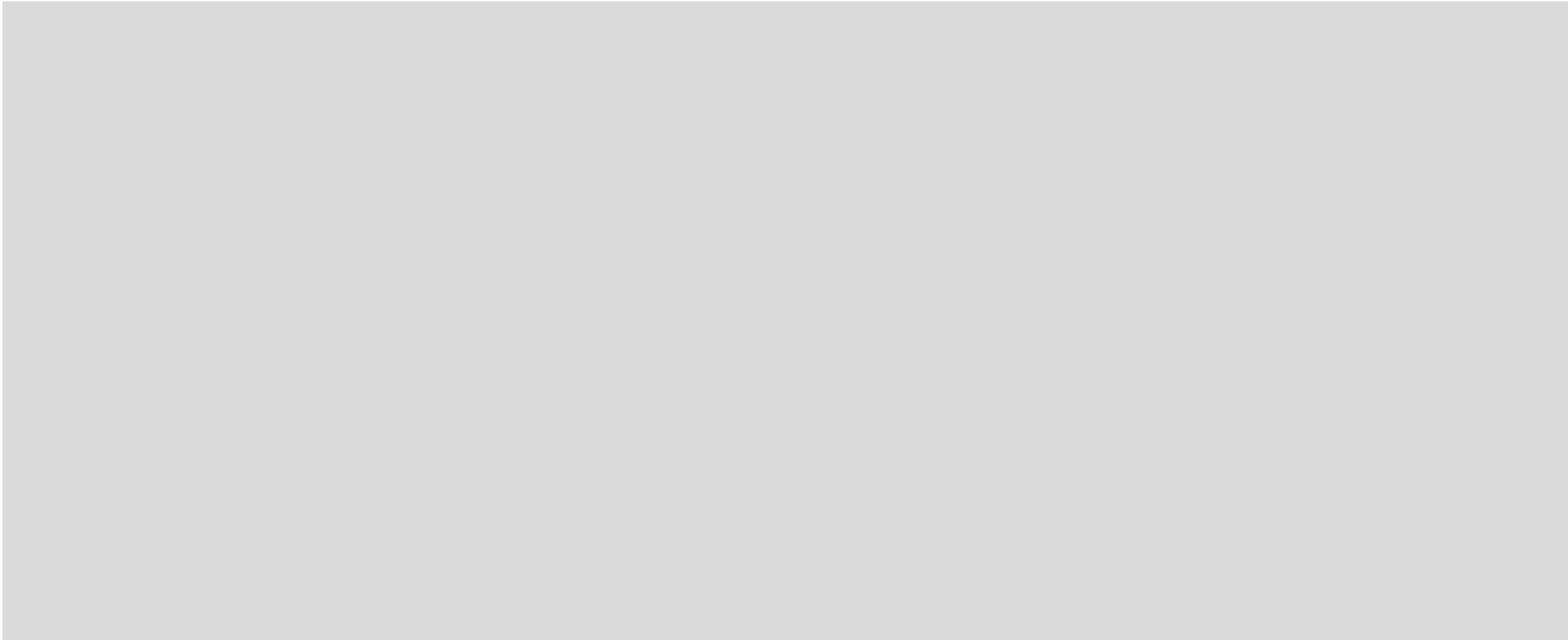
http://rcin.org.pl



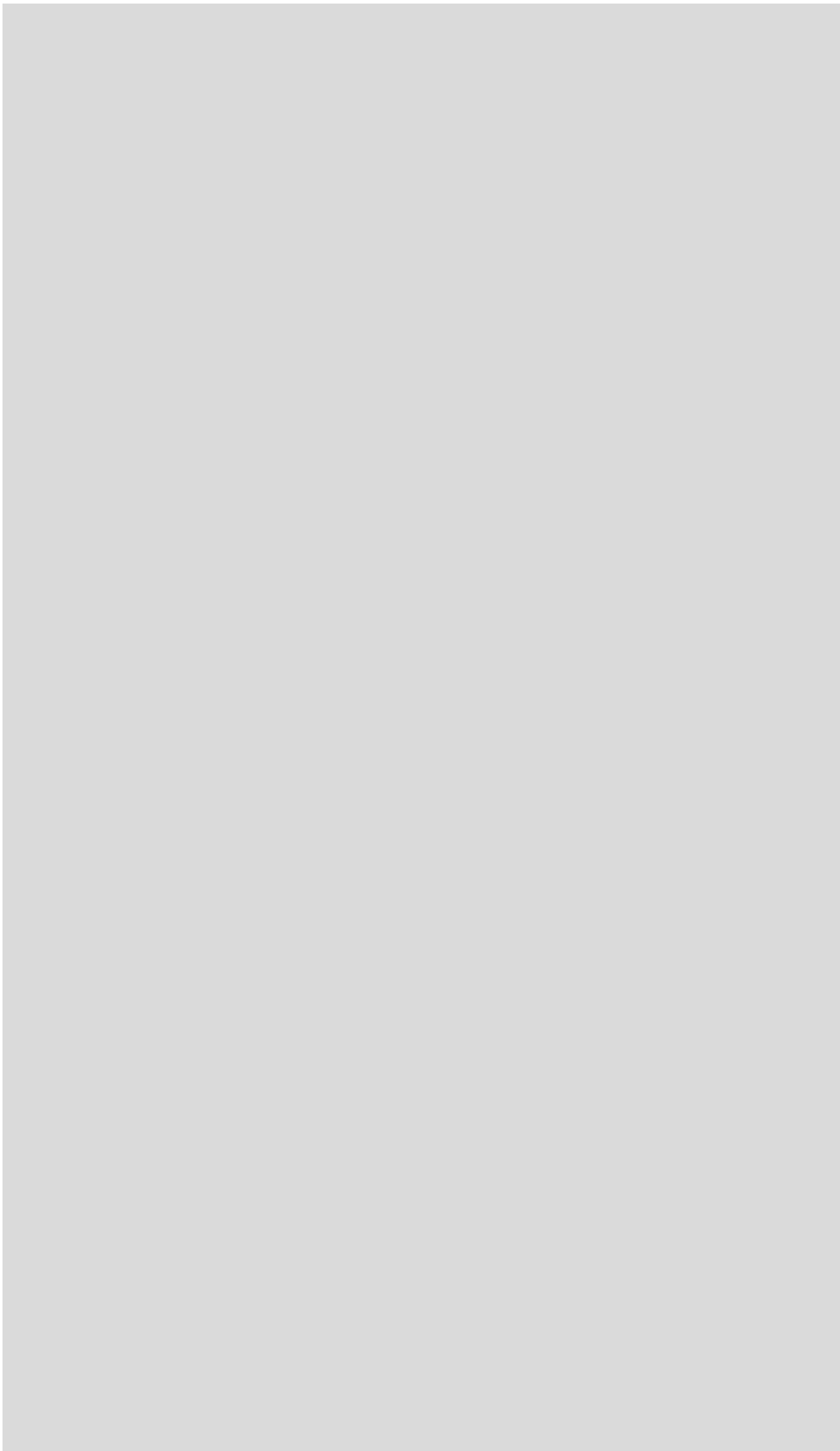
http://rcin.org.pl



http://rcin.org.pl



http://rcin.org.pl



http://rcin.org.pl

Publication 3

Żukowska WB, Boratyńska K, Wachowiak W

Comparison of range-wide chloroplast microsatellite and needle trait variation patterns in *Pinus mugo* Turra (dwarf mountain pine)

iForest – Biogeosciences and Forestry

doi: 10.3832/ifor1860-009

Impact factor: 1.070

MNiSW points: 25

http://rcin.org.pl

Research Article
doi: 10.3832/ifor1860-009
(Early View)

Comparison of range-wide chloroplast microsatellite and needle trait variation patterns in *Pinus mugo* Turra (dwarf mountain pine)

Weronika Barbara Zukowska [(1)],
Krystyna Boratynska [(1)],
Witold Wachowiak [(1-2)]

http://rcin.org.pl

http://rcin.org.pl

http://rcin.org.pl

http://rcin.org.pl

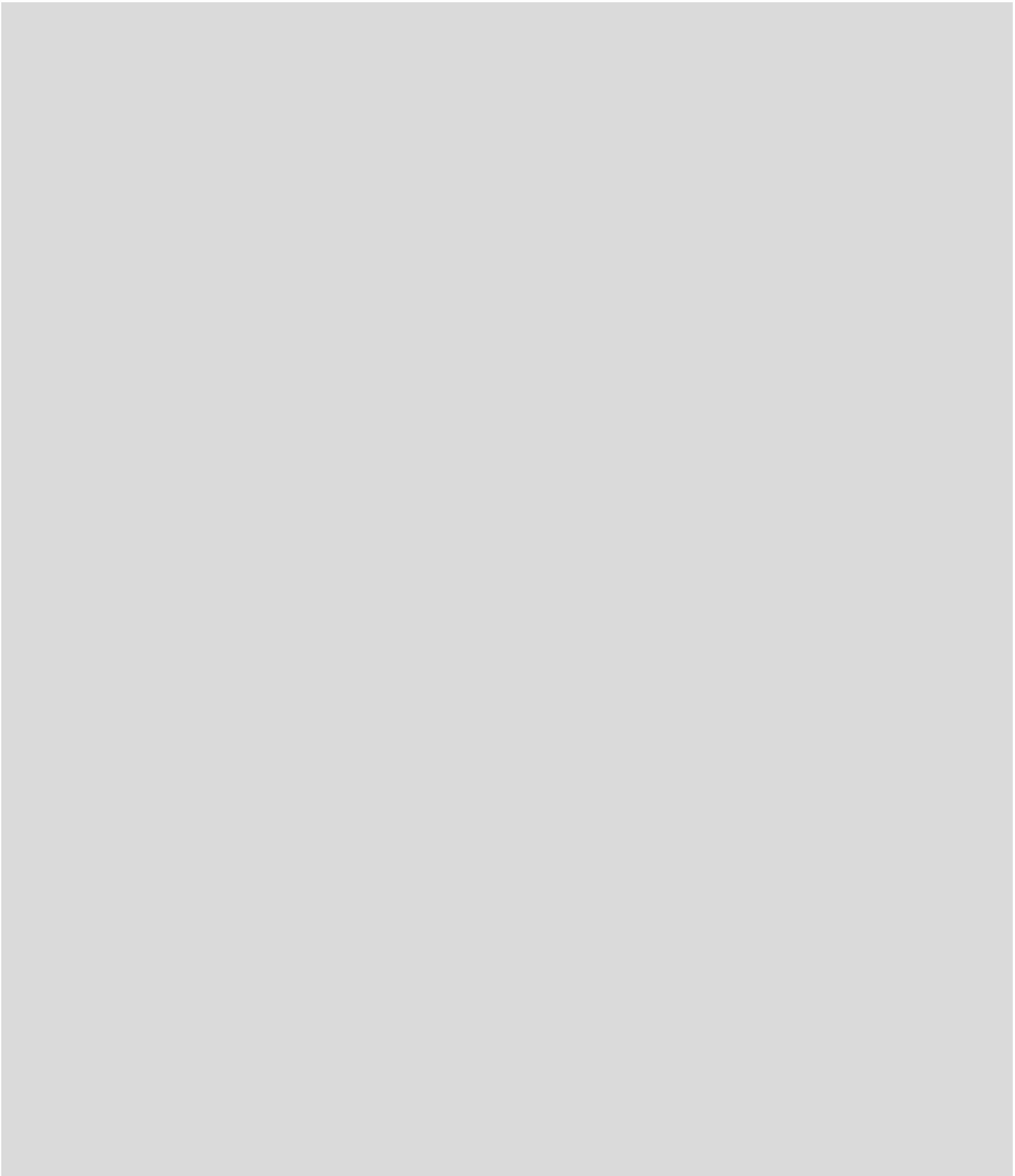
http://rcin.org.pl

http://rcin.org.pl

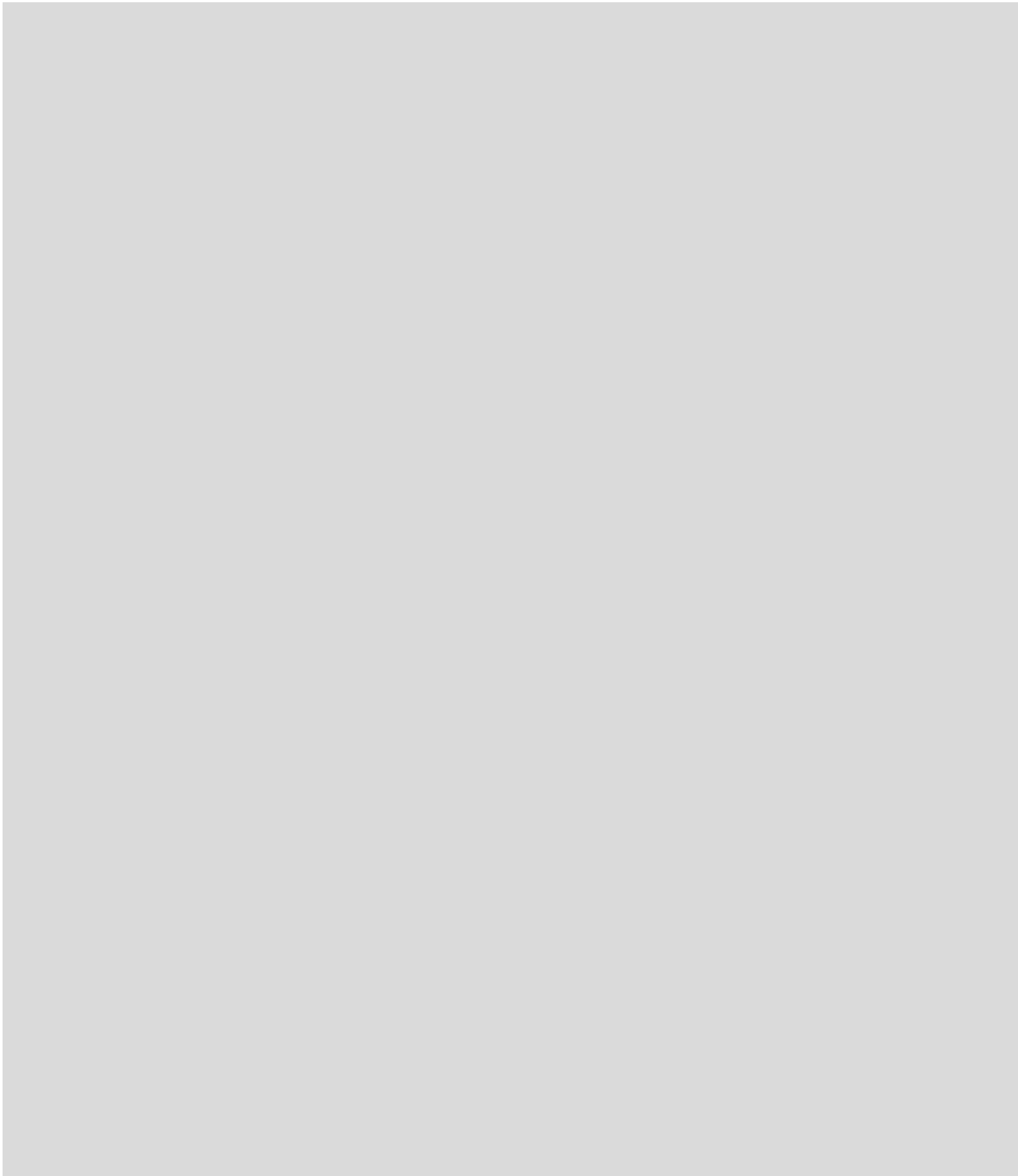
http://rcin.org.pl

http://rcin.org.pl

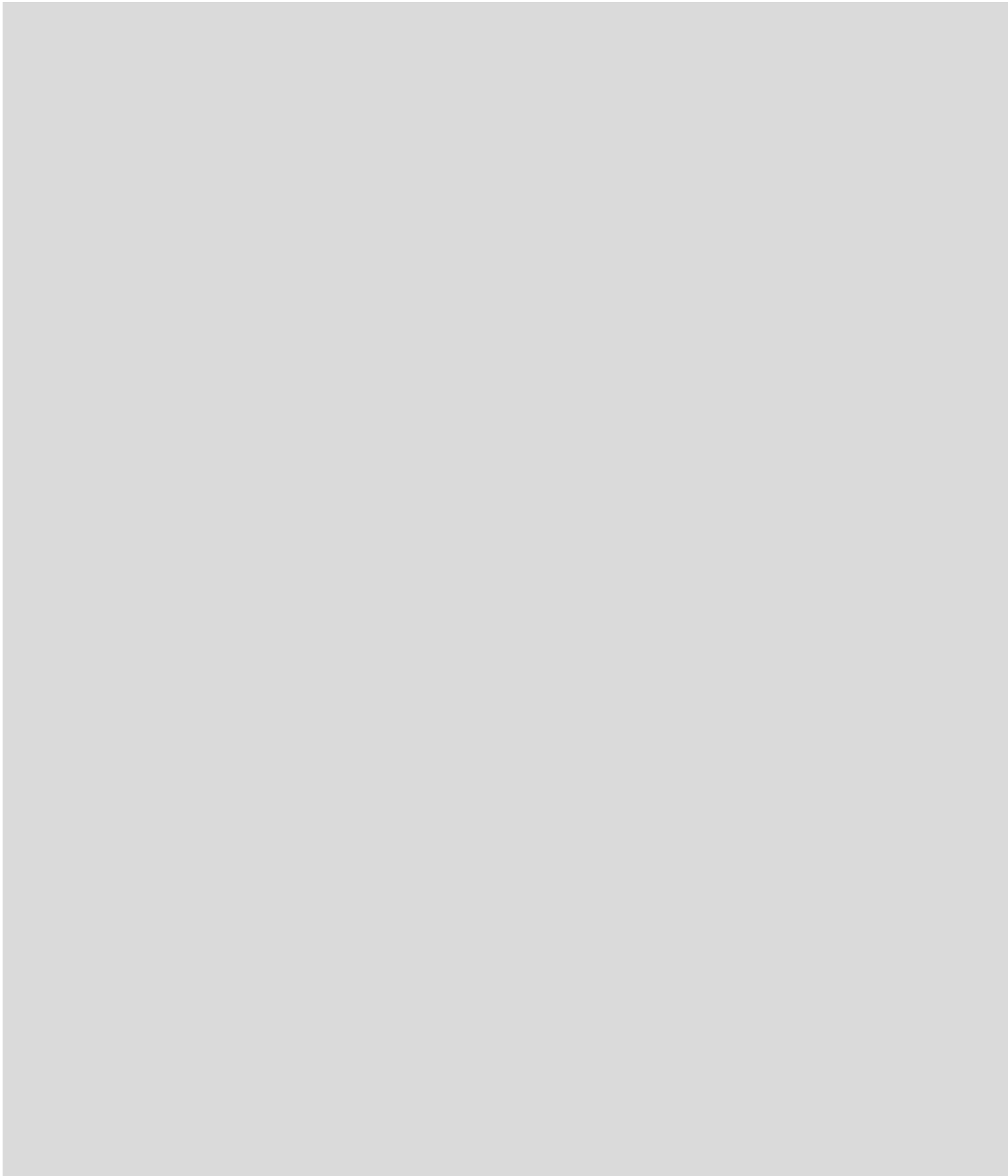
http://rcin.org.pl



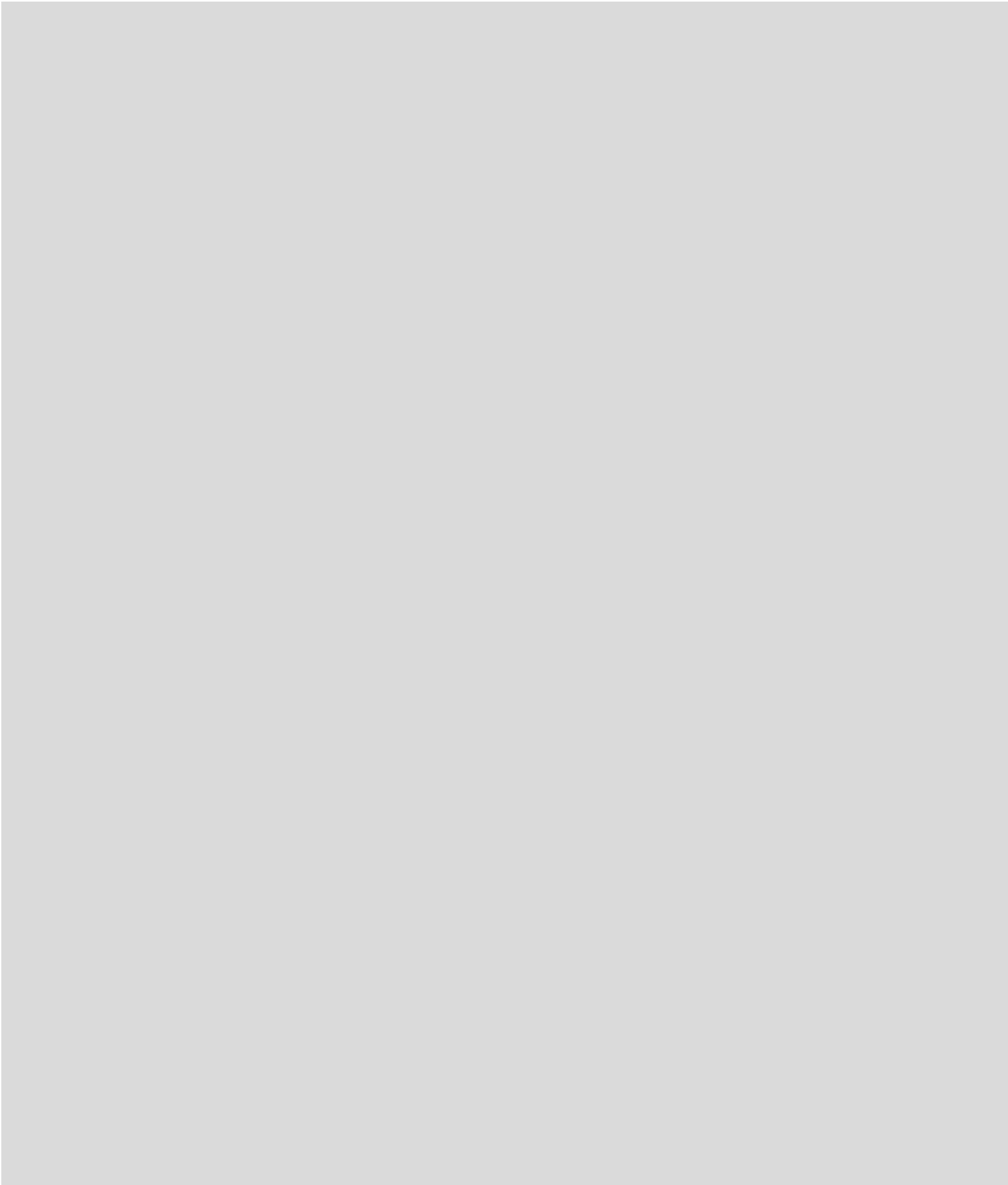
http://rcin.org.pl



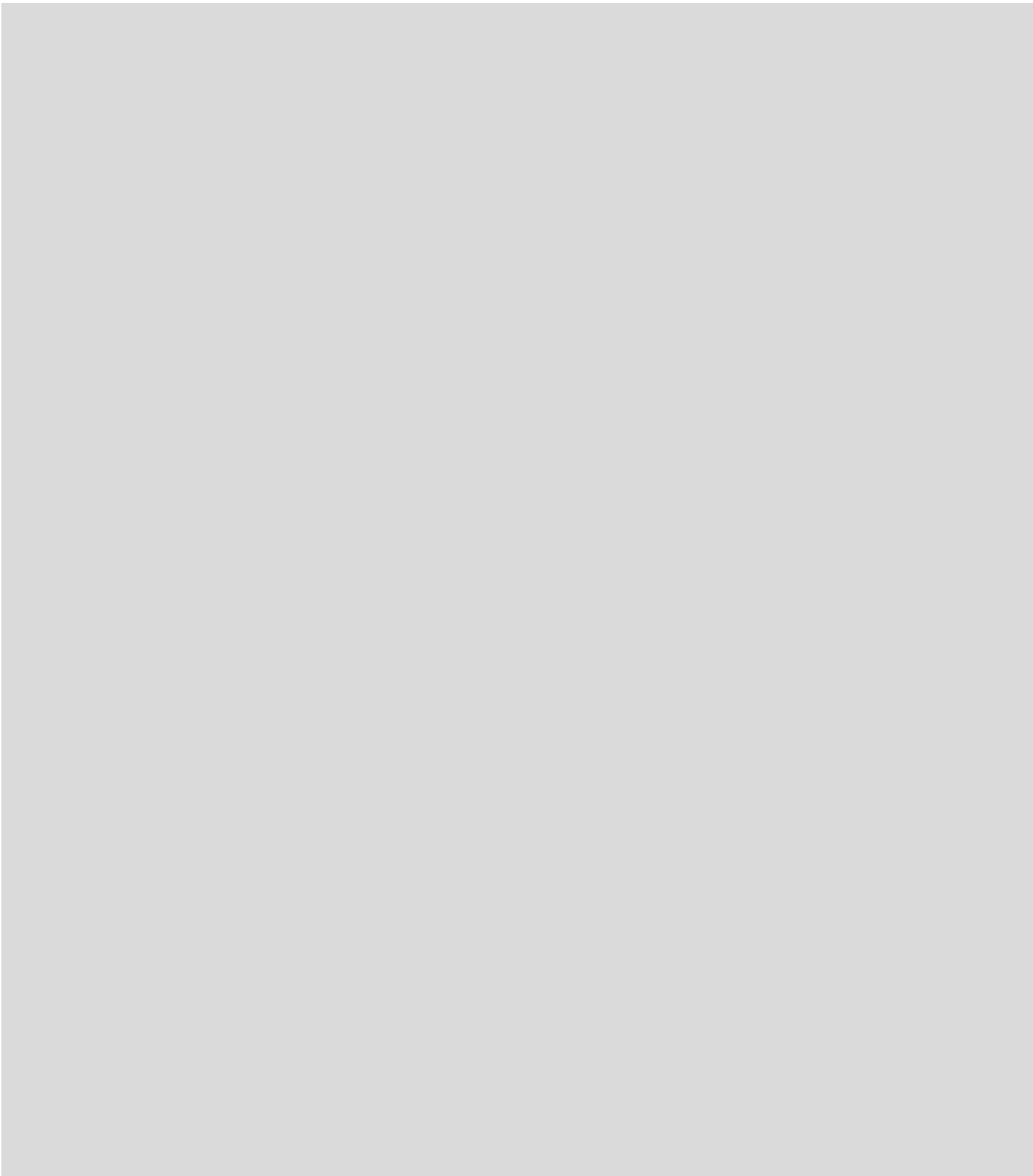
http://rcin.org.pl



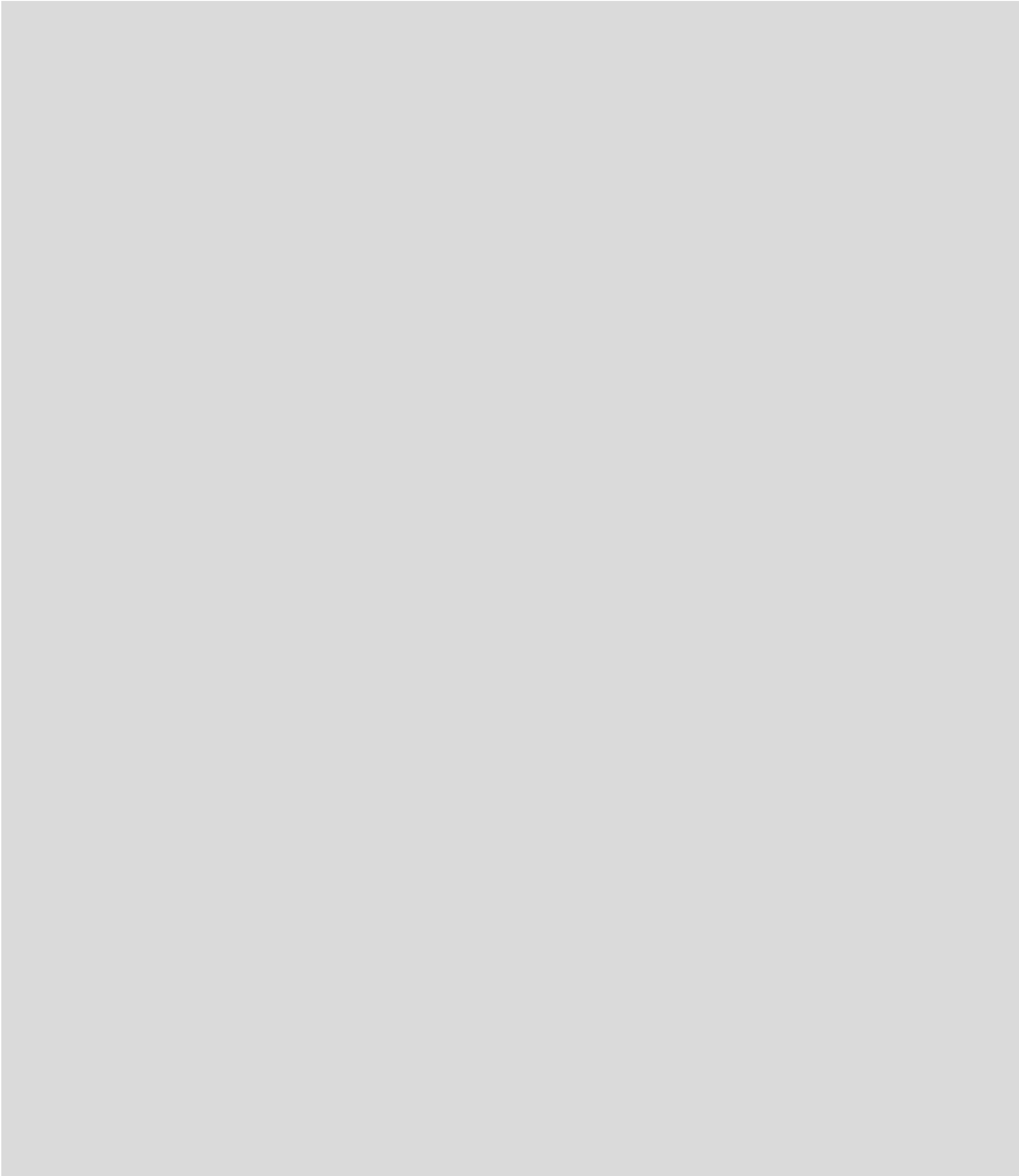
http://rcin.org.pl



http://rcin.org.pl



http://rcin.org.pl



http://rcin.org.pl

Publication 4

Żukowska WB, Wachowiak W

Utility of closely related taxa for genetic studies of adaptive variation and speciation: current state and perspectives in plants with focus on forest tree species

Journal of Systematics and Evolution 54: 17-28

Impact factor: 1.134

MNiSW points: 25

http://rcin.org.pl

JSE Journal of Systematics and Evolution

doi: 10.1111/jse.12177

Review

Utility of closely related taxa for genetic studies of adaptive variation and speciation: Current state and perspectives in plants with focus on forest tree species

Weronika B. Żukowska[1]* and Witold Wachowiak[1,2]

[1]Institute of Dendrology, Polish Academy of Sciences, Parkowa 5, 62-035 Kornik, Poland
[2]Institute of Environmental Biology, Faculty of Biology, Adam Mickiewicz University, Umultowska 89, 61-614 Poznan, Poland
*Author for correspondence. E-mail: wzukowska@man.poznan.pl. Tel.: 48-618170033. Fax: 48-618170166.
Received 12 April 2015; Accepted 22 August 2015; Article first published online 29 September 2015

Abstract Studies of adaptation and speciation have greatly benefited from rapid progress of DNA sequencing and genotyping technologies. Comparative genomics of closely related taxa has great potential to advance evolutionary research on genetic architecture of adaptive traits and reproductive isolation. Such studies that utilized closely related plant species and ecotypes have already provided some important insights into genomic regions and/or genes that are potentially involved in local adaptation and speciation. The choice of an appropriate species model for such research is crucial. The paper discusses current approaches used to reveal the patterns of intra- and interspecific divergence due to natural selection. Its outcomes in herbaceous plants and forest trees are briefly summarized and compared to reveal general regularities concerning evolutionary processes. We then highlight the importance of multispecies studies and discuss the utility of several related pine taxa as fine candidates for evolutionary inferences. Genetically similar but ecologically and phenotypically diverged taxa seem a promising study system to search for genomic patterns of speciation and adaptive variation.

Key words: adaptation, divergence, genetic variation, natural selection, *Pinus mugo*, *P. sylvestris*.

Local adaptation to different selective regimes (i.e., water availability, soil type, photoperiod and temperature) may result in various ecotypes, exhibiting phenotypic differences under specific environmental conditions. The emergence of ecotypes might be an outcome of a number of factors including, e.g., limited migration, density dependent viability and environmental boundaries, and selection on phenotypic plasticity (de Jong, 2005). Adaptation may even lead to ecological speciation, causing reproductive isolation between populations previously connected by gene flow (Rundle & Nosil, 2005). Currently, many closely related plant species are distinguished on the basis of their morphology, but genetic divergence responsible for the observed diversity is usually poorly understood. The relationship between a genotype and phenotype in trees is known even less than in other plant species. The main reason behind this situation is that the forest tree research community is indeed incomparably smaller than the group that studies other plant species such as *Arabidopsis thaliana* or crops. Also, the research is predominantly focused on a few temperate species of high economic value, whereas tree species of ecological importance are commonly ignored (Neale & Kremer, 2011). Although studies of local adaptation in trees have a long tradition of common-garden experiments (provenance trials), such research gives primarily phenotypic information but cannot identify particular genes involved in adaptation (Gonzalez-Martínez et al., 2006). Furthermore, trees have large population size and long generation time which make them challenging to study. Most tree genomes are large and therefore costly to sequence. The genomics of adaptation in trees is thus still mostly unknown.

Studies utilizing closely related taxa to decipher genomic signatures of adaptive divergence and speciation are now emerging. Recent work in this area is described later in this paper. So far, it appears that adaptive traits are generally polygenic (e.g., Sork et al., 2013), though there are many examples of traits controlled by a single gene, e.g., plant resistance to biotic stresses depends mainly on monogenic traits (Vinocur & Altman, 2005). Additionally, Strasburg et al. (2012) compared studies utilizing genome scans in both closely and more distantly related hybridizing plant populations and concluded that there was only little evidence for large genomic islands of divergence as predicted by divergence hitchhiking (Via, 2012). In point of fact, highly divergent regions were rather small and spread throughout the genome, i.e., they did not form large clusters (Strasburg et al., 2012). At present, an increasing number of markers and loci in various species are being thoroughly examined thanks to rapid development of next generation sequencing (NGS) and genotyping technologies. As it seems that it is only a question of time before numerous tree genomes are sequenced, challenges lie rather in the lack of appropriate statistical and bioinformatics methods as well as study systems that will accurately reflect various evolutionary phenomena (Manel et al., 2010).

In this paper we first review current approaches and outcomes of studies addressing fundamental questions about

© 2015 Institute of Botany, Chinese Academy of Sciences

http://rcin.org.pl

the genetic basis of adaptive divergence and speciation in closely related plant taxa: How does adaptation evolve in different species and populations? What is the relative importance of ecological divergence in allopatric vs. sympatric speciation? Do adaptive traits involve many loci with small effects or fewer loci with larger effects? What is the role of introgression through hybridization in development of new adaptive variation? Are species boundaries genic or chromosomal? We then discuss the benefits of using multispecies approach in evolutionary studies as compared to single-species approach. Finally, we focus on forest tree species as promising study systems in genetic assessments of phenotypic and adaptive variation. Specifically, we point out a pine complex comprising several related taxa as particularly noteworthy for studies of various evolutionary mechanisms.

Genomic patterns of divergence: Current approaches

Patterns of polymorphism found at the genomic level in outcrossing populations result both from demographic processes related to population history and evolutionary factors. Such processes as mutation, drift, recombination, migration, gene flow and selection interact to shape genetic variation currently observed in natural populations (Nosil & Feder, 2013). Across the environmental range occupied by plant species local adaptation occurs as a balance between the divergent pressure of natural selection in different populations and gene flow among them (Savolainen et al., 2007). In naturally outbreeding populations, selection is expected to have highly localized effects on the genome, whilst demographic events related to population history should affect all regions in a similar way (Storz, 2005). Quantitative genetic studies have provided clear data on phenotypic differentiation between natural populations of plant species. Traits related to adaptation are usually highly heritable, as shown in several quantitative genetic studies (e.g., Howe et al., 2003). However, the genetic basis underlying phenotypic and adaptive variation is mostly unknown. So far, efforts to detect loci under selection in plants have mostly focused on single species. However, assuming that selection acting on variation within a species may eventually lead to speciation, comparative studies of intra- and interspecific genetic variation at closely related but highly differentiated taxa can improve the power to detect genes involved in adaptation (Nosil & Feder, 2013). Predictions of environmental change provide a strong impetus to improve our understanding of plant adaptation as currently optimal phenotypes may suffer a fitness deficit in growth and survival. Therefore, it is important to quantify the relative contributions of different forces acting to maintain adaptive genetic diversity and structure in natural populations.

There are several approaches used to detect genomic regions that are under selection and involved in adaptive divergence (Table 1). They include different though complementary fields of study such as population genomics and quantitative genetics (Fig. 1). Overall, the ‘bottom-up’ and ‘top-down’ approach can be distinguished (Barrett & Hoekstra, 2011). In the first case, genetic data is used without a priori assumptions about a given phenotype. A set of individuals is screened for genetic markers to find loci under selection. Such genome scans make use of the principle of genetic hitchhiking (Smith & Haigh, 1974). In this case, signs of speciation and adaptation may be detected at the genomic level in loci which patterns of nucleotide variation significantly differ from the standard neutral model of evolution (Nielsen, 2005). Individual non-neutral loci are then statistically evaluated and qualified as outliers affected by local adaptation if they show reduced genetic variability at the within-population level and increased differentiation between phenotypically diverged populations (Stinchcombe & Hoekstra, 2008). Similarly, a rapidly-developing field of landscape genetics uses such large genome scans to reveal spatial patterns of genetic diversity in heterogeneous and fragmented habitats and evolutionary processes that influence patterns of variation. A branch of landscape genetics, landscape genomics, aims to identify environmental factors that shape adaptive genetic variation (Manel & Holderegger, 2013).

By contrast, the top-down approach starts from phenotypic observations leading to identification of underlying genomic regions using either quantitative trait loci (QTL), admixture or association mapping. QTL mapping provides information on the genetic basis of phenotypic traits using a dense linkage map created with many genetic markers (Nielsen et al., 2009). It is primarily applicable to complex traits, but it is also useful to explore adaptive variation in forest trees (Gonzalez-Martínez et al., 2006). QTL mapping appears to be rather demanding since it requires a pedigreed mapping population with a large number of progeny (e.g., Sork et al., 2013). In contrast, association mapping approach does not require segregating families and can thus constitute a valuable alternative to QTL studies. It also makes use of linkage disequilibrium (LD) between a large set of genetic markers to find relationships among genetic variants and phenotypes (genome-wide association study, GWAS), but it may equally work with fewer candidate genes (Nielsen et al., 2009). In GWASs large single nucleotide polymorphism (SNP) datasets are usually used, because SNPs are abundant and located in coding and regulatory regions that have a direct impact on a phenotype (Sork et al., 2013). In turn, admixture mapping relies on far fewer markers in natural admixed populations, e.g., in hybrid zones, using a high degree of LD (Nielsen et al., 2009).

Newly emerging methods are now going beyond the gene level and focus on large-scale microarray-based studies of gene expression and protein composition. Functional genomics may prove useful as adaptation is likely to occur through changes in levels and patterns of gene expression or divergence of proteins, rather than via DNA variation in non-coding regions (Fay & Wittkopp, 2008). Loci under selection may be detected using outlier analyses similar to those applied to genome scans (Rice et al., 2011).

Genomic patterns of divergence: Case studies in plants

Ecotypic and interspecific comparisons of related taxa are of great value to investigate various evolutionary processes. Comprehensive genome-wide research on plants has given

http://rcin.org.pl

Table 1 A list of some evolutionary questions that can be addressed by genomic approaches

Approach	Corresponding questions	Limitations
Genome scans	➢ How does adaptation and speciation affect patterns of genetic variation? ➢ Can we find any correlation between the pattern of genome-wide polymorphism and environmental variables? ➢ Is adaptation primarily polygenic or does it involve fewer loci? ➢ Does adaptation evolve predominantly from new mutations or standing genetic variation? ➢ Do some loci show clinal variation? ➢ Does adaptation to similar habitats yield a similar genetic pattern? ➢ What is the amount and direction of gene flow between populations/species? ➢ In which situations the breakdown of species boundaries results in adaptive introgression? ➢ How do the genomic islands of divergence change in the course of speciation?	× Most markers uncover only neutral genetic variation × Typically, a whole-genome scan with many markers is required × The method's capability to detect functionally relevant loci is limited × The method is prone to false positives × The approach is less effective in identifying adaptation arising from standing genetic variation × The scan itself does not provide any information on an association of the identified outliers with environmental factors
Landscape genetics	➢ How do landscape features affect adaptation and selection? ➢ How do geographical barriers influence gene flow, migration and the spread of adaptive variants? ➢ Are fitness trade-offs at individual loci dependent on landscape features?	× The method has analytical and methodological shortcomings × There is an insufficient number of detailed and comprehensive databases × There are difficulties with replication across various landscapes
QTL mapping	➢ How does local adaptation affect the phenotype? ➢ Which loci are involved in phenotypic divergence and species isolation? ➢ What is the genomic location of the causative variants?	× A dense linkage map, measurable phenotypes and controlled crosses with numerous progeny are required × The approach has limited mapping resolution × A relevant part of phenotypic variation is rarely explained × The method is ineffective for genes with a minor phenotypic effect
Association mapping	➢ How does local adaptation affect the phenotype? ➢ Which loci are involved in phenotypic divergence and species isolation? ➢ What is the genomic location of the causative variants?	× Mapping resolution is affected by the rate of LD decay
Admixture mapping	➢ Which genomic regions are responsible for adaptive divergence and reproductive isolation in hybrid zones? ➢ What is the amount and direction of interspecific gene flow? ➢ In which situations the breakdown of species boundaries results in adaptive introgression?	× Species with strong reproductive boundaries and localities with recent hybrid generations require large sample size
Transcriptomics & Proteomics	➢ What is the effect of genetic variation on gene expression and protein composition? ➢ What is the link between gene expression and a particular phenotype? ➢ How does natural selection affect molecular changes? ➢ Are patterns of gene expression similar for distinct populations occupying comparable habitats? ➢ What is the physiological response to environmental changes and how fast is it? ➢ What is the difference in molecular composition of locally adapted and non-adapted populations/wild and domesticated species/hybrids and parental taxa/closely related and more diverged taxa?	× The choice of tissue, developmental stage, etc., has a huge influence on an obtained set of transcripts/proteins × Low expressed genes are under-represented

insight into genomic architecture of adaptation and speciation phenomena (summarized in Table 2). Many studies clearly point to the correlation between patterns of genome-wide polymorphism and environmental factors, but in many cases it is hard to recognize the exact selective forces responsible for it. The functional and phenotypic relevance of the identified loci also remain largely unknown. For example, Lee & Mitchell-Olds (2012) concluded that genomic patterns of polymorphism in *Arabidopsis thaliana* are associated with environmental adaptation. Genes that responded to selective pressure in heterogeneous landscapes were mostly of unknown functions. A similar correlation was also found, e.g., in the Alpine plant *Arabis alpina* L. from the Swiss and French Alps. The authors used a large genome scan of 825 amplified fragment length polymorphism loci (AFLP) and environmental variables related to precipitation, temperature and topography. 9% of the loci exhibited signals of ecological relevance, and they mostly corresponded with mean annual minimum temperature (Poncet et al., 2010).

More detailed answers have been received from studies comparing plant populations occupying distinct habitats, where the selective force limiting the plant growth is apparent, e.g., in the case of adaptation to harsh soil conditions. Such local adaptation may be a result of antagonistic pleiotropy explaining occurrence of fitness trade-offs, where alleles beneficial in one environment are

http://rcin.org.pl

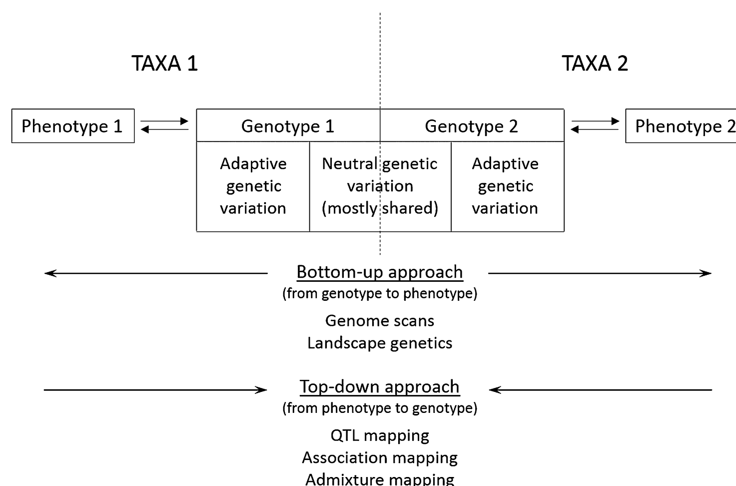


Fig. 1. A schematic representation of a study system that takes advantage of closely related taxa for studies of adaptive variation and speciation. Closely related yet phenotypically and ecologically distinct species may serve as an excellent experimental system for studies of the genetic background of adaptive variation and speciation. Since they share most of the neutral genetic variation, regions under selection can be identified in comparative studies once whole-genome resources are developed. Generated sets of genetic markers can be used with ('top-down' approach) or without ('bottom-up' approach) any prior phenotypic information.

Table 2 Main outcomes of studies reviewed in this contribution

Species studied	Conclusion	References
Arabidopsis lyrata *Arabis alpina* *Helianthus* spp. *Picea glauca*	The proportion of loci identified as outliers is usually small.	Kane & Rieseberg, 2007; Namroud et al., 2008; Poncet et al., 2010; Scascitelli et al., 2010; Turner et al., 2010; Andrew & Rieseberg, 2013
Arabidopsis lyrata *Helianthus* spp. *Picea glauca* *Populus* spp. *Quercus* spp.	Genomic islands of divergence are generally small and spread throughout a genome.	Gömöry et al., 2001; Scotti-Saintagne et al., 2004; Kane & Rieseberg, 2007; Namroud et al., 2008; Lexer et al., 2010; Scascitelli et al., 2010; Turner et al., 2010; Andrew & Rieseberg, 2013
Arabidopsis spp. *Helianthus annuus* *Mimulus guttatus* *Silene vulgaris* *Picea glauca*	Most adaptive traits have polygenic nature.	Bratteler et al., 2006; Kane & Rieseberg, 2007; Namroud et al., 2008; Lowry et al., 2009; Turner et al., 2010; Lee & Mitchell-Olds, 2012
Arabis alpina *Senecio* spp. *Picea abies* *Populus tremula* *Quercus* spp.	Patterns of genome-wide polymorphism often have clinal distribution.	Ducousso et al., 1996; Ingvarsson et al., 2006, 2008; de Carvalho et al., 2010; Neophytou et al., 2010; Poncet et al., 2010; Chen et al., 2012; Muir et al., 2013
Helianthus spp. *Senecio* spp. *Picea* abies *Pinus sylvestris* *Populus* spp.	Adaptation may evolve from new mutations, standing genetic variation or introgression.	Kane & Rieseberg, 2007; Rieseberg et al., 2007; de Carvalho et al., 2010; Lexer et al., 2010; Scascitelli et al., 2010; Chen et al., 2012; Andrew et al., 2013; Muir et al., 2013; Wachowiak et al., 2014b
Helianthus spp. *Silene* spp. *Populus* spp. *Quercus* spp.	Interspecific boundaries seem to be 'porous' with only a modest number of loci contributing to reproductive isolation.	Gömöry et al., 2001; Scotti-Saintagne et al., 2004; Minder & Widmer, 2008; Kane et al., 2009; Lexer et al., 2010; Neophytou et al., 2010; Scascitelli et al., 2010; Andrew & Rieseberg, 2013

http://rcin.org.pl

detrimental in another (Mitchell-Olds et al., 2007). In a close relative of *A. thaliana*, the US subspecies of *A. lyrata*, Turner et al. (2010) found candidate loci for serpentine adaptation, using *A. thaliana* homologs as a reference. Three of the identified loci were additionally sequenced in the Scottish subspecies of *A. lyrata*. As it turned out, there was the valine-to-glycine substitution fixed in the serpentine populations both in Scotland and in the US at one of the loci. In contrast, the polymorphisms in the two other loci differed between the European and the US subspecies. Furthermore, most of the polymorphisms that were associated with the soil type experienced trade-offs. This was consistent with the authors' expectations: trade-offs may be a reason why a lot of plants are endemic to serpentine soils, because mutations responsible for serpentine adaptation are probably deleterious on other sites. In a different experiment employing a sunflower *Helianthus annuus* L. selective sweeps revealed potential genes for adaptation to drought and salt tolerance, although, less likely, these genes may only represent loci which are in close proximity to the actual genes under selection. Their products are probably involved in DNA repair, DNA binding, formation of cell walls or in some biosynthetic and metabolic reactions, suggesting the polygenic nature of drought and salt tolerance. Moreover, 18% of them exhibit homology to known stress-responsive genes. However, the influence of the identified genes on the phenotype of *H. annuus* remains unknown (Kane & Rieseberg, 2007). Local adaptive divergence was also found, e.g., in the common monkey-flower *Mimulus guttatus* DC. between the coastal perennial and inland annual ecotypes concerning salt tolerance, with no trade-offs detected in field experiments (Lowry et al., 2009), or in the serpentine and non-serpentine forms of a campion *Silene vulgaris* (Moench) Garcke, with both, major and minor QTLs, contributing to the observed phenotypes (Bratteler et al., 2006). It seems, therefore, that comparisons of populations from contrasting environments are useful for finding genes controlling traits allowing them to survive in unfavorable conditions. Evidently, local adaptation may be a result of multiple mechanisms, so it is hard to make any generalizations.

Another important finding drawn from genomic studies is that environmental heterogeneity leads to formation of ecotypes adapted to local conditions, even in the presence of gene flow. For instance, Sambatti & Rice (2006) assessed the local adaptation of distinct forms of *H. exilis* A. Gray, with reciprocal transplant trials at serpentine and riparian habitats in northern California. They concluded that each site exerts different selective pressure and that strong selection allows maintaining ecotypic variation within *H. exilis*, but considerable gene flow between populations prevents local speciation. Conversely, Muir et al. (2013) analyzed the interspecific hybrid zone between ragworts *Senecio aethnensis* Jan ex DC. and *S. chrysanthemifolius* Poir., forming an altitudinal cline on Mount Etna (Sicily). These species are distinct in their morphology and physiology, but they are connected by gene flow sufficient enough to prevent divergence at the nucleotide level. However, the study suggests that there is diversifying selection on genes, differentially expressed in those two taxa, that are somewhat involved in maintaining altitude-related adaptation. Consequently, *S. aethnensis* and *S. chrysanthemifolius* should rather be considered not as separate species but as ecotypes adapted to a steep altitudinal gradient.

Adaptive introgression and speciation in the presence of gene flow

Studies of hybrid zones have greatly contributed to our understanding of genomic architecture of different evolutionary phenomena. Research involving congeners from admixed sites enabled to elucidate the intensity, direction and extent of gene flow (e.g., Scascitelli et al., 2010), genomics of reproductive isolation (e.g., Minder & Widmer, 2008; Lexer et al., 2010) or showed that hybridization may have a significant role in speciation and adaptation (e.g., Rieseberg et al., 2007). Admixed populations are convenient to explore genetic background of speciation and adaptation for several reasons. First of all, they consist of various genotypes resulting from natural recombination events. Moreover, they give an opportunity to work on long-lived or hard to breed taxa. Ultimately, all aspects of interspecific gene flow and introgression are present in hybrid zones, as opposed to controlled crosses (Rieseberg et al., 1999).

Integration of genetic material from one species into another may increase fitness of hybrid genotypes in a given environment that will be favored by natural selection as compared to parental species. The gene flow can proceed in the populations leading to propagation of hybrids and adaptive divergence even if admixed individuals initially have reduced fertility or viability (Gross & Rieseberg, 2005). Various species and ecotypes of hybridizing sunflowers turned out particularly useful for assessing the possible effects of intra- and interspecific gene flow. They proved that adaptive introgression may facilitate, e.g., range expansion (Rieseberg et al., 2007), herbivore resistance (Whitney et al., 2006) and abiotic tolerance (Whitney et al., 2010). Furthermore, as shown in the study of ecotypes of *Helianthus petiolaris* Nutt., some environments may exert pressure strong enough to initiate speciation despite ongoing gene flow between the considered taxa (Andrew & Rieseberg, 2013). Andrew & Rieseberg (2013) concluded that divergence between the dune and non-dune forms of *H. petiolaris* growing at Great Sand Dunes National Park and Preserve (Colorado, USA) was restricted rather to a few large genomic regions, widely associated with environmental factors. Selective sweeps were primarily found in the dune ecotype, which may have adapted via new mutations, standing genetic variation or migration of beneficial alleles from *H. anomalus* S.F. Blake (Andrew et al., 2013). Such an asymmetric pattern may be general during the early stage of ecological speciation. On the basis of these outcomes, it seems that the Great Sand Dune sunflowers are changing from ecotypes to incipient species, resulting from strong edaphic selection (Andrew & Rieseberg, 2013). In turn, research on broadly sympatric and hybridizing *H. petiolaris* and *H. annuus* proved that these species are little differentiated and isolated only to small extent, leaving much of their genomes prone to gene flow (Kane et al., 2009). As opposed to the ecotypic comparison of *H. petiolaris*, there are more 'islands of divergence' that are more dispersed throughout the genomes of *H. petiolaris* and *H. annuus*, which is consistent with models of ecological speciation (Andrew & Rieseberg,

http://rcin.org.pl

2013). These results also support the genic view of speciation and correspond with studies of genome-wide divergence patterns in hybridizing campions *Silene latifolia* Poir. and *S. dioica* (L.) Clairv. (Minder & Widmer, 2008). The two campions are closely related, yet morphologically and phenologically distinct. They also differ in pollinator and habitat preference (Karrenberg & Favre, 2008). Similar to *H. petiolaris* and *H. annuus*, interspecific boundaries between *S. latifolia* and *S. dioica* seem to be 'porous' but maintained by divergent selection acting on particular adaptive genes (Minder & Widmer, 2008).

Single vs. multispecies approach

Despite a variety of studies that have been conducted so far to draw conclusions regarding mechanisms of natural selection and adaptation, it is still hard to make general conclusions about these continuous processes. Many studies pay attention to most environmental variables, but they primarily focus on a single factor which has clinal distribution or compare phenotypes from two distinct habitats (Lowry, 2010). Furthermore, it is relatively easy to omit very complex relationships between distinct species occupying the same area. As long as research is small-scale and based mainly on single taxa, an informative geographical information system (GIS) or growing genomic and ecological datasets for model and non-model species are helpful but insufficient. To maximize the power of evolutionary studies, it is required to conduct comprehensive multispecies research that mirrors actual conditions and tests various hypotheses in different environments, occupied by both sympatric and allopatric populations. Lawrie & Petrov (2014) demonstrated that comparative genomics would indeed have greater detection power once supplemented with ultra-deep population sampling of closely related species. Such multispecies data may be then used to create computer simulation models useful to study evolutionary mechanisms. Obviously, such models also have to incorporate detailed information on landscape features (Lowry, 2010), be able to test various hypotheses, pay attention to stochastic processes and take into account the specific properties of the species studied (Manel et al., 2010). Cautious choice of ecologically relevant species (=keystone species, i.e., pivotal to ecosystem functioning) is crucial in this context as it may allow identifying some general patterns that are true also in other ecosystems (Manel & Holderegger, 2013). Once the model is tested for multiple species in distinct landscapes, it should be able to accurately predict how various environmental changes will affect the populations in question (Manel & Holderegger, 2013).

There are also other benefits of multispecies study systems. Inter- and intraspecific comparisons of organisms living in similar environments give an opportunity to study the likeliness of convergent evolution at the genetic level, namely parallel evolution acting through mutations that emerge in independent lineages or collateral evolution either of introgressed alleles or alleles present in an ancestral population (Stern, 2013). Considering samples from refugial areas, closely related and outgroup species could also help to resolve the role of new mutations and standing genetic variation in development of ecological, phenotypic and physiological differentiation between locally adapted populations. Such differentiation is often a result of postglacial recolonization, but the genetics underlying variation at adaptive traits is mostly unknown. What is more, much can be discovered from research involving species which are interfertile. As mentioned earlier, they can be subjects of admixture mapping that enables mapping of the loci involved in reproductive isolation, speciation and adaptation in natural populations (Lexer et al., 2010). Since speciation is rather a process than a single event (Nosil et al., 2009), it would be equally important to study related taxa at various stages of the whole speciation continuum to see how divergent genomic regions evolve over time and in different environmental conditions (Andrew & Rieseberg, 2013). However, for some species with long generation times, like for instance many forest trees, it is impossible to obtain a few generations of pedigrees in a reasonable time scale. Another challenges include the possibility that adaptive traits might be reflected differently in genomes of distinct species (García-Gil et al., 2003; Ingvarsson et al., 2006). Furthermore, current observations may not explain the actual situation since they may have resulted from adaptation to conditions that do not exist anymore (Manel et al., 2010).

Closely related forest tree species as a model for studies of adaptive variation and speciation

Trees are undoubtedly of key importance to many terrestrial and coastal environments. Their ecological and economical value is obvious: they sequester carbon dioxide, improve air, water and soil quality, provide us with oxygen, energy, food, timber, paper and many more. Many organisms are dependent on niches created by trees, and considering species richness, woody plants can be regarded primary landscape modulators (Shachak et al., 2008). Longevity and immobility of trees force them to cope with environmental pressure in a different way than animal species do. Populations must either migrate, evolve or be plastic to survive (Reusch & Wood, 2007), in which case the first possibility is impossible for plant individuals. Genetic diversity of woody plants is greater than of non-woody species with comparable life history traits (Hamrick & Godt, 1996). This feature greatly facilitates adaptation via standing genetic variation, which is particularly important considering environmental changes. Forest trees are good objects to study adaptive variation and speciation processes, because they are undomesticated unlike crop plants which genomes have been influenced by many domestication bottlenecks (Neale & Savolainen, 2004). Most species are highly outcrossing, thus extensive gene flow by pollen spreads adaptive genetic variants across their populations. It also lowers differentiation at neutral loci (Aitken et al., 2008). Nonetheless, genetic variation in adaptive loci remains visible, because directional selection is supposed to lead to a loss of ancestral polymorphisms, leaving more fixed differences at adaptive loci or at linked sites, as compared to background genetic variation (Broughton & Harrison, 2003; Via, 2009). Thus, genetic diversity of each local

http://rcin.org.pl

population should be reduced at certain loci, responsible for their fitness-related adaptation. Keeping in mind that adaptive traits are often correlated with climatic variables, methods implemented by landscape genomics may come in useful. High outcrossing rates, ability of long-distance dispersal and large effective population size eventually lower the level of LD in trees which makes them feasible for employing landscape genomics approach to identify genes underlying complex traits (Gonzalez-Martínez et al., 2006; Sork et al., 2013). The rapid decay of LD requires genotyping of a large number of markers, but it is advantageous once a marker-trait association is confirmed, because it is probable that this marker is in close proximity to the gene responsible or is the functional variant itself (Neale & Kremer, 2011). Such analyses may lead to better understanding of the relationship between gene flow, speciation and adaptation in trees, because they facilitate specific localization of genetic variants standing behind the observed phenotypes.

Emerging study systems of closely related forest tree species

Comparative genomic studies have proved useful for unveiling various evolutionary mechanisms (e.g., Sharma et al., 2014; Zhang et al., 2014). As for trees, there is no single model species, and attention is focused rather on the ones of regional importance. However, comparative genomics of closely related trees has a great potential to serve as a link in studies of the genetic basis of intra- but also interspecific differentiation.

So far, the research utilizing closely related tree species for evolutionary inferences has focused on few genera. The most prominent study system consists of poplars, which have relatively small genome size, can be transformed with *Agrobacterium* (Neale & Kremer, 2011) and form hybrid zones suitable for admixture mapping (Lexer et al., 2010). *Populus trichocarpa* Torr. & A. Grey was the first sequenced tree genome (Tuskan et al., 2006), and so it facilitated further work on poplars. For instance, Lexer et al. (2010) studied three hybrid zones of ecologically divergent *P. alba* L. and *P. tremula* L. They discovered that even though reproductive isolation between these two species is strong, introgression of neutral or advantageous alleles is still possible. This finding is consistent with other studies indicating that species boundaries are porous (mentioned earlier in the paper). In another example, 70 mapped microsatellite loci were used to study patterns of neutral and adaptive population divergence in *P. tremula*. The authors detected the genetic signature indicating postglacial admixture between divergent lineages of *P. tremula* in Scandinavia. They concluded that admixture facilitated adaptation via standing genetic variation by contributing to formation of phenotypic cline for bud set across a latitudinal gradient in Sweden (de Carvalho et al., 2010). Similar evidence of clinal variation in candidate genes for timing of bud set in *P. tremula* was also confirmed with SNPs (Ingvarsson et al., 2006, 2008).

There are only a few study systems similar to poplars. Parallel studies have been carried out dealing with interfertile oaks that also differ in ecology and form hybrid zones. It turned out that the results greatly corresponded with findings obtained for poplars and other species. Outcomes of many studies addressing differentiation between *Quercus patraea* (Matt.) Liebl. and *Q. robur* L. indicate that most of their genomes is little differentiated with only some markers showing greater differentiation and likely contributing to species divergence (e.g., Gömöry et al., 2001; Scotti-Saintagne et al., 2004; Neophytou et al., 2010). These results again confirm that species integrity is maintained through small genome regions, leaving the rest of them prone to gene flow. In one of such work the authors were able to group the examined microsatellite loci into ‘species’ and ‘provenance discriminant’. Consequently, the former may represent genomic regions influenced by directional selection maintaining species identity, whereas ‘provenance specific’ loci would rather be a part of regions affected by interspecific gene flow and adaptation to local environmental conditions (Neophytou et al., 2010). Research on oak species also proved the evident clinal phenological variation in trees (e.g., Ducousso et al., 1996).

In comparison with angiosperms, gymnosperms have genomes of very great size ($>1.0 \times 10^{10}$ bp for conifers; Neale & Savolainen, 2004). However, conifers’ megagametophyte is a haploid tissue that can be used to directly determine the haplotype sequence. Conifers are obviously of great ecological importance. They dominate in many ecosystems, especially in boreal forests. They have low level or even no population structure, which is hardly seen in other species, so their diversity reflects natural selection acting throughout their evolution (Neale & Savolainen, 2004). Comparative evolutionary studies in conifers identified candidate genes for adaptation to local environmental conditions (e.g., Namroud et al., 2008) and confirmed clinal variation at SNPs in candidate genes for timing of bud set (e.g., Chen et al., 2012). However, there are a few examples of works that focused on closely related species. For instance, Hamilton et al. (2013) examined a spruce hybrid zone between *Picea sitchensis* (Bong.) Carr. and *P. glauca*. Seeds from the contact zone of the focal species had been planted in a common garden within the hybrid zone. 384 candidate-gene SNPs were subsequently genotyped in 721 individual trees, which were also assessed for their height and autumn cold hardiness at the age of ten. The outcomes of the study pointed to weak intrinsic interspecific boundaries, but introgression seemed asymmetric and shifted towards *P. sitchensis*. There were no obvious phenotypic clines, but variation in the proportion of ancestry along the zone appeared to be shaped by both geographic and climatic factors.

Pine relatives as a promising study system

Despite progress in genomic studies, research in conifers has been limited due to the lack of a reference genome. Fortunately, this situation has recently changed as the genomes of Norway spruce, *Picea abies* (L.) H. Karst (Nystedt et al., 2013), white spruce, *P. glauca* (Moench) Voss (Birol et al., 2013) and loblolly pine, *Pinus taeda* L. (Neale et al., 2014) have been sequenced. Furthermore, at least two other pine genomes (*P. sylvestris* L. and *P. pinaster* Aiton) should be available soon due to ongoing genome sequencing initiatives (http://www.procogen.eu/). With the fast-growing number of sequenced genomes it seems that pines may soon advance comparative evolutionary studies in plants. Pines differ in morphology and ecology showing clear patterns of adaptive variation across large geographical areas of the Northern

http://rcin.org.pl

Hemisphere as well as some tropical and subtropical territories. There are over a hundred pine species with several very closely related taxa. One of such excellent candidates for comparative evolutionary studies consists of Scots pine (*P. sylvestris*) and the taxa from the *P. mugo* complex. Scots pine has the largest Euroasian distribution of all pine species (Critchfield & Little, 1971), and at present it occupies mostly allopatric locations as compared to the members of the *P. mugo* complex. The complex includes typical dwarf mountain pine, *P. mugo* Turra, that grows in high elevation habitats, e.g., in the Eastern Alps and in the Carpathian Mountains, and mountain pine, *P. uncinata* (Ramond) Domin, that occurs in the Western Alps and in the Pyrenees (Monteleone et al., 2006). The group also incorporates other pines including *P. uliginosa* Neumann that grows on peat-bogs in Central Europe. Particular taxa within the complex are closely related, but they differ largely in growth forms and occupied environments. The mentioned pine species were a subject of several biometric, biochemical and molecular studies that show that they are a very suitable model system for comparative genomics studies of the demographic and evolutionary forces driving divergence in natural populations (Table 3). For instance, they form randomly mating natural populations of considerable ecological diversity that are suitable to effectively contrast genetic variation at neutral alleles with variation at genomic regions under selection. Provenance trials and common garden experiments show high level of heritability and differentiation of the species at quantitative traits related to gradients of water availability, photoperiod and temperature. High differentiation at the traits including growth forms is accompanied by low genetic differentiation at neutral marker loci indicating that quantitative traits are subject to diversifying selection. Most recent studies on Scots pine and the taxa from the *P. mugo* complex delivered new genomic resources for the species including the reference *de-novo* transcriptome sequence for Scots pine and a large database of over 200 000 SNPs for each species (Wachowiak et al., 2015). With the fast progress of NGS methodology, those novel genomic resources can be directly used to genotype many samples at many genetic markers to provide high resolution of adaptive and neutral genetic variation within the pine species. This will certainly help to build a more complete picture of genetic architecture of adaptive traits to advance forest management, conservation and breeding strategies in pines and other plant species.

Conclusions

Complementary approaches providing information on both genotypes and phenotypes and their ecological relevance are needed to successfully address questions about the genetic

Table 3 Distinctive features of *Pinus sylvestris* and the *Pinus mugo* complex as a potential model system for evolutionary studies

Feature	Conclusions and relevance	References
Close relatedness as evident from studies using morphological, biochemical and molecular approaches.	The focal pines share the same evolutionary history and have likely diverged only recently. ➢ Useful for studies of early stages of speciation.	Lewandowski et al., 2000; Bogunic et al., 2003; Grotkopp et al., 2004; Heuertz et al., 2010; Boratyńska et al., 2011; Wachowiak et al., 2011, 2013, 2015
Differences in phenotype, ecology and geographical distribution despite uniform genetic background.	Morphological differences seem to result from adaptation to different environments. ➢ Useful for finding loci responsible for local adaptation.	
Pinus sylvestris is already a subject of numerous research and constantly growing genomic resources for pines.	➢ Gives a fine background for comparative studies.	Hurme et al., 2000; Komulainen et al., 2003; Yin et al., 2003; Cheddadi et al., 2006; Joosen et al., 2006; Wachowiak et al., 2014b, 2015; The Conifer Genome Network (http://www.pinegenome.org/)
Incomplete reproductive isolation.	*Pinus sylvestris* and members of the *Pinus mugo* complex may form hybrid zones. ➢ Offers a possibility of admixture mapping to map loci involved in reproductive isolation, speciation and local adaptation. ➢ Offers an opportunity to study gene flow, introgression and influence of hybridization on species evolution. ➢ Microevolutionary processes that can be inferred from studying such hybrid zones are very likely to be true for hybridizing populations of other species forming similar aggregates.	Wachowiak et al., 2005; Kormutak et al., 2008; Wachowiak & Prus-Głowacki, 2008; Jasińska et al., 2010, 2014a
High genetic differentiation of *Pinus uliginosa* populations and diversity comparable to other pines despite its disjunctive locations and small population size.	Genetic diversity of *Pinus uliginosa* may have been formed when its populations were much more numerous. The possible reasons of great interpopulation differentiation include: different origin, long-term isolation and influence of hybridization processes. ➢ Such disjunctive stands may serve as useful objects to infer whether adaptation to replicate and isolated environments yields convergent genomic and/or phenotypic changes. ➢ They are also valuable to assess the influence of range fragmentation and drift on genetic variability of populations.	Lewandowski et al., 2002; Wachowiak & Prus-Głowacki, 2009; Wachowiak et al., 2011, 2013

http://rcin.org.pl

basis of local adaptation and speciation. Generating a vast amount of genomic data is no longer a limiting factor thanks to the development of efficient NGS and genotyping methods. Still, we are far from understanding a link between genetic variation and phenotypic differences, not to mention their ecological importance. Most of all, there is certainly a need for comparative genomic research on different species to build a broader picture of evolutionary processes and a particular group of genes involved in local adaptation and speciation. Assuming that many closely related plant species, including Scots pine and the taxa from the *P. mugo* complex, show high ecological and phenotypic differentiation that is usually accompanied by low background genetic variation, such experimental systems may constitute very informative models useful for searching for patterns of speciation and adaptive variation across genomes (Wachowiak et al., 2011, 2015). Studies of closely related species have already proved valuable, but more informative experimental systems are needed to better understand evolutionary processes shaping present adaptive divergence and to come up with effective molecular strategies for management of natural resources in the face of ongoing environmental changes.

Acknowledgements

This work was financially supported by the Polish National Science Centre (Grant No. DEC-2012/05/E/NZ9/03476).

Conflict of Interest

The authors declare that they have no conflict of interest.

References

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

Andrew RL, Kane NC, Baute GJ, Grassa CJ, Rieseberg LH. 2013. Recent nonhybrid origin of sunflower ecotypes in a novel habitat. *Molecular Ecology* 22: 799–813.

Andrew RL, Rieseberg LH. 2013. Divergence is focused on few genomic regions early in speciation: Incipient speciation of sunflower ecotypes. *Evolution* 67: 2468–2482.

Barrett RD, Hoekstra HE. 2011. Molecular spandrels: Tests of adaptation at the genetic level. *Nature Reviews: Genetics* 12: 767–780.

Birol I, Raymond A, Jackman SD, Pleasance S, Coope R, Taylor GA, Yuen MM, Keeling CI, Brand D, Vandervalk BP, Kirk H, Pandoh P, Moore RA, Zhao Y, Mungall AJ, Jaquish B, Yanchuk A, Ritland C, Boyle B, Bousquet J, Ritland K, Mackay J, Bohlmann J, Jones SJ. 2013. Assembling the 20 Gb white spruce (*Picea glauca*) genome from whole-genome shotgun sequencing data. *Bioinformatics* 29: 1492–1497.

Bogunic F, Muratovic E, Brown SC, Siljak-Yakovlev S. 2003. Genome size and base composition of five *Pinus* species from the Balkan region. *Plant Cell Reports* 22: 59–63.

Boratynska K, Jasinska AK, Marcysiak K, Sobierajska K. 2011. *Pinus uliginosa* from Czarne Bagno peat-bog (Sudetes) compared morphologically to related *Pinus* species. *Dendrobiology* 65: 17–28.

Bratteler M, Baltisberger M, Widmer A. 2006. QTL analysis of intraspecific differences between two *Silene vulgaris* ecotypes. *Annals of Botany* 98: 411–419.

Broughton RE, Harrison RG. 2003. Nuclear gene genealogies reveal historical, demographic and selective factors associated with speciation in field crickets. *Genetics* 163: 1389–1401.

Cheddadi R, Vendramin GG, Litt T, Francois L, Kageyama M, Lorentz S, Laurent J-M, de Beaulieu J-L, Sadori L, Jost A, Lunt D. 2006. Imprints of glacial refugia in the modern genetic diversity of *Pinus sylvestris*. *Global Ecology and Biogeography* 15: 271–282.

Chen J, Källman T, Ma X, Gyllenstrand N, Zaina G, Morgante M, Bousquet J, Eckert A, Wegrzyn J, Neale D, Lagercrantz U, Lascoux M. 2012. Disentangling the roles of history and local selection in shaping clinal variation of allele frequencies and gene expression in Norway spruce (*Picea abies*). *Genetics* 191: 865–881.

Critchfield WB, Little EL. 1971. *Geographic distribution of the pines of the world*. Washington: U.S. Department of Agriculture.

de Carvalho D, Ingvarsson PK, Joseph J, Suter L, Sedivy C, Macaya-Sanz D, Cottrell J, Heinze B, Schanzer I, Lexer C. 2010. Admixture facilitates adaptation from standing variation in the European aspen (*Populus tremula* L.), a widespread forest tree. *Molecular Ecology* 19: 1638–1650.

de Jong G. 2005. Evolution of phenotypic plasticity: Patterns of plasticity and the emergence of ecotypes. *New Phytologist* 166: 101–117.

Ducousso A, Guyon JP, Kremer A. 1996. Latitudinal and altitudinal variation of bud burst in western populations of sessile oak (*Quercus petraea* (Matt) Liebl). *Annals of Forest Science* 53: 775–782.

Fay JC, Wittkopp PJ. 2008. Evaluating the role of natural selection in the evolution of gene regulation. *Heredity* 100: 191–199.

García-Gil MR, Mikkonen M, Savolainen O. 2003. Nucleotide diversity at two phytochrome loci along a latitudinal cline in *Pinus sylvestris*. *Molecular Ecology* 12: 1195–1206.

Gömöry D, Yakovlev I, Zhelev P, Jedinaková J, Paule L. 2001. Genetic differentiation of oak populations within the *Quercus robur*/ *Quercus petraea* complex in Central and Eastern Europe. *Heredity* 86: 557–563.

Gonzalez-Martínez SC, Krutovsky KV, Neale DB. 2006. Forest-tree population genomics and adaptive evolution. *New Phytologist* 170: 227–238.

Gross BL, Rieseberg LH. 2005. The ecological genetics of homoploid hybrid speciation. *Journal of Heredity* 96: 241–252.

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

Hamilton JA, Lexer C, Aitken SN. 2013. Genomic and phenotypic architecture of a spruce hybrid zone (*Picea sitchensis* x *P. glauca*). *Molecular Ecology* 22: 827–841.

Hamrick JL, Godt MJW. 1996. Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society B: Biological Sciences* 351: 1291–1298.

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

Howe GT, Aitken SN, Neale DB, Jermstad KD, Wheeler NC, Chen THH. 2003. From genotype to phenotype: Unraveling the complexities of cold adaptation in forest trees. *Canadian Journal of Botany* 81: 1247–1266.

http://rcin.org.pl

Hurme P, Sillanp¨a ¨a MJ, Arjas E, Repo T, Savolainen O. 2000. Genetic basis of climatic adaptation in Scots pine by Bayesian quantitative trait locus analysis. *Genetics* 156: 1309–1322.

Ingvarsson PK, Garcı́a MV, Hall D, Luquez V, Jansson S. 2006. Clinal variation in *phyB2*, a candidate gene for day-length-induced growth cessation and bud set, across a latitudinal gradient in European aspen (*Populus tremula*). *Genetics* 172: 1845–1853.

Ingvarsson PK, Garcia MV, Luquez V, Hall D, Jansson S. 2008. Nucleotide polymorphism and phenotypic associations within and around the *phytochrome B2* Locus in European aspen (*Populus tremula*, Salicaceae). *Genetics* 178: 2217–2226.

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

Joosen RVL, Lammers M, Balk PA, Brønnum P, Konings MCJM, Perks M, Stattin E, van Wordragen MF, van der Geest AHM. 2006. Correlating gene expression to physiological parameters and environmental conditions during cold acclimation of *Pinus sylvestris*, identification of molecular markers using cDNA microarrays. *Tree Physiology* 26: 1297–1313.

Kane NC, King MG, Barker MS, Raduski A, Karrenberg S, Yatabe Y, Knapp SJ, Rieseberg LH. 2009. Comparative genomic and population genetic analyses indicate highly porous genomes and high levels of gene flow between divergent *Helianthus* species. *Evolution* 63: 2061–2075.

Kane NC, Rieseberg LH. 2007. Selective sweeps reveal candidate genes for adaptation to drought and salt tolerance in common sunflower, *Helianthus annuus*. *Genetics* 175: 1823–1834.

Karrenberg S, Favre A. 2008. Genetic and ecological differentiation in the hybridizing campions *Silene dioica* and *S. latifolia*. *Evolution* 62: 763–773.

Komulainen P, Brown GR, Mikkonen M, Karhu A, Garcı́a-Gil MR, O'Malley D, Lee B, Neale DB, Savolainen O. 2003. Comparing EST-based genetic maps between *Pinus sylvestris* and *Pinus taeda*. *Theoretical and Applied Genetics* 107: 667–678.

Kormutak A, Demankova B, Gömöry D. 2008. Spontaneous hybridization between *Pinus sylvestris* L. and *P. mugo* Turra in Slovakia. *Silvae Genetica* 57: 76–82.

Lawrie DS, Petrov DA. 2014. Comparative population genomics: Power and principles for the inference of functionality. *Trends in Genetics* 30: 133–139.

Lee CR, Mitchell-Olds T. 2012. Environmental adaptation contributes to gene polymorphism across the *Arabidopsis thaliana* genome. *Molecular Biology and Evolution* 29: 3721–3728.

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

Lewandowski A, Samocko J, Boratynska K, Boratynski A. 2002. Genetic differences between two Polish populations of *Pinus uliginosa*, compared to *P. sylvestris* and *P. mugo*. *Dendrobiology* 48: 51–57.

Lexer C, Joseph JA, van Loo M, Barbara T, Heinze B, Bartha D, Castiglione S, Fay MF, Buerkle CA. 2010. Genomic admixture analysis in European *Populus* spp. reveals unexpected patterns of reproductive isolation and mating. *Genetics* 186: 699–712.

Lowry DB. 2010. Landscape evolutionary genomics. *Biology Letters* 6: 502–504.

Lowry DB, Hall MC, Salt DE, Willis JH. 2009. Genetic and physiological basis of adaptive salt tolerance divergence between coastal and inland *Mimulus guttatus*. *New Phytologist* 183: 776–788.

Manel S, Holderegger R. 2013. Ten years of landscape genetics. *Trends in Ecology and Evolution* 28: 614–621.

Manel S, Joost S, Epperson BK, Holderegger R, Storfer A, Rosenberg MS, Scribner KT, Bonin A, Fortin M-J. 2010. Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. *Molecular Ecology* 19: 3760–3772.

Minder AM, Widmer A. 2008. A population genomic analysis of species boundaries: Neutral processes, adaptive divergence and introgression between two hybridizing plant species. *Molecular Ecology* 17: 1552–1563.

Mitchell-Olds T, Willis JH, Goldstein DB. 2007. Which evolutionary processes influence natural genetic variation for phenotypic traits? *Nature Reviews: Genetics* 8: 845–856.

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

Muir G, Osborne OG, Sarasa J, Hiscock SJ, Filatov DA. 2013. Recent ecological selection on regulatory divergence is shaping clinal variation in *Senecio* on Mount Etna. *Evolution* 67: 3032–3042.

Namroud M-C, Beaulieu J, Juge N, Laroche J, Bousquet J. 2008. Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. *Molecular Ecology* 17: 3599–3613.

Neale DB, Kremer A. 2011. Forest tree genomics: Growing resources and applications. *Nature Reviews: Genetics* 12: 111–122.

Neale DB, Savolainen O. 2004. Association genetics of complex traits in conifers. *Trends in Plant Science* 9: 325–330.

Neale DB, Wegrzyn JL, Stevens KA, Zimin AV, Puiu D, Crepeau MW, Cardeno C, Koriabine M, Holtz-Morris AE, Liechty JD, Martínez-García PJ, Vasquez-Gross HA, Lin BY, Zieve JJ, Dougherty WM, Fuentes-Soriano S, Wu LS, Gilbert D, Març ais G, Roberts M, Holt C, Yandell M, Davis JM, Smith KE, Dean JF, Lorenz WW, Whetten RW, Sederoff R, Wheeler N, McGuire PE, Main D, Loopstra CA, Mockaitis K, deJong PJ, Yorke JA, Salzberg SL, Langley CH. 2014. Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome Biology* 15: R59.

Neophytou C, Aravanopoulos FA, Fink S, Dounavi A. 2010. Detecting interspecific and geographic differentiation patterns in two interfertile oak species (*Quercus petraea* (Matt.) Liebl. and *Q. robur* L.) using small sets of microsatellite markers. *Forest Ecology and Management* 259: 2026–2035.

Nielsen EE, Hemmer-Hansen J, Larsen PF, Bekkevold D. 2009. Population genomics of marine fishes: Identifying adaptive variation in space and time. *Molecular Ecology* 18: 3128–3150.

Nielsen R. 2005. Molecular signatures of natural selection. *Annual Review of Genetics* 39: 197–218.

Nosil P, Feder JL. 2013. Genome evolution and speciation: Toward quantitative descriptions of pattern and process. *Evolution* 67: 2461–2467.

Nosil P, Funk DJ, Ortiz-Barrientos D. 2009. Divergent selection and heterogeneous genomic divergence. *Molecular Ecology* 18: 375–402.

Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin YC, Scofield DG, Vezzi F, Delhomme N, Giacomello S, Alexeyenko A, Vicedomini R, Sahlin K, Sherwood E, Elfstrand M, Gramzow L, Holmberg K, Hällman J, Keech O, Klasson L, Koriabine M, Kucukoglu M, Käller M, Luthman J, Lysholm F, Niittylä T, Olson A, Rilakovic N, Ritland C, Rossello JA, Sena J, Svensson T, Talavera-Lopez C, Theissen G, Tuominen H, Vanneste K, Wu ZQ, Zhang B, Zerbe P, Arvestad L, Bhalerao R, Bohlmann J, Bousquet J, Garcia Gil R, Hvidsten TR, de Jong P, MacKay J, Morgante M, Ritland K, Sundberg B, Thompson

http://rcin.org.pl

SL, Van de Peer Y, Andersson B, Nilsson O, Ingvarsson PK, Lundeberg J, Jansson S. 2013. The Norway spruce genome sequence and conifer genome evolution. *Nature* 497: 579–584.

Poncet BN, Herrmann D, Gugerli F, Taberlet P, Holderegger R, Gielly L, Rioux D, Thuiller W, Aubert S, Manel S. 2010. Tracking genes of ecological relevance using a genome scan in two independent regional population samples of *Arabis alpina*. *Molecular Ecology* 19: 2896–2907.

Reusch TB, Wood TE. 2007. Molecular ecology of global change. *Molecular Ecology* 16: 3973–3992.

Rice AM, Rudh A, Ellegren H, Qvarnström A. 2011. A guide to the genomics of ecological speciation in natural animal populations. *Ecology Letters* 14: 9–18.

Rieseberg LH, Kim S-C, Randell RA, Whitney KD, Gross BL, Lexer C, Clay K. 2007. Hybridization and the colonization of novel habitats by annual sunflowers. *Genetica* 129: 149–165.

Rieseberg LH, Whitton J, Gardner K. 1999. Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics* 152: 713–727.

Rundle HD, Nosil P. 2005. Ecological speciation. *Ecology Letters* 8: 336–352.

Sambatti JBM, Rice KJ. 2006. Local adaptation, patterns of selection, and gene flow in the Californian serpentine sunflower (*Helianthus exilis*). *Evolution* 60: 696–710.

Savolainen O, Pyhäjärvi T, Knürr T. 2007. Gene flow and local adaptation in trees. *Annual Review of Ecology, Evolution, and Systematics* 38: 595–619.

Scascitelli M, Whitney KD, Randell RA, King M, Buerkle CA, Rieseberg LH. 2010. Genome scan of hybridizing sunflowers from Texas (*Helianthus annuus* and *H. debilis*) reveals asymmetric patterns of introgression and small islands of genomic differentiation. *Molecular Ecology* 19: 521–541.

Scotti-Saintagne C, Mariette S, Porth I, Goicoechea PG, Barreneche T, Bodenès C, Burg K, Kremer A. 2004. Genome scanning for interspecific differentiation between two closely related oak species [*Quercus robur* L. and *Q. petraea* (Matt.) Liebl.]. *Genetics* 168: 1615–1626.

Shachak M, Boeken B, Groner E, Kadmon R, Lubin Y, Meron E, Ne'Eman G, Perevolotsky A, Shkedy Y, Ungar ED. 2008. Woody species as landscape modulators and their effect on biodiversity patterns. *Bioscience* 58: 209–221.

Sharma A, Li X, Lim YP. 2014. Comparative genomics of Brassicaceae crops. *Breeding Science* 64: 3–13.

Smith JM, Haigh J. 1974. The hitch-hiking effect of a favourable gene. *Genetical Research* 23: 23–35.

Sork VL, Aitken SN, Dyer RJ, Eckert AJ, Legendre P, Neale DB. 2013. Putting the landscape into the genomics of trees: Approaches for understanding local adaptation and population responses to changing climate. *Tree Genetics & Genomes* 9: 901–911.

Stern DL. 2013. The genetic causes of convergent evolution. *Nature Reviews: Genetics* 14: 751–764.

Stinchcombe JR, Hoekstra HE. 2008. Combining population genomics and quantitative genetics: Finding the genes underlying ecologically important traits. *Heredity* 100: 158–170.

Storz JF. 2005. Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology* 14: 671–688.

Strasburg JL, Sherman NA, Wright KM, Moyle LC, Willis JH, Rieseberg LH. 2012. What can patterns of differentiation across plant genomes tell us about adaptation and speciation? *Philosophical Transactions of the Royal Society B: Biological Sciences* 367: 364–373.

Turner TL, Bourne EC, Von Wettberg EJ, Hu TT, Nuzhdin SV. 2010. Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nature Genetics* 42: 260–263.

Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen G-L, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Dejardin A, depamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjärvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leple J-C, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouze P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai C-J, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, Van de Peer Y, Rokhsar D. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313: 1596–1604.

Via S. 2009. Natural selection in action during speciation. *Proceedings of the National Academy of Sciences USA* 106 (Suppl 1): 9939–9946.

Via S. 2012. Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philosophical Transactions of the Royal Society B: Biological Sciences* 367: 451–460.

Vinocur B, Altman A. 2005. Recent advances in engineering plant tolerance to abiotic stress: Achievements and limitations. *Current Opinion in Biotechnology* 16: 123–132.

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

Wachowiak W, Cavers S, Żukowska WB. 2014a. Interspecific gene flow and ecological selection in a pine (*Pinus* sp.) contact zone. *Plant Systematics and Evolution* 301: 1643–1652.

Wachowiak W, Celinski K, Prus-Głowacki W. 2005. Evidence of natural reciprocal hybridisation between *Pinus uliginosa* and *P. sylvestris* in the sympatric population of the species. *Flora - Morphology, Distribution, Functional Ecology of Plants* 200: 563–568.

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

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

Wachowiak W, Prus-Głowacki W. 2009. Different patterns of genetic structure of relict and isolated populations of endangered peat-bog pine (*Pinus uliginosa* Neumann). *Journal of Applied Genetics* 50: 329–339.

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

Wachowiak W, Wojkiewicz B, Cavers S, Lewandowski A. 2014b. High genetic similarity between Polish and North European Scots pine (*Pinus sylvestris* L.) populations at nuclear gene loci. *Tree Genetics & Genomes* 10: 1015–1025.

http://rcin.org.pl

Whitney KD, Randell RA, Rieseberg LH. 2006. Adaptive introgression of herbivore resistance traits in the weedy sunflower *Helianthus annuus*. *The American Naturalist* 167: 794–807.

Whitney KD, Randell RA, Rieseberg LH. 2010. Adaptive introgression of abiotic tolerance traits in the sunflower *Helianthus annuus*. *New Phytologist* 187: 230–239.

Yin TM, Wang XR, Andersson B, Lerceteau-Köhler E. 2003. Nearly complete genetic maps of *Pinus sylvestris* L. (Scots pine) constructed by AFLP marker analysis in a full-sib family. *Theoretical and Applied Genetics* 106: 1075–1083.

Zhang DY, Li C, Li Q, Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, Storz JF, Antunes A, Greenwold MJ, Meredith RW, Ödeen A, Cui J, Zhou Q, Xu L, Pan H, Wang Z, Jin L, Zhang P, Hu H, Yang W, Hu J, Xiao J, Yang Z, Liu Y, Xie Q, Yu H, Lian J, Wen P, Zhang F, Li H, Zeng Y, Xiong Z, Liu S, Zhou L, Huang Z, An N, Wang J, Zheng Q, Xiong Y, Wang G, Wang B, Wang J, Fan Y, da Fonseca RR, Alfaro-Nuñez A, Schubert M, Orlando L, Mourier T, Howard JT, Ganapathy G, Pfenning A, Whitney O, Rivas MV, Hara E, Smith J, Farre M, Narayan J, Slavov G, Romanov MN, Borges R, Machado JP, Khan I, Springer MS, Gatesy J, Hoffmann FG, Opazo JC, Håstad O, Sawyer RH, Kim H, Kim KW, Kim HJ, Cho S, Li N, Huang Y, Bruford MW, Zhan X, Dixon A, Bertelsen MF, Derryberry E, Warren W, Wilson RK, Li S, Ray DA, Green RE, O'Brien SJ, Griffin D, Johnson WE, Haussler D, Ryder OA, Willerslev E, Graves GR, Alström P, Fjeldså J, Mindell DP, Edwards SV, Braun EL, Rahbek C, Burt DW, Houde P, Zhang Y, Yang H, Wang J, Consortium AG, Jarvis ED, Gilbert MT, Wang J. 2014. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* 346: 1311–1320.

http://rcin.org.pl

Publication 5

Wachowiak W, Cavers S, **Żukowska WB**

Interspecific gene flow and ecological selection in a pine (*Pinus* sp.) contact zone

Plant Systematics and Evolution 301: 1643-1652

Impact factor: 1.361

MNiSW points: 20

http://rcin.org.pl

Plant Syst Evol (2015) 301:1643–1652
DOI 10.1007/s00606-014-1182-x

CrossMark

ORIGINAL ARTICLE

Interspecific gene flow and ecological selection in a pine (*Pinus* sp.) contact zone

Witold Wachowiak · Stephen Cavers · Weronika B. Żukowska

Received: 4 April 2014 / Accepted: 18 November 2014 / Published online: 10 December 2014
© The Author(s) 2014. This article is published with open access at Springerlink.com

Abstract Nucleotide polymorphisms in a set of nuclear genes were studied in a sympatric population of pines *Pinus mugo* and *Pinus sylvestris* that includes trees classified as pure species and polycormic (multi-stemmed) individuals of potentially hybrid origin. Patterns of genetic diversity were compared between those groups of samples and to the reference allopatric populations of the species in Europe. Polymorphisms at the gene loci clearly distinguished pure parental species as measured by conventional frequency-based statistics and Bayesian assignment of samples into separate genetic clusters. Most individuals classified based on phenotypic assessments as putative hybrids were genetically very similar to *P. mugo* showing no existing average net divergence and genetic assignment to the same genetic cluster. On the other hand, individuals of *P. sylvestris* showed homogenous genetic background to the reference populations of the species from Central and Northern Europe. Ten individuals of admixed genetic composition were found in all three groups of samples; however, the majority of hybrids except one individual were identified across the samples classified as *P. mugo* and polycormic pines. Those trees that contained a mixture of nuclear gene haplotypes observed in the reference populations of pure species and *cp*DNA from *P. mugo*, most likely represent the first generation of hybrids. Analysis of the allelic frequency spectra and compound neutrality tests identified deviations from neutrality at several genes. This contact zone seems suitable for selection of a mapping population both in hybrid and parental species for admixture mapping to effectively search for polymorphisms that may play role in species adaptive variation and speciation.

Keywords Nucleotide polymorphisms · Hybridization · Natural selection · Divergence · *Pinus mugo* · *Pinus sylvestris*

Handling editor: Sylvain Razafimandimbison.

Electronic supplementary material The online version of this article (doi:10.1007/s00606-014-1182-x) contains supplementary material, which is available to authorized users.

W. Wachowiak
Faculty of Biology, Institute of Environmental Biology, Adam Mickiewicz University, Umultowska 89, 61-614 Poznań, Poland

W. Wachowiak · S. Cavers
Centre for Ecology and Hydrology Edinburgh, Bush Estate, Penicuik, Midlothian EH26 0QB, UK

W. Wachowiak (✉) · W. B. Żukowska
Institute of Dendrology, Polish Academy of Sciences, Parkowa 5, 62-035 Kórnik, Poland
e-mail: witoldw@man.poznan.pl

Introduction

Natural hybridisation is an important process that creates recombinants from interspecific mating between divergent parental taxa where they come into geographic contact (Arnold and Martin 2010). Hybridization occurs in roughly 10 % of animal species and 25 % of plant species and it may have various evolutionary consequences for the taxa involved (Baack and Rieseberg 2007). For instance, it may cause the swamping of the species with the smaller effective population size by gene flow from the more abundant species, integration of genetic material from one species into another through repeated back-crossing (introgression), homoploid hybrid speciation in which the new hybrid lineages become reproductively isolated from parental populations, and finally, the transfer of adaptive

http://rcin.org.pl

traits across species boundaries (Baack and Rieseberg 2007). There are well-documented examples which show that natural selection favours hybrid genotypes that may have equivalent or even higher fitness as compared to parental species due to environmental selection (Arnold et al. 2004; Minder and Widmer 2008). Even in the case of initially reduced fertility or viability of hybrids from early generations, gene flow can proceed in the populations leading to the propagation of hybrids and adaptive divergence (Gross and Rieseberg 2005).

Natural hybridisation was postulated between closely related Scots pine (*Pinus sylvestris* L.) and the taxa from the *Pinus mugo* complex including dwarf mountain pine (*P. mugo* T.) (Christensen 1987). Despite close phylogenetic relationships, the species are highly differentiated in phenotype (tree/shrub), geographical range (widespread/restricted) and ecology (generalist/specialist). *P. sylvestris* is the most widespread and economically important forest tree species in Europe and Asia, whereas *P. mugo* is an endemic species typical to the mountain regions of Central and Southern Europe. The present distribution of Scots pine is a result of postglacial migration from several glacial refugia (Pyhäjärvi et al. 2008). It is supposed that recolonisation created zones of secondary contacts between isolated local populations from ice-free regions which survived the last glacial maximum with populations from southern refugia. As the ranges of *P. sylvestris* and the taxa from the *P. mugo* complex overlapped in some part of their distribution, hybridisation between the species has likely contributed to high diversification observed especially within the *P. mugo* complex.

At present, those closely related but ecologically differentiated taxa form several contact zones in Central Europe that create unique environments for comparative studies of interspecific hybridization, introgression and the maintenance of species differences in the presence of gene flow. One of them is a sympatric population of *P. sylvestris* and *P. mugo* at the 'Bór na Czerwonem' peatbog in the Nowotarska Valley, Poland. This population contains a mixture of individuals that could be classified as both pure species and polycormic (multi-stemmed) trees of untypical morphology. The sympatric occurrence of phenotypically differentiated taxa in a very diverse habitat of the peatbog complexes provides a unique opportunity for genomic studies of the role of introgressive hybridization and ecological selection on the species adaptive divergence and evolution. However, nucleotide polymorphisms at nuclear genomes of individuals from contact zones of *P. sylvestris* and the taxa from the *P. mugo* complex have not been studied so far.

Here, we evaluated hybridization patterns and the role of interspecific gene flow in shaping genetic variation in a contact zone of dwarf mountain pine (*P. mugo*) and Scots pine (*P. sylvestris*). Using nucleotide sequence variation in a multilocus nuclear gene dataset and a set of the reference allopatric populations of the species, we looked at the patterns of population divergence through ecological selection and adaptation in the presence of interspecific gene exchange. Specifically, we tested for the patterns of neutral and adaptive variation at the loci and assessed the role of hybridization and selection in generating the genomic patterns of diversity in the specific peatbog habitats of the species contact zones as compared to the reference allopatric populations of the species in Europe.

Materials and methods

Sampling and DNA extraction

Seeds from 60 individual trees were collected from Bór na Czerwonem reserve in Nowotarska Valley, Poland (Table 1). This stand is represented by three major phenotypic groups of individuals including (1) bushy *P. mugo*-like pines, (2) oligo- and polycormic (multi-stemmed) pines of untypical morphology and a height of over 2 m that cover most of the peatbog and (3) monocormic (single-stemmed) *P. sylvestris*-like individuals that are dominant at some central and north-eastern parts of the peatbog. Selected phenotypic traits, i.e. growth form, bark colour of the upper part of a trunk and main branches, colour and shape of needles and setting angle of conelet from the previous year were used for preliminary taxonomic classification (Christensen 1987). In total, 20 trees from each of the three groups were sampled. In the course of analyses (see below), 10 putative hybrids were found among those 60 samples from the studied contact zone and they were grouped separately in most analyses. In addition, several reference populations from the European distribution of the species were used for the comparisons of the level of nucleotide variation due to hybridization and selection in the hybrid zone vs. genetic differentiation in the allopatric populations of the species not affected by interspecific gene flow (Fig. 1; Table 1). In different analysis, the reference populations were treated separately but also the nearest allopatric populations of the pure species were grouped together to compare a similar number of samples relative to the putative hybrid zone (Table 2). In total, 194 samples were analysed from the contact zone and the reference locations (Table 2). Genomic DNA was extracted from megagametophytes from germinated seeds using DNeasy Plant Mini Kit (Qiagen).

PCR amplification and DNA sequencing

Nucleotide diversity patterns were studied in a set of eight nuclear gene loci related to cellular metabolism, transport, signal transduction and transcription regulation (Online

http://rcin.org.pl

Table 1 Geographical location of the analysed sympatric stand of *P. mugo* and *P. sylvestris* and the reference allopatric populations

Acronym	Location	Region	Longitude	Latitude	Altitude
Sympatric *P. mugo* and *P. sylvestris* stand					
BOR_M	*P. mugo* Bór na Czerwonem	EU_C	20°02′20″	49°27′35″	620
BOR_S	*P. sylvestris* Bór na Czerwonem	EU_C	20°02′20″	49°27′35″	620
BOR_PC	Polycormic Bór na Czerwonem	EU_C	20°02′20″	49°27′35″	620
Reference *P. mugo* populations					
M1	Poland_Śląskie Kamienie	EU_C	15°36′80″	50°46′35″	1,300
M12	Austria_Karwendel Alps	EU_C	11°17′45″	47°22′42″	1,400
M4	Romania_Eastern Carpathians	CARP	24°48′00″	47°34′03″	1,720
M5	Romania_Southern Carpathians	CARP	25°27′06″	45°25′55″	2,070
M7	Bulgaria_Pirin	BALK	23°25′22″	41°46′07″	2,000
M8	Montenegro_Durmitor Mts.	BALK	19°05′27″	43°09′33″	2,100
M14	Italy_Carnic Alps	EU_S	13°08′50″	46°32′40″	1,300
M16	Italy_Abruzzi	EU_S	13°58′30″	41°46′20″	2,200
Reference *P. sylvestris* populations					
PS43	Poland_Jarocin	EU_C	17°28′40″	51°58′20″	120
PS36	Austria_Pernitz	EU_C	16°00′00″	47°54′50″	500
PS39	Finland_Punkaharju	FIN	29°23′21″	61°45′33″	80
PS40	Finland_Kolari	FIN	24°30′00″	67°11′00″	190
PS44	Sweden_Krp.Tjärnbergsheden	SWE	20°48′00″	64°37′12″	110
PS45	Sweden_Väster Mjőingenn	SWE	13°34′48″	62°45′00″	640

For most between population analyses, samples of the reference populations were divided into corresponding geographical regions

resource 1) (Ersoz et al. 2010). In addition, the species diagnostic *cp*DNA marker for *P. sylvestris* vs. *P. mugo* in *trnF*-*trnL* region (Taberlet et al. 1991) was screened in the samples. This DraI restriction enzyme PCR–RFLP marker was developed based on a single nucleotide polymorphism that leads to an undigested PCR product for *P. sylvestris* and a digested in one place (two bands) for *P. mugo* (Wachowiak et al. 2000). As *cp*DNA is paternally inherited in pines and transmitted by pollen, the comparative analysis of the phenotypes and composition of the chloroplast genomes in each individual may be useful to identify hybrids. PCR amplification was performed with Thermo MBS thermal cyclers and carried out in a total volume of 15 µl containing about 15 ng of haploid template DNA, 10 µM of each of dNTP, 0.2 µM of each forward and reverse primers, 0.15U of Taq DNA polymerase, 1× BSA, 1.5 mM of $MgCl_2$ and 1× PCR buffer (BioLabs). Standard amplification procedures were used with initial denaturation at 94 °C for 3 min. followed by 35 cycles with 30 s. denaturation at 94 °C, 30 s. annealing at 60 °C for nuclear loci and 53 °C for *trnF*-*trnL* region and 1 min. 30 s. extension at 72 °C, and a final 5 min. extension at 72 °C. PCR fragments were purified using ExoI-Sap (exonuclease I, Shrimp Alkaline Phosphatase) enzymatic treatment. About 20 ng of the PCR product was used as a template in 10 µl sequencing reactions with the Big Dye Terminator DNA Sequencing Kit (Applied Biosystems, Carlsbad, CA, USA) performed by the Genomed (Warsaw, Poland) sequencing service. Multilocus haplotypes were determined by direct sequencing of haploid DNA from megagametophyte (maternally derived haploid tissue surrounding embryo, which in gymnosperms has the same genotype as the egg cell). CodonCode Aligner software ver. 3.7.1 (Codon Code Corporation, Dedham, MA, USA) was used for editing of the chromatograms, visual inspection of all polymorphic sites detected and alignment and some insertion/deletions were manually adjusted across the samples using GenDoc. The reference sequence from *Pinus taeda* was used for outgroup comparisons. Haplotype sequence data at the nuclear loci analyzed are deposited in GenBank (NCBI accession number: KM277840-KM277893).

Tests for interspecific gene flow

We tested for introgressive hybridization and admixture patterns in the pine species by comparing the level of nucleotide and haplotype polymorphisms, divergence and difference in the allelic frequency spectra between different groups of samples. The samples included the reference pure species populations, hybrids identified in this contact zone and the remaining groups of samples from Bór na Czerwonem including *P. mugo*-like, *P. sylvestris*-like and oligo- and polycormic pines (Table 2). Nucleotide diversity was measured as the average number of nucleotide differences per site (π) between two sequences (Nei 1987). Multilocus estimates of population mutation parameter, theta (θ_W, equal to $4N_e\mu$, where N_e is the effective population size and μ is the mutation rate per nucleotide site per

http://rcin.org.pl

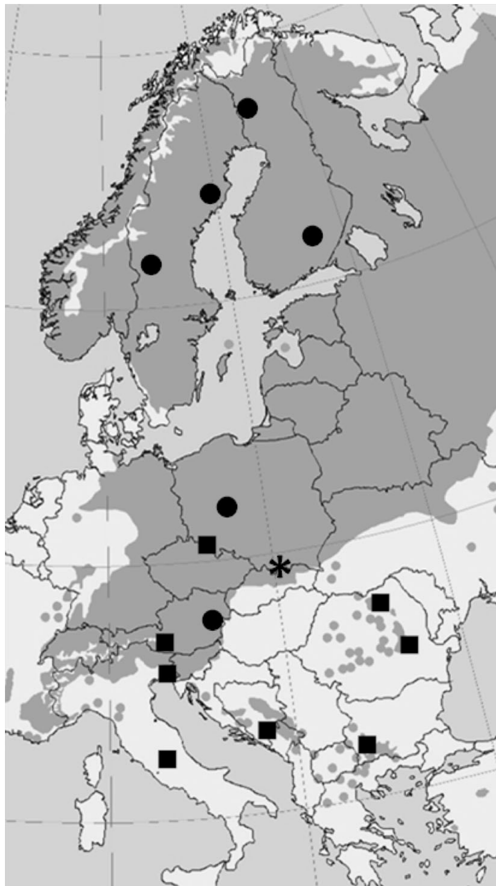


Fig. 1 Geographical location of the analysed hybrid population (*asterisk*) and the reference allopatric populations of *P. mugo* (*filled square*) and *P. sylvestris* (*filled circle*). Distribution map of Scots pine marked in *grey*

generation) (Watterson 1975), were computed based on the number of total and/or silent segregating sites and the length of each locus. The number of haplotypes (N_e) and haplotype diversity (H_d) were computed for each gene using DnaSP v.5. The number and frequency of unique and shared haplotypes in pairwise comparisons between species were calculated with Arlequin v.3 (Excoffier et al. 2005). Locus-by-locus estimates of net divergence between groups of samples (Nei 1987), the number of shared, exclusive and fixed polymorphic sites and haplotypes for each locus were determined using SITES 1.1. Clustering analysis based on a Bayesian assignment of samples to different groups was applied to look at the relationships between samples from the contact zone and the reference populations of the species using BAPS 6.0 software (Corander and Tang 2007). In the genetic mixture analysis, each locus was input separately as a fasta file using the MLST format and ten independent runs were conducted for each K (1–30) to estimate the number of clusters for all samples combined. The codon linkage model was used, the number of iterations used to estimate admixture coefficients for the individuals was set to 100, the number of the reference individuals was set to 100 and the number of iterations used to estimate admixture coefficients for the reference individuals was 10. The number of populations was inferred from the combined maximum likelihood and the highest posterior probability estimates over all runs. The software was also used for Bayesian admixture analysis that uses genotype information for each marker to estimate admixture parameters. A relationship between groups of samples defined was further evaluated based on the mean genetic distance. The number of base differences per sequence from averaging over all sequence pairs between groups was calculated using MEGA software (Tamura et al. 2011). Genetic differentiation in pairwise comparisons between populations was measured as Wright's fixation index (Weir and Cockerham 1984), F_{ST} over all polymorphic sites detected and tested for significance by 1,000 permutations of the samples between populations (Excoffier et al. 2005). We also performed the analysis of the genomic composition of paternally transmitted *cp*DNA in samples from the contact zone of the species. In this analysis, PCR products of diagnostic *trnF-trnL* marker were digested with DraI restriction enzyme and scored after electrophoresis on 2 % agarose gel as species-specific to *P. sylvestris* (an undigested product) and species-specific to *P. mugo* (a digested product with two bands).

Tests for natural selection

We looked if natural selection due to local adaptation to specific peatbog environments affected genes studied in both parental species and hybrids. The loci were examined for the evidence of selection based on the analysis of the allelic frequency spectra as compared to the genetic background of the reference populations and departures from neutral expectations of polymorphisms vs. divergence at the interspecific level. Deviations from the frequency distribution spectrum expected under the standard neutral model of evolution were assessed using the frequency spectrum test and coalescence-based approaches (Tajima 1989). The distribution of Tajima's *D* test statistics was investigated for each population or regional groups of populations. The significance of multilocus estimates of the test statistics was evaluated by comparison to a distribution generated by 1,000 coalescent simulations using the HKA programme. Orthologous sequences from the outgroup species were used in the Hudson-Kreitman-Aguadé (HKA) test (Jiggins et al. 2008) to look for overall departures from neutral expectations by assessing the level of multilocus polymorphism and divergence. Deviations of particular genes from the allelic and polymorphic sites frequency distribution spectra expected under the standard neutral

http://rcin.org.pl

Table 2 Summary statistics of nucleotide and haplotype variation and frequency distribution spectra in the hybrid and reference populations

Regional group	n	L	*SNPs* (Sing.)	Nucleotide diversity			D^c	Haplotype diversity		
				π_{tot}	θ^a_{tot}	CI (95 %)[b]		N	N %	H_d (SD)
BOR_M	15	3,222	41 (15)	0.0041	0.0038	0.0025–0.0057	−0.059	29	23.5 (3.0)	0.556 (0.098)
BOR_S	19	3,227	29 (11)	0.0026	0.0025	0.0016–0.0039	−0.502	29	22.7 (3.8)	0.435 (0.078)
BOR_PC	16	3,226	47 (13)	0.0035	0.0042	0.0028–0.0062	−0.041	35	28.0 (6.1)	0.611 (0.098)
BOR_hybrids	10	3,198	39 (12)	0.0043	0.0044	0.0028–0.0068	−0.063	28	23.5 (0.8)	0.647 (0.118)
M1_M12	20	3,203	46 (16)	0.0043	0.0039	0.0026–0.0056	−0.061	40	26.5 (3.8)	0.575 (0.079)
M4_M5	20	3,200	49 (18)	0.0044	0.0040	0.0027–0.0058	−0.233	34	25.0 (6.1)	0.536 (0.092)
M7_M8	20	3,193	53 (12)	0.0044	0.0044	0.0030–0.0063	−0.121	40	27.3 (6.1)	0.511 (0.074)
M14_M16	19	3,203	51 (17)	0.0045	0.0043	0.0030–0.0063	0.070	43	29.5 (6.8)	0.643 (0.079)
EU_C	20	3,178	37 (18)	0.0030	0.0033	0.0022–0.0049	−0.618*	34	26.5 (6.1)	0.506 (0.077)
FIN	20	3,227	34 (11)	0.0029	0.0030	0.0019–0.0,045	−0.637*	35	27.3 (4.5)	0.475 (0.098)
SWE	15	3,226	41 (25)	0.0032	0.0037	0.0024–0.0056	−0.810*	37	26.5 (9.1)	0.541 (0.083)

n number of samples analysed per locus, L average length of the sequences in base pairs excluding indels, *SNPs* number of polymorphic sites detected (number of singleton mutations), π nucleotide diversity (Nei 1987), N total number of haplotypes detected, N% percentage out of total number of 132 haplotypes identified across all loci and samples (percentage of unique haplotypes detected across all loci and samples), H_d haplotype diversity (standard deviation)

[*] $P < 0.05$

[a] Median multilocus θ for all sites

[b] 95 % credibility intervals for θ

model of evolution were investigated using two compound neutrality tests including HEW and DHEW (Zeng et al. 2007). Significance levels of the above tests were determined by carrying out 10,000 coalescent simulations based on Watterson's estimator of theta as implemented in dh package. For neutrality test that needs a species outgroup, we used orthologous GenBank sequences of *P. taeda* to contrast the level of intraspecific polymorphisms with interspecific divergence that should be positively correlated for neutrally evolving loci (Hudson et al. 1987). The genetic differentiation at the loci was measured as fixation index (F_{ST}) and its significance was evaluated by 1,000 permutations of the samples between different groups using Arlequin v.3 software (Excoffier et al. 2005).

Results

Genetic variation

About ~3.2 kbp of nuclear DNA sequence was aligned across all nuclear genes and samples providing a set of 146 polymorphic sites. Similar average total nucleotide diversity ($\pi_{tot} = \sim 0.005$) and multilocus estimates of nucleotide diversity ($\theta_{tot} = 0.004$) were found in the group of the reference *P. mugo* populations, *P. mugo* from Bór na Czerwonem and ten individuals of admixed ancestry classified as hybrids (Table 2). Diversity of samples classified as *P. sylvestris* and polycormic pines was slightly lower and similar to the reference *P. sylvestris* population (Table 2). In the species contact zone, no net divergence was observed between *P. mugo* vs. polycormic pines and divergence to the reference *P. mugo* populations was marginal (0.0002–0.0005) and several times lower as compared to *P. sylvestris* form Bór na Czerwonem and the reference populations (0.0016–0.0018). Similarly, no divergence was found between *P. sylvestris* from that population and the reference samples of the species (Table 3). Hybrids showed overall slightly lower divergence to *P. sylvestris* than *P. mugo*. Out of all 132 haplotypes detected across 8 loci in all 194 samples analysed, the majority (56 %) were unique (present only once). Similar proportion of haplotypes (23–29 % out of total identified) was found in each of the group of samples defined (Table 2). Only one unique haplotype was found in the group of hybrids. The average haplotype diversity (H_d) ranged between 0.43 and 0.64 (Table 2).

Population structure and differentiation

At all polymorphic sites combined, high differentiation ($F_{ST} = 0.28$–0.38) was found between the reference populations of *P. mugo* and *P. sylvestris* ($P < 0.01$). Hybrids showed significant differentiation to all other analysed groups of samples; however, the absolute values were slightly lower in reference to Scots pine than other groups

http://rcin.org.pl

Table 3 Net divergence in pairwise comparisons between the defined groups of samples

	BOR_M	BOR_S	BOR_PC	BOR_hybrids	M1_M12	M4_M5	M7_M8	M14_16	EU_C	FIN	SWE
BOR_M											
BOR_S	0.0017										
BOR_PC	0.0000	0.0017									
BOR_hybrids	0.0006	0.0003	0.0006								
M1_M12	0.0005	0.0024	0.0004	0.0015							
M4_M5	0.0002	0.0025	0.0003	0.0013	0.0002						
M7_M8	0.0003	0.0021	0.0003	0.0010	0.0001	0.0000					
M14_16	0.0002	0.0021	0.0001	0.0011	0.0002	0.0001	0.0001				
EU_C	0.0017	0.0000	0.0017	0.0003	0.0023	0.0024	0.0019	0.0020			
FIN	0.0018	0.0000	0.0017	0.0003	0.0025	0.0026	0.0021	0.0021	0.0000		
SWE	0.0016	0.0000	0.0016	0.0005	0.0021	0.0023	0.0018	0.0018	0.0000	0.0001	
PT	0.0200	0.0196	0.0199	0.0198	0.0201	0.0207	0.0203	0.0202	0.0197	0.0198	0.0199

Pinus taeda (PT) was used as an out group

Table 4 F_{ST} at all polymorphic sites combined between geographical groups of the hybrid and reference populations

	BOR_M	BOR_S	BOR_PC	BOR_hybrids	M1_M12	M4_M5	M7_M8	M14_M16	EU_C	FIN
BOR_M										
BOR_S	0.362*									
BOR_PC	−0.044	0.353*								
BOR_hybrids	0.136*	0.115*	0.127*							
M1_M12	0.050	0.385*	0.050	0.208*						
M4_M5	0.045	0.392*	0.054	0.212*	−0.019					
M7_M8	0.043	0.338*	0.052	0.154*	−0.009	−0.023				
M14_M16	0.027	0.349*	0.020	0.168*	0.010	0.009	0.006			
EU_C	0.348*	−0.025	0.341*	0.103*	0.372*	0.383*	0.328*	0.337*		
FIN	0.337*	0.011	0.325*	0.089*	0.371*	0.373*	0.320*	0.327*	−0.004	
SWE	0.307*	0.010	0.306*	0.125*	0.332*	0.338*	0.285*	0.292*	−0.014	0.031

* $P < 0.01$

(Table 4). Samples of polycormic pines from Bór na Czerwonem showed significant differentiation to *P. sylvestris* ($F_{ST} = 0.353$) from that area and all the reference populations of Scots pine ($F_{ST} = 0.31$–0.34). However, they showed no differentiation to *P. mugo* from Bór na Czerwonem and the reference *P. mugo* populations (Table 4). Two genetic clusters were detected in mixture analysis using BAPS software. A close genetic similarity was found between polycormic pines from Bór na Czerwonem, *P. mugo* from that area and the reference *P. mugo* populations that all belonged to one cluster. The other group was formed by the samples of *P. sylvestris* from Bór na Czerwonem and the reference Scots pine populations (Fig. 2). No admixture was found between the reference populations of *P. mugo* and *P. sylvestris* indicating strong discriminating power of the nucleotide polymorphisms at the nuclear loci to distinguish each species. Evidence of admixture was found at ten individuals in total including four samples from the group of polycormic pines, five from *P. mugo* and one from *P. sylvestris* (Fig. 2). The group of admixed samples contained a mixture of haplotypes unique for pure *P. sylvestris* and *P. mugo* populations. The only unique allele in hybrids was found in two samples at locus Pr4-21 and it resulted from a single mutation. Based on the genetic distance analyses, the admixed samples showed closer genetic relationship with *P. sylvestris* (Supplementary Fig. 1). All analysed samples phenotypically characterised as peatbog forms of *P. sylvestris* had *cp*DNA haplotypes species-specific to *P. sylvestris* except one individual that had *cp*DNA of *P. mugo*. The same individual was also shown to have admixed origin in clustering analysis at nuclear gene loci (Fig. 2). All individuals from the other groups of samples defined including *P. mugo*, polycormic pines and hybrids had *cp*DNA haplotypes species-specific to *P. mugo*.

http://rcin.org.pl

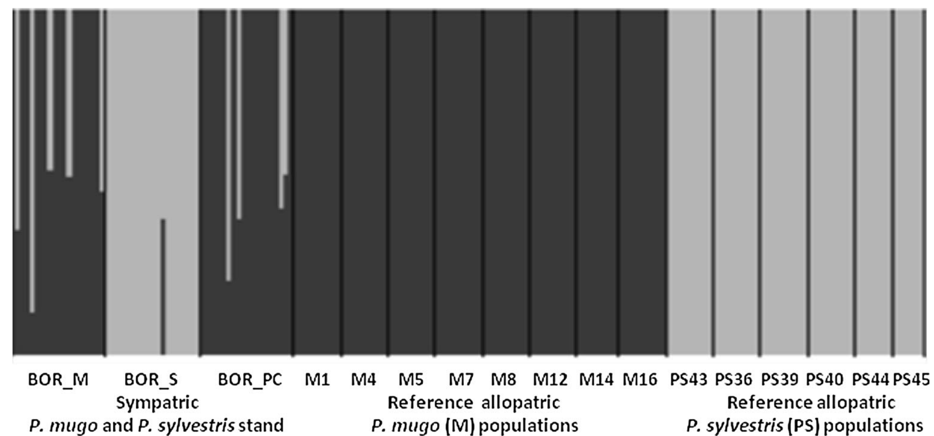


Fig. 2 Bar plot from cluster analysis (BAPS) in the group of pines from Bór na Czerwonem and the reference samples indicating two genetic clusters ($K = 2$). The *grey* and *black* colours represent proportional assignment to each cluster, the *black vertical line* separate the corresponding groups of populations. Evidence of admixture ($P < 0.01$) was found only across the samples from the sympatric population of the species from Bór na Czerwonem including four trees from the group of samples preliminarily classified based on phenotypic traits as polycormic pines (BOR_PC), five from *P. mugo*-like (BOR_M) and one from *P. sylvestris*-like (BOR_S) individuals

Selection

An excess of singleton mutations as compared to expectations under the standard neutral model was detected by significantly negative multilocus Tajima's *D* only in Scots pine from the reference populations of the species ($D = -0.618$ to -0.810, $P < 0.05$) (Table 2). At individual loci contrasting values of Tajima's *D* were found at Pr4-12 with significantly negative values for *P. sylvestris* ($D = -2.046$, $P < 0.01$) from Bór na Czerwonem vs. significantly positive value for *P. mugo* from that area ($D = 1.724$, $P < 0.05$). At Pr4-17 Tajima's *D* was significantly negative in ten hybrid individuals ($D = -1.667$, $P < 0.05$). Both compound neutrality tests provided evidence on selection at locus Pr4-12 ($P < 0.01$) and DHEW test at Pr4-21 ($P < 0.05$) in *P. sylvestris* from Bór na Czerwonem. Evidence on selection was also found at locus Pr4-4 in polycormic pines from that area in HEW test ($P < 0.05$).

In a multilocus HKA test, overall positive correlation between intraspecific polymorphism and interspecific divergence to the outgroup species at eight loci was found in all defined groups of samples including hybrids. Hybrids showed significant differentiation to *P. mugo* samples in the allelic frequency spectra at two loci including Pr4-5 and Pr4-19 and at one locus (Pr4-10) as compared to *P. sylvestris*. In the group of hybrids, alleles specific and observed only in the allopatric populations of *P. sylvestris* were found at eight samples at locus Pr4-5 and haplotypes specific to *P. mugo* at seven samples at locus Pr4-10. The remaining alleles at those two loci were common for both parental species. The group of hybrids showed no differentiation ($P < 0.01$) to any of the parental species at five loci (including Pr4-4, Pr4-12, Pr4-17, Pr4-21, Pr4-27). There was clear differentiation between *P. mugo* and polycormic pines vs. *P. sylvestris* at most loci (Supplementary Table 2). *P. mugo* and polycormic pines from Bór na Czerwonem showed significant variation to some reference populations of *P. mugo* at four loci (Pr4-5, Pr4-17, Pr4-19, Pr4-27). No evidence of differentiation was found between *P. sylvestris* from Bór na Czerwonem and the reference populations of the species.

Discussion

In our research, nuclear gene loci were sequenced and analysed for intra- and interspecific nucleotide variation in a panel of individuals derived from the contact zone and the allopatric reference stands of the two pine species. The aim was to evaluate the role of introgressive hybridization and selection on nucleotide diversity patterns of the analysed population. High genetic identity was observed between most samples from the group of oligo- and polycormic pines and *P. mugo* from Bór na Czerwonem reserve as evident from very similar nucleotide diversity ($\pi_{tot} = \sim 0.004$; $\theta_{tot} = 0.004$), non-existing net divergence and no significant differentiation in the allelic frequency spectra at all polymorphic sites combined and most individual loci. Those two groups also formed a uniform genetic cluster in a Bayesian mixture analysis, showed marginal divergence (0.0002–0.0006) to the reference *P. mugo* populations and shared higher proportion of haplotypes and SNPs as compared to the monocormic pines from

http://rcin.org.pl

that area classified as *P. sylvestris*. In contrast, *P. sylvestris* from Bór na Czerwonem showed a high genetic similarity to the reference *P. sylvestris* populations. This genetic similarity of the corresponding groups of samples to the allopatric populations of the species indicates that the majority of analysed individuals from that area represent pure *P. mugo* and *P. sylvestris* samples. In the previous studies, the variety of morphological forms observed on this area was explained in biometric and biochemical studies as either the result of intensive hybridisation and introgression that changed the population into a hybrid swarm (Bobowicz 1990) or as a mixture of mostly pure pine species from the *P. mugo* complex and *P. sylvestris*, which phenotypes were influenced by specific growing conditions of the peatbog environments (Odrzykoski 2002). As the polymorphism at the genomic regions used in our study clearly distinguishes both putative parental species, our genetic data support the suggestion that exceptional morphology of some oligo- and polycormic individuals from peatbog populations may be due to environmental variation but they most likely represent *P. mugo* (Wachowiak et al. 2006).

However, in addition to pure species growing on this peatbog, we detected ten individuals in total that clearly result from admixture between *P. mugo* and *P. sylvestris*. The majority of hybrid individuals were identified in a group of samples classified initially based on phenotypic traits as *P. mugo* and/or oligo- and polycormic trees except one monocormic individual classified based on phenotypic assessments as *P. sylvestris*. Therefore, our preliminary phenotypic classification of the samples based on some basic biometric traits failed to distinguish hybrids. All the hybrids had *cp*DNA of *P. mugo* and they contained a mixture of nuclear gene haplotypes observed in the reference allopatric populations of both parental species. The only unique haplotype found in two hybrid trees resulted from a single point mutation. That group of hybrids showed closer genetic similarity to *P. sylvestris* evident from the higher number of specific *P. sylvestris* alleles at the loci and lower net divergence. Previous nucleotide diversity studies in pines indicated a high intragenic recombination rate (González-Martinez et al. 2006; Wachowiak et al. 2009). Considering the genetic composition of hybrids and lack of recombining genotypes, it seems that those trees most likely represent first generation hybrids with *P. sylvestris* as a maternal species.

Our results correspond with some previous observations. Barriers against interspecific hybridisation and no evidence of bidirectional gene flow between *P. sylvestris* and *P. mugo* were suggested in some previous research that indicated hybrid seeds derived only from *P. sylvestris*-like individuals pollinated with *P. mugo* but not from reciprocal crossings (Wachowiak et al. 2005b). Lack of hybrids resulting from hybridization between *P. mugo* as a maternal and *P. sylvestris* as a paternal tree and putative hybrid individuals from reverse crossing combinations were found based on a joint analysis of *cp*DNA, izozymes and phenotypic characteristics of trees (Wachowiak and Prus-Głowacki 2008). So far, the only evidence of reciprocal hybridization was found in a sympatric population of *P. sylvestris* and peatbog pine (*Pinus uliginosa* Neumann), a taxon from the *P. mugo* complex (Wachowiak et al. 2005a). Analyses of the genetic composition of seeds derived from hybrid trees would be useful to assess other possible hybridization and/or introgression trajectories of those individuals. However, the presence of hybrid embryos would not necessarily mean that such hybrids succeed and exist in peatbog environments, as far as we can conclude from our results. It will also be necessary to grow hybrid seedlings to look at the phenotypic variation and underlying genetic variability of morphological forms. Our results suggest that the first generation hybrids may express extreme phenotypic variability as compared to parental species.

Our study provides evidence on selection at some of the analysed loci. Natural selection can cause fixation of advantageous alleles that have a positive fitness effect and potential to speed up adaptation in new genetic background of hybrids (De Carvalho et al. 2010; Martin et al. 2006). Two loci in our hybrid dataset showed increased frequency towards alleles specific to *P. sylvestris* at calcium-dependent protein kinase (Pr4-5) and alleles specific to *P. mugo* at mys transcription factor (Pr4-10). Such increase of frequency of alleles unique to one of the parental species and not observed at other loci suggests that they are under selection in the hybrids' genetic background and potentially increase their fitness in a peatbog environment. In the case of parental species, strong directional selection at some loci due to local adaptation in ecologically diverged peatbog environments should increase differentiation between the peatbog and the reference allopatric populations of the species as a result of selection for different alleles in different populations. In presence of no population structure within parental species observed in our dataset, significant difference in the allelic frequency spectra was found at a few loci in *P. mugo*. For instance, at calcium-dependent protein kinase (Pr4-5) and cytochrome P450 reductase (Pr4-17), only a subset of alleles (two in each case) was found in *P. mugo* samples as compared to the reference allopatric populations of the species. In contrast, no evidence of allelic frequency difference to the reference populations was found across *P. sylvestris* samples. However, both Scots pine samples from the hybrid zone and the reference populations showed evidence on selection at two loci including proton myo-inositol transporter and receptor protein kinase (Pr4-12 and Pr4-21) in compound neutrality tests. This departure from neutrality

http://rcin.org.pl

most likely reflects the species-wide pattern of selection at the genes in the European range that, however, cannot be directly linked to adaptive variation in peatbog environments. Our study reports a set of new genes with patterns of selection in the hybrid zone of two closely related pine species that contribute to so far a few such loci detected in pines (e.g. Eveno et al. 2008; Kujala and Savolainen 2012; Wachowiak et al. 2009).

Conclusions

Polymorphisms at the analysed genomic regions can discriminate both studied pine species. These polymorphisms could be used for tracking interspecific gene flow and evaluation of species composition in other contact zones of the species where individuals with mixed morpho-anatomical characteristics were described [e.g. in the Alps (Christensen 1987), Rila Mts. (Yurukov and Tashev 1992)]. Our study shows that the examined contact zone includes the majority of pure parental species individuals and some proportion of hybrids (∼17 %). Considering the species composition and environmental gradients not optimal for either of the parental species, the investigated and potentially similar hybrid zones seem suitable to study the influence of a local habitat on natural selection at the genes involved in local adaptation of hybrids and parental species from contrasting environments. We identified several genes that may be under natural selection as evident from the pattern of nucleotide polymorphisms in the samples from the hybrid zone and the reference parental populations. Our study shows that it will be possible to select a suitable mapping population of a sufficient size both in hybrid and parental species for admixture mapping to effectively genotype and search for polymorphisms at many genomic regions that may play role in species adaptive variation and speciation.

Acknowledgments The research was financially supported by the Polish National Science Centre (Grant No. 2011/01/B/NZ8/01634). WBŻ acknowledge financial support from Polish National Science Centre (Grant No. DEC-2012/05/E/NZ9/03476). We thank Euforgen network for providing distribution map of Scots pine.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

Arnold ML, Martin NH (2010) Hybrid fitness across time and habitats. Trends Ecol Evol 25:530–536

Arnold ML, Bouck AC, Scott Cornman R (2004) Verne Grant and Louisiana Irises: is there anything new under the sun? New Phytol 161:143–149. doi:10.1046/j.1469-8137.2003.00856.x

Baack EJ, Rieseberg LH (2007) A genomic view of introgression and hybrid speciation. Curr Opin Genet Dev 17:513–518. doi:10.1016/j.gde.2007.09.001

Bobowicz MA (1990) Hybrids between *Pinus mugo* Turra × *Pinus sylvestris* L. from ''Bór na Czerwonem'' reserve in Novotarska Valey [in Polish]. Wydawnictwo Naukowe UAM, Poznań

Christensen KI (1987) A morphometric study of the *Pinus mugo* Turra complex and its natural hybridization with *P. sylvestris* L. (*Pinaceae*). Feddes Repert 98:623–635. doi:10.1002/fedr.19870981113

Corander J, Tang J (2007) Bayesian analysis of population structure based on linked molecular information. Math Biosci 205:19–31

De Carvalho D et al (2010) Admixture facilitates adaptation from standing variation in the European aspen (*Populus tremula* L.), a widespread forest tree. Molec Ecol 19:1638–1650. doi:10.1111/j.1365-294X.2010.04595.x

Ersoz ES, Wright MH, González-Martínez SC, Langley CH, Neale DB (2010) Evolution of disease response genes in loblolly pine: insights from candidate genes. PLoS One 5:e14234. doi:10.1371/journal.pone.0014234

Eveno E et al (2008) Contrasting patterns of selection at *Pinus pinaster* Ait. Drought stress candidate genes as revealed by genetic differentiation analyses. Molec Biol Evol 25:417–437. doi:10.1093/molbev/msm272

Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evol Bioinform Online 1:47–50

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

Gross BL, Rieseberg LH (2005) The ecological genetics of homoploid hybrid speciation. J Heredity 96:241–252. doi:10.1093/jhered/esi026

Hudson RR, Kreitman M, Aguade M (1987) A test of neutral molecular evolution based on nucleotide data. Genetics 116:153–159

Jiggins CD, Salazar C, Linares M, Mavarez J (2008) Hybrid trait speciation and *Heliconius* butterflies. Philos Trans, Ser B 363:3047–3054. doi:10.1098/rstb.2008.0065

Kujala S, Savolainen O (2012) Sequence variation patterns along a latitudinal cline in Scots pine (*Pinus sylvestris*): signs of clinal adaptation? Tree Genet Genomes 8:1451–1467

Martin NH, Bouck AC, Arnold ML (2006) Detecting adaptive trait introgression between *Iris fulva* and *I. brevicaulis* in highly selective field conditions. Genetics 172:2481–2489. doi:10.1534/genetics.105.053538

Minder AM, Widmer A (2008) A population genomic analysis of species boundaries: neutral processes, adaptive divergence and introgression between two hybridizing plant species. Molec Ecol 17:1552–1563. doi:10.1111/j.1365-294X.2008.03709.x

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Odrzykoski IJ (2002) Genetic variation study of Dwarf mountain pine (*Pinus mugo*) with the use of molecular and biochemical markers [in Polish]. Wydawnictwo Naukowe UAM, Poznań

Pyhäjärvi T, Salmela MJ, Savolainen O (2008) Colonization routes of *Pinus sylvestris* inferred from distribution of mitochondrial DNA variation. Tree Genet Genomes 4:247–254

Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. Pl Molec Biol 17:1105–1109. doi:10.1007/bf00037152

http://rcin.org.pl

Tajima F (1989) Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molec Biol Evol 28:2731–2739. doi:10.1093/molbev/msr121

Wachowiak W, Prus-Głowacki W (2008) Hybridisation processes in sympatric populations of pines *Pinus sylvestris* L., *P. mugo* Turra and *P. uliginosa* Neumann. Pl Syst Evol 271:29–40

Wachowiak W, Leśniewicz K, Odrzykoski I, Augustyniak H, Prus-Głowacki W (2000) Species specific cpDNA markers useful for studies on the hybridization between *Pinus mugo* and *P. sylvestris*. Acta Soc Bot Poloniae 69:273–276

Wachowiak W, Celiński K, Prus-Głowacki W (2005a) Evidence of natural reciprocal hybridisation between *Pinus uliginosa* and *P. sylvestris* in the sympatric population of the species. Flora 200:563–568. doi:10.1016/j.flora.2005.06.007

Wachowiak W, Lewandowski A, Prus-Głowacki W (2005b) Reciprocal controlled crosses between *Pinus sylvestris* and *P. mugo* verified by a species-specific cpDNA marker. J Appl Genet 46:41–43

Wachowiak W, Odrzykoski I, Myczko Ł, Prus-Głowacki W (2006) Lack of evidence on hybrid swarm in the sympatric population of *Pinus mugo* and *P. sylvestris*. Flora 201:307–316. doi:10.1016/j.flora.2005.08.004

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

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. Theor Popul Biol 7:256–276

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370

Yurukov S, Tashev A (1992) Studies of natural hybrids between Scots pine (*Pinus sylvestris* L.) and Mountain pine (*Pinus mugo* Turra) in the South-east Rila Mts. Nauka za Gorata 29:39–43

Zeng K, Shi S, Wu C-I (2007) Compound tests for the detection of hitchhiking under positive selection. Molec Biol Evol 24:1898–1908

http://rcin.org.pl

Supplementary material

Supplementary Table 1. Analysed loci.

Locus Acronym	PCR Primer F	PCR Primer R	Gene description	Gene Bank Acc. Number[3]
Pr4-4[1]	TGTCACTGCCCAGAGCTATTC	ATCACAGCCGCTCCAAAAC	putative aquaporin	KM277840-277848 KC980645-980675
Pr4-5[1]	CATCTCCTTCAAACCTCTTATTTCC	GATGCTTGAACATGATCCC	calcium-dependent protein kinase	KM277849-277852 KC980680-980694
Pr4-10[1]	CATTGCCTACGATTTCC	CTTTTGAGATGAACCAGAC	myc transcription factor	KM277853-277855 KC980696-980701
Pr4-12[1]	CTGCTCAAGTGAAAGG	CTGATTGTGGATTCTGTG	proton myo-inositol transporter	KM277856-277861 KC980725-980743
Pr4-17[1]	CTGGAAGCTGATTCTTTTG	CCTCTAGTTCTGGTTG	cytochrome P450 reductase	KM277862-277867 KC980746 - 980761
Pr4-19[1]	CTCTACCACATCATCTCC	TTTCACTCTCGTGTCTTTCACC	laccase	KM277868-277872 KC980770 - 980781
Pr4-21[1]	ACATGGTGTTTGGCAGG	AATGAGGAGGGTGGTAGAG	receptor protein kinase	KM277873-277885 KC980786 - 980807
Pr4-27[1]	TAGCAGACGGTATTCACACAGTCC	CCACAACCACCTTGCATCATTATTT	putative auxin induced - transcription factor	KM277886-277893 KC980811 - 980828
trnF-trnL[2]	CGAAATCGGTAGACGCTACG	ATTTGAACTGGTGACACGAG	intergenic region of cpDNA	AF543753 - 543754

Primers derived from [1] Ersoz ES, Wright MH, González-Martínez SC, Langley CH, D.B. N (2010) Evolution of Disease Response Genes in Loblolly Pine: Insights from Candidate Genes. PLoS One 5 (12):e14234. doi:10.1371/journal.pone.0014234; [2] Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Molecular Biology 17 (5):1105-1109. doi:10.1007/bf00037152. [3] Accession numbers of the haplotype sequence data of the loci at the NCBI repository (http://www.ncbi.nlm.nih.gov).

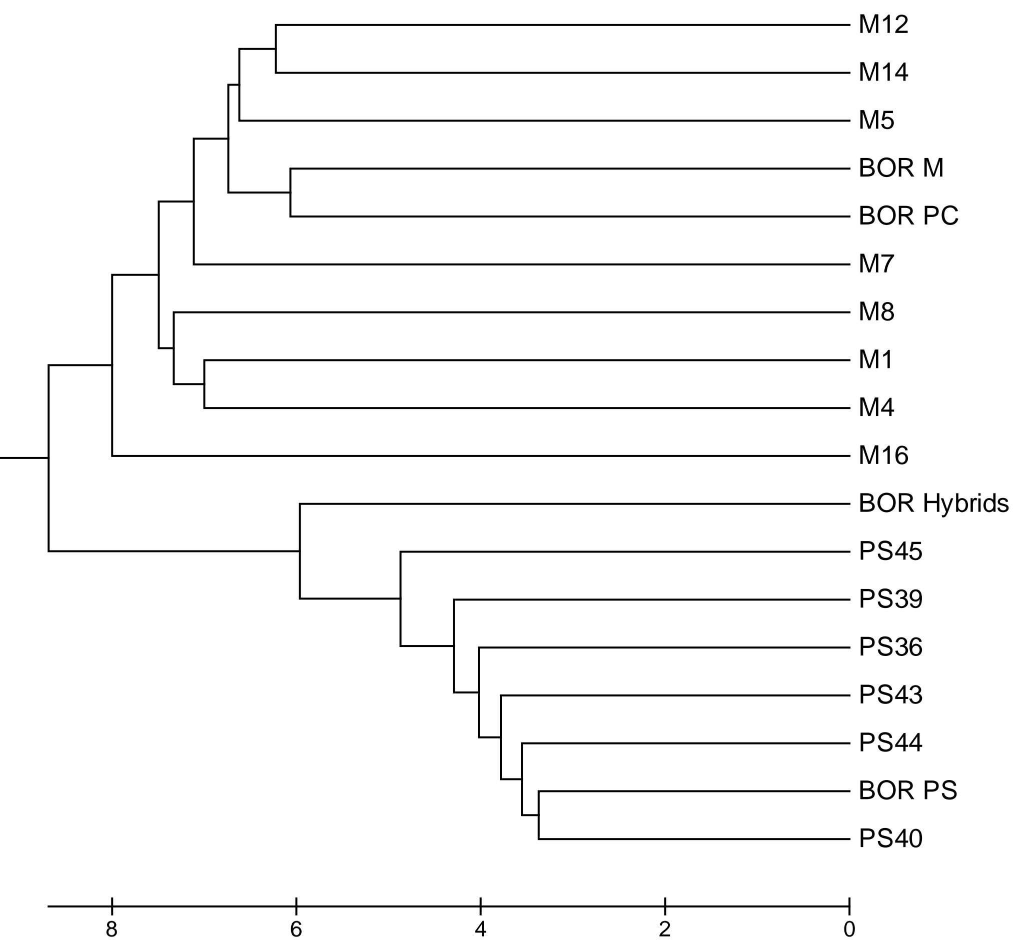
http://rcin.org.pl

Supplementary Table 2. Significant allelic frequency differentiation (F_{ST}, $p<0.01$) between defined groups of samples at individual loci (marked in superscripts). Any differentiation was found between the reference *P. sylvestris* groups (EU_C, FIN, SWE).

	BOR_M	BOR_S	BOR_PC	BOR_Hybrids	M1_M12	M4_M5	M7_M8	M14_16
BOR_S	0.361[Pr4_4],0.330[Pr4_5] 0.661[Pr4_10],0.522[Pr4_12] 0.452[Pr4_19]							
BOR_PC		0.269[Pr4_4], 0.549[Pr4_10] 0.589[Pr4_12] 0.441[Pr4_19]						
BOR_Hybrids	0.468[Pr4_5]	0.580[Pr4_10]						
M1_M12	0.235[Pr4_17]	0.276[Pr4_4], 0.656[Pr4_5] 0.784[Pr4_10], 0.317[Pr4_12] 0.288[Pr4_17] 0.602[Pr4_19]	0.202[Pr4_17]	0.821[Pr4_5], 0.305[Pr4_19]				
M4_M5		0.271[Pr4_4], 0.611[Pr4_5] 0.747[Pr4_10], 0.327[Pr4_12] 0.215[Pr4_17],0.559[Pr4_19]	0.133[Pr4_19]	0.749[Pr4_5], 0.276[Pr4_19]				
M7_M8	0.228[Pr4_27]	0.219[Pr4_4], 0.487[Pr4_5] 0.866[Pr4_10], 0.263[Pr4_12] 0.235[Pr4_17], 0.499[Pr4_19] 0.162[Pr4_27]	0.182[Pr4_17] 0.237[Pr4_27]	0.604[Pr4_5], 0.209[Pr4_19]				
M14_16	0.228[Pr4_5] 0.159[Pr4_17]	0.200[Pr4_4], 0.606[Pr4_5] 0.891[Pr4_10], 0.396[Pr4_12] 0.222[Pr4_17] 0.507[Pr4_19]	0.219[Pr4_5]	0.726[Pr4_5]	0.318[Pr4_5]		0.164[Pr4_5]	
EU_C	0.171[Pr4_4] 0.668[Pr4_10] 0.512[Pr4_12] 0.446[Pr4_19] 0.165[Pr4_21]		0.157[Pr4_4] 0.557[Pr4_10] 0.578[Pr4_12] 0.433[Pr4_19]	0.590[Pr4_10]	0.188[Pr4_4], 0.548[Pr4_5] 0.789[Pr4_10], 0.314[Pr4_12] 0.240[Pr4_17], 0.596[Pr4_19] 0.122[Pr4_27]	0.184[Pr4_4], 0.511[Pr4_5] 0.752[Pr4_10], 0.324[Pr4_12] 0.166[Pr4_17], 0.556[Pr4_19] 0.161[Pr4_21], 0.129[Pr4_27]	0.145[Pr4_4], 0.388[Pr4_5] 0.870[Pr4_10], 0.262[Pr4_12] 0.200[Pr4_17], 0.497[Pr4_19] 0.170[Pr4_27]	0.124[Pr4_4], 0.516[Pr4_5] 0.895[Pr4_10], 0.392[Pr4_12] 0.197[Pr4_17], 0.500[Pr4_19] 0.104[Pr4_21], 0.125[Pr4_27]
FIN	0.251[Pr4_4] 0.379[Pr4_5] 0.606[Pr4_10] 0.412[Pr4_12] 0.459[Pr4_19] 0.194[Pr4_27]		0.213[Pr4_4] 0.314[Pr4_5] 0.497[Pr4_10] 0.485[Pr4_12] 0.447[Pr4_19] 0.232[Pr4_27]	0.514[Pr4_10]	0.233[Pr4_4], 0.701[Pr4_5] 0.727[Pr4_10], 0.220[Pr4_12] 0.200[Pr4_17], 0.607[Pr4_19] 0.223[Pr4_27]	0.229[Pr4_4], 0.653[Pr4_5] 0.709[Pr4_10], 0.238[Pr4_12] 0.565[Pr4_19], 0.220[Pr4_27]	0.184[Pr4_4], 0.529[Pr4_5] 0.822[Pr4_10], 0.169[Pr4_12] 0.169[Pr4_17], 0.506[Pr4_19] 0.255[Pr4_27]	0.167[Pr4_4], 0.644[Pr4_5] 0.838[Pr4_10], 0.295[Pr4_12] 0.186[Pr4_17], 0.511[Pr4_19] 0.219[Pr4_27]
SWE	0.266[Pr4_4] 0.609[Pr4_10] 0.534[Pr4_12] 0.426[Pr4_19]		0.211[Pr4_4] 0.494[Pr4_10] 0.598[Pr4_12] 0.412[Pr4_19]	0.511[Pr4_10]	0.232[Pr4_4], 0.325[Pr4_5] 0.752[Pr4_10], 0.344[Pr4_12] 0.245[Pr4_17], 0.577[Pr4_19]	0.221[Pr4_4], 0.289[Pr4_5] 0.711[Pr4_10], 0.346[Pr4_12] 0.536[Pr4_19]	0.174[Pr4_4], 0.844[Pr4_10] 0.286[Pr4_12], 0.482[Pr4_19] 0.136[Pr4_27]	0.167[Pr4_4], 0.326[Pr4_5] 0.873[Pr4_10], 0.415[Pr4_12] 0.181[Pr4_17], 0.491[Pr4_19]

http://rcin.org.pl

Supplementary Figure 1. Estimates of evolutionary divergence over sequence pairs between sympatric stand of *P. mugo* and *P. sylvestris* and the reference allopatric populations. The number of base differences per sequence from averaging over all sequence pairs between groups are shown. The analysis involved 146 positions in the final dataset after removal of ambiguous positions for each sequence pair. M – *Pinus mugo*, PS – *Pinus sylvestris*, BOR – groups of samples defined at "Bór na Czerwonem" peat-bog (see Material and Methods for details).



http://rcin.org.pl

Publication 6

Wachowiak W, **Żukowska WB**, Wójkiewicz B,
Cavers S, Litkowiec M

Hybridization in contact zone between temperate European pine species

Tree Genetics & Genomes 12: 48

Impact factor: 2.132

MNiSW points: 35

http://rcin.org.pl

Tree Genetics & Genomes (2016) 12: 48
DOI 10.1007/s11295-016-1007-x

ORIGINAL ARTICLE

Hybridization in contact zone between temperate European pine species

Witold Wachowiak[1,2] · **Weronika B. Żukowska**[1] · **Błażej Wójkiewicz**[1] · **Stephen Cavers**[3] · **Monika Litkowiec**[1]

Received: 27 August 2015 / Revised: 13 April 2016 / Accepted: 15 April 2016 / Published online: 4 May 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract Hybridization studies are important to advance our understanding of the interspecific gene flow and its evolutionary consequences in closely related species. Hybridization and admixture patterns were assessed in a contact zone and reference populations of European pine species using sequence data from 26 nuclear genes and a species-diagnostic cpDNA marker. Reference populations formed three distinct genetic clusters comprising *Pinus sylvestris*, *Pinus mugo*/*Pinus uliginosa*, and *Pinus uncinata*. Evidence of population structure was found only in *P. uliginosa*. Based on phenotypic characteristics and molecular data, we identified five groups of individuals in the contact zone in Poland, comprising forms of the parental species and intermediates that were most probably the result of interspecific crosses. A combination of nuclear gene sequence data and a diagnostic organelle marker were used to show that hybridization is frequent in the contact zone and results in hybrid trees with distinct phenotypic identity. The influence of selection in maintaining hybrid phenotypes in environments unsuited to parental species was inferred from nucleotide polymorphism data. A lack of admixture in reference populations suggests that hybridization has not occurred during post-glacial migration and so the contact zone represents a distinct, active example of ongoing evolution. Pine populations in this zone will be a valuable system for studying the genetic basis of hybrid advantage in environmental conditions untypical of pure parental species.

Keywords Hybridization · Molecular markers · Natural selection · Local adaptation · Speciation

Communicated by F. Gugerli

Electronic supplementary material The online version of this article (doi:10.1007/s11295-016-1007-x) contains supplementary material, which is available to authorized users.

✉ Witold Wachowiak
witoldw@man.poznan.pl

[1] Institute of Dendrology, Polish Academy of Sciences, Parkowa 5, 62-035 Kórnik, Poland

[2] Institute of Environmental Biology, Faculty of Biology, Adam Mickiewicz University, Umultowska 89, 61-614 Poznań, Poland

[3] Centre for Ecology and Hydrology Edinburgh, Bush Estate, Penicuik Midlothian EH26 0QB, UK

Introduction

Natural hybridization creates recombinants from interspecific mating between divergent parental taxa when they come into geographic contact (Arnold and Martin 2010; Lexer and Widmer 2008). Hybridization has various evolutionary consequences for the taxa involved (Baack and Rieseberg 2007). For instance, it may cause swamping of the species with the smaller effective population size by gene flow from the more abundant species, integration of genetic material from one species into another through repeated back-crossing (introgression), homoploid hybrid speciation—in which the new hybrid lineages become reproductively isolated from parental populations, or transfer of adaptive traits across species boundaries (Baack and Rieseberg 2007). Well-documented examples show that hybrid genotypes may have equivalent or higher fitness relative to parental species due to environmental selection (Andrew and Rieseberg 2013; Rieseberg et al. 2007; Whitney et al. 2010). Even where hybrid fertility or viability is reduced in early generations, gene flow can nevertheless allow propagation of hybrids and adaptive divergence (Gross and Rieseberg 2005). Contact among species during major population movements or range shifts, such as those associated with the post-glacial recolonization of

http://rcin.org.pl

Europe, may therefore have played an important role in the evolution of closely related contemporary species.

Hybridization was postulated to have occurred between the closely related Scots pine (*Pinus sylvestris* L.) and taxa from the *Pinus mugo* complex including dwarf mountain pine (*Pinus mugo* Turra), peat bog pine (*Pinus uliginosa* Neumann), and mountain pine (*Pinus uncinata* Ramond ex DC.) (Jasińska et al. 2010; Wachowiak and Prus-Głowacki 2008). These taxa differ from each other in phenotype, geographical distribution, and ecology, in particular for traits related to dehydrative stress and temperature (Critchfield and Little 1966). *Pinus sylvestris* is the most widespread and economically important forest tree species in Europe and Asia and shows adaptive variation in response to environmental gradients, e.g., in timing of bud set and cold hardiness (González-Martínez et al. 2006; Hurme et al. 2000). *Pinus mugo* is an endemic species typical of the mountainous regions of Europe (Critchfield and Little 1966). *Pinus uliginosa* was described in the Central Sudetes where it grows mainly on peat bogs, and *Pinus uncinata* is adapted to mountain environments and is most abundant in the Alps and Pyrenees. The latter, together with *P. mugo* and *P. uliginosa*, comprise the *P. mugo* complex (Businsky and Kirschner 2006; Christensen 1987b). The present distribution of these species and *P. sylvestris* is mostly allopatric and results from postglacial migration and range shifts following changes in environmental conditions and competition with other forest tree species. However, recolonization created zones of contact between species as their ranges overlapped in some parts of their distributions. At present, contact zones exist in several places in the mountains of southern and central Europe (Christensen 1987a). As the species were interfertile in controlled crosses (Wachowiak et al. 2005a), natural hybridization has been suggested as a main driver of phenotypic divergence of several taxa described within the *P. mugo* complex (Christensen 1987a). By comparing patterns of genetic variation and admixture in contact zones and in reference populations, the role of introgressive hybridization and ecological selection on adaptive divergence can be assessed. Such studies should also allow the investigation of the genetic consequences of hybridization and the potential use of hybrid zones for admixture mapping of adaptively important traits.

Here, we tested for evidence of interspecific admixture and selection in a contact zone between Scots pine and the taxa from the *P. mugo* complex. Using nucleotide polymorphism data from nuclear and chloroplast genomes, we compared genetic variation at intra- and interspecific level in the contact zone and in reference populations of the species across Europe. We quantified admixture outside the contact zone to test whether historic interspecific gene flow has played a role in species divergence during postglacial recolonization and to evaluate the role of introgressive hybridization and selection on patterns of genetic divergence and evolution in the focal species.

Materials and methods

Experimental design and sample grouping

Samples from a contact zone and 30 reference populations of *P. sylvestris* and the taxa from the *P. mugo* complex from their European distribution were used in the study (Fig. 1, Table 1). Previous biometric studies suggested the existence of relict *P. uncinata* populations in the Silesian Lowlands. Those locations are geographically distant from the contemporary range of the species (Marcysiak and Boratyński 2007) but close to the contact zone. Therefore, we included *P. uncinata* in our analysis.

The contact zone comprised sympatric populations of the pine species at the Zieleniec reserve which is the largest peat bog complex in the Sudety Mountains, southwest Poland. For taxonomic classification of trees in the contact zone, we used several phenotypic traits, i.e., growth form, bark color of the upper part of stem and main branches, color and shape of needles, and setting angle of conelet from the previous year (when available) (Christensen 1987b). Based on these traits, we identified pure bushy *P. mugo* individuals (PMZ), monocormic (single-stemmed) *P. sylvestris* individuals (PSZ), mono- and oligocormic (one-three stemmed) *P. uliginosa* individuals (PUZ), and a group of oligo- and polycormic (multi-stemmed) *P. uliginosa*-like pines of a height of over 2 m and untypical morphology that could not be classified as either of parental species (HBZ). Needles from 40 to 60 trees representing each phenotypic group were sampled. The samples were collected over about 50 ha. Trees of different phenotypes were intermixed across the area and the distance between sampled trees was from 5 to 100 m.

The sampled trees were screened with a diagnostic chloroplast DNA marker (cpDNA) from the *trn*F–*trn*L region (Wachowiak et al. 2000). This PCR-RFLP marker (using *Dra*I as a restriction enzyme) is paternally inherited and dispersed through pollen and seeds (Wachowiak et al. 2006a). It was previously developed based on a single-nucleotide polymorphism that leads to an undigested PCR product for *P. sylvestris* (S marker) and a digested PCR product for *P. mugo* (M marker) giving two bands after electrophoresis on agarose gel. Analysis of many populations from across the distribution range of the pine species in Europe confirmed that this marker discriminates *P. sylvestris* from taxa of the *P. mugo* complex (Wachowiak et al. 2000). PCR-RFLP analysis of the samples from Zieleniec reserve indicated the presence of the M marker in all *P. mugo*, *P. uliginosa*, and the group of taxonomically unclassified polycormic (multi-stemmed) *P. uliginosa*-like pines of untypical morphology (HBZ). However, the M marker was also found in some individuals that had been initially classified as *P. sylvestris*. Therefore, we distinguished another group of samples at Zieleniec reserve, namely monocormic *P. sylvestris*-like individuals carrying cpDNA diagnostic for the *P. mugo* complex (HPSZ) (Table 1).

http://rcin.org.pl

A subset of 16 to 20 individuals from each group was selected for molecular analyses at 26 nuclear gene loci (Supplementary Table S1). We compared levels of nucleotide variation in the contact zone versus reference populations from allopatric zones of the species. The reference populations were treated either separately or were grouped by geographical location to compare a similar number of samples relative to the putative hybrid zone. *Pinus sylvestris* reference populations were grouped into Northern Europe (Finland and Sweden), Central Europe (Poland and Austria), Southern Europe (Spain), and Northwest Europe (Scotland; Table 1). *Pinus mugo* reference populations were grouped into Central Europe (Sudetes and Alps), Carpathians (Eastern and Southern), and Balkans (Pirin and Durmitor Mts.), and Italy (Carnic Alps and Abruzzi). Seeds from at least ten trees separated at a distance of a minimum 50 m were sampled from each reference population providing a set of 384 trees (Table 2). Using the QIAGEN DNeasy Plant Mini Kit, DNA was extracted from either haploid megagametophyte of seeds (samples from reference populations) or from needles (samples from the contact zone).

PCR amplification and sequencing

Twenty-six nuclear gene loci related to regulation of gene expression, metabolism, signal transduction, and transport were analyzed (Supplementary Table S1). PCR primers are given in Supplementary Table S1. PCRs were performed with Thermo MBS thermal cyclers and carried out in a total volume of 15 μl containing 15 ng of template DNA, 10 μM of each dNTP, 0.2 μM each of forward and reverse primers, 0.15 U *Taq* DNA polymerase, 1× BSA, 1.5 μM of $MgCl_2$, and 1× PCR buffer (BioLabs). Standard amplification procedures were used with initial denaturation at 94 °C for 3 min followed by 35 cycles with 30 s denaturation at 94 °C, 30 s annealing at 60 °C and 1 min 30 s extension at 72 °C, and a final 5 min extension at 72 °C. PCR fragments were purified using Exonuclease I-Shrimp Alkaline Phosphatase enzymatic treatment. About 20 ng of PCR products were used as templates in 10 μl sequencing reactions with the BigDye Terminator v3.1. DNA Sequencing Kit (Applied Biosystems) performed by Genomed S.A. (Warsaw, Poland) on a 3130xl Genetic Analyzer (Applied Biosystems). CodonCode Aligner software ver. 3.7.1 (Codon Code Corporation, Dedham, MA,

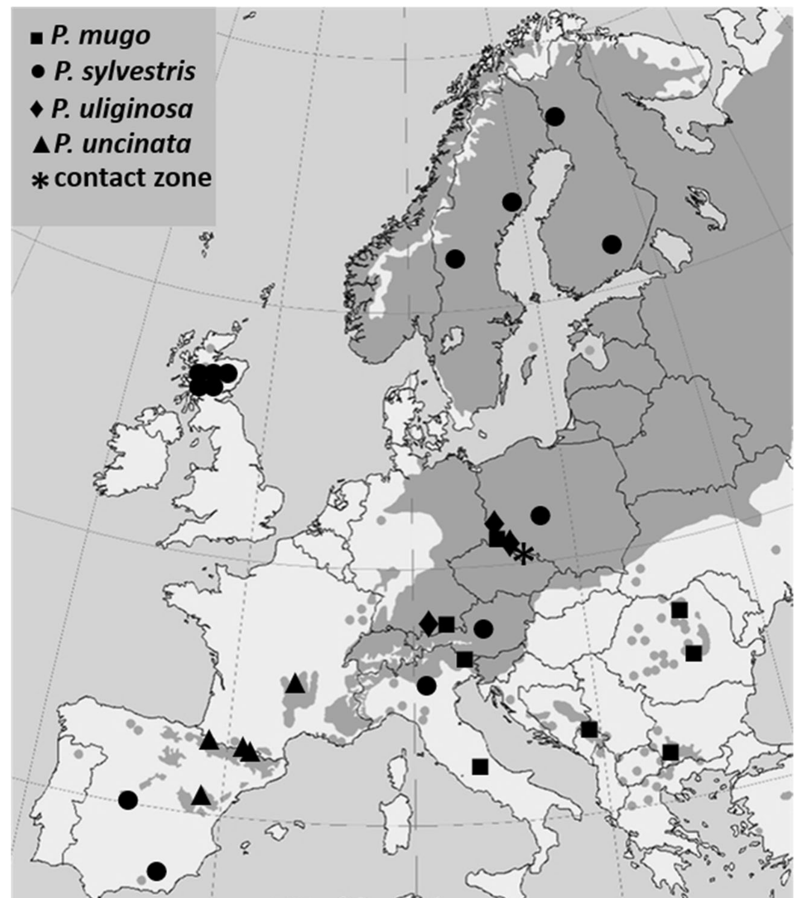


Fig 1 Geographical location of the contact zone of pine species from Zieleniec reserve and reference allopatric populations of *Pinus sylvestris* and the taxa from the *P. mugo* complex. Distribution range of *P. sylvestris* provided by Euforgen network is marked in *dark gray*

http://rcin.org.pl

Table 1 Location of populations from Zieleniec reserve and the four reference pine species

Acronym	Location	Long. [°E]	Lat. [°N]	Alt. [m]	Acronym	Location	Long. [°E]	Lat. [°N]	Alt. [m]
Contact zone					18. PS34	Scotland_Black Wood of Rannoch	−4.320	56.670	275
1. PMZ—*P. mugo*	Zieleniec reserve	16.425	50.341	760	19. PS36	Austria_Pernitz	16.000	47.914	500
2. HPSZ—*P. sylvestris*-like	Zieleniec reserve	16.425	50.341	760	20. PS37	Spain_Trevenque	3.547	37.096	1170
3. PSZ—*P. sylvestris*	Zieleniec reserve	16.425	50.341	760	21. PS38	Spain_Valsain	−4.035	40.865	1350
4. PUZ—*P. uliginosa*	Zieleniec reserve	16.425	50.341	760	22. PS39	Finland_Punkaharju	29.358	61.730	80
5. HBZ—*P. uliginosa*-like	Zieleniec reserve	16.425	50.341	760	23. PS40	Finland_Kolari	24.050	67.183	190
Pinus mugo					24. PS43	Poland_Jarocin	17.477	51.972	120
6. PM1	Poland_Śląskie Kamienie	15.600	50.783	1420	25. PS44	Sweden_Krp. Tjärnbergsheden	20.738	64.62	110
7. PM4	Romania_Carpathians E	24.800	47.567	1720	26. PS45	Sweden_Väster Mjöingenn	13.581	62.750	640
8. PM5	Romania_Carpathians S	25.452	45.432	2070	27. PS46	Italy, Cella di Palmia	10.165	44.633	180
9. PM7	Bulgaria_Pirin	23.423	41.769	2000	*Pinus uliginosa*				
10. PM8	Montenegro_Durmitor Mts.	19.091	43.159	2100	28. PUG1	Poland_Węgliniec reserve	15.227	51.294	190
11. PM12	Austria_Karwendel Mts.	11.296	47.378	1400	29. PUG2	Germany_Mittelwald	11.274	47.480	860
12. PM14	Italy_Carnic Alps	13.147	46.544	1300	30. PUG3	Poland_Batorów reserve	16.384	50.459	710
13. PM16	Italy_Abruzzi	13.975	41.772	2200	*Pinus uncinata*				
Pinus sylvestris					31. PUN17	Andorra_Val de Ransol	1.594	42.625	2025
14. PS30	Scotland_Shieldaig	−5.641	57.512	81	32. PUN18	Andorra_San Miguel de Engolasters	1.570	42.524	2000
15. PS31	Scotland_Glen Tanar	−2.856	57.056	160	33. PUN23	Spain_Castiello de Jaca	−0.537	42.636	1720
16. PS32	Scotland_Rothiemurcus	−3.770	57.150	318	34. PUN24	Spain_Sierra de Gudar	−0.606	40.391	2000
17. PS33	Scotland_Glen Affric	−4.920	57.270	256	35. PUN28	France_Col de la Croix de Morand	3.054	45.684	1400

Groups of populations defined at the Zieleniec reserve include *Pinus mugo* samples (PMZ), hybrid *P. sylvestris*-like pines with cp DNA of the *P. mugo* complex (HPSZ), *P. sylvestris* samples (PSZ), *Pinus uliginosa* samples (PUZ), and a group of *P. uliginosa*-like polycormic hybrids with cpDNA of the *P. mugo* complex (HBZ). Pure species reference samples include pines of *P. mugo* from Central Europe (PM1, PM12), the Carpathians (PM4, PM5), Balkans (PM7, PM8), and Italy (PM14, PM16); *P. sylvestris* from Northern Europe (PS39-40, PS44-45), Central Europe (PS36, PS43), Southern Europe (PS37-38), and Northwest Europe (PS30-34); *P. uliginosa* from Poland and Germany and *P. uncinata* from the Iberian Peninsula and Massif Central

http://rcin.org.pl

Table 2 Nucleotide and haplotype diversity of the groups of pines defined at Zieleniec reserve (Table 1) and reference populations of pure taxa of *Pinus mugo*, *P. sylvestris*, *P. uliginosa*, and *P. uncinata* (average values across 26 nuclear genes are reported)

Species/group of samples				Nucleotide polymorphisms				Haplotype diversity	
	N	n	L	SNPs	S	π_{total}	Tajima's D	N_h	H_d (SD)
Contact zone of the species									
PMZ—*P. mugo*	20	40	404	7.3	2.0	0.0044	−0.148	7.2	0.650 (0.054)
HPSZ—*P. sylvestris*-like	20	40	404	8.1	1.8	0.0055	0.206	7.7	0.710 (0.047)
PSZ—*P. sylvestris*	20	40	404	7.2	2.2	0.0039	−0.292	6.7	0.541 (0.062)
PUZ—*P. uliginosa*	20	40	404	8.2	1.9	0.0049	−0.164	8.1	0.695 (0.053)
HBZ—*P. uliginosa*-like	16	40	404	7.0	2.1	0.0045	−0.138	6.9	0.658 (0.058)
Reference populations									
PM_CE—*P. mugo*	30	30	449	6.9	1.8	0.0039	−0.039	5.4	0.609 (0.067)
PS_CE—*P. sylvestris*	20	20	448	7.1	2.9	0.0042	−0.301	5.3	0.625 (0.080)
PUG_BM—*P. uliginosa*	14	14	441	6.4	2.6	0.0044	−0.108	4.3	0.636 (0.099)
PUN_FR—*P. uncinata*	10	10	440	6.4	2.2	0.0053	0.267	4.2	0.704 (0.117)
Reference populations _All									
PM—*P. mugo*	79	79	450.9	10.7	3.6	0.0039	−0.536	9.0	0.603 (0.042)
PS—*P. sylvestris*	135	135	450.9	13.7	5.2	0.0042	−0.859	12.3	0.590 (0.031)
PUG—*P. uliginosa*	24	24	450.9	8.4	2.9	0.0046	−0.254	5.8	0.650 (0.073)
PUN—*P. uncinata*	50	50	450.9	9.5	2.0	0.0048	0.170	6.8	0.665 (0.049)

Reference populations: CE—Central Europe; BM—populations from Batorów reserve and Mittelwalde; FR—Col de la Croix de Morand (France) (Table 1)

N sample size, n number of gene copies analyzed at each locus, L length of sequence in base pairs, S number of segregating sites, π nucleotide diversity (Nei 1987), Tajima's D test (Tajima 1989), N_h number of haplotypes, H_d haplotype diversity (standard deviation)

USA) was used for editing of the chromatograms, visual inspection of all polymorphic sites detected, and alignments. GenBank reference sequences from *Pinus taeda* were used as an outgroup in neutrality tests (Supplementary Table S1). The haplotype phase (combined multi-locus polymorphism) of samples amplified from needles was determined using PHASE software and reference haplotype sequences of the pure species derived from DNA sequencing of the haploid megagametophyte tissue. Haplotype sequence data are deposited in the NCBI repository (Supplementary Table S1).

Nucleotide variation

We looked at the patterns of nucleotide and haplotype polymorphism among groups of samples from the contact zone, and between the samples from the contact zone and reference pure species populations. Nucleotide diversity was measured as the average number of differences per site (π) between two sequences (Nei 1987). The number of haplotypes (N_h) and haplotype diversity (H_d) were computed for each gene using DnaSP v.5 (Librado and Rozas 2009). The number and frequency of unique and shared haplotypes in pairwise comparisons between species was calculated with Arlequin v.3.5 (Excoffier and Lischer 2010). Locus-by-locus estimates of net divergence between the groups of samples (Nei 1987) were determined using SITES 1.1 (Hey and Wakeley 1997). Genetic relationships between reference pure species populations and groups of samples from the contact zone were assessed based on pairwise net average genetic distance at all polymorphic sites detected at 26 loci using Unweighted Pair Group Method with Arithmetic Mean (UPGMA) as implemented in Mega 6.06 (Tamura et al. 2013). Standard errors of the interior branches were calculated using 1000 bootstrap procedure and the Maximum Composite Likelihood model as implemented in Mega.

Population structure and taxonomic relationships

Population structure was studied using Bayesian clustering methods and analysis of molecular variance. In addition, we assessed genetic differentiation between groups of samples from the contact zone and the reference populations of the species. Clustering of individuals and populations and estimates of admixture were explored using BAPS 6.0 (Corander and Tang 2007). Ten independent runs were conducted for each K (1–35) to estimate the number of clusters for all samples combined, for the three taxa from the *P. mugo* complex (*P. mugo*, *P. uliginosa*, and *P. uncinata*), *P. sylvestris*, and samples from the contact zone based on all polymorphic sites. The linear linkage model was used, the

http://rcin.org.pl

number of iterations used to estimate admixture coefficients was set to 100, the number of reference individuals was set to 100, and the number of iterations used to estimate admixture coefficients for the reference individuals was 10. Genetic differentiation in pairwise comparisons between populations was measured as Wright's fixation index (Weir and Cockerham 1984) over all polymorphic sites and tested for significance by 1000 permutations of the samples between populations using Arlequin v.3.5 software (Excoffier and Lischer 2010). To further assess among-population differentiation, we used principal coordinate analysis (PCoA) based on the mean net genetic distance between populations at all polymorphic sites.

Tests for natural selection

We tested for the effects of selection on genes in parental species at both reference and contact zone populations. Tests were based on a comparison of allelic frequency spectra between different groups of populations, and departures from neutral expectations of polymorphisms versus divergence at the interspecific level. Deviations from the standard neutral model of evolution were assessed using the frequency spectrum test and coalescence-based approaches. Tajima's *D* (Tajima 1989) was computed using the difference between two distinct estimates of the scaled mutation parameter theta for each locus, and statistical significance was evaluated by a comparison to a distribution generated by 1000 coalescent simulations using DnaSP 5.1 (Librado and Rozas 2009). Species may undergo different population histories that can influence the assumptions of standard neutrality tests. Therefore, deviations from neutrality were also tested using two compound neutrality tests that are robust to demographic processes, HEW and DHEW (Zeng et al. 2007). The HEW and DHEW tests are a compound of Fay and Wu's *H* and Tajima's *D*/Fay and Wu's *H* with the Ewens-Watterson neutrality test, respectively (Zeng et al. 2007). Significance levels were determined by 10,000 coalescent simulations based on Watterson's estimator of theta as implemented in the *dh* package (http://zeng-lab.group.shef.ac.uk/wordpress/?page_id=28). The distribution of the test statistic was investigated for each locus in all samples combined from each species. Genetic differentiation in pairwise comparisons within and between species was studied locus by locus, and significance thresholds of the F_{ST} values were set at 99 % and estimated by 1000 permutations of the samples between populations, regional groups, and species using Arlequin v.3.5. The false discovery rate (FDR) adjustment for multiple testing (lambda = 0.15, FDR level = 0.01) was conducted using QVALUE software based on the distribution of *P* values for the set of F_{ST} statistics (Storey and Tibshirani 2003).

Results

Nucleotide and haplotype diversity

Across the 26 genes, ~10 kbp were aligned providing a set of 579 polymorphic sites (Supplementary material, Appendix file). Average nucleotide diversity ranged between $\pi_{tot} = 0.0036$ and $\pi_{tot} = 0.0055$, and it was slightly higher in samples from the contact zone versus pure species populations (Table 2, Supplementary Table S2). Out of 789 haplotypes detected across 26 loci in all 384 samples analyzed, the majority (57 %) were unique (present only once). Exclusive haplotypes were found in all species and the contact zone groups. The within-population percentage of unique haplotypes ranged from 2.3 % (*P. uliginosa*-like individuals from the contact zone) to 0.2 % (*P. uliginosa* from Batorów reserve) (Supplementary Table S3). The average haplotype diversity was similar for each species ($H_d = 0.59–0.66$), regional groups of populations ($H_d = 0.61–0.70$), and groups of samples from the contact zone ($H_d = 0.54–0.71$) (Table 2). In the contact zone, no net divergence was observed between *P. mugo*, *P. uliginosa*, and polycormic *P. uliginosa*-like pines (Fig. 2, Supplementary Table S4 and S5). Divergence of these three groups was lower from the reference *P. uliginosa* (0.004–0.013) and *P. mugo* populations (0.005–0.013) and slightly higher from *P. uncinata* (0.009–0.023) (Fig. 2). *Pinus sylvestris* from Zieleniec and the reference populations of the species showed low divergence (0.001–0.031), and they were grouped together with *P. sylvestris*-like trees with M cpDNA haplotype (Supplementary Table S4 and S5, Fig. 2). The group of *P. sylvestris*-like trees from Zieleniec showed a similar level of divergence (0.010–0.015) from *P. mugo*, *P. sylvestris*, and *P. uliginosa* from that area (Supplementary Table S4).

Population structure

Of the reference populations, *P. mugo* and two of the *P. uliginosa* populations formed one group in the cluster analysis, while *P. sylvestris* and *P. uncinata* formed separate groups (total $K = 3$; Fig. 3a). The *P. uliginosa* population from Węgliniec showed greater similarity to *P. uncinata*. Two *P. sylvestris* individuals from Central and Northern Europe, two *P. mugo*, and several *P. uliginosa* and *P. uncinata* samples were identified as potentially admixed (Fig. 3a).

In the contact zone at Zieleniec, three clusters were identified: one included *P. mugo* (PMZ), *P. uliginosa* (PUZ), and polycormic *P. uliginosa*-like pines (HBZ) from that area; the second contained *P. sylvestris* (PSZ); and the third contained *P. sylvestris*-like samples (HPSZ) that were admixed with *P. mugo*/*P. uliginosa* and *P. uncinata* (Fig. 3a). There was evidence of admixture in a group containing *P. mugo*, *P. uliginosa*, and polycormic *P. uliginosa*-like pines (HBZ)

http://rcin.org.pl

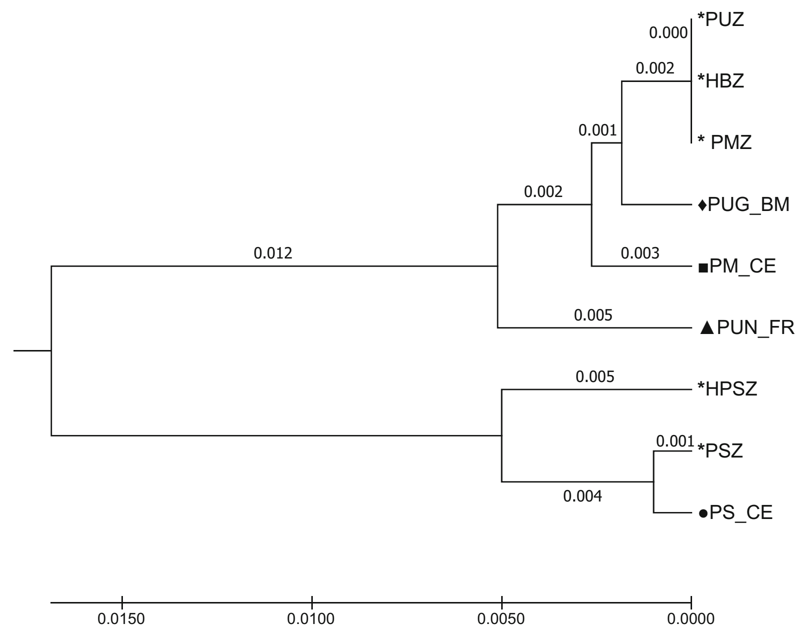


Fig 2 Unweighted pair group method with arithmetic mean (UPGMA) tree based on average net genetic distances between groups of samples from a contact zone at Zieleniec reserve (*asterisks*) and reference populations of the species including *Pinus mugo* (*filled square*), *P. sylvestris* (*filled circle*), *P. uliginosa* (*filled diamond*), and *P. uncinata* (*filled triangle*) (see Table 1 for details). *Numbers* indicate branch length. Genetic distance and its standard errors for each pairwise comparisons are shown in Supplementary Table S4

that appears to result from hybridization with *P. sylvestris* and *P. uncinata*. Only M and S haplotypes of the *trn*F–*trn*L cpDNA region were observed across all samples analyzed (Fig. 3b). Clear division between species was found in the PCoA analysis based on the genetic distance between populations (Fig. 4). Groups of *P. mugo* pines from Zieleniec reserve (PMZ and PUZ) and oligo- and polycormic *P. uliginosa*-like pines (HBZ) were placed between *P. mugo* and *P. uliginosa* populations. *Pinus sylvestris*-like hybrids (HPSZ) were placed between *P. mugo*/*P. uliginosa* groups and *P. sylvestris*. *Pinus sylvestris* from the contact zone clustered with allopatric populations of the species (Fig. 4).

Significant differentiation between each pair of taxa was found across all polymorphic sites with *P. mugo* and *P. sylvestris* from Zieleniec reserve as the most diverged ($F_{ST}=0.40$, $P<0.01$) (Table 3). Populations in regional groups of *P. mugo* and *P. sylvestris* showed low but significant differentiation across all polymorphic sites except *P. sylvestris* populations from Central and Northern Europe. No differentiation was found between *P. mugo, P. uliginosa*, and the group of polycormic pines from Zieleniec reserve (Table 3). The other groups showed similar levels of differentiation as between populations of allopatric zones of *P. mugo* and *P. sylvestris.*

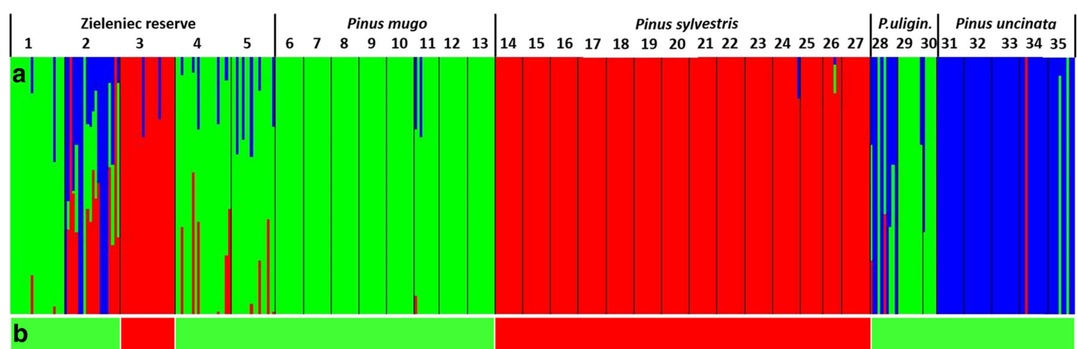


Fig 3 **a** Bar plot from Bayesian assignment (BAPS, Corander and Tang 2007) of the group of pines from Zieleniec reserve and the reference samples indicating three genetic clusters ($K=3$) corresponding to pure *Pinus mugo*, *P. sylvestris*, and *P. uncinata* species. Different *colors* represent proportional assignment of each sample to individual clusters, the *black vertical lines* separate the corresponding populations, and *numbers* above *vertical bars* refer to populations given in Table 1. Evidence of admixture ($P<0.01$) was found across the samples from the sympatric population of the species from Zieleniec reserve. At the intraspecific level, heterogeneous population structure was observed between *P. uliginosa* populations. **b** Assignment of samples according to *trn*F–*trn*L cpDNA marker diagnostic to the *P. mugo* complex (M marker—*green*) and *P. sylvestris* (S marker—*red*)

http://rcin.org.pl

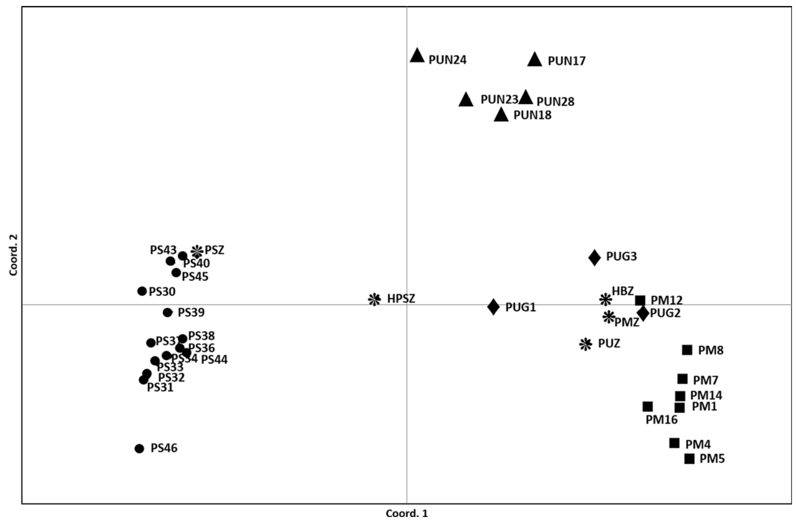


Fig 4 Plot of the first two axes of a principal coordinates analysis (PCoA) based on a genetic distance matrix for groups of samples from Zieleniec reserve and reference populations of the *Pinus mugo*, *P. sylvestris*, *P. uliginosa*, and *P. uncinata* at all polymorphic sites from 26 nuclear genes. Population acronyms as in Table 1

Neutrality tests

An excess of singleton mutations across genes was evident as multilocus Tajima's *D* was negative in most groups ($D = -0.859$ to -0.039) except *P. sylvestris*-like samples from Zieleniec reserve and reference *P. uncinata* (Table 2). Significant deviations from neutrality at Tajima's *D* and/or HEW and DHEW compound neutrality tests were found at eight loci in *P. sylvestris* and *P. mugo*, seven at *P. uliginosa* and at one locus in *P. uncinata* (Supplementary Table S6). In the samples from Zieleniec reserve, significant excess of singleton mutations was found at locus Pr4_5 in a group of oligo- and polycormic *P. uliginosa*-like pines, at Pr4_17 in *P. sylvestris*-like individuals, at Pr1_28, Pr2_17, Pr4_17 in *P. sylvestris*, and Pr2_28 in *P. mugo*. An excess of intermediate-frequency variants was found at Pr2_47 in *P. sylvestris* and Pr4_19 in polycormic pines from Zieleniec reserve (Supplementary Table S6).

Locus-specific genetic differentiation and selection

The loci studied showed high divergence between species and very low differentiation between populations within species (Supplementary Table S7). No significant differentiation at any locus was observed between samples from Zieleniec reserve defined as *P. mugo*/*P. uliginosa* and a group of taxonomically unclassified *P. uliginosa*-like polycormic pines. These three groups of samples were highly diverged from *P. sylvestris* and *P. sylvestris*-like samples at most of the loci (Supplementary Table S8). Divergence of populations from Zieleniec from the reference populations (Supplementary Table S9) was higher than divergence between pure species populations (Supplementary Table S7). At the SNP level, several loci were significantly differentiated between *P. sylvestris* and the *P. mugo* complex: 46 SNPs were found in 22 genes from all categories between *P. sylvestris* and *P. mugo*, 40 SNPs in 16 genes between *P. sylvestris* versus *P. uliginosa*, and 24 SNPs in 11 genes between *P. sylvestris* versus *P. uncinata*.

Table 3 Pairwise F_{ST} between groups of *Pinus* samples from Zieleniec reserve and the reference populations of pure taxa at all polymorphic sites (significant values at $P < 0.01$ are marked in bold)

	1	2	3	4	5	6	7	8
1. PMZ—*P. mugo*								
2. HPSZ—*P. sylvestris*-like	**0.117**							
3. PSZ—*P. sylvestris*	**0.331**	**0.094**						
4. PUZ—*P. uliginosa*	−0.001	**0.099**	**0.306**					
5. HBZ—*P. uliginosa*-like	0.001	**0.110**	**0.326**	−0.001				
6. PM_CE—*P. mugo*	**0.055**	**0.191**	**0.400**	**0.064**	**0.056**			
7. PS_CE—*P. sylvestris*	**0.325**	**0.090**	**0.025**	**0.290**	**0.314**	**0.396**		
8. PUG_BM—*P. uliginosa*	**0.045**	**0.140**	**0.360**	**0.045**	**0.033**	**0.067**	**0.354**	
9. PUN_FR—*P. uncinata*	**0.088**	**0.101**	**0.255**	**0.096**	**0.084**	**0.140**	**0.256**	**0.079**

http://rcin.org.pl

Within the *P. mugo* complex, *P. mugo* versus *P. uliginosa* and *P. uncinata* were highly diverged at 5 and 12 SNPs, respectively, and *P. uliginosa* and *P. uncinata* at 6 SNPs (Supplementary Table S10). These polymorphic sites and genes form a valuable set of new markers for tracking hybrid genotypes. The most diverged genes were those involved in regulation of gene expression, signal transduction, transport, and cellular metabolism and appear to be of high importance in speciation and local adaptation (Supplementary Table S1 and S10).

Discussion

Here, we used nucleotide polymorphism data to compare patterns of genetic variation in pine trees from the contact zone with those from allopatric zones of four species from across Europe. We focused on the contact zone between species at a peat bog complex located in the Sudety Mts. in Poland. The aim was to evaluate the role of introgressive hybridization and selection on nucleotide diversity in the analyzed populations. To date, such studies have been limited by the lack of diagnostic biometric and biochemical characters for tracking interspecific gene flow and the identification of individual hybrid trees. For instance, many anatomical traits of needles show overlapping frequency distributions among taxa (Boratyńska et al. 2015). In our study, we used polymorphisms at nuclear genes and a cpDNA marker to identify interspecific admixture. Nuclear markers that experience high rates of intraspecific gene flow are especially relevant for species delimitation (Petit and Excoffier 2009). Our data clearly indicate that the markers applied in our study can accurately discriminate pure parental species. These markers provide a large resource of SNP information for future use in tracking interspecific gene flow and evaluating species composition in other contact zones where individuals with mixed morpho-anatomical characteristics have been described (e.g., in the Alps (Christensen 1987a), Slovakia (Kormutak et al. 2014)).

Hybrids can exhibit intermediate trait values, combine traits from both parents, and/or exhibit extreme trait values as compared to the parental species (Gross and Rieseberg 2005). In our study, high genetic similarity was observed between samples classified as *P. mugo*, *P. uliginosa*, and a group of polycormic *P. uliginosa*-like pines (nucleotide diversity $\pi_{tot}=\sim 0.0044–0.0049$, zero net divergence, no significant differentiation in allele frequency spectra). The three groups formed a uniform genetic cluster and showed low divergence (about 1 %) from pure-species populations of the taxa from the *P. mugo* complex. In contrast, they differed clearly from *P. sylvestris* from Zieleniec reserve and from reference populations of the species. Those oligo- and polycormic *P. uliginosa*-like pines that had cpDNA diagnostic to the *P. mugo* complex should be considered as hybrids between *P. mugo* and *P. uliginosa* with no evidence of a substantial contribution from *P. sylvestris*. In addition, we identified a group of *P. sylvestris*-like trees that had cpDNA markers characteristic of the *P. mugo* complex. This group of monocormic pines showed clear evidence of admixture between *P. mugo*/*P. uliginosa*, *P. sylvestris*, and also *P. uncinata*. These trees represent a second group of phenotypically distinct hybrids with *P. sylvestris* acting as a mother tree at an early stage of hybridization, and cpDNA delivered by pollination from the polycormic pines of the *P. mugo* complex. All hybrids shared some of the nuclear gene haplotypes observed in the reference populations of parental species. Different levels of introgression of those individuals, as evident from variable assignment to the parental species, indicate that this group of pines includes F1 and subsequent generations of hybrids, although always fertilized by pollen carrying the cpDNA diagnostic for the *P. mugo* complex. This asymmetric introgression between taxa may be the result of selection, but different mechanisms of incompatibility between hybrids cannot be excluded (Currat et al. 2008).

Pinus uliginosa had a heterogeneous genetic background, with one population from Węgliniec reserve showing close genetic similarity to *P. uncinata*. This population showed patterns of admixture not observed in *P. uliginosa* from the species *locus classicus* at Batorów reserve (Neumann 1837) and the population from Mittelwalde. This suggests that some individuals classified as *P. uliginosa* from Węgliniec but also Zieleniec reserve may represent remnants of marginal populations of *P. uncinata*. Interestingly, the pattern of admixture observed in our study is in line with some biochemical and biometric data on cone and needle morphology (Marcysiak and Boratyński 2007 and references therein). These studies showed that *P. uliginosa* from Węgliniec shares some biometric traits with *P. sylvestris*, *P. mugo*, and *P. uncinata* and is distinct as compared to other allopatric stands of the species (Lewandowski et al. 2000; Prus-Głowacki et al. 1998). Therefore, the existence of remnant populations of *P. uncinata* in the Silesian Lowlands seems possible and may result from the expansion of west-European refugia across the northwestern pre-Alpine territories during the late Dryas/early Holocene (Marcysiak and Boratyński 2007). Alternatively, considering the large number of shared ancestral alleles still segregating in the species from the *P. mugo* complex (Wachowiak et al. 2013), interspecific gene flow could create combinations of alleles in individual hybrid trees clustered in our analysis as different taxonomic units. These results provide another dimension to the very complex demographic history of the taxa within the *P. mugo* complex

Our data contribute to the assessment of the genetic relationships of the taxa from the *P. mugo* complex showing evidence of close genetic identity of *P. uliginosa* as compared to *P. mugo* and *P. uncinata* (net divergence of 0.8–0.9 %, respectively). *Pinus uliginosa* does not harbor a distinct gene pool as compared to *P. mugo* (Fig. 3a), and therefore it should not be

http://rcin.org.pl

considered as a separate species but rather subspecies within the *P. mugo* complex. The taxon was originally described from the peat bog population at Batorów reserve in the Sudety Mts. (Neumann 1837). In a biometric revision of the complex, it was proposed as a subspecies of *P. uncinata* (Businsky and Kirschner 2006) or as a synonym of *Pinus rotundata*—a pine species forming high-altitude populations in the Alps and Northern Carpathians and small isolated populations in the Pyrenees and Massif Central (Christensen 1987b). Our molecular analysis did not aim to resolve the taxonomic position of *P. uliginosa*. However, the data showed closer genetic similarity of *P. uliginosa* to *P. mugo* than *P. uncinata*. It is clear that more molecular studies are needed to clarify the taxonomic status of several taxa described within the *P. mugo* complex and to evaluate the role of interspecific gene flow with *P. sylvestris* (Christensen 1987b).

Barriers to interspecific hybridization and a lack of evidence for bidirectional gene flow between *P. sylvestris* and *P. mugo* were suggested in previous studies that found hybrid seeds derived only from *P. sylvestris*-like individuals pollinated with *P. mugo* but not from reciprocal crossings (Wachowiak et al. 2005a). A lack of hybrids from crossings between *P. mugo* as a maternal and *P. sylvestris* as a paternal tree, but putative hybrid individuals from reverse crossing combinations (with *P. mugo* as a pollen donor), were found based on a joint analysis of cpDNA, isozymes, and phenotypic characteristics of trees (Wachowiak and Prus-Głowacki 2008) and at nuclear genes in a *P. sylvestris* and *P. mugo* population (Kormutak et al. 2014). Cryptic hybrids between *P. sylvestris* and *P. uncinata* were found in the sympatric populations of the species (Jasińska et al. 2010). So far, the only evidence of reciprocal hybridization was found in a sympatric population of *P. sylvestris* and *P. uliginosa* (Wachowiak et al. 2005b).

Our results suggest that hybrids express distinct phenotypic variability as compared to parental species. In previous biometric and biochemical studies, the variety of morphological forms observed in sympatric populations of *P. sylvestris* and the *P. mugo* complex was explained as either the result of intensive hybridization and introgression that changed the population into a hybrid swarm or as a mixture of mostly pure pine species from the *P. mugo* complex and *P. sylvestris* (Bobowicz 1990; Odrzykoski 2002; Wachowiak et al. 2006b). Our data indicate that hybridization takes places in contact zones of the species and leads to propagation of viable hybrid trees. They exist together with pure parental species and maintain their phenotypic identity.

Natural selection can cause the fixation of advantageous alleles (or chromosomal segments) in ecologically diverged hybrids (Beaumont and Balding 2004; Buerkle and Lexer 2008; Lexer et al. 2003). Introgressed alleles may often have a positive fitness effect in their new genetic background and traits responsible for adaptation can be transferred between species (Martin et al. 2006). Foreign alleles in different genetic or ecological backgrounds will show a range of fitness outcomes, but only those that increase the adaptive optimum in a given environment will effectively introgress. Consequently, introgression of alleles derived from other species has the potential to speed adaptation (Gompert and Buerkle 2010), which may be particularly influential in populations undergoing spatial or temporal transitions into new environments. Our study provides evidence of successful hybridization within the sympatric study population but no evidence for interspecific gene flow outside the contact zone. It is possible that hybrids have reduced fitness in environmental conditions occupied by parental species, and they may persist best in new habitats. Indeed, our results show that hybrid genotypes have succeeded in peat bogs close to mountain regions, which are environments untypical of either parental species.

Co-existence of morphologically variable taxa and hybrids together with asymmetric gene flow indicates the role of selection in maintaining certain phenotypes. Strong directional selection on loci underlying fitness-related adaptation in the ecologically diverged hybrids should increase the frequency of advantageous alleles. We expect, however, that the effect of selection should be localized in the genome and the genetic background of a species should not be affected by the spatial expansion of an advantageous allele (Currat et al. 2008). Two distinct groups of hybrids seem to have maintained their phenotypic differentiation, and we found signatures of selection at some loci as compared to background variation. *Pinus sylvestris*-like hybrids showed increased frequency of alleles specific to *P. sylvestris* and alleles specific to *P. mugo* at different genes. Similarly, at one polyol transporter gene, oligo- and polycormic hybrids showed differentiation from *P. mugo* but not from *P. uliginosa*. An increase in frequency of alleles unique to one of the parental species at some loci may result from selection of particular alleles in the hybrids that increase their fitness in the peat bog environment. In contrast, no evidence of differentiation at any locus was found between *P. mugo* and *P. uliginosa*, and between *P. uliginosa* and a group of polycormic hybrids from Zieleniec reserve. In the case of parental species, strong directional selection at some loci due to local adaptation in ecologically diverged peat bog environments should increase differentiation between the peat bog and the reference pure-species populations. In the absence of population structure in the parental species, significant differences in allele frequency spectra and/or departures from neutrality between reference and contact zone populations were found at nine loci in *P. mugo* and eight in *P. sylvestris*. These loci showed no evidence of differentiation between pure-species populations. Assuming different patterns of diversity at selectively influenced loci relative to background genetic variation, this increased population differentiation suggests selection in response to specific peat bog environments not optimal for either of the parental species (e.g.,

http://rcin.org.pl

Eveno et al. 2008; Kujala and Savolainen 2012; Wachowiak et al. 2009).

In many cases, as shown in theoretical models and experimental studies of contact zones, introgression is highly asymmetric (Currat et al. 2008; Petit et al. 2004) and may go from the local to the invading species. If this scenario holds for the investigated pine taxa, then we could consider *P. sylvestris* and possibly *P. uliginosa* as an invading taxon as compared to the local *P. mugo*. At present, however, this contact zone is isolated from the continuous range of any of the parental taxa. In the case of invading species, the pattern of introgression at some neutral loci resulting from the range expansion of a species into an already occupied territory may mimic the effect of selection. Therefore, more studies of the contact zone dynamics (e.g., Cinget et al. 2015) at Zieleniec reserve and its neighborhood are needed. Such studies would help to test invasion models and evaluate the role of demographic processes on the patterns of genome-wide nucleotide sequence variation.

With the presence of different taxa and hybrid groups in environmental conditions not optimal for either of the putative parental species, the contact zone at Zieleniec reserve is a relevant biological system for studying the role of hybridization on adaptation to new environments at the genetic level. This approach has recently been developed in several model plant species demonstrating the role of hybridization and adaptive introgression in the evolution of irises (*Iris*; Arnold et al. 2004), ecological divergence of sunflowers (*Helianthus*; Rieseberg et al. 2007), and the signatures of divergent and balancing selection in campions (*Silene*; Minder and Widmer 2008) and poplar (*Populus*; Lexer et al. 2010). So far, a few genes involved in adaptation or speciation have been identified in plants including hybrid sterility loci (Lexer and Widmer 2008), determinants of flower color-linked pollinator shifts (Hoballah et al. 2007), and genes involved in hybrid necrosis (Bomblies and Weigel 2007). Our study shows that the investigated taxa maintain genetic and phenotypic differentiation in the presence of extensive gene flow. Considering the abundance of trees growing on peat bog in our focal populations, both *P. sylvestris*-like hybrids and oligo- and polycormic *P. uliginosa*-like pines could serve as suitable mapping populations in the search for loci underlying local adaptation and genetic and phenotypic differentiation between taxa.

Acknowledgments The research was financially supported by the Polish National Science Centre (grant no. 2011/01/B/NZ8/01634). WŻ and BW acknowledge financial support from Polish National Science Centre (grant no. DEC-2012/05/E/NZ9/03476). WW and SC acknowledge financial support from NERC (grant no. NE/H003959/1). We thank Euforgen network for providing the distribution map of Scots pine. We gratefully acknowledge the comments by the Associate Editor and two anonymous reviewers who helped us to improve the earlier version of the manuscript.

Data Archiving Statement The nucleotide sequences of the studied loci were submitted to NCBI repositories (accession nos. KC979194–KC980906; see Supplementary material for details).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

Andrew RL, Rieseberg LH (2013) Divergence is focused on few genomic regions early in speciation: incipient speciation of sunflower ecotypes. Evolution 67:2468–2482

Arnold ML, Bouck AC, Scott Cornman R (2004) Verne Grant and Louisiana irises: is there anything new under the sun? New Phytol 161:143–149

Arnold ML, Martin NH (2010) Hybrid fitness across time and habitats. Trends Ecol Evol 25:530–536

Baack EJ, Rieseberg LH (2007) A genomic view of introgression and hybrid speciation. Curr Opin Genet Dev 17:513–518

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. Mol Ecol 13:969–980

Bobowicz MA (1990) Hybrids between *Pinus mugo* Turra x *Pinus sylvestris* L. from "Bór na Czerwonem" reserve in Novotarska Valley [in Polish]. Wyd Nauk UAM, Poznań

Bomblies K, Weigel D (2007) Hybrid necrosis: autoimmunity as a potential gene-flow barrier in plant species. Nat Rev Genet 8:382–393

Boratyńska K, Jasińska AK, Boratyński A (2015) Taxonomic and geographic differentiation of *Pinus mugo* complex on the needle characteristics. Sys Biodiv:1-15

Buerkle CA, Lexer C (2008) Admixture as the basis for genetic mapping. Trends Ecol Evol 23:686–694

Businsky R, Kirschner J (2006) Nomenclatural notes on the *Pinus mugo* complex in Central Europe. Phyton Ann Rei Bot A 46(1):129–139

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

Christensen KI (1987b) Taxonomic revision of the *Pinus mugo* complex and *P. x rhaetica* (*P. mugo* x *sylvestris*) (Pinaceae). Nord J Bot 7: 383–408

Cinget B, de Lafontaine G, Gérardi S, Bousquet J (2015) Integrating phylogeography and paleoecology to investigate the origin and dynamics of hybrid zones: insights from two widespread North American firs. Mol Ecol 24:2856–2870

Corander J, Tang J (2007) Bayesian analysis of population structure based on linked molecular information. Math Biosci 205:19–31

Critchfield WB, Little E (1966) Geographic distribution of the Pines of the world. U.S. Department of Agriculture

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

Eveno E, Collada C, Guevara MA et al (2008) Contrasting patterns of selection at *Pinus pinaster* Ait. drought stress candidate genes as revealed by genetic differentiation analyses. Mol Biol Evol 25: 417–437

http://rcin.org.pl

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Res 10:564–567

Gompert Z, Buerkle AC (2010) Introgress: a software package for mapping components of isolation in hybrids. Mol Ecol 10:378–384

González-Martínez SC, Krutovsky KV, Neale DB (2006) Forest-tree population genomics and adaptive evolution. New Phytol 170:227–238

Gross BL, Rieseberg LH (2005) The ecological genetics of homoploid hybrid speciation. J Hered 96:241–252

Hey J, Wakeley J (1997) A coalescent estimator of the population recombination rate. Genetics 145:833–846

Hoballah ME, Gübitz T, Stuurman J et al (2007) Single gene–mediated shift in pollinator attraction in Petunia. Plant Cell Online 19:779–790

Hurme P, Sillanpää MJ, Arjas E, Repo T, Savolainen O (2000) Genetic basis of climatic adaptation in Scots pine by Bayesian quantitative trait locus analysis. Genetics 156:1309–1322

Jasińska AK, Wachowiak W, Muchewicz E, Boratyńska K, Montserrat JM, Boratyński A (2010) Cryptic hybrids between *Pinus uncinata* and *P. sylvestris*. Bot J Linn Soc 163:473–485

Kormutak A, Brana M, Manka P, Galgoci M, Libantova J, Camek V, Bolecek P, Gömöry D (2014) Hybridization processes in putative hybrid swarms of Scots pine and mountain dwarf pine as revealed by chloroplast DNA. Acta Biol Cracov Bot 56(2):61–66

Kujala S, Savolainen O (2012) Sequence variation patterns along a latitudinal cline in Scots pine (*Pinus sylvestris*): signs of clinal adaptation? Tree Genet Genome 8:1451–1467

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

Lexer C, Widmer A (2008) The genic view of plant speciation: recent progress and emerging questions. Philos Trans R Soc Lond B Biol Sci 363:3023–3036

Lexer C, Welch ME, Raymond O, Rieseberg LH (2003) The origin of ecological divergence in *Helianthus paradoxus* (Asteraceae): selection on transgressive characters in a novel hybrid habitat. Evolution 57:1989–2000

Lexer C, Joseph JA, van Loo M et al (2010) Genomic admixture analysis in European *Populus* spp. reveals unexpected patterns of reproductive isolation and mating. Genetics 186:699–712

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

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

Martin NH, Bouck AC, Arnold ML (2006) Detecting adaptive trait introgression between *Iris fulva* and *I. brevicaulis* in highly selective field conditions. Genetics 172:2481–2489

Minder AM, Widmer A (2008) A population genomic analysis of species boundaries: neutral processes, adaptive divergence and introgression between two hybridizing plant species. Mol Ecol 17:1552–1563

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Neumann C (1837) Über eine auf den Seefeldern der Reinerz und einigen ähnlichen Gebirgsmooren der königl. Oberforsterei Karlsberg in der Grafschaft Glatz vorkommende noch unbeschriebene Form der Gattung Pinus [in German]. Jahresb Schlesische Ges

Odrzykoski IJ (2002) Genetic variation study of Dwarf mountain pine (*Pinus mugo*) with the use of molecular and biochemical markers [in Polish]. Wyd Nauk UAM, Poznań

Petit RJ, Bodénès C, Ducousso A, Roussel G, Kremer A (2004) Hybridization as a mechanism of invasion in oaks. New Phytol 161:151–164

Petit RJ, Excoffier L (2009) Gene flow and species delimitation. Trends Ecol Evol 24:386–393

Prus-Głowacki W, Bujas E, Ratyńska H (1998) Taxonomic position of *Pinus uliginosa* Neumann as related to other taxa of *Pinus mugo* complex. Acta Soc Bot Pol 67:269–274

Rieseberg LH, Kim S-C, Randell RA, Whitney KD, Gross BL, Lexer C, Clay K (2007) Hybridization and the colonization of novel habitats by annual sunflowers. Genetica 129:149–165

Storey JD, Tibshirani R (2003) Statistical significance for genome wide studies. Proc Natl Acad Sci U S A 100:9440–9445

Tajima F (1989) Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729

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

Wachowiak W, Leśniewicz K, Odrzykoski I, Augustyniak H, Prus-Głowacki W (2000) Species specific cpDNA markers useful for studies on the hybridization between *Pinus mugo* and *P. sylvestris*. Acta Soc Bot Pol 69:273–276

Wachowiak W, Lewandowski A, Prus-Głowacki W (2005a) Reciprocal controlled crosses between *Pinus sylvestris* and *P. mugo* verified by a species-specific cpDNA marker. J Appl Genet 46:41–43

Wachowiak W, Celiński K, Prus-Głowacki W (2005b) Evidence of natural reciprocal hybridisation between *Pinus uliginosa* and *P. sylvestris* in the sympatric population of the species. Flora 200: 563–568

Wachowiak W, Stephan BR, Schulze I, Prus-Głowacki W, Ziegenhagen B (2006a) A critical evaluation of reproductive barriers between closely related species using DNA markers—a case study in *Pinus*. Plant Syst Evol 257:1–8

Wachowiak W, Odrzykoski I, Myczko Ł, Prus-Głowacki W (2006b) Lack of evidence on hybrid swarm in the sympatric population of *Pinus mugo* and *P. sylvestris*. Flora 201:307–316

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

Wachowiak W, Boratyńska K, Cavers S (2013) Geographical patterns of nucleotide diversity and population differentiation in three closely related European pine species in the Pinus mugo complex. Bot J Linn Soc 172:225–238

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370

Whitney KD, Randell RA, Rieseberg LH (2010) Adaptive introgression of abiotic tolerance traits in the sunflower *Helianthus annuus*. New Phytol 187:230–239

Zeng K, Shi S, Wu C-I (2007) Compound tests for the detection of hitchhiking under positive selection. Mol Biol Evol 24:1898–1908

http://rcin.org.pl

Supplemental material

Title: Hybridization in contact zone between temperate European pine species.

Authors: Wachowiak Witold, Żukowska Weronika B., Wójkiewicz Błażej, Cavers Stephen, Litkowiec Monika

Supplementary Table S1. Analysed loci and Gene Bank Accession numbers of the reference haplotype sequences.

Locus	PCR Primer Forward	PCR Primer Reverse	Gene function [Category[3]]	Gene Bank Acc. Nr.	Reference seq.[4]
Pr1_5	[1]GATCATTCTAGGCACAGCACAAG	[2]CCTGTACCGTGTTCATCAATTTAGCAAG	Seryl-tRNA synthetase [E]	KC979194 - KC979222	FJ054308
Pr1_9	[1]GGAATTAAAACCAGACGAAC	[2]GAAGAGGAATGACCAG	transcription factor [E]	KC979223 - KC979245	FJ059669
Pr1_11	[1]GACCAGGCAAGGAAACAAAAG	[2]TTGGCAATCGGTTGATGGGGAG	putative glucuronidase 3 [M]	KC979259 - KC979288	FJ063196
Pr1_12	[1]TCCCCATTCTCCAAAC	[2]CCATCCAATCCTTCATC	NAD(P)-linked oxidoreductase- like protein [M]	KC979289 - KC979325	FJ064802
Pr1_13	[1]AACGCTGTGACTTGTCC	[2]CTTGTAATTCAGATATTCACGGGG	hypothetical protein [UN]	KC979326 - KC979347	JQ018340
Pr1_19	[1]CCGTATGCAAAGCATTTC	[2]ACCTGATCGTGTTGTG	Glycosyltransferase [M]	KC979461- KC979482	FJ070807
Pr1_22	[1]TGAAGGGAGAGGACTAC	[2]ACCCAGAAACACAAAGAGGAAAC	hypothetical protein [UN]	KC979506 - KC979528	FJ073702
Pr1_28	[1]GCAACTTCCCCTTTTTC	[2]ACAGTGTGAGAGACGAG	translation initiation factor- 4G-like [E]	KC979604 - KC979621	FJ045321
Pr1_40	[1]TACAATATGTTCCTGTCAAGGC	[2]TACGCTCACATGGTCTCCTTC	Homeobox domain containing - protein [ST]	KC979697 - KC979721	FJ092868
Pr2_3	[1]GAGGAGATGGTTTTGTATG	[2]ATCCAGGTGCTACTTC	2-3 ethylene-responsive- transcription factor 1B[E]	KC979828 - KC979847	FJ049919
Pr2_4	[1]GACTGGGATACTCTTTTGG	[2]AGTGAATAACACTTGGCTGCTATAC	f-box family protein [ST]	KC979848 - KC979854	FJ050118
Pr2_17	[1]GCATTAGTCTGTCTGTTC	[2]GTGTTTCTAGGGCAATC	putative polyol transporter [T]	KC979947 - KC979961	FJ066481
Pr2_23	[1]GCCCAAATGGTTATACATAACACTC	[2]CCATTCCATCGGCACAGTCATC	Transcribed locus [UN]	KC979985 - KC980011	FJ073096
Pr2_28	[1]CCTCCCATCATTCTTTCTTCC	[2]GAATTGCAGCCCTTGCACAAGAC	basic leucine zipper transcription-factor-like protein [E]	KC980025 - KC980041	FJ082481
Pr2_38	[1]CCATCATACAACTATCCAC	[2]ACAGAGAATAATGGGGCAC	hypothetical protein [UN]	KC980123 - KC980143	JQ020451
Pr2_42	[1]GCATAGCCATCCATATC	[2]GGGTGTGAATTTTTTTGGTG	transducin/WD40 domain- containing protein [ST]	KC980161 - KC980175	FJ105246
Pr2_47	[1]TTCATAAAGCCCCCCATCC	[2]TCTGATTTCAAAGTCGCC	hexokinase 1 [ST]	KC980219 - KC980234	FJ140440
Pr4-4	[1]TGTCACTGCCCAGAGCTATTC	[2]ATCACAGCCGCTCCAAAAC	Putative aquaporin [e] [T]	KC980645 - KC980679	DQ370110
Pr4-5	[1]CATCTCCTTCAAACCTCTTATTTCC	[2]GATGCTTGAACATGATCCC	calcium dependent proteokinase [e] [ST]	KC980680 - KC980695	AY874681
Pr4-10	[1]CATTGCCTACGATTTCC	[2]CTTTTGAGATGAACCAGAC	mys transcription factor [e] [E]	KC980696 - KC980702	FJ085365
Pr4-12	[1]CTGCTCAAGTGAAAGG	[2]CTGATTGTGGATTCTGTG	proton myo-inositol transporter [e] [T]	KC980725 - KC980745	FJ082525
Pr4-17	[1]CTGGAAGCTGATTCTTTTG	[2]CCTCTAGTTCTGGTTG	cytochrome P450 reductase [e] [ST]	KC980746 - KC980762	FJ079163
Pr4-19	[1]CTCTACCACATCATCTCC	[2]TTTCACTCTCGTGTCTTTCACC	laccase [e] [M]	KC980770 - KC980785	FJ062700
Pr4-21	[1]ACATGGTGTTTGGCAGG	[2]AATGAGGAGGGTGGTAGAG	Receptor protein kinase [e] [ST]	KC980786 - KC980810	FJ053757
Pr4-27	[1]TAGCAGACGGTATTCACACAGTCC	[2]CCACAACCACCTTGCATCATTATTT	putative auxin induced - transcription factor [e] [E]	KC980811 - KC980829	AY289601
Pr4-41	[1]TGCAAGCTGTAAGGTAAAACCCTCAT	[2]CAACATCAAAACTGAAACCACCAGTC	ethylene responsive element- binding protein [e] [ST]	KC980864 - KC980906	EU394012

[1,2] – vector sequence (1=GTAAAACGACGGCCAGT and 2=CAGGAAACAGCTATGACC) was present as a part of PCR primers used for amplification of the loci studied; [3] – E-gene expression regulation; M-metabolisms; ST-signal transduction; T-transport; UN-unknown; [4] – outgroup Gene Bank reference sequence used for divergence estimates.

http://rcin.org.pl

Supplementary Table S2. Multilocus average nucleotide and haplotype diversity of the groups of pines defined at Zieleniec reserve and reference populations of pure *Pinus mugo*, *P. sylvestris*, *P. uliginosa* and *P. uncinata* species. Populations and groups acronyms as explained in Table 1 & 2 of the main text.

Groups	N	L	SNPs	S	π_{total}	Tajima's D	N_h	H_d	SD
PMZ	18.5	404.2	7.3	2.0	0.0044	-0.148	7.2	0.650	0.054
HPSZ	19.7	404.2	8.1	1.8	0.0055	0.206	7.7	0.710	0.047
PSZ	19.5	404.2	7.2	2.2	0.0039	-0.292	6.7	0.541	0.062
PUZ	19.2	404.2	8.2	1.9	0.0049	-0.164	8.1	0.695	0.053
HBZ	15.4	404.2	7.0	2.1	0.0045	-0.138	6.9	0.658	0.058
PM_CE	28.4	449.1	6.9	1.8	0.0039	-0.039	5.4	0.609	0.067
PM_CARP	19.4	449.0	6.2	2.3	0.0036	-0.295	4.6	0.553	0.088
PM_BAL	19.7	448.8	6.6	2.3	0.0038	-0.253	4.7	0.589	0.082
PS_SCO	48.5	449.8	8.3	2.4	0.0039	-0.339	6.5	0.575	0.051
PS_CE	19.1	448.2	7.1	2.9	0.0042	-0.301	5.3	0.625	0.080
PS_SP	19.7	449.0	6.1	1.7	0.0037	-0.300	4.5	0.553	0.084
PS_SCA	32.4	450.0	8.8	3.3	0.0044	-0.515	6.8	0.638	0.062
P. mugo	77.2	450.9	10.7	3.6	0.0039	-0.536	9.0	0.603	0.042
P. sylvestris	128.8	450.9	13.7	5.2	0.0042	-0.859	12.3	0.590	0.031
P. uliginosa	23.3	450.9	8.4	2.9	0.0046	-0.254	5.8	0.650	0.073
P. uncinata	49.8	450.9	9.5	2.0	0.0048	0.170	6.8	0.665	0.049

PM – *P. mugo*, PS – *P. sylvestris*, CE – Central Europe, CARP – Carpathians, BAL – Balkans, SCO – Scotland, SP – Spain, SCA – Scandinavia

N – sample size; L – length of sequence in base pairs; S – number of segregating sites; π – nucleotide diversity (Nei 1987); *D* – Tajima's *D* test (Tajima 1989); N_h – number of haplotypes; H_d – haplotype diversity (standard deviation); $^{*}P<0.05$; $^{***}P<0.01$

http://rcin.org.pl

Supplementary Table S3. Numbers of common and unique haplotypes at the groups of samples analysed. Population's acronyms as explained in Table 1 of the main text.

		Pine contact zone				***Pinus mugo***				***Pinus sylvestris***					***Pinus uliginosa***			***Pinus uncinata***		
Groups	PMZ	HPSZ	PSZ	PUZ	HBZ	CARP	BALK	CE	IT	SCO	CE	SP	SCA	IT	PUG1	PUG2	PUG3	Andorra	Spain	France
Samples size	20	20	20	20	16	20	20	30	20	50	20	20	34	20	12	12	6	20	20	10
Number of haplotypes	205	213	185	211	193	126	126	126	129	151	186	192	123	70	105	91	71	128	123	109
%	10.3	10.7	9.3	10.6	12.1	6.3	6.3	4.2	6.5	3.0	9.3	9.6	3.6	3.5	8.8	7.6	11.8	6.4	6.2	10.9
Number of unique haplotypes	35	31	36	35	37	20	19	26	20	29	33	42	30	8	13	12	1	10	10	7
%	1.8	1.6	1.8	1.8	2.3	1.0	1.0	0.9	1.0	0.6	1.7	2.1	0.9	0.4	1.1	1.0	0.2	0.5	0.5	0.7

CARP-Carpathians, BAL-Balkans, CE-Central Europe, IT-Italy, SCO-Scotland, SP-Spain

http://rcin.org.pl

Supplementary Table S4. Net genetic distance (standard error in brackets) between groups of samples from the hybrid zone and reference populations of *Pinus mugo* (Central Europe), *P. sylvestris* (Central Europe), *P. uliginosa* (Poland and Germany) and *P. uncinata* (France). Divergence levels < 0.01 are shaded. Populations and groups acronyms as explained in Table 1 & 2 of the main text.

	1	2	3	4	5	6	7	8
1. PMZ - *P. mugo*	0							
2. HPSZ - *P. sylvestris*-like	0.015 [0.002]							
3. PSZ - *P. sylvestris*	0.043 [0.006]	0.010 [0.002]						
4. PUZ - *P. uliginosa*	0.000 [0.000]	0.012 [0.002]	0.040 [0.005]					
5. HBZ - *P. uliginosa*-like	0.000 [0.000]	0.013 [0.002]	0.042 [0.006]	0.000 [0.000]				
6. PM_CE - *P. mugo*	0.005 [0.001]	0.023 [0.003]	0.054 [0.007]	0.006 [0.001]	0.005 [0.001]			
7. PS_CE - *P. sylvestris*	0.044 [0.006]	0.010 [0.001]	0.002 [0.001]	0.039 [0.005]	0.041 [0.005]	0.055 [0.007]		
8. PUG_BM - *P. uliginosa*	0.004 [0.001]	0.017 [0.002]	0.047 [0.007]	0.004 [0.001]	0.003 [0.001]	0.005 [0.001]	0.047 [0.006]	
9. PUN_FR - *P. uncinata*	0.009 [0.001]	0.012 [0.002]	0.031 [0.005]	0.011 [0.001]	0.009 [0.001]	0.014 [0.002]	0.032 [0.005]	0.008 [0.001]

http://rcin.org.pl

Supplementary Table S5. Net genetic distance (lower-left matrix) and standard error (upper-right) between groups of samples from Zieleniec reserve and reference populations of *Pinus mugo*, *P. sylvestris*, *P. uliginosa* and *P. uncinata*. Divergence levels < 0.01 are shadowed. Populations acronyms as explained in Table 1 of the main text.

PMZ	HPSZ	PSZ	PUZ	HBZ	PM1	PM4	PM5	PM7	PM8	PM12	PM14	PM16	PS30	PS31	PS32	PS33	PS34	PS36	PS37	PS38	PS39	PS40	PS43	PS44	PS45	PS46	PUG 1	PUG 2	PUG 3	PUN 17	PUN 18	PUN 23	PUN 24	PUN 28
	0.002	0.006	0	0	0.001	0.002	0.002	0.001	0.001	0.001	0.001	0.002	0.007	0.007	0.007	0.007	0.006	0.006	0.008	0.007	0.006	0.007	0.006	0.006	0.006	0.009	0.002	0.001	0.002	0.003	0.002	0.003	0.003	0.002
0.015		0.002	0.002	0.002	0.004	0.004	0.004	0.003	0.004	0.003	0.004	0.004	0.002	0.003	0.002	0.002	0.002	0.002	0.003	0.002	0.002	0.002	0.002	0.002	0.002	0.005	0.002	0.003	0.003	0.003	0.002	0.002	0.003	0.002
0.043	0.01		0.005	0.006	0.008	0.008	0.008	0.008	0.008	0.007	0.008	0.008	0.001	0.002	0.002	0.001	0.001	0.002	0.002	0.002	0.002	0.001	0.001	0.002	0.001	0.005	0.005	0.007	0.006	0.006	0.004	0.004	0.004	0.005
0	0.012	0.04		0	0.002	0.002	0.002	0.001	0.001	0.001	0.001	0.002	0.006	0.006	0.005	0.006	0.005	0.005	0.006	0.006	0.005	0.006	0.005	0.005	0.005	0.008	0.002	0.001	0.002	0.003	0.002	0.002	0.003	0.002
0	0.013	0.042			0.002	0.002	0.002	0.001	0.001	0.001	0.001	0.002	0.006	0.007	0.006	0.006	0.006	0.005	0.007	0.006	0.006	0.006	0.005	0.005	0.005	0.009	0.002	0.001	0.002	0.002	0.002	0.002	0.003	0.001
0.011	0.028	0.063	0.011	0.01		0.001	0.001	0.001	0.002	0.002	0.002	0.002	0.009	0.009	0.008	0.008	0.008	0.008	0.009	0.008	0.008	0.009	0.008	0.007	0.008	0.011	0.004	0.002	0.002	0.004	0.004	0.004	0.006	0.003
0.009	0.029	0.064	0.01	0.009	0.006		0.001	0.001	0.001	0.002	0.001	0.002	0.009	0.009	0.008	0.008	0.008	0.008	0.009	0.008	0.008	0.009	0.008	0.008	0.008	0.01	0.004	0.002	0.003	0.005	0.005	0.005	0.006	0.003
0.012	0.032	0.068	0.013	0.012	0.006	0.004		0.002	0.003	0.003	0.002	0.002	0.009	0.009	0.009	0.009	0.009	0.008	0.01	0.009	0.009	0.009	0.008	0.008	0.008	0.011	0.004	0.002	0.003	0.005	0.004	0.005	0.006	0.003
0.006	0.027	0.062	0.006	0.005	0.005	0.003	0.007		0.001	0.001	0.001	0.002	0.009	0.009	0.008	0.008	0.008	0.008	0.009	0.008	0.008	0.008	0.008	0.008	0.008	0.01	0.003	0.001	0.002	0.004	0.004	0.004	0.005	0.002
0.007	0.028	0.062	0.008	0.006	0.006	0.007	0.01	0.004		0.002	0.001	0.002	0.009	0.009	0.008	0.008	0.008	0.008	0.009	0.008	0.008	0.008	0.008	0.007	0.008	0.011	0.003	0.001	0.003	0.004	0.003	0.004	0.005	0.002
0.006	0.023	0.051	0.007	0.006	0.014	0.011	0.014	0.009	0.009		0.001	0.003	0.008	0.009	0.007	0.007	0.007	0.007	0.009	0.008	0.007	0.008	0.007	0.007	0.007	0.01	0.003	0.002	0.003	0.003	0.002	0.003	0.004	0.002
0.008	0.028	0.063	0.008	0.007	0.006	0.007	0.01	0.003	0.004	0.008		0.002	0.009	0.008	0.008	0.008	0.008	0.008	0.009	0.008	0.008	0.008	0.008	0.008	0.008	0.01	0.004	0.001	0.003	0.004	0.004	0.005	0.005	0.003
0.011	0.025	0.057	0.01	0.01	0.011	0.01	0.007	0.008	0.013	0.013	0.011		0.008	0.009	0.008	0.008	0.008	0.008	0.009	0.008	0.008	0.007	0.008	0.007	0.007	0.011	0.004	0.002	0.003	0.004	0.004	0.004	0.007	0.004
0.052	0.015	0.004	0.05	0.053	0.073	0.074	0.08	0.073	0.075	0.064	0.073	0.068		0.002	0.001	0.001	0.001	0.002	0.002	0.002	0.002	0.001	0.001	0.001	0.001	0.004	0.005	0.008	0.007	0.006	0.005	0.005	0.005	0.006
0.055	0.019	0.01	0.049	0.054	0.071	0.073	0.078	0.072	0.074	0.069	0.071	0.069	0.006		0.001	0.001	0.001	0.002	0.001	0.001	0.002	0.002	0.003	0.001	0.003	0.003	0.006	0.009	0.008	0.007	0.006	0.005	0.005	0.006
0.051	0.016	0.009	0.046	0.051	0.071	0.069	0.074	0.069	0.072	0.065	0.071	0.066	0.004	0.003		0.001	0.001	0.001	0.002	0.001	0.001	0.002	0.002	0.001	0.002	0.004	0.005	0.008	0.007	0.007	0.005	0.005	0.005	0.006
0.051	0.016	0.007	0.045	0.049	0.068	0.068	0.073	0.068	0.069	0.061	0.067	0.065	0.005	0.003	0.002		0.001	0.001	0.002	0.002	0.001	0.002	0.002	0.001	0.002	0.004	0.005	0.008	0.008	0.007	0.005	0.005	0.005	0.006
0.047	0.015	0.007	0.043	0.046	0.067	0.065	0.071	0.065	0.066	0.058	0.064	0.062	0.005	0.006	0.003	0.002		0.001	0.002	0.002	0.001	0.002	0.002	0.001	0.002	0.004	0.005	0.007	0.007	0.006	0.005	0.005	0.004	0.005
0.047	0.013	0.007	0.041	0.043	0.064	0.064	0.067	0.062	0.066	0.057	0.065	0.061	0.01	0.007	0.005	0.005	0.005		0.002	0.001	0.001	0.002	0.002	0.001	0.002	0.004	0.005	0.007	0.007	0.006	0.005	0.004	0.004	0.005
0.059	0.023	0.013	0.052	0.057	0.073	0.074	0.08	0.074	0.077	0.071	0.073	0.072	0.01	0.005	0.01	0.009	0.011	0.01		0.001	0.002	0.002	0.002	0.001	0.003	0.004	0.006	0.008	0.008	0.007	0.006	0.005	0.005	0.006
0.049	0.015	0.008	0.044	0.047	0.067	0.065	0.069	0.064	0.068	0.059	0.065	0.062	0.01	0.008	0.007	0.011	0.01	0.006	0.007		0.001	0.002	0.002	0.001	0.002	0.005	0.005	0.008	0.007	0.007	0.005	0.005	0.005	0.005
0.048	0.015	0.008	0.043	0.046	0.068	0.067	0.07	0.065	0.069	0.059	0.068	0.063	0.009	0.008	0.004	0.007	0.006	0.001	0.012	0.007		0.001	0.002	0.001	0.002	0.004	0.005	0.008	0.007	0.006	0.005	0.004	0.004	0.005
0.046	0.013	0.003	0.045	0.046	0.071	0.07	0.073	0.066	0.067	0.056	0.068	0.061	0.002	0.01	0.006	0.006	0.006	0.008	0.015	0.01	0.005		0.001	0.002	0.001	0.004	0.006	0.008	0.007	0.006	0.005	0.005	0.005	0.006
0.045	0.011	0.001	0.042	0.043	0.064	0.065	0.071	0.064	0.063	0.054	0.063	0.059	0.002	0.009	0.008	0.006	0.005	0.008	0.01	0.011	0.009	0.004		0.002	0.001	0.004	0.005	0.007	0.007	0.005	0.004	0.004	0.004	0.005
0.045	0.014	0.005	0.041	0.044	0.062	0.061	0.065	0.061	0.063	0.057	0.063	0.059	0.006	0.006	0.003	0.004	0.003	0.003	0.01	0.008	0.004	0.008	0.008		0.002	0.005	0.004	0.007	0.007	0.006	0.004	0.004	0.004	0.005
0.047	0.012	0.003	0.044	0.045	0.066	0.068	0.07	0.067	0.066	0.054	0.065	0.059	0.003	0.01	0.008	0.006	0.007	0.008	0.014	0.012	0.008	0.003	0	0.006		0.005	0.005	0.007	0.006	0.006	0.005	0.004	0.004	0.005
0.073	0.042	0.031	0.067	0.072	0.093	0.085	0.094	0.086	0.09	0.085	0.087	0.089	0.028	0.02	0.02	0.02	0.022	0.022	0.026	0.029	0.022	0.026	0.026	0.027	0.03		0.008	0.01	0.011	0.01	0.008	0.008	0.008	0.009
0.013	0.015	0.033	0.01	0.01	0.025	0.022	0.023	0.018	0.022	0.02	0.024	0.024	0.039	0.039	0.037	0.038	0.037	0.031	0.04	0.032	0.034	0.037	0.035	0.03	0.035	0.06		0.002	0.003	0.003	0.003	0.003	0.004	0.002
0.004	0.021	0.053	0.004	0.004	0.011	0.009	0.014	0.006	0.007	0.01	0.005	0.012	0.062	0.063	0.06	0.058	0.054	0.055	0.064	0.057	0.057	0.057	0.054	0.051	0.057	0.083	0.013		0.002	0.004	0.003	0.004	0.004	0.002
0.011	0.016	0.043	0.012	0.01	0.012	0.018	0.018	0.013	0.012	0.017	0.015	0.018	0.055	0.058	0.055	0.056	0.053	0.049	0.061	0.051	0.051	0.049	0.048	0.048	0.047	0.086	0.016	0.012		0.004	0.003	0.003	0.005	0.002
0.015	0.019	0.037	0.016	0.013	0.028	0.03	0.035	0.023	0.021	0.018	0.026	0.027	0.048	0.055	0.051	0.05	0.047	0.047	0.055	0.046	0.045	0.04	0.036	0.045	0.04	0.077	0.018	0.017	0.019		0.001	0.001	0.002	0.001
0.012	0.013	0.027	0.013	0.01	0.026	0.029	0.029	0.023	0.022	0.015	0.024	0.025	0.039	0.046	0.042	0.039	0.036	0.036	0.045	0.036	0.035	0.033	0.029	0.034	0.034	0.066	0.016	0.014	0.017	0.004		0.001	0.002	0.001
0.017	0.012	0.026	0.014	0.013	0.028	0.031	0.035	0.024	0.025	0.021	0.03	0.027	0.036	0.039	0.036	0.035	0.035	0.03	0.037	0.033	0.03	0.032	0.028	0.032	0.03	0.061	0.015	0.021	0.017	0.006	0.007		0.002	0.001
0.023	0.018	0.029	0.023	0.021	0.042	0.043	0.05	0.035	0.036	0.031	0.038	0.045	0.034	0.039	0.036	0.034	0.032	0.03	0.038	0.035	0.028	0.03	0.029	0.031	0.032	0.061	0.025	0.027	0.026	0.015	0.014	0.01		0.001
0.009	0.012	0.031	0.011	0.009	0.019	0.022	0.026	0.016	0.017	0.015	0.017	0.023	0.038	0.044	0.042	0.039	0.037	0.037	0.043	0.038	0.036	0.034	0.031	0.035	0.035	0.068	0.015	0.011	0.01	0.006	0.006	0.008	0.006	

http://rcin.org.pl

Supplementary Table S6. Loci with an excess of low-frequency variants and significant deviation from neutrality based on Tajima's *D* (TD), HEW and DHEW tests. Significance: *P<0.05, **P<0.01

Gene	*P. mugo*	*P. sylvestris*	*P. uliginosa*	*P. uncinata*	PMZ	HPSZ	PSZ	PUZ	HBZ
Pr1_5	DHEW*	HEW**, DHEW*	HEW**, DHEW**	HEW*					
Pr1_9			HEW*, DHEW*						
Pr1_11	-1.809^{TD}**, HEW**, DHEW**								
Pr1_12		-2.000^{TD}**, HEW*, DHEW**							
Pr1_28	HEW*	-1.983^{TD}**	HEW*				HEW*		
Pr1_40	HEW*, DHEW*								
Pr2_3		HEW*							
Pr2_17							HEW*, DHEW*		
Pr2_28	-1.715^{TD}*, HEW*, DHEW**		-1.654^{TD}*		-2.016^{TD}**, HEW**, DHEW**				
Pr2_42	HEW*								
Pr2_47							1.761^{TD}*		
Pr4_4			-1.701^{TD}*						
Pr4_5	-1.736^{TD}*, HEW**, DHEW**		HEW*						-1.889^{TD}**, HEW*, DHEW**
Pr4_12		-1.960^{TD}**, HEW**, DHEW**							
Pr4_17		-2.181^{TD}**				-1.928^{TD}**, HEW**, DHEW**	-2.111^{TD}**		
Pr4_19		-1.633^{TD}*, HEW*, DHEW*							DHEW*
Pr4_27	HEW*, DHEW*		-1.729^{TD}*, HEW*, DHEW*						
Pr4_41		-1.640^{TD}*							

http://rcin.org.pl

Supplementary Table S7. Pairwise F_{ST} values between taxa and between groups of populations of the reference *Pinus mugo*, *P. sylvestris*, *P. uliginosa* and *P. uncinata* species. Statistically significant values (P<0.01) after correction for multiple testing are marked in bold.

	Between-species comparisons						***P. mugo* regional groups**			***P. sylvestris* regional groups**							
Gene	**PM-PS**	**PM-PUG**	**PM-PUN**	**PS-PUG**	**PS-PUN**	**PUG-PUN**	**PMCE-PMCP**	**PMCE-PMB**	**PMCP-PMB**	**PSCO-PCE**	**PSCO-PSSP**	**PSCO-PSSCN**	**PCE-PSSP**	**PCE-PSSCN**	**PSSP-PSSCN**	**PUG**	**PUN**
Pr1_5	**0.081**	-0.018	0.006	0.031	**0.091**	0.025	0.020	0.014	0.016	0.042	-0.020	0.053	0.055	-0.039	0.056	-0.035	0.001
Pr1_9	**0.086**	-0.013	0.000	0.080	**0.096**	-0.011	-0.006	-0.004	-0.020	**0.344**	0.100	**0.050**	0.206	0.117	0.005	-0.005	0.018
Pr1_11	0.019	-0.015	**0.027**	-0.008	0.033	0.067	-0.023	-0.024	-0.050	-0.016	0.004	0.020	0.050	-0.032	0.078	-0.122	0.003
Pr1_12	**0.057**	-0.003	0.015	0.058	**0.051**	0.059	-0.023	0.033	-0.012	-0.020	-0.019	0.053	-0.052	-0.014	-0.014	0.249	0.054
Pr1_13	0.006	-0.002	0.021	0.015	0.025	0.076	0.048	0.048	-0.047	-0.007	-0.005	-0.006	0.020	0.030	0.031	0.013	0.032
Pr1_19	0.018	-0.004	**0.058**	0.019	**0.066**	**0.220**	0.070	0.014	-0.001	0.050	-0.014	0.024	0.006	-0.034	-0.001	0.091	-0.005
Pr1_22	0.042	-0.016	**0.058**	0.006	0.064	**0.087**	**0.083**	0.077	-0.042	0.038	0.051	-0.007	-0.050	-0.022	-0.012	-0.021	0.055
Pr1_28	0.007	-0.021	0.012	-0.017	0.015	0.032	0.199	-0.013	0.157	0.040	-0.016	-0.006	0.052	0.092	0.002	-0.026	0.070
Pr1_40	**0.068**	**0.174**	**0.371**	0.011	0.029	**0.139**	0.063	-0.030	0.048	0.010	0.122	**0.109**	0.047	0.053	-0.038	0.197	0.007
Pr2_3	**0.160**	-0.015	0.007	**0.206**	**0.143**	**0.066**	-0.011	-0.014	-0.009	-0.026	0.144	-0.020	**0.195**	-0.021	0.120	0.060	-0.005
Pr2_4	0.016	0.060	**0.346**	-0.005	0.001	**0.205**	-0.011	0.000	0.000	-0.006	-0.007	-0.023	0.020	-0.016	-0.018	0.123	0.093
Pr2_17	0.051	-0.018	0.001	0.038	**0.055**	-0.016	0.017	0.048	0.003	-0.035	0.008	0.067	-0.001	0.048	0.129	-0.129	0.010
Pr2_23	0.031	-0.008	0.024	-0.014	-0.003	-0.006	0.137	0.017	0.004	-0.019	0.041	-0.023	0.000	-0.014	0.056	0.107	-0.024
Pr2_28	0.037	-0.002	0.022	0.042	0.029	**0.089**	-0.040	0.017	0.001	-0.021	-0.019	-0.007	-0.001	-0.024	0.026	0.011	0.054
Pr2_38	0.004	0.006	-0.011	0.002	0.005	-0.003	0.022	-0.035	-0.004	0.022	0.028	-0.001	0.134	-0.019	0.009	0.082	-0.001
Pr2_42	**0.070**	**0.334**	**0.396**	0.013	0.029	-0.015	0.051	-0.018	0.038	-0.033	0.024	-0.021	0.051	-0.026	0.009	-0.014	0.035
Pr2_47	0.003	-0.017	-0.001	-0.010	0.018	0.024	-0.036	-0.012	0.014	-0.024	-0.028	-0.024	-0.026	-0.025	-0.038	0.066	-0.002
Pr4_4	**0.083**	-0.010	**0.032**	0.024	0.051	0.008	0.006	-0.024	-0.011	-0.034	0.050	-0.022	0.050	-0.033	0.074	-0.016	0.012
Pr4_5	**0.053**	-0.009	**0.023**	0.022	0.024	0.000	0.116	0.022	0.003	-0.035	-0.021	0.074	-0.038	0.028	0.026	0.092	0.172
Pr4_10	**0.062**	**0.081**	**0.253**	-0.018	0.036	0.031	0.075	0.054	-0.030	-0.007	-0.007	0.044	0.053	0.080	0.080	-0.038	0.302
Pr4_12	0.052	-0.013	**0.062**	0.035	0.020	0.095	0.017	-0.006	-0.033	-0.003	-0.012	**0.113**	-0.006	0.054	0.055	0.040	0.119
Pr4_17	0.022	0.039	**0.089**	0.003	0.005	0.095	0.007	-0.023	0.001	0.052	0.001	0.045	0.001	-0.040	0.007	0.004	-0.013
Pr4_19	**0.146**	0.057	**0.227**	0.047	0.022	0.053	-0.020	-0.015	-0.030	0.033	-0.035	**0.016**	0.027	-0.017	-0.013	0.098	0.062
Pr4_21	0.006	-0.021	0.010	-0.020	0.028	0.027	0.016	0.016	0.008	-0.002	-0.004	0.016	0.052	-0.039	0.054	-0.037	0.018
Pr4_27	0.000	0.076	0.016	0.077	0.010	**0.153**	-0.040	-0.011	0.000	-0.017	**0.220**	**0.110**	**0.196**	0.054	0.058	-0.002	0.131
Pr4_41	**0.075**	0.001	**0.033**	0.065	**0.121**	0.031	0.093	0.018	0.019	-0.016	-0.016	-0.018	0.050	-0.017	-0.007	0.117	0.006

http://rcin.org.pl

Supplementary Table S8. Pairwise *Fst* values between groups of populations from Zieleniec reserve at 26 genes. Statistically significant values (P<0.01) after correction for multiple testing are marked in bold.

Gene	PMZ-HPSZ	PMZ-PSZ	PMZ-PUZ	PMZ-HBZ	HPSZ-PSZ	HPSZ-PUZ	HPSZ-HBZ	PSZ-PUZ	PSZ-HBZ	PUZ-HBZ
Pr1_5	**0.179**	**0.71**	0.012	0.015	**0.378**	**0.083**	**0.205**	**0.606**	**0.736**	0.011
Pr1_9	**0.352**	**0.674**	0.057	0.006	**0.132**	**0.152**	**0.226**	**0.476**	**0.573**	-0.012
Pr1_11	**0.092**	**0.159**	-0.007	-0.004	-0.013	**0.064**	**0.123**	**0.119**	**0.191**	0.02
Pr1_12	**0.075**	**0.384**	-0.004	-0.004	**0.187**	0.021	**0.045**	**0.313**	**0.378**	-0.012
Pr1_13	0.101	**0.216**	-0.019	-0.012	0.016	**0.08**	**0.056**	**0.196**	**0.158**	-0.023
Pr1_19	-0.013	**0.12**	0.041	-0.009	0.062	**0.094**	**0.025**	**0.318**	**0.213**	0
Pr1_22	0.101	**0.216**	-0.019	-0.012	0.016	**0.08**	**0.056**	**0.196**	**0.158**	-0.023
Pr1_28	**0.26**	**0.486**	-0.022	-0.022	**0.071**	**0.235**	**0.228**	**0.451**	**0.446**	-0.02
Pr1_40	**0.072**	**0.161**	-0.021	-0.005	**0.144**	**0.036**	0.025	**0.125**	**0.159**	-0.009
Pr2_3	**0.279**	**0.628**	0.017	-0.003	**0.145**	**0.17**	**0.197**	**0.525**	**0.567**	-0.017
Pr2_4	**0.157**	**0.061**	-0.019	-0.067	0	**0.111**	**0.132**	**0.029**	0.031	-0.074
Pr2_17	**0.177**	**0.305**	-0.006	0.125	0.07	**0.177**	**0.222**	**0.32**	**0.429**	0.043
Pr2_23	**0.146**	**0.382**	0.002	0.042	**0.098**	**0.078**	**0.029**	**0.293**	**0.213**	-0.005
Pr2_28	**0.202**	**0.508**	0	0.038	**0.188**	**0.134**	**0.109**	**0.422**	**0.373**	-0.008
Pr2_38	0.031	-0.012	0.008	0.01	0.032	**0.067**	**0.03**	**0.045**	-0.015	0.08
Pr2_42	**0.083**	**0.25**	0	-0.022	0.033	**0.135**	**0.082**	**0.32**	**0.257**	-0.005
Pr2_47	0.043	**0.148**	-0.011	0.001	0.025	0.035	**0.075**	**0.111**	**0.192**	-0.004
Pr4_4	**0.048**	**0.119**	-0.018	0.018	**0.063**	**0.058**	**0.07**	**0.148**	**0.244**	0.009
Pr4_5	**0.311**	**0.206**	0.022	-0.024	0.004	**0.172**	**0.335**	**0.071**	**0.237**	0.052
Pr4_10	**0.24**	**0.536**	-0.015	0.002	**0.293**	**0.218**	**0.22**	**0.55**	**0.617**	-0.022
Pr4_12	0.048	**0.306**	0.015	-0.008	**0.138**	**0.182**	**0.136**	**0.495**	**0.432**	-0.013
Pr4_17	**0.157**	**0.177**	0.034	0.001	0.007	**0.049**	**0.187**	**0.065**	**0.198**	0.076
Pr4_19	**0.196**	**0.49**	-0.014	-0.002	**0.129**	**0.166**	**0.136**	**0.453**	**0.483**	-0.017
Pr4_21	-0.001	0.025	0.024	-0.002	0.05	0.018	-0.016	**0.121**	**0.026**	0.03
Pr4_27	0.045	**0.059**	-0.023	0.016	0.007	0.028	**0.094**	**0.054**	**0.094**	0.006
Pr4_41	**0.099**	**0.296**	-0.01	0.05	**0.093**	**0.113**	**0.1**	**0.317**	**0.328**	0.014

http://rcin.org.pl

Supplementary Table S9. Pairwise *Fst* values between groups of pine samples from Zieleniec reserve and the reference *Pinus mugo*, *P. sylvestris*, *P. uliginosa* and *P. uncinata* species at 26 genes. Statistically significant values (P<0.01) are marked in bold.

Gene	PMZ-MCE	PMZ-SCE	PMZ_uliginosa	MZ-uncinata	HPSZ-mugo	HPSZ-sylvestris	HPSZ_uliginosa	HPSZ-uncinata	PSZ-mugo	PSZ-sylvestris
Pr1_5	-0.013	**0.521**	0.007	**0.047**	**0.244**	**0.178**	**0.093**	**0.142**	**0.722**	0.094
Pr1_9	0.020	**0.689**	0.022	0.012	**0.222**	**0.144**	**0.354**	**0.253**	**0.552**	-0.018
Pr1_11	**0.301**	**0.368**	**0.248**	**0.238**	**0.301**	**0.197**	**0.184**	**0.221**	**0.338**	**0.179**
Pr1_12	**0.074**	**0.437**	0.020	**0.131**	**0.099**	**0.249**	0.025	0.037	**0.353**	0.003
Pr1_13	0.062	**0.297**	**0.073**	**0.069**	**0.147**	**0.103**	**0.065**	**0.115**	**0.273**	0.063
Pr1_19	0.018	0.031	0.048	**0.075**	0.009	0.018	**0.092**	0.028	0.079	**0.115**
Pr1_22	0.062	**0.297**	**0.073**	**0.069**	**0.147**	**0.103**	0.065	**0.115**	**0.273**	0.063
Pr1_28	-0.005	**0.334**	0.008	**0.470**	**0.281**	**0.07**	**0.276**	**0.071**	**0.515**	0.073
Pr1_40	0.028	**0.177**	**0.056**	**0.137**	0.021	**0.2**	0.006	**0.230**	**0.212**	0.004
Pr2_3	**0.078**	**0.584**	**0.119**	**0.111**	**0.31**	**0.147**	**0.264**	**0.089**	**0.648**	0.046
Pr2_4	0.051	**0.222**	0.043	**0.115**	**0.308**	-0.001	**0.290**	-0.016	**0.205**	0.058
Pr2_17	**0.154**	**0.364**	**0.098**	**0.089**	**0.216**	**0.056**	**0.129**	**0.082**	**0.389**	0.000
Pr2_23	0.027	**0.363**	0.075	**0.347**	**0.275**	**0.089**	0.029	**0.067**	**0.53**	0.034
Pr2_28	0.012	**0.48**	0.003	**0.122**	**0.172**	**0.209**	**0.183**	**0.096**	**0.472**	0.007
Pr2_38	-0.004	**0.184**	**0.133**	-0.001	0.029	0.039	**0.088**	0.031	-0.025	**0.111**
Pr2_42	**0.122**	**0.193**	0.014	**0.055**	**0.32**	0.018	0.024	0.017	**0.502**	0.022
Pr2_47	0.015	**0.083**	0.039	**0.140**	**0.053**	0.040	0.035	**0.126**	**0.186**	0.004
Pr4_4	**0.115**	**0.152**	**0.178**	**0.064**	**0.15**	**0.081**	**0.179**	**0.049**	**0.264**	0.019
Pr4_5	0.136	**0.338**	0.021	**0.081**	**0.429**	-0.011	**0.172**	0.058	**0.36**	**0.053**
Pr4_10	0.066	**0.588**	**0.089**	**0.055**	**0.31**	**0.325**	0.021	**0.145**	**0.748**	0.017
Pr4_12	0.018	**0.192**	0.035	0.008	**0.139**	0.003	**0.205**	0.026	**0.409**	0.069
Pr4_17	**0.093**	**0.189**	**0.130**	**0.145**	**0.237**	-0.002	**0.229**	**0.059**	**0.24**	0.009
Pr4_19	**0.163**	**0.507**	-0.006	**0.093**	**0.42**	**0.148**	**0.138**	0.015	**0.63**	0.083
Pr4_21	**0.447**	**0.463**	**0.475**	**0.531**	**0.47**	**0.482**	**0.515**	**0.557**	**0.493**	**0.477**
Pr4_27	0.051	**0.126**	0.013	**0.152**	**0.093**	**0.109**	**0.072**	**0.178**	0.053	**0.075**
Pr4_41	0.018	**0.273**	**0.125**	**0.182**	**0.17**	**0.08**	**0.139**	**0.184**	**0.389**	0.001

http://rcin.org.pl

Gene	PSZ_uliginosa	PSZ-uncinata	PUZ-mugo	PUZ-sylvestris	PUZ_uliginosa	PUZ-uncinata	HBZ-mugo	HBZ-sylvestris	HBZ_uliginosa	HBZ-uncinata
Pr1_5	**0.617**	**0.625**	0.032	**0.432**	-0.014	-0.002	-0.005	**0.586**	0.017	0.011
Pr1_9	**0.683**	**0.550**	0.007	**0.504**	0.073	0.004	-0.020	**0.596**	0.013	-0.018
Pr1_11	**0.203**	**0.247**	**0.281**	**0.336**	**0.218**	**0.214**	**0.309**	**0.378**	**0.269**	**0.260**
Pr1_12	**0.314**	**0.222**	**0.089**	**0.37**	-0.001	**0.095**	0.094	**0.437**	-0.010	**0.119**
Pr1_13	**0.122**	**0.163**	0.033	**0.341**	**0.059**	**0.087**	**0.047**	**0.309**	0.048	**0.076**
Pr1_19	**0.314**	0.009	**0.14**	**0.121**	0.002	**0.250**	0.046	0.062	0.023	**0.157**
Pr1_22	**0.122**	**0.163**	0.033	**0.341**	**0.059**	**0.087**	**0.047**	**0.309**	0.048	**0.076**
Pr1_28	**0.478**	**0.040**	-0.008	**0.324**	0.008	**0.438**	-0.001	**0.312**	0.014	**0.447**
Pr1_40	**0.062**	**0.068**	0.023	**0.15**	0.022	**0.135**	**0.070**	**0.186**	0.037	**0.157**
Pr2_3	**0.585**	**0.328**	0.050	**0.483**	**0.063**	0.036	0.069	**0.522**	**0.078**	**0.062**
Pr2_4	**0.187**	-0.004	0.062	**0.181**	0.056	**0.084**	**0.149**	**0.194**	-0.006	**0.086**
Pr2_17	**0.301**	**0.204**	0.048	**0.313**	0.039	0.051	-0.030	**0.472**	0.000	0.037
Pr2_23	**0.202**	0.015	**0.092**	**0.275**	0.014	**0.252**	**0.16**	**0.204**	-0.011	**0.163**
Pr2_28	**0.482**	**0.284**	0.016	**0.414**	0.004	0.061	**0.033**	**0.36**	0.015	0.024
Pr2_38	**0.087**	-0.012	0.112	**0.281**	**0.232**	0.051	-0.025	**0.102**	0.043	-0.006
Pr2_42	**0.140**	**0.095**	**0.144**	**0.28**	0.081	**0.129**	**0.101**	**0.201**	0.021	**0.066**
Pr2_47	**0.168**	**0.229**	0.026	0.051	0.058	**0.108**	-0.010	**0.135**	0.042	**0.090**
Pr4_4	**0.431**	**0.029**	0.081	**0.168**	**0.138**	**0.072**	0.053	**0.145**	0.115	**0.055**
Pr4_5	**0.074**	0.004	**0.202**	**0.25**	-0.021	0.015	0.071	**0.392**	0.041	**0.096**
Pr4_10	**0.272**	**0.268**	0.023	**0.599**	**0.078**	**0.083**	-0.004	**0.666**	0.074	**0.101**
Pr4_12	**0.475**	**0.205**	-0.011	**0.41**	-0.018	**0.092**	0.024	**0.325**	-0.007	**0.056**
Pr4_17	**0.223**	**0.066**	**0.158**	**0.073**	**0.174**	**0.078**	0.057	**0.215**	0.045	**0.131**
Pr4_19	**0.371**	**0.159**	**0.202**	**0.474**	0.009	**0.079**	**0.261**	**0.485**	0.018	**0.050**
Pr4_21	**0.523**	**0.553**	**0.437**	**0.47**	**0.485**	**0.550**	**0.443**	**0.454**	**0.483**	**0.535**
Pr4_27	**0.085**	**0.172**	0.049	**0.124**	0.009	**0.151**	0.009	**0.161**	0.015	**0.150**
Pr4_41	**0.361**	**0.355**	-0.003	**0.295**	0.074	**0.130**	0.047	**0.315**	0.017	**0.075**

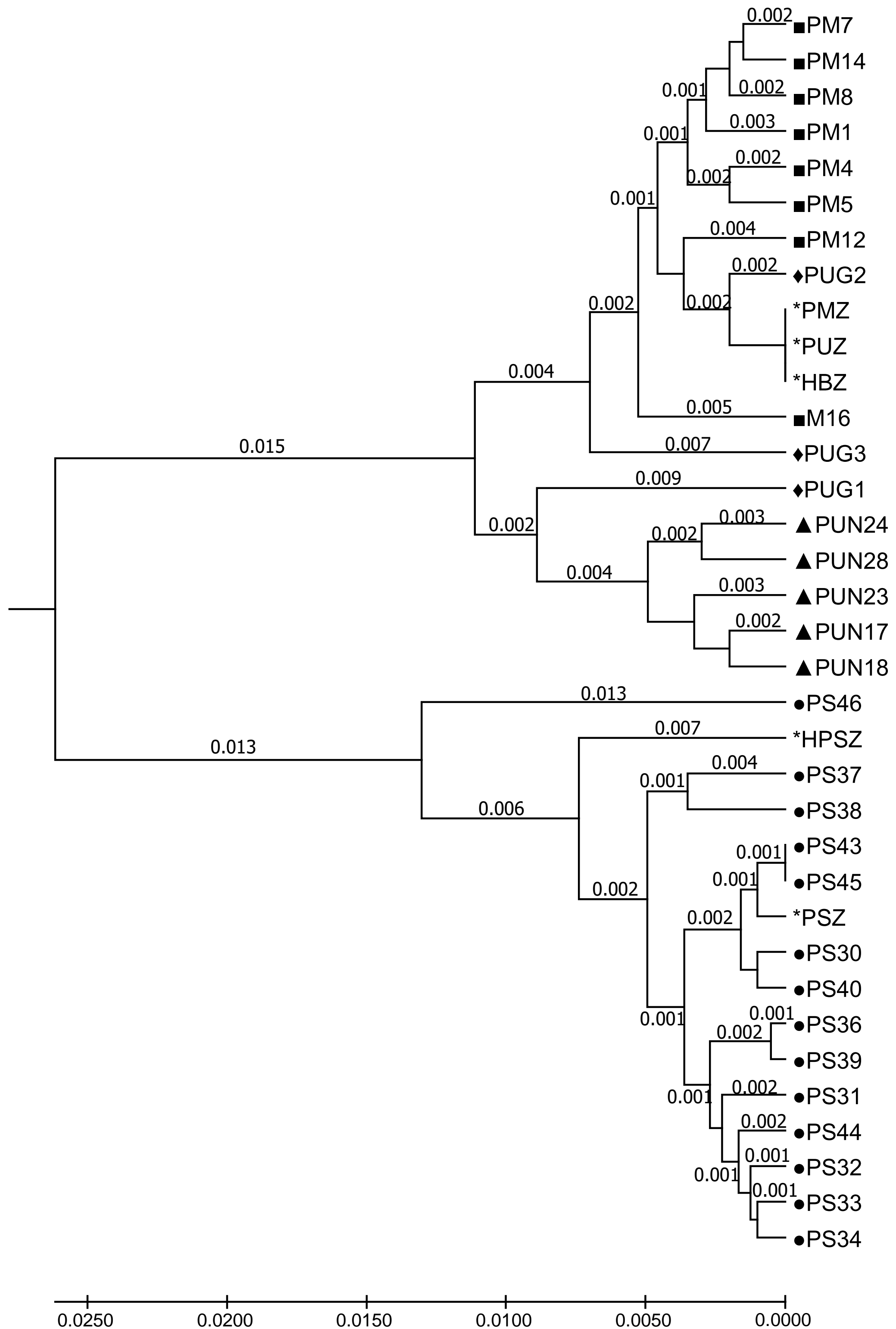
http://rcin.org.pl

Supplementary Table S10. Number of most diverged SNPs and indels of highest frequency difference (>0.75) between species.

Gene	*P. mugo* vs. *P. sylvestris*	*P. mugo* vs. *P. uliginosa*	*P. mugo* vs. *P. uncinata*	*P. sylvestris* vs. *P. uliginosa*	*P. sylvestris* vs. *P. uncinata*	*P. uliginosa* vs. *P. uncinata*
Pr1_5	6			5	5	
Pr1_9	3			3	2	
Pr1_11	1				1	
Pr1_12	2			4	2	
Pr1_13	1			1		
Pr1_19	1			1	1	
Pr1_22	2		2	2	1	
Pr1_28	1		2			3
Pr1_40	2		1			
Pr2_3	3				2	
Pr2_17	2			2		
Pr2_23	4		3	2		
Pr2_28	2			2		
Pr2_38	2				2	
Pr2_42	1	4	4			
Pr2_47	1			1	2	
Pr4_4	3			3		3
Pr4_5	3					
Pr4_10	1	1		1	1	
Pr4_12				4		
Pr4_17	1			2		
Pr4_19	2			3		
Pr4_41	2			4	5	

http://rcin.org.pl

Supplementary Figure S1.
Unweighted Pair Group Method with Arithmetic Mean (UPGMA) tree based on net genetic distance between groups of samples from Zieleniec reserve (*) and reference populations of the species including *Pinus mugo* (■), *P. sylvestris* (●), *P. uliginosa* (♦), *P. uncinata* (▲) (see Table 1 for details). Numbers indicate branch length. Genetic distance and its standard errors for each pairwise comparisons are shown in Supplementary Table S5.



http://rcin.org.pl

Appendix file

FASTA format file of the polymorphic sites detected at 26 genes in the groups of pines defined at Zieleniec reserve and reference populations of pure *Pinus mugo*, *P. sylvestris*, *P. uliginosa* and *P. uncinata* species. See Table 1 for details.

Available online:
https://static-content.springer.com/esm/art%3A10.1007%2Fs11295-016-1007-x/MediaObjects/11295_2016_1007_MOESM2_ESM.txt

last accessed: 22/02/17

http://rcin.org.pl