Article (refereed) - postprint

Wachowiak, Witold; Wojkiewicz, Blazej; Cavers, Stephen; Lewandowski, Andrzej. 2014. **High genetic similarity between Polish and North European Scots pine (Pinus sylvestris L.) populations at nuclear gene loci**. *Tree Genetics & Genomes*, 10 (4). 10.1007/s11295-014-0739-8

Witold Wachowiak[1,2,3*], Błażej Wójkiewicz[2], Stephen Cavers[3], Andrzej Lewandowski[2]

High genetic similarity between Polish and North European Scots pine (*Pinus sylvestris* L.) populations at nuclear gene loci

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

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

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

***Corresponding authors:** Witold Wachowiak (witoldw@man.poznan.pl), Institute of Dendrology, Polish Academy of Sciences, Parkowa 5, 62-033 Kórnik

Abstract

Nucleotide polymorphisms in a set of 32 nuclear genes was studied in nineteen mountain, peatbog and lowland populations of Scots pine representing known phenotypic races and populations of presumably relict character for the species in Poland. At 29 genes, the pattern of genetic variation was compared to 11 reference populations from Northern, Western and Southern Europe. Similar levels of nucleotide polymorphism and excess of low frequency mutations was observed in Polish populations (π_{tot} =0.0055, D= -0.308) and as compared to the reference samples (π_{tot} =0.0054, D= -0.170). Bayesian assignment and conventional frequency based statistics indicate that Polish populations share the same genetic background at the analysed nuclear gene markers. However, the populations showed a much closer genetic relationship with North European samples than other regional groups of populations. Across the very uniform genetic background of the populations we identified several genes with outlier patterns of haplotype, polymorphism frequency variation and departures from compound neutrality tests. Our data indicate that the Central and North European parts of the Scots pine distribution seem particularly suitable for association genetic studies to link phenotypic and genetic variation at a large geographical scale.

Keywords: nucleotide polymorphisms, population structure, natural selection, recolonization

INTRODUCTION

During their evolution, plants have coped with environmental changes by migration and local adaptation to new environments. Polymorphisms at nuclear loci provide highly valuable insights into evolutionary history of the species (Wright and Andolfatto 2008). Different forces influence patterns of nucleotide variation including mutation, recombination and random genetic drift. Across a species' distribution range natural selection at loci involved in adaptation can result in deviations at some regions observed for instance as a reduced diversity or difference in allele frequency spectra as compared to species background genetic variation (Savolainen and Pyhäjärvi 2007; Neale and Ingvarsson 2008; Achaz 2009). Historical processes related to population size changes and range shifts are also known to drive deviations from neutrality that may mimic the effect of selection. However, as demographic processes influence genome-wide patterns of genetic variation, effects of selection can be detected at individual loci as outlier patterns of nucleotide variation. This is particularly expected of highly outcrossing, wind-pollinated forest tree species which experience the homogenizing effect of efficient gene flow on the pattern of genetic diversity across large geographical ranges.

Scots pine is the most widely distributed conifer in the world with a huge distribution range in Europe and Asia, and is of high ecological and economic importance. The present day distribution of this species and other plants in Europe has been largely shaped by recolonization from refugial populations after last glacial maximum (25-18 000 years ago (ya)) (Petit et al. 2003). So far palynological and phylogeographic data indicate that Scots pine survived in southern European peninsulas (Iberia, Italy and the Balkans) and probably in some parts of eastern and central Europe (Cheddadi et al. 2006; Blarquez et al. 2010; Willis and van Andel 2004) and only reached their most northerly limits around 9000 ya. However, not all putative southern refugial populations participated in recolonization as suggested by studies of some mitochondrial DNA polymorphisms (Soranzo et al. 2000; Pyhäjärvi et al. 2008; Naydenov et al. 2007). Nucleotide sequence analysis at several nuclear loci also provided little evidence for the effects of recent (post-glacial) population size changes during migration and suggested bottlenecks in the mid-to-late Pleistocene (Pyhäjärvi et al. 2007). Nucleotide variation in Scottish populations was not compatible with a simple recolonization model (Wachowiak et al. 2010) and some evidence for possible northern refugia has been suggested by ancient DNA studies (Parducci et al. 2012). Regardless of migration trajectories, postglacial expansion of the species was accompanied by adaptation to local environments as evident from quantitative genetic studies in Scots pine that showed large variation of morphological (growth dynamics, wood quality), physiological, growth traits and stress response across the environmental gradient related to photoperiod and temperature (Salmela et al. 2013; Hurme et al. 1997). This phenotypic differentiation is accompanied by relatively low variation in neutral markers, e.g. between Finnish populations

(Karhu et al. 1996), between Scandinavian and eastern parts of the range (Wang et al. 1991) and at several genes among populations along a latitudinal cline in continental Europe (Dvornyk et al. 2002; Garcia-Gil et al. 2003). However, considering the low resolution of the markers developed so far, the species population history in Europe is far from fully described. The same is true for our understanding of the genetic basis of phenotypic and adaptive variation, where evidence for selection has so far been found in only several gene fragments (Wachowiak et al. 2009; Kujala and Savolainen 2012).

In Poland Scots pine is the most important forest-forming tree species. In the lowlands, the proportion of pine in the forests is about 70% (Boratyński 1993). In the highlands, pine is used as an admixed and a pioneer crop or nurse species especially in fir forests and it constitutes about 24% of total forest area in Carpathian and Sudeten Mountains. Due to their high ecological and economic importance, populations of the species were extensively studied for their breeding value. Provenance trial experiments of Polish populations from different geographical locations showed high diversity in adaptive morphological and physiological traits, growth dynamics and stress resistance (Boratyński 1993; Giertych and Oleksyn 1992; Sabor 2006). Based on morphological and growth performance traits, several local ecotypes of the species were distinguished across the distribution range (Boratyński 1993). So far genetic studies of the species have applied some molecular and biochemical markers to estimate genetic variation within and among populations in relation to the mating systems (Burczyk et al. 2000) and natural hybridization in contact zones of the species with the taxa from the *P. mugo* complex (Wachowiak and Prus-Głowacki 2008). More recently, a single population from Poland was included in studies that used nucleotide polymorphism to test historical demographic scenarios (Pyhäjärvi et al. 2007) and selection at candidate genes (Wachowiak et al. 2009; Kujala and Savolainen 2012). Due to the poorly resolved recolonization history of the species, the origin of most Polish stands, including in particular the putatively relict mountain populations, is unknown. Furthermore, genetic relationships between presumed native populations and other parts of the European distribution, and the influence of demographic and selection processes on patterns of differentiation have not been studied so far.

In this study we looked at the genetic variation in populations of Scots pine from across Poland. We focused on putative native populations and used nucleotide sequence variation from multiple loci to analyse them in reference to populations from Northern, Western and Southern Europe. We aimed to assess whether Polish populations showed patterns of genetic differentiation locally or relative to other European samples, and whether selection was evident.

MATERIAL AND METHODS

Study locations and DNA extraction

Seeds of Scots pine were sampled in nineteen natural populations of the species in Poland and eleven reference populations of the species in Europe (Figure 1, Table 1.). Sampling of Polish populations was focused on most likely native stands represented by samples of known relict character including many mountain and peatbog locations from National Parks and natural reserves. Although much effort was made to choose natural stands, the non-native origin of some western lowland populations of the species cannot be completely ruled out. A century ago those territories were under German forest management and the use of non-local seed stocks is possible (Krakau et al. 2013). However, there are no available DNA markers that would allow precise verification of the Scots pine populations in reference to their recolonization routes and postglacial origin (Naydenov et al. 2007; Pyhäjärvi et al. 2008). Reference populations of the species included stands from Northern Europe, Scotland and Spain. In total, 297 samples were analysed, comprising of ten different trees from most locations (Table 2&3). Genomic DNA was extracted from megagametophytes (maternally derived haploid tissue surrounding the embryo, which in gymnosperms has the same genotype as the egg cell) from germinated seeds using DNeasy Plant Mini Kit (Qiagen).

A total of thirty two gene fragments including loci related to cellular metabolism, transport, signal transduction and transcription regulation were used for sequencing (Supplementary Table S1). PCR-amplification was performed with Thermo MBS thermal cyclers and carried out in a total volume of 15µl containing about 15ng of haploid template DNA, 10µM of each dNTP, 0.2µM each of forward and reverse primers, 0.15U Taq DNA polymerase, 1x BSA, 1.5 µM of $MgCl_2$ and 1x PCR buffer (BioLabs). Standard amplification procedures were used with initial denaturation at 94°C for 3 min. followed by 35 cycles with 30 sec. denaturation at 94°C, 30 sec. annealing at 60°C and 1 min. 30 sec. 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 20ng of PCR product was used as a template in 10µl sequencing reactions with the Big Dye Terminator DNA Sequencing Kit (Applied Biosystems, Carlsbad, CA, USA) performed by the Genomed (Poland, Warsaw) and GenePool (UK, Edinburgh) sequencing service. Multilocus haplotypes were determined by direct sequencing of haploid DNA; samples were sequenced in one direction for each locus. As the rate of low frequency mutations is comparable to other reports in conifers (see below), we believe that potential sequencing errors resulting from one direction sequencing are unlikely to bias the results. 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. Some insertion/deletions were manually adjusted across the samples using GenDoc. The haplotypes at

each locus reported in this paper were submitted to NCBI sequence database (Acc. Nr. KJ502673-KJ503230). Reference sequences from several species including *P. pinaster, P. taeda, P. mugo* and *P. uliginosa* were used for outgroup comparisons.

Nucleotide and haplotype polymorphisms

About 15 kbp of nuclear DNA sequence was aligned across genes. The pattern of nucleotide polymorphism was analysed in individual populations and regional groups from Poland, and in comparison to three reference geographical locations from Northern Europe, Scotland and Spain. Nucleotide diversity was measured as the average number of nucleotide differences per site (π) between two sequences (Nei 1987). Multilocus estimates of Watterson's (1975) 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 generation) were computed based on the number of total and/or silent segregating sites and the length of each locus. The non-linear least-squares estimate of ρ ($\rho=4N_ec$, were N_e is effective population size, c is the recombination rate) was fitted by the *nls*-function implemented in *R* (www.r-project.org) based on the correlation coefficient r^2 (Hill and Robertson 1968) between polymorphic informative sites and the distance in base pairs (bp) between sites. The ratio of recombination to mutation rates (ρ/θ) was calculated to estimate the relative influence of those factors on patterns of diversity observed in geographical regions defined for the species. The number of haplotypes (N_e) 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 et al. 2005). Locus-by-locus estimates of net divergence between Scots pine geographical regions (Nei 1987), the number of shared, exclusive and fixed polymorphic sites and haplotypes for each locus were determined using SITES 1.1. (www.bio.cst.temple.edu/~hey/program_files/Sites). Deviations from the frequency distribution spectrum expected under the standard neutral model of evolution were assessed using frequency spectrum test and coalescence based approaches (Tajima 1989). The distribution of Tajima's *D* test statistics was investigated for each population and regional group. The significance of multilocus estimates of the test statistics were evaluated by comparison to a distribution generated by 1000 coalescent simulations using the HKA program (www.bio.cst.temple.edu/~hey/program_files/HKA). Deviations of particular genes from site frequency distribution spectra expected under the standard neutral model of evolution were investigated using two compound neutrality tests, HEW and DHEW (Zeng et al. 2007). Significance levels of these tests were determined by carrying out 10 000 coalescent simulations based on Watterson's estimator of theta as implemented in dh package. Orthologous sequences from the outgroup species were used in the Hudson-Kreitman-Aguadé (HKA) test (Hudson et al. 1987) to look

for overall departures from neutral expectations by assessing the level of multilocus polymorphism and divergence.

Population structure

Genetic differentiation was measured as Wright's fixation index (Weir and Cockerham 1984), F_{ST}, over all polymorphic sites and tested for significance by 1000 permutations of the samples between populations and regional groups (Excoffier et al. 2005). The hierarchical distribution of multilocus genetic variation among populations and regional groups was estimated using an analysis of molecular variance (AMOVA) in Arlequin v.3. 5. Clustering analysis was applied to look at the relationships between individual Polish populations, regional groups and reference geographical locations 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 and 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. The number of populations was inferred as the combined maximum likelihood and highest posterior probability estimates over all runs. To further assess among-population differentiation we used Principal Coordinate Analysis (PCoA) based on the F_{ST} matrix of pairwise comparisons between populations across all polymorphic sites.

Tests for outlier pattern of differentiation at the loci

Genetic differentiation in pairwise comparisons between individual populations and geographical groups of the species was studied locus by locus at both haplotype and SNP level. The significance of genetic differentiation at the loci was evaluated by 1000 permutations of the samples between populations and regional groups using Arlequin ver. 3.5 (Excoffier et al. 2005). Among population and group haplotype diversity was estimated by the nearest neighbour S_{nn} statistics (Hudson et al. 1992; Lynch and Crease 1990) and tested for significance using 1000 permutations, where samples were randomly assigned into different groups (Hudson 2000). The full SNP dataset was used to test for loci under selection using the hierarchical analysis of Excoffier et al. (2009) in Arlequin v3.5. Coalescent simulations were used to estimate a null distribution and confidence intervals around the observed values, which allowed identification of outliers among locus-specific F_{ST} values conditioned on the multilocus distribution of F_{ST} values. Each simulated group consisted of 100 subpopulations, and 20,000 replicates of the coalescent were used to identify the expected distribution of F_{ST}. The significance thresholds of the F_{ST} values were set at 95% and 99% of the F_{ST} values from the simulated data (Excoffier and Lischer 2010).

RESULTS

Genetic variation in Polish populations

From the 32 genes studied in 19 Polish populations ~15 kbp were aligned across the loci providing a set of 606 polymorphic sites. Very similar average nucleotide diversity (π) was found in the populations with total π_{tot} =~0.0051-0.0060 and average π_{tot} = 0.0055 (Table 2). No significant difference between populations was found for multilocus nucleotide diversity, which ranged between θ_{tot}=0.0052 to θ_{tot}=0.0073 (average θ_{tot}=0.0058) (Table 2). The regional groups of Polish populations showed very similar diversity to the reference populations with π_{sil} =~0.0083-0.0068 and θ_{silt}=0.0053 to θ_{silt}=0.0068 (Table 3, Supplementary Figure 1). No divergence was observed between Polish regional groups of Scots pine (Supplementary Table S2). Net divergence was marginal for Polish and Northern European populations (0.0001) and it was lower as compared to Scottish and Spanish groups (0.0002-0.0005; Supplementary Table S2).

Between 63-77% of all SNPs detected were shared in pairwise comparisons between the 19 Polish populations. There was a large and similar proportion of the polymorphic sites shared across the regional groups of populations in Poland (68-77%) and as compared to the reference populations (61-73%) (Supplementary Table S3). Polish populations shared slightly higher proportion of SNPs with Northern European samples. As for individual SNPs, each of the Polish populations and regional groups of Scots pine had similar proportions of shared haplotypes out of total number detected (~21-26% and 41-52%), except Spanish group (~25%). Unique haplotypes were found in each Polish population and regional groups of the species (Table 2&3). The average haplotype diversity was similar for 19 Polish populations (H_d =0.59-0.71, average 0.66) and regional groups of Polish and reference populations (H_d=0.60-0.68, average 0.65) (Table 2&3). Relative to neutral expectations, an overall excess of singleton mutations was detected, as indicated by significantly negative multilocus Tajima's *D* in most Polish groups and North European samples (*D*= -0.415 to -0.772; P<0.05) (Table 3). Both compound neutrality tests provided evidence for selection at 6 loci including Pr1_5, Pr1_46, PhytP, Pr4-12, Pr4-19, Pr4_6. Overall, a positive correlation between intraspecific polymorphism and interspecific divergence at the loci was found in pairwise comparisons between each regional group and outgroup species in a multilocus HKA test.

The decay of linkage disequilibrium between pairs of parsimony informative sites at 32 loci in *P. sylvestris* from five regions in Poland was rapid with the expected r^2 values of 0.2-0.4 at a distance of about ~200 bp (Supplementary Figure S2). The decay of LD in western Polish populations (ρ=0.0122±0.0013, r2=0.2 at ~200bp) was slightly more rapid as compared to other locations. In reference populations, the slowest decay of LD was observed in *P. sylvestris* from Spain (ρ=0.0007±0.0004) as compared to other regions (Table 3, Supplementary Figure S2).

Correspondingly, the absolute values of the recombination to diversity ratio were highest ($\rho / \theta > 1$) for Polish populations from southern and western part of the country, similar for other Polish populations and Scotland and slightly lower ($\rho / \theta < 0.1$) for populations from Northern Europe and Spain.

Population structure and differentiation

The close genetic similarity between Polish populations was evident in the clustering analysis using BAPS software, which suggested the presence of one genetic cluster. Similarly, only one cluster was found when Polish populations were analysed together with reference Scots pine samples. Greater similarity between Polish and North European populations as compared to other reference samples from Scotland and Spain was found in the Principal Coordinate Analysis (Figure 2). At all polymorphic sites combined differentiation between Polish populations was found only between Hajnówka and Gubin populations (F_{ST}=0.07, P<0.01). No significant differentiation measured as average F_{ST} across all polymorphic sites was found between any of the Polish geographical regions (Table 4). Polish populations were also not differentiated from Northern European samples (except only south-western region with low but significant F_{ST}=0.018, p<0.05). In contrast, all regional Polish populations showed significant differentiation as compared to Scottish and Spanish populations (F_{ST}=0.032-0.086, p<0.05) (Table 4). In a hierarchical AMOVA of Polish Scots pine samples, 98.6% and 99.9 % of differentiation was found within individual populations and regional groups, respectively (Supplementary Table S4). As compared to the reference samples, most variation was found within populations ranging from 91.7 % to 98% for Polish vs. Spanish and Northern Europe samples, respectively (Supplementary Table S4).

Variation at individual loci and polymorphic sites

In contrast with the lack of population genetic structure, clear differentiation in pairwise comparisons between some Polish populations was found at several loci (Supplementary Table S5). The highest number of significant pairwise differentiations was found for three mountain populations (Tarnawa, Koryciska Wielkie and Szczeliniec) and single lowland population from Gubin. Differentiation in the allelic frequency spectra between regional groups of Polish populations was found only between southern vs. south-west and western populations, and between northern and eastern populations (Supplementary Table S6). A consistent pattern of significant population differentiation based on haplotype structure, allelic frequency spectra and outlier polymorphisms was found at Pr 4_6 locus (Nucleus Dicer) (Supplementary Table S7). This locus showed also departures from neutrality based on compound neutrality tests**.**

Strong outlier patterns of variation for both haplotype structure and allelic frequency spectra (p<0.01) were found at six loci in pairwise comparisons between Polish populations and reference samples

(Table 5). These genes are involved in plant metabolism (alpha-N-acetylglucosaminidase, DEAD-box ATP-dependent RNA- helicase), signal transduction (U-box domain-containing protein, transducin/WD40 domain-containing protein, phytochrome) and expression regulation (putative auxin induced-transcription factor). Four of the loci also showed departures from neutrality based on compound neutrality tests (Supplementary Table S7). At PhytP Polish populations lack 11bp deletion and two associated SNPs present in all other reference populations of the species.

Despite shared genetic background between Polish and North European samples we identified 11 outlier SNPs across four loci (Pr1_4, Pr11_4, Pr27_4, PhytP), for which frequency significantly differed between regions (Supplementary Table S8). Using the outlier detection approach, 35 SNPs were identified in 13 genes in Polish vs. Scottish group and 23 SNPs in 11 genes in Polish vs. Spanish populations (Supplementary Table S9). Most (74%) of the total of 49 outlier SNPs detected were unique to corresponding group comparisons. These loci represent genes from all categories including regulation of gene expression, signal transduction, cellular metabolism and transport.

DISCUSSION

Nucleotide polymorphisms and population structure

In this study, we sampled populations of Scots pine from a wide distribution range across Europe, with a focus on native Polish populations derived from contrasting mountain, peatbog and lowland environments likely to force adaptive responses. We studied nucleotide polymorphism at a set of nuclear loci to investigate patterns of genetic variation between Polish populations, and in reference to samples from North, West and Southern Europe. Our data provide no evidence of any difference in the amounts of nucleotide diversity in Polish populations. Despite differences in geographical location and census size of the populations (small, relict mountain and peatbog populations vs. large lowland stands) they showed very similar levels of nucleotide polymorphism to those observed in other continental Scots pine populations (Pyhäjärvi et al. 2007; Wachowiak et al. 2009). Similar high nucleotide and haplotype diversity levels were accompanied by a very similar genetic background, as shown by many shared polymorphic sites and alleles resulting in zero net divergence. Nucleotide polymorphism at the vast majority of the loci was therefore mostly due to neutral evolutionary processes and population history of the species as evident from the random distribution of nucleotide variation across populations and overall positive correlation between intraspecific polymorphism and divergence from the outgroup species. Genetic variation in Polish populations fitted expectations of recent population expansion, i.e. an excess of rare alleles and low frequency mutations, presence of unique polymorphic sites in a similar proportion in every population and equally high nucleotide

diversity across different locations. Such genetic similarity suggests common population history accompanied by efficient gene flow.

Overall, we observed very low genetic variation with no evidence of any outlier Polish population of significant difference in haplotype or SNP frequency spectra across many genes. This pattern of genetic similarity is different to what was observed at much smaller geographical scale in Scottish populations, which showed high among population heterogeneity at many individual nuclear gene loci due to high spatial environmental heterogeneity and likely admixture of populations of different postglacial history (Wachowiak et al. 2013b). The high similarity between Polish populations follows the pattern expected for large continuous populations exposed to consistent high gene flow between populations. This pattern also supports previous observations of generally low genetic variation at neutral markers between populations from the continuous range (Wachowiak et al. 2009). The environmental heterogeneity among the Polish stands (mountain, peatbog and lowland populations) does not seem to have any effect on the level of genetic diversity within populations. Within the scope of this dataset, we can conclude that the populations share the same postglacial history.

In reference to the other European samples studied, we observed no heterogeneity in the pattern of nucleotide diversity among Polish and North European populations. Considering the very close genetic relationship between these regions, it seems likely that these parts of the species distribution have experienced similar demographic history. More differentiated were the other two parts of the Scots pine distribution as evident from an overall significant difference in allelic frequency spectra and higher divergence as compared to northern group. In case of populations from north-western part of the distribution, previous work using candidate nuclear genes has indicated high diversity in Scottish populations and patterns of allelic frequency incompatible with a simple colonisation and expansion model (Wachowiak et al. 2010). Some distinct characters have been observed in Spanish populations, for instance at *mt*DNA fragments and some isozyme loci, which suggests they have been isolated and contributed little to the recolonization of the rest of Europe (Soranzo et al. 2000; Prus-Glowacki et al. 2012). Our data provide additional evidence of the distinct character of both of these groups. However, assuming that their population history is different (predating at least partially the last glacial period of >100k years), we still observed that the vast majority of genetic diversity was common across populations, with many shared SNPs and haplotypes. Consequently we found no overall differentiation across populations in clustering analysis at all nuclear gene loci. This lack of large scale population structure at nuclear loci is supported by recent comparative sequencing studies of nuclear genes of Scots pine and its close relatives from *P. mugo* complex that showed very similar genetic background across species with many shared polymorphisms between species that diverged up to ~5mya (Wachowiak et al. 2011; Wachowiak et al. 2013a). Therefore it seems that life history

characteristics of pine trees, such as longevity and overlapping generations in large populations, slow down fixation of new mutations and divergence at evolutionarily neutral loci (Austerlitz et al. 2000; Excoffier 2004; Ray et al. 2003). Despite Scots pine's substantial range expansion and population size fluctuation across Europe during the last ~20ky, simulation studies based on nucleotide sequence variation at several nuclear gene loci suggest that patterns of nucleotide polymorphism at a set of nuclear genes could be best explained by an ancient (>1 mya) bottleneck (Pyhäjärvi et al. 2007; Kujala and Savolainen 2012). Our sequencing data supports the observation that the large population sizes of forest trees and their capability for maintenance of high levels of genetic variation within populations is likely to buffer against rapid changes in genetic diversity across the species distribution due to population size and range shifts.

Effects of natural selection

Across the species distribution range and environmental gradients, local adaptation is seen as a balance between diversifying selection and gene flow (Savolainen et al. 2007). At mutation-drift equilibrium, genetic drift and gene flow influence the level of differentiation between populations for selectively neutral markers (Savolainen et al. 2007; Kawecki and Ebert 2004). Considering the homogenizing effect of gene flow on patterns of nucleotide diversity, evident as low differentiation between populations in our dataset, the outlier pattern of differentiation at some loci may indicate the influence of selection. In our study we included loci related to the regulation of gene expression, signal transduction and stress response, all of which may potentially be important in local adaptation across environmental gradients. Against the very uniform genetic background of the Polish populations, strong evidence for selection was found at one regulatory locus (Pr4_6). This locus also showed deviations in compound neutrality tests that have high specificity for detecting positive selection and are robust against demographic processes (such as population size changes) and other non-equilibrium forces that may cause deviations in the patterns of nucleotide polymorphism (Zeng et al. 2007). A few other loci were found to deviate from neutrality based on haplotype structure or allelic frequency spectra between Polish and reference Scots pine populations including alpha-N-acetylglucosaminidase, DEAD-box ATP-dependent RNA- helicase, U-box domain-containing protein, transducin/WD40 domain-containing protein, phytochrome and auxin induced-transcription factor. Comparing the lack of differentiation between geographical regions at the majority of the loci in our dataset with the known high divergence at quantitative traits for characters of adaptive importance, it is possible that difference in the frequency and distribution of polymorphism, which is accompanied in some cases by deviation from neutrality in coalescence simulations tests, may be due to diversifying selection across the species range. The genes showing deviations from neutrality are involved in regulation of gene expression, signal transduction and cellular metabolism and appear to be important in local adaptation. Interestingly, Polish populations were fixed for an 11bp deletion and

associated SNPs at PhytP that was segregating in reference populations. This result may potentially reflect some sampling bias or the polymorphisms could be associated with the population differentiation in some adaptive traits driven by photoperiod. Previous studies of other Scots pine phytochromes did not link variation at the loci with latitudinal clines in the species phenology, however association of polymorphisms in phytochrome locus PhyB2 was found to correlate with clinal variation in timing of growth in *Populus* (Ingvarsson et al. 2006).

So far attempts to detect loci under selection in Scots pine have identified a set of a few candidate genes including dehydrin (Wachowiak et al. 2009) and ft/tfl1 -like and pseudo response regulator 1 genes (Kujala and Savolainen 2012). The capability to detect selection at individual loci is heavily dependent on the strength and time since selection, and the number of loci involved in screening (Wright and Gaut 2005). Although quantitative trait studies provide ample evidence of adaptive phenotypic differentiation across the Scots pine range in phenology, resistance to stress factors, growth and wood quality (Boratyński 1993; Giertych and Oleksyn 1992; Sabor 2006; Hurme et al. 1997; Oleksyn et al. 1992), the genetic basis of that variation remains unknown. Some quantitative trait studies have indicated that genomic regions of large phenotypic effect may exist in forest tree species (Kinloch et al. 1970; Howe et al. 2003). However, there is growing evidence from most recent association genetic studies that the majority of quantitative traits in trees are controlled by many genes with small individual effects (Eckert et al. 2009; González-Martinez et al. 2007; Neale and Kremer 2011). Therefore, the signature of selection may be more readily detectable as covariance of allele frequencies at multiple loci (Latta 2004; Le Corre and Kremer 2003; Derory et al. 2010). Considering the high recombination rate in conifers and the fact that only a small proportion of the genome has so far been tested for selection, the question remains as to whether there are genes of major effect underlying some adaptive traits in these species. Considering the low divergence between populations observed in our study, it seems likely that adaptive divergence in recently colonized areas results from selection acting on standing genetic variation segregating in natural populations.

Conclusions

Our study has shown the very high genetic similarity between Polish populations of Scots pine across mountain, peatbog and lowland locations. Polymorphism at the set of nuclear genes we focussed on indicated that they most likely share a common postglacial history with Northern European Scots pine populations, distinct from the more isolated North-West and South European distribution range of the species. Several genes with outlier patterns of differentiation suggest the role of diversifying selection acting across environmental gradients. Considering that the North and Central continental range spans very steep environmental gradients related to latitudinal differences in photoperiod and temperature,

this part of Scots pine distribution seems particularly suitable for association genetic studies to link phenotypic and genetic variation at a large geographical scale.

Acknowledgements

The research was financially supported by Polish Ministry of Science and Higher Education (grant nr. 2975/B/P01/2010/39). BW acknowledges financially support from Polish National Science Centre (grant nr. DEC-2012/05/E/NZ9/03476). We thank Euforgen network for providing distribution map of Scots pine.

Data Archiving Statement

The nucleotide sequences of the studied loci were submitted to NCBI repositories (Acc. Nr. KJ502673-KJ503230).

Table 1. Geographical location of the analysed stands of Scots pine in Poland and reference populations. Samples were divided into geographical regions including Poland (SW: south-west, W: west, S: south, N: north, E: east), Scotland (SCO), Spain (SP) and Norther Europe (EU_N).

Nr.	Location (stand)	Region	Longitude	Latitude	Altitude	Nr.	Location	Region	Longitude	Latitude	Altitude
1	**Wegliniec peatbog**	PL_SW	15^{0}14′	51^{0}17′	200	16	Miłomłyn	PL_N	19^{0}50′	53^{0}45′	130
2	**Chojnik**	PL_SW	15^{0}38′	50^{0}50′	600	17	Wyszków	PL_E	21^{0}29′	52^{0}40′	120
3	**Szczeliniec**	PL_SW	16^{0}14′	50^{0}26′	850	18	Hajnówka	PL_E	23^{0}57′	52^{0}74′	170
4	**Tarnowskie Góry**	PL_SW	18^{0}56′	50^0 30′	300	19	Rezerwat Liski	PL_E	23^{0}52′	50^{0}56′	23
5	**Gubin**	PL_W	14^{0}43′	51^{0}57′	70	20	Scotland_Shieldaig	SCO	-5^{0}38′	57^{0}30′	81
6	**Barlinek**	PL_W	15^{0}14′	52^{0}59′	70	21	Scotland_Glen Tanar	SCO	-2^{0}51′	57^{0}2′	160
7	Bolewice	PL_W	16^{0}07′	52^{0}23′	80	22	Scotland_Rothiemurcys	SCO	-3^{0}46′	57^{0}8′	318
8	Syców	PL_W	17^{0}43′	51^{0}18′	180	23	Scotland_Glen Affric	SCO	-4^{0}55′	57^{0}16′	256
9	**Pusta Wielka**	PL_S	20^{0}27′	49^{0}28′	600	24	Scotland_Black Wood	SCO	-4^{0}19′	56^{0}40′	275
10	**Koryciska Wielkie**	PL_S	19^{0}48′	49^{0}16′	1100	25	Spain_Trevenque	SP	3^{0}32′	37^{0}05′	1170
11	Pieniński Park Narodowy (PPN)	PL_S	20^{0}33′	49^{0}42′	700	26	Spain_Valsain	SP	4^{0}02′	40^{0}51′	1350
12	Tarnawa **peatbog**	PL_S	22^{0}49′	49^{0}06′	660	27	Finland_Punkaharju	EU_N	29^{0}23′	61^{0}45′	80
13	Woziwoda	PL_N	17^{0}55′	53^{0}40′	120	28	Finland_Kolari	EU_N	24^{0}3′	67^{0}11′	190
14	Pisz	PL_N	21^{0}46′	53^{0}36′	130	29	Sweden_Krp.Tjärnbergsheden	EU_N	20^{0}48′	64^{0}37′	110
15	Tabórz	PL_N	20^{0}02′	53^{0}45′	130	30	Sweden_Väster Mjőingenn	EU_N	13^{0}34′	62^{0}45′	640

Table 2. Summary statistics of nucleotide and haplotype variation and frequency distribution spectra across 32 nuclear genes in Polish populations of Scots pine.

					Nucleotide diversity				Haplotype diversity		
Population	**n**	**L**	**SNPs**	**Sing**	π_{total}	θ^{a}	**C.I. (95%)[b]**	**D**	**N**	**N%**	**H_d (SD)**
Wegliniec peatbog	12	15077	221	74	0.0053	0.0057	0.0047-0.0069	-0.411*	139	26.4 (2.8)	0.644 (0.115)
Chojnik	10	15046	207	88	0.0052	0.0062	0.0051-0.0076	-0.489*	129	23.6 (2.6)	0.668 (0.114)
Szczeliniec	10	15057	213	93	0.0057	0.0063	0.0051-0.0077	-0.657*	126	23.2 (3.5)	0.655 (0.124)
Tarnowskie Góry	10	15058	220	110	0.0055	0.0073	0.0059-0.0088	-1.00*	137	24.4 (3.3)	0.708 (0.131)
Gubin	10	15083	219	87	0.0054	0.0056	0.0045-0.0068	-0.017	130	23.9 (2.6)	0.680 (0.122)
Barlinek	10	15080	217	72	0.0057	0.0057	0.0047-0.0070	-0.157	133	24.4 (3.0)	0.681 (0.120)
Bolewice	10	15063	248	113	0.0060	0.0064	0.0053-0.0078	-0.371*	136	25.3 (1.8)	0.680 (0.122)
Syców	10	15029	236	100	0.0059	0.0060	0.0049-0.0074	-0.278	133	24.8 (2.6)	0.671 (0.120)
Pusta Wielka	10	15061	216	84	0.0055	0.0063	0.0052-0.0077	-0.451*	132	23.9 (3.2)	0.674 (0.119)
Koryciska Wielkie	10	15069	195	80	0.0051	0.0053	0.0043-0.0065	-0.50*	118	22.0 (1.2)	0.590 (0.121)
PPN	10	15068	214	96	0.0052	0.0056	0.0046-0.0069	-0.259	124	22.7 (1.9)	0.627 (0.116)
Tarnawa **peatbog**	10	15093	206	69	0.0054	0.0054	0.0044-0.0066	-0.163	125	23.0 (1.8)	0.639 (0.139)
Woziwoda	10	15060	215	87	0.0055	0.0057	0.0047-0.0070	-0.319	135	26.2 (2.3)	0.687 (0.113)
Pisz	10	15048	219	79	0.0055	0.0061	0.0050-0.0075	-0.269	134	25.1 (3.5)	0.645 (0.145)
Tabórz	10	15074	202	77	0.0052	0.0053	0.0043-0.0065	-0.221	126	23.6 (1.8)	0.655 (0.120)
Miłomłyn	10	15077	233	114	0.0056	0.0061	0.0050-0.0074	-0.252	133	24.8 (2.5)	0.675 (0.128)
Wyszków	10	15034	201	65	0.0053	0.0052	0.0042-0.0064	0.010	127	23.6 (2.8)	0.667 (0.115)
Hajnówka	10	15083	213	72	0.0056	0.0052	0.0042-0.0063	0.045	128	23.4 (2.1)	0.655 (0.108)
Rezerwat Liski	10	15049	210	81	0.0055	0.0056	0.0045-0.0068	-0.088	121	21.6 (2.5)	0.628 (0.112)
Total/Mean	**192**	**15064**	**216**	**86**	**0.0055**	**0.0058**	**0.0048-0.0071**	**-0.308**	**130**	**24.0 (2.5)**	**0.659 (0.121)**

n- number of samples analysed per locus; L – average length of the sequences in base pairs excluding indels; SNPs- number of polymorphic sites detected; Sing – number of singleton mutations; π – nucleotide diversity (Nei 1987); [a] median multilocus θ for all sites calculated using Markov chain Monte Carlo (MCMC) simulation under a Bayesian model (Pyhäjärvi et al. 2007); [b] 95% credibility intervals for θ; D – multilocus Tajima's D statistics (Tajima 1989); N – number of haplotypes; N% - percentage out of total number of haplotypes identified across loci (percentage of unique haplotypes); H_d – haplotype diversity (standard deviation); * P<0.01

Table 3. Summary statistics of nucleotide and haplotype variation and frequency distribution spectra across 29 nuclear genes in regional groups of Polish populations of Scots pine and reference samples.

Regional					Nucleotide diversity							Haplotype diversity		
group	n	L	SNPs	Sing.	π_{total}	π_{silent}	θ^a	C.I. (95%)[b]	ρ^c	ρ/θ	D^d	N	N%	H_d (SD)
PL_SW	42	13099	325	112	0.0059	0.0081	0.0067	(0.0056-0.0080)	0.0074	1.105	-0.772*	237	51.55 (0.06)	0.674 (0.057)
PL_W	40	13138	327	90	0.0059	0.0083	0.0068	(0.0056-0.0081)	0.0122	1.794	-0.487*	226	47.52 (0.05)	0.679 (0.057)
PL_S	40	13153	301	92	0.0057	0.0078	0.0060	(0.0050-0.0072)	0.0025	0.417	-0.320	207	47.00 (0.03)	0.648 (0.058)
PL_N	40	13130	318	104	0.0056	0.0079	0.0067	(0.0056-0.0080)	0.0055	0.821	-0.495*	219	43.34 (0.05)	0.668 (0.056)
PL_E	30	13112	290	96	0.0058	0.0079	0.0060	(0.0049-0.0073)	0.0023	0.383	-0.272	198	41.33 (0.04)	0.651 (0.062)
SCO	50	12997	298	72	0.0052	0.0076	0.0061	(0.0051-0.0073)	0.0010	0.164	-0.147	210	47.83 (0.08)	0.621 (0.046)
SP	20	13015	227	43	0.0052	0.0068	0.0053	(0.0042-0.0065)	0.0004	0.075	0.053	137	24.77 (0.04)	0.601 (0.081)
EU_N	35	12525	300	95	0.0057	0.0080	0.0067	(0.0056-0.0081)	0.0005	0.075	-0.415*	202	45.67 (0.05)	0.660 (0.060)
Total/Mean	297	13021	298	88	0.0056	0.0078	0.0063	0.0052-0.0076	0.0040	0.604	-0.357	205	43.63 (0.05)	0.650 (0.059)

n- number of samples analysed per locus; L – average length of the sequences in base pairs excluding indels; SNPs- number of polymorphic sites detected; Sing – number of singleton mutations; π – nucleotide diversity (Nei 1987); [a] median multilocus θ for silent sites; [b] 95% credibility intervals for θ; [c] ρ – average recombination rate estimates for a set of 29 loci ; D – multilocus Tajima's D statistics (Tajima 1989); N – number of haplotypes; N% - percentage out of total number of haplotypes identified across loci (percentage of unique haplotypes); H_d – haplotype diversity (standard deviation); *P<0.01

Table 4. F_{ST} at all polymorphic sites combined between geographical groups of Scots pine from Poland and reference populations.

	1	2	3	4	5	6	7
1. PL_SW							
2. PL_W	-0.011						
3. PL_S	0.007	-0.001					
4. PL_N	-0.013	-0.003	-0.006				
5. PL_E	-0.010	0.003	0.005	-0.003			
6. SCO	0.043*	0.032*	0.035*	0.033*	0.041*		
7. SP	0.086*	0.075*	0.073*	0.067*	0.079*	0.034*	
8. EU_N	0.018*	0.010	0.012	0.011	0.009	0.024*	0.053*

Table 5. Loci of outlier F_{ST} and S_{nn} patterns between Polish and reference populations of Scots pine.

	PL-SCO		**PL-SP**		**PL-EUN**	
Gene	**F_{ST}**	**S_{nn}**	**F_{ST}**	**S_{nn}**	**F_{ST}**	**S_{nn}**
Pr1_46	0.085**	0.216***	0.041	0.191**	0.008	0.172
Pr2_30	0.074**	0.189**	0.107**	0.187	0.030	0.176
Pr2_35	0.091**	0.187**	0.128**	0.189	-0.006	0.166
Pr2_42	0.002	0.182	0.100**	0.199**	-0.012	0.176
PhytP	0.628**	0.800***	0.865**	0.935***	0.483**	0.816***
Pr4-27	0.025	0.191***	0.326**	0.203**	0.078**	0.182

Figure 1. Location of the studied Polish and reference Scots pine populations (Table 1) on the species distribution range in Europe (dark grey).

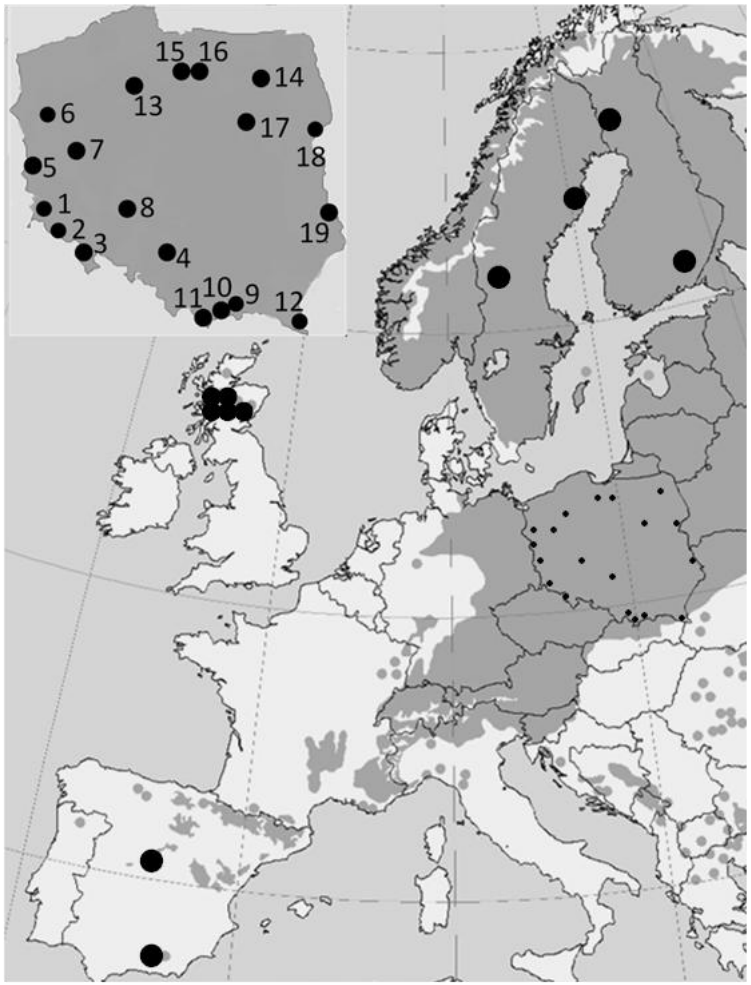
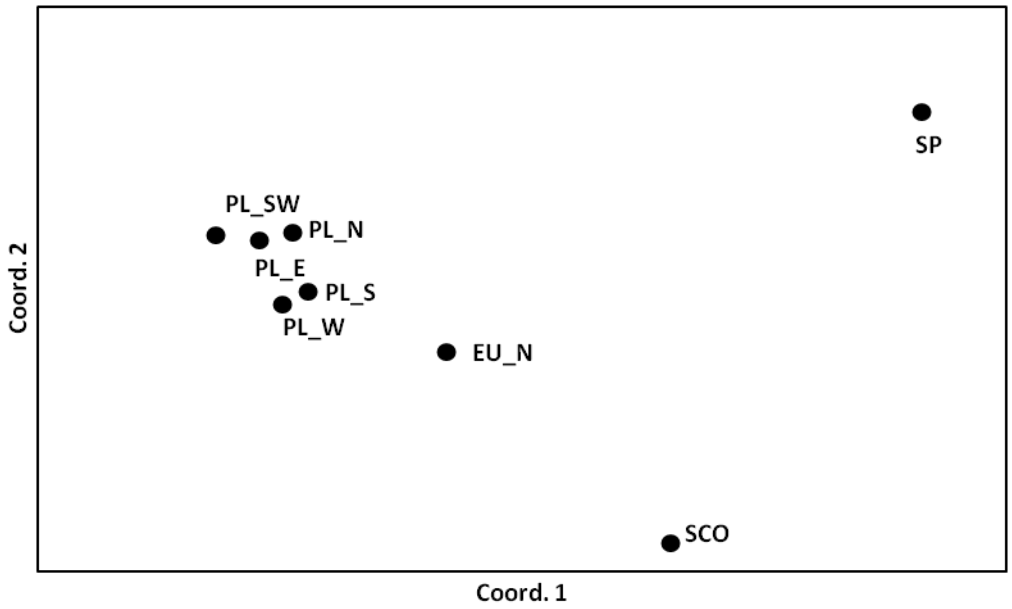


Figure 2. Principal Coordinates analysis (PCoA) based on pairwise F_{ST} matrix at all polymorphic sites detected (695 SNPs) showing genetic relationships between five geographical regions in Poland (PL) and reference Scots pine populations from Northern Europe (EU_N), Scotland (SCO) and Spain (SP).



SUPPLEMENTARY MATERIAL

Witold Wachowiak[1,2,3*], Błażej Wójkiewicz[2], Stephen Cavers[3], Andrzej Lewandowski[2]

High genetic similarity between Polish and North European Scots pine (*Pinus sylvestris* L.) populations at nuclear gene loci

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

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

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

***Corresponding authors:** Witold Wachowiak (witoldw@man.poznan.pl), Institute of Dendrology, Polish Academy of Sciences, Parkowa 5, 62-033 Kórnik, Phone: +48 618170033, Fax: +48 618170166

SUPPLEMENTARY TABLES

Supplementary Table S1. Analysed loci

Locus Acronym	PCR Primers (F-upper, and R-lower)	Gene function [Category[3]]	Base pairs screened [4] Total	Indels [5]
Pr1_5	[1]GATCATTCTAGGCACAGCACAAG [2]CCTGTACCGTGTTCATCAATTTAGCAAG	Seryl-tRNA synthetase [E]	532	0
Pr1_11	[1]GACCAGGCAAGGAAACAAAAG [2]TTGGCAATCGGTTGATGGGGAG	putative glucuronidase 3 [M]	721	1 (7)
Pr1_15	[1]CATTATTATCCAAGGGCGAG [2]GAGGCTTTGAGTCACCGTTAC	mitogen-activated protein-kinase [ST]	593	3 (28)
Pr1_18	[1]AAGCGACTCAAAAGGGG [2]TCGGCTGTATTGTCTC	alpha 1,3-glucosidase [M]	516	2 (12)
Pr1_22	[1]TGAAGGGAGAGGACTAC [2]ACCCAGAAACACAAAGAGGAAAC	hypothetical protein [UN]	301	0
Pr1_26	[1]CCCATTTTAGCAAACCC [2]GAAGTGAAGATGAGCATAAG	putative pre-mRNA branch site-protein p14 [E]	409	1 (3)
Pr1_28	[1]GCAACTTCCCCTTTTTC [2]ACAGTGTGAGAGACGAG	translation initiation factor-4G-like [E]	522	0
Pr1_31	[1]TGTGAAGCAGAGGAAC [2]GATGGTAAGGATGGTG	putative patellin-4 [T]	315	1 (3)
Pr1_36	[1]GCGTTCATCATCTCAAGCC [2]CTAATCTCTCTTATTGTCATCTCCACC	transcribed locus [UN]	434	5 (13)
Pr1_43	[1]GGACATTGTACTGTTGG [2]GGGTAAATGGAAAGAGTATTGG	beta-galactosidase [M]	591	0
Pr1_46	[1]ATCCAGTCCCTTCTCCACCTATCC [2]AAGTGCTAAGCATCAAGCAGATAATCC	alpha-N-acetylglucosaminidase [M]	511	2 (12)
Pr2_5	[1]ATTCATCCACTTCCCCC [2]GTGTATTTGGTGTTGCAG	transcribed locus [UN]	299	1 (3)
Pr2_16	[1]TCACTTGGCAGAAGAC [2]GAGAGATTCTTTGGAGAC	glutamyl-tRNA reductase [M]	403	0
Pr2_30	[1]CACTTGTCATCTGCTC [2]CTTGGAAGGATAGAATCTG	U-box domain-containing [ST] protein	402	0
Pr2_35	[1]ACCCACAAATTGCCAG [2]GCCGTGATTATCGAAGAG	DEAD-box ATP-dependent RNA- helicase [M]	420	1 (3)
Pr2_41	[1]GAAAAGGATCAAATTGTGGG [2]GCTAACATTGGCTGTGG	F-box protein GID2 [ST]	378	-
Pr2_42	[1]GCATAGCCATCCATATC [2]GGGTGTGAATTTTTTTGGTG	transducin/WD40 domain-containing protein [ST]	511	-
Pr2_47	[1]TTCATAAAGCCCCCCATCC	hexokinase 1 [ST]	549	2 (3)

	[2]TCTGATTTCAAAGTCGCC			
Pr2_48	[1]GCTATGCGTTACTTGG	S-methyl-5-thioribose kinase [ST]	709	5 (21)
	[2]TGAGTTGAGCTGCTTG			
dhn2PP	CTGCAGAGACTGTGCCTGAGC	dehydrin [a] [M]	426	1 (8)
	CCAGGGAGCTTTTCCTTGATCT			
PhytP	TGCAGGGCATAGACGAACTTAGC	phytochrome P [b] [E]	715	4 (86)
	GGCAATCTGCAAGAAACAAAACGAC			
Pr4-1	[1]CCGATGCTGACTCTTCTAAC	Peroxidase [c] [ST]	341	4 (15)
	[2]AGCGAATTTGGAGGATGA			
Pr4-4	[1]TGTCACTGCCCAGAGCTATTC	Putative aquaporin [c] [T]	640	5 (57)
	[2]ATCACAGCCGCTCCAAAAC			
Pr4-6	[1]AATTCCTTTGGTCTTGGAG	Nucleas Dicer [E]	538	0
	[2]AGTTTGCCGTGTGAAAGTTTG			
Pr4-8	[1]GCCAGAAAATATGTGATTAGAAGCC	lanatoside 15oacetylesterase [M]	642	1 (1)
	[2]AGTTTTCCACCATCATCAAG			
Pr4-9	[1]CATTTTGCAGAGGCAAG	methionine syntase [M]	707	2 (1)
	[2]TGCGTAAGGCAGAACAG			
Pr4-11	[1]CCTTCTATTTGAATCCCTTG	scl1 protein [c] [E]	439	2 (19)
	[2]CATAGTAACAGCCTACAG			
Pr4-12	[1]CTGCTCAAGTGAAAGG	proton myo-inositol transporter [c] [T]	535	2 (3)
	[2]CTGATTGTGGATTCTGTG			
Pr4-19	[1]CTCTACCACATCATCTCC	laccase [c] [M]	333	0
	[2]TTTCACTCTCGTGTCTTTCACC			
Pr4-21	[1]ACATGGTGTTTGGCAGG	Receptor protein kinase [c] [ST]	394	1 (2)
	[2]AATGAGGAGGGTGGTAGAG			
Pr4-27	[1]TAGCAGACGGTATTCACACAGTCC	putative auxin induced -	403	0
	[2]CCACAACCACCTTGCATCATTATTT	transcription factor [c] [E]		
Pr4-34	[1]ACCCTGTATCGATGGGTATGGAGAT	transcription factor bHLH62-	360	1 (3)
	[2]TTTCATGTGGTTTGTTGGTACAGAACCTGCAATCA	like- gene [c] [E]		

[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]-average across all samples, [5]-number of indels and total length in parenthesis; DNA regions described in: [a] Eveno et al., 2007; [b] Pyhäjärvi et al., 2007; [c] Ersoz et al. 2010.

Supplementary Table S2. Net divergence between Scots pine samples from Poland and reference populations.

	PL_SW	PL_W	PL_S	PL_N	PL_E	SCO	SP
PL_SW							
PL_W	0.0000						
PL_S	0.0000	0.0000					
PL_N	0.0000	0.0000	0.0000				
PL_E	0.0000	0.0000	0.0000	0.0000			
SCO	0.0003	0.0002	0.0002	0.0002	0.0003		
SP	0.0005	0.0004	0.0004	0.0004	0.0005	0.0002	
EU_N	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0003

Supplementary Table S3. Percentage (%) of shared SNPs of Scots pine samples from Poland and reference populations.

	PL_SW	PL_W	PL_S	PL_N	PL_E	SCO	SP
PL_SW							
PL_W	72						
PL_S	71	77					
PL_N	68	75	75				
PL_E	71	73	76	74			
SCO	67	70	70	69	68		
SP	61	66	67	65	65	70	
EU_N	69	72	69	72	73	70	66

Supplementary Table S4. AMOVA based on all polymorphic sites detected.

A. Source of variation	Sum of squares	Variance components	Percentage variation
PL_19 populations	**average over 606 SNPs**		
Among populations	857.38	0.59	1.39
Within populations	6833.25	42.25	98.61
Total	7690.63	42.85	
PL _ 5 regions	**average over 560 SNPs**		
Among populations	167.84	0.05	0.12
Within populations	7103.16	40.43	99.88
Total	7271.00	40.48	
PL all vs. EU_N	**average over 609 SNPs**		
Among populations	87.39	0.83	1.97
Within populations	8661.16	41.04	98.03
Total	8748.55	41.87	
PL vs. SCO	**average over 615 SNPs**		
Among populations	185.570	1.94945	4.62
Within populations	9121.367	40.32275	95.39
Total	9306.937	42.27219	
PL vs. SP	**average over 594 SNPs**		
Among populations	168.70	3.64	8.30
Within populations	7974.07	40.22	91.70
Total	8142.77	43.86	

Supplementary Table S5. Significant values of F_{ST} statistics (P<0.01) for corresponding loci (marked in superscript) in pairwise comparisons between 19 Polish populations of Scots pine

Population	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. Wegliniec peatbog																		
2. Chojnik																		
3. Szczeliniec																		
4. Tarnowskie Góry	0.38 Pr1_22																	
5. Gubin			0.16 Pr2_16															
6. Barlinek	0.42 Pr4_6			0.32 Pr1_31 0.32 Pr2_47														
7. Bolewice																		
8. Syców				0.32 Pr1_31	0.25 Pr2_5													
9. Pusta Wielka	0.3 Pr2_30	0.42 Pr2_30																
10. Koryciska Wielkie		0.35 Pr2_30	0.71 Pr2_48		0.39 Pr2_48	0.6 Pr4_6 0.32 dhn2	0.40 Pr2_48	0.38 Pr2_48 0.34 dhn2										
11. PPN			0.3 Pr4_4			0.38 Pr4_6 0.25 dhn2		0.28 dhn2										
12. Tarnawa **peatbog**	0.36 Pr2_42	0.36 Pr4_9	0.33 Pr1_43 0.4 Pr4_4 0.19 Pr4_9			0.30 Pr4_6 0.36 Pr4_9	0.20 Pr4_9	0.37 Pr1_43 0.35 Pr4_9	0.28 Pr4_9	0.26 Pr4_9 0.28 dhn2	0.25 Pr4_9							
13. Woziwoda			0.19 phytP	0.35 Pr4_9 0.34 Pr4_11				0.43 Pr4_11				0.73 Pr4_9						
14. Pisz			0.20 Pr1_28					0.27 Pr2_48				0.29 Pr4_9						
15. Tabórz			0.18 Pr1_28															
16. Miłomłyn										0.44 Pr4_6 0.21 dhn2								
17. Wyszków			0.18 Pr1_28			0.18 Pr2_48						0.5 Pr1_43 0.35 Pr4_9						
18. Hajnowka		0.35 Pr4_4			0.47 Pr1_31	0.45 Pr4_6					0.4 Pr4_4	0.52 Pr4_4 0.38 Pr4_9			0.28 Pr1_26 0.4 Pr4_4			
19. Rezerwat Liski																0.24 Pr1_5		

Supplementary Table S6.
Significant values of F_{ST} statistics (P<0.01) for corresponding loci (marked in superscript) in pairwise comparisons between groups of Polish populations defined and reference populations.

	1	2	3	4	5	6	7
1: **PL_SW**							
2: **PL_W**							
3: **PL_S**	0.08 $^{Pr5\_1}$ 0.05 $^{Pr16\_2}$ 0.06 $^{Pr28\_1}$ 0.09 $^{Pr30\_2}$	0.08 $^{Pr4\_4}$ 0.170 $^{Pr6\_4}$					
4: **PL_N**							
5: **PL_E**				0.13 $^{Pr19\_4}$			
6: **SCO**	0.42 ^{phytP}, 0.08 $^{Pr1\_4}$ 0.12 $^{Pr11\_4}$ 0.06 $^{Pr28\_1}$ 0.19 $^{Pr30\_2}$ 0.08 $^{Pr35\_2}$	0.48 ^{phytP} 0.12 $^{Pr35\_2}$ 0.10 $^{Pr48\_2}$	0.32 ^{phytP} 0.1 $^{Pr11\_1}$ 0.16 $^{Pr11\_4}$ 0.11 $^{Pr46\_1}$ 0.08 $^{Pr48\_2}$	0.41 ^{phytP} 0.24 $^{Pr11\_4}$ 0.08 $^{Pr30\_2}$ 0.14 $^{Pr46\_1}$	0.3 ^{phytP} 0.11 $^{Pr48\_2}$		
7: **SP**	0.70 ^{phytP} 0.11 $^{Pr15\_1}$ 0.26 $^{Pr16\_2}$ 0.11 $^{Pr21\_4}$ 0.38 $^{Pr27\_4}$ 0.14 $^{Pr28\_1}$ 0.12 $^{Pr35\_2}$ 0.12 $^{Pr42\_2}$ 0.15 $^{Pr48\_2}$	0.73 ^{phytP} 0.30 $^{Pr27\_4}$ 0.16 $^{Pr28\_1}$ 0.14 $^{Pr30\_2}$ 0.17 $^{Pr35\_2}$ 0.10 $^{Pr42\_2}$ 0.21 $^{Pr47\_2}$ 0.17 $^{Pr48\_2}$	0.65 ^{phytP} 0.30 $^{Pr27\_4}$ 0.24 $^{Pr28\_1}$ 0.16 $^{Pr30\_2}$ 0.14 $^{Pr47\_2}$ 0.16 $^{Pr48\_2}$	0.70 ^{phytP} 0.12 $^{Pr11\_4}$ 0.21 $^{Pr16\_2}$ 0.42 $^{Pr27\_4}$ 0.19 $^{Pr28\_1}$ 0.12 $^{Pr46\_1}$ 0.13 $^{Pr48\_2}$	0.6 ^{phytP} 0.13 $^{Pr15\_1}$ 0.15 $^{Pr16\_2}$ 0.39 $^{Pr27\_4}$ 0.16 $^{Pr28\_1}$ 0.16 $^{Pr30\_2}$ 0.12 $^{Pr35\_2}$ 0.20 $^{Pr48\_2}$	0.22 $^{Pr16\_2}$ 0.22 $^{Pr27\_4}$ 0.19 $^{Pr28\_1}$ 0.07 $^{Pr36\_1}$	
8: **EU_N**	0.28 ^{phytP} 0.08 $^{Pr27\_4}$ 0.10 $^{Pr30\_2}$	0.35 ^{phytP}	0.04 $^{Pr12\_4}$	0.26 ^{phytP} 0.10 $^{Pr27\_4}$ 0.07 $^{Pr36\_1}$	0.21 $^{Pr19\_4}$	0.15 $^{Pr11\_4}$ 0.08 $^{Pr19\_4}$ 0.13 $^{Pr35\_2}$ 0.05 $^{Pr36\_1}$	0.19 ^{phytP} 0.15 $^{Pr16\_2}$ 0.12 $^{Pr28\_1}$ 0.16 $^{Pr30\_2}$ 0.17 $^{Pr35\_2}$

Supplementary Table S7. Pairwise F_{ST} and S_{nn} values between analysed groups of populations defined for *P. sylvestris*. Statistically significant values: ** 0.001<P<0.01; *** P<0.001. [1] – significant HEW test, [2] - significant DHEW test

	PL-SCO		**PL-SP**		**PL-EUN**		**PL_5 regional groups**	
Gene	**F_{ST}**	**S_{nn}**	**F_{ST}**	**S_{nn}**	**F_{ST}**	**S_{nn}**	**F_{ST}**	**S_{nn}**
Pr1_5[1,2]	0.003	0.175	0.023	0.184	-0.011	0.159	0.018**	0.200
Pr1_11	0.047	0.185	0.018	0.177	-0.010	0.158	-0.013	0.189
Pr1_15	0.021	0.183	0.065	0.185	0.010	0.168	0.001	0.204
Pr1_18	0.000	0.169	0.030	0.162	-0.013	0.152	0.005	0.187
Pr1_22	-0.003	0.178	0.075	0.189	-0.007	0.176	-0.017	0.203
Pr1_26	0.027	0.232***	-0.015	0.210	0.009	0.172	-0.009	0.206
Pr1_28	0.001	0.180	**0.165****	0.198	0.025	0.178	0.020**	0.206
Pr1_31	0.024	0.181	-0.022	0.169	0.041	0.172	0.001	0.196
Pr1_36	0.015	0.182	0.035	0.181	0.022	0.172	-0.005	0.203
Pr1_43	0.034	0.190	0.005	0.194	0.065	0.184	-0.005	0.220*
Pr1_46[1,2]	**0.085****	0.216***	0.041	0.191**	0.008	0.172	0.009	0.205
Pr2_5	-0.012	0.183	-0.007	0.183	0.003	0.176	0.014	0.214
Pr2_16	0.014	0.170	**0.167****	0.182	-0.008	0.165	0.012	0.200
Pr2_30	**0.074****	0.189**	**0.107****	0.187	0.030	0.176	0.020*	0.208
Pr2_35	**0.091****	0.187**	**0.128****	0.189	-0.006	0.166	0.002	0.204
Pr2_41	0.023	0.176	0.008	0.179	-0.011	0.164	0.007	0.206
Pr2_42	0.002	0.182	**0.100****	0.199**	-0.012	0.176	0.000	0.210
Pr2_47	0.007	0.177	**0.127****	0.178	0.020	0.168	0.004	0.200
Pr2_48	**0.106****	0.177	**0.195****	0.188	0.030	0.175	0.010**	0.203
dhn2PP	0.019	0.235***	0.019	0.217**	-0.012	0.199**	0.005	0.238*
PhytP[1,2]	**0.628****	0.800***	**0.865****	0.935***	**0.483****	0.816***	0.004	0.234*
Pr4-1[1]	0.032	0.194**	0.057	0.194	0.000	0.166	0.006	0.213
Pr4-4	-0.002	0.169	-0.024	0.170	-0.004	0.166	0.005	0.199
Pr4-11	**0.122****	0.174	0.039	0.173	-0.004	0.169	0.003	0.192
Pr4-12[1,2]	-0.007	0.163	-0.015	0.168	0.009	0.159	0.011*	0.194
Pr4-19[1,2]	0.007	0.185**	0.002	0.181	0.046	0.175	0.013	0.210*
Pr4-21	0.001	0.197**	0.026	0.204	-0.008	0.191**	0.011	0.234**
Pr4-27	0.025	0.191***	**0.326****	0.203**	**0.078****	0.182	-0.002	0.203
Pr4-34	-0.006	0.169	0.006	0.168	-0.011	0.165	-0.002	0.197
Pr4_6[1,2]	na	na	na	na	na	na	0.044***	0.217*
Pr4_8	na	na	na	na	na	na	-0.011	0.191
Pr4_9	na	na	na	na	na	na	0.014*	0.216

Supplementary Table S8. Outlier SNPs detected across loci in pairwise comparisons between five geographical groups determined for *P. sylvestris* from Poland.

Locus	Pos.	SNPs	F_{ST}	*p*	Locus	Pos.	SNPs	F_{ST}	*p*
Pr22_1	186	G/A	0.044	0.040	Pr48_2	148	A/G	0.041	0.039
Pr28_1	289	G/A	0.071	0.005	Pr48_2	386	A/G	0.049	0.028
Pr28_1	295	T/A	0.085	0.004	Pr48_2	609	A/G	0.041	0.041
Pr5_2	250	T/C	0.058	0.020	dhn2	219	C/T	0.063	0.020
Pr16_2	40	G/T	0.040	0.029	dhn2	256	T/G	0.036	0.050
Pr16_2	144	G/A	0.045	0.020	dhn2	314	T/C	0.055	0.028
Pr16_2	287	C/G	0.045	0.022	PL1_4	74	T/G	0.038	0.038
Pr16_2	317	G/C	0.045	0.017	PL1_4	169	C/T	0.038	0.044
Pr16_2	346	T/G	0.036	0.045	Pr6_4	157	G/C	0.073	0.007
Pr35_2	350	A/G	0.048	0.025	Pr19_4	8	T/C	0.045	0.023

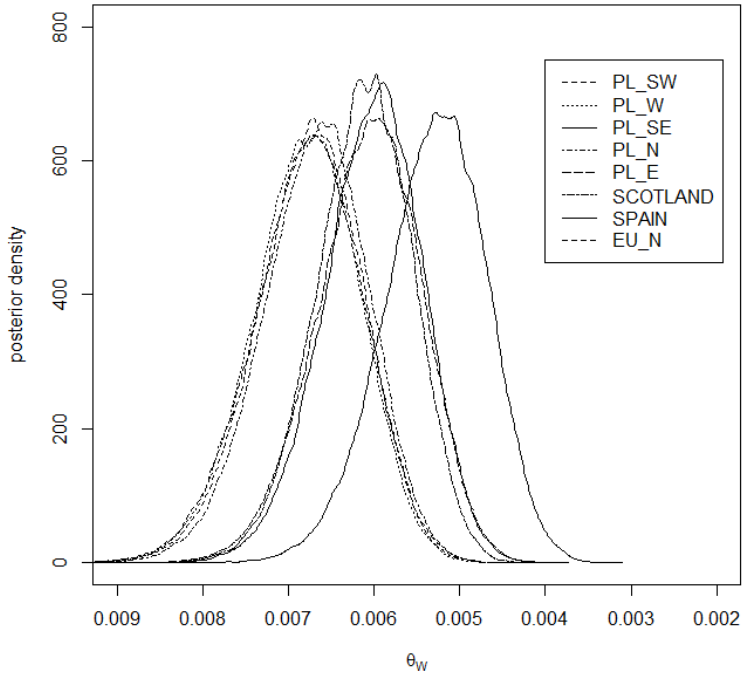
Supplementary Table S9. Outlier SNPs detected across loci in pairwise comparisons between Polish and reference populations.

PL vs. EU_N									
Locus	**Pos.**	**SNPs**	**F_{ST}**	***p***	**Locus**	**Pos.**	**SNPs**	**F_{ST}**	***p***
Pr1_4	303	C/G	0.091	0.0049	PhytP	304	C/A	0.568	p<0.0001
Pr1_4	316	A/G	0.067	0.0078	PhytP	396	G/-	0.632	p<0.0001
Pr11_4	25	T/C	0.157	0.0010	PhytP	408	C/-	0.600	p<0.0001
Pr27_4	209	C/G	0.336	p<0.0001	PhytP	421	T/-	0.261	p<0.0001
Pr27_4	224	T/C	0.201	0.0010	PhytP	425	A/-	0.542	p<0.0001
PhytP	130	T/A	0.632	p<0.0001					
PL vs. SCO									
Pr11_1	74	T/C	0.113	0.0020	Pr48_2	221	A/G	0.129	0.0029
Pr11_1	307	C/A	0.092	0.0039	Pr48_2	250	C/-	0.074	0.0029
Pr15_1	386	A/6	0.191	0.0039	Pr48_2	251	C/-	0.074	0.0029
Pr26_1	273	C/A	0.165	p<0.0001	Pr48_2	253	A/-	0.074	0.0029

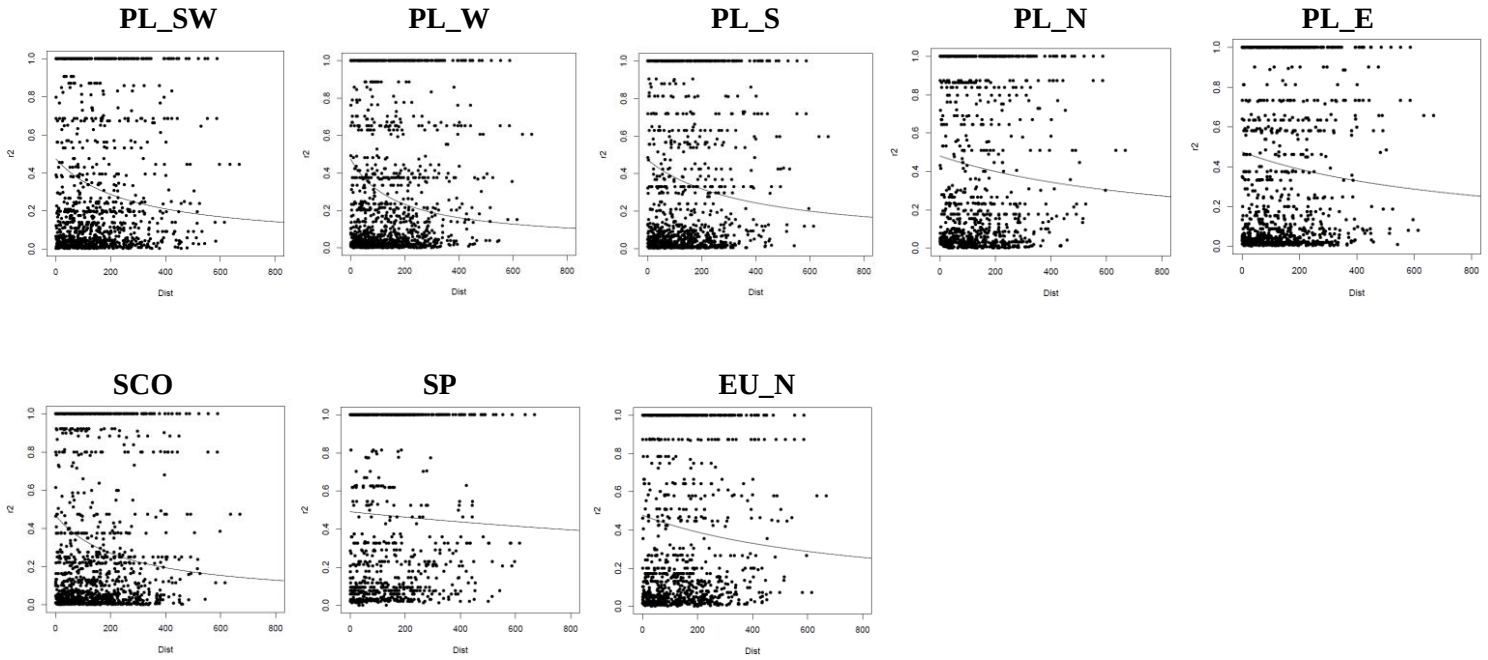
Pr26_1	350	G/A	0.202	p<0.0001	Pr48_2	527	T/A	0.129	0.0039
Pr26_1	173	G/A	0.100	0.0020	Pr48_2	249	T/-	0.074	0.0049
Pr26_1	326	T/A	0.124	0.0068	Pr1_4	303	C/G	0.063	0.0068
Pr46_1	88	G/A	0.114	p<0.0001	Pr1_4	316	A/T	0.066	0.0068
Pr46_1	143	C/A	0.114	p<0.0001	Pr27_4	224	T/C	0.074	0.0098
Pr46_1	304	C/A	0.456	p<0.0001	dhn2	277	C/T	0.085	0.0029
Pr46_1	369	T/C	0.212	p<0.0001	dhn2	22	T/A	0.154	p<0.0001
Pr46_1	301	C/T	0.114	0.0010	PhytP	130	T/A	0.705	p<0.0001
Pr30_2	7	G/A	0.148	0.0010	PhytP	304	C/A	0.705	p<0.0001
Pr35_2	178	T/A	0.139	0.0020	PhytP	396	G/-	0.723	p<0.0001
Pr35_2	350	A/G	0.103	0.0029	PhytP	408	C/-	0.703	p<0.0001
Pr41_2	185	G/A	0.108	0.0020	PhytP	421	T/-	0.438	p<0.0001
Pr47_2	22	G/A	0.158	0.0010	PhytP	425	A/-	0.668	p<0.0001
Pr47_2	316	T/C	0.107	0.0059					
PL vs. SP									
Pr28_1	70	G/T	0.355	p<0.0001	Pr48_2	611	A/T	0.175	0.0088
Pr46_1	369	T/C	0.194	0.0059	Pr1_4	161	C/A	0.503	p<0.0001
Pr16_2	292	G/A	0.150	0.0068	Pr11_4	25	T/C	0.183	0.0059
Pr16_2	302	G/C	0.375	p<0.0001	Pr27_4	224	T/C	0.698	p<0.0001
Pr35_2	350	A/G	0.168	0.0049	Pr27_4	284	C/T	0.223	0.0010
Pr42_2	239	G/A	0.367	0.0020	PhytP	130	T/A	0.905	p<0.0001
Pr42_2	265	G/A	0.468	0.0010	PhytP	304	C/A	0.905	p<0.0001
Pr47_2	375	T/C	0.202	0.0049	PhytP	396	G/-	0.924	p<0.0001
Pr47_2	381	A/C	0.202	0.0010	PhytP	408	C/-	0.905	p<0.0001
Pr47_2	22	G/A	0.302	p<0.0001	PhytP	421	T/-	0.633	p<0.0001
Pr48_2	158	T/G	0.173	0.0029	PhytP	425	A/-	0.869	p<0.0001
Pr48_2	513	T/C	0.439	p<0.0001					

SUPPLEMENTARY FIGURES

Supplementary Figure 1. Multilocus silent theta for five geographical regions in Poland and reference Scots pine populations from Northern Europe, Scotland and Spain. See Table 3 for details.



Supplementary Figure 2.
Scatter plot of the squared correlation coefficient of allele frequencies (r^2) as a function of distance in base pairs between pairs of polymorphic sites in five geographical regions in Poland and reference Scots pine populations from Northern Europe, Scotland and Spain at 29 nuclear loci combined. Decline in linkage disequilibrium is shown by nonlinear fitting curve of the mutation-recombination-drift model. Recombination rate parameter ρ (standard error in parenthesis) for PL_SW is ρ=0.0074 ± 0.0009, PL_W is ρ=0.0122±0.0013, PL_S is ρ=0.0025±0.000, PL_N is ρ=0.0055±0.0008, PL_E is ρ=0.0023±0.0005, SCO is ρ=0.0082±0.0010, SP is ρ=0.0007±0.0004, EU_N is ρ=0.0025±0.0005.



References

Achaz G (2009) Frequency Spectrum Neutrality Tests: One for All and All for One. Genetics 183 (1):249-258. doi:10.1534/genetics.109.104042

Austerlitz F, Mariette S, Machon N, Gouyon PH, Godelle B (2000) Effects of colonization processes on genetic diversity: Differences between annual plants and tree species. Genetics 154 (3):1309-1321

Blarquez O, Carcaillet C, Bremond L, Mourier B, Radakovitch O (2010) Trees in the subalpine belt since 11 700 cal. BP: origin, expansion and alteration of the modern forest. Holocene 20 (1):139-146

Boratyński A (1993) Systematics and geographical distribution. [In:] Biology of Scots pine. Edit Białobok S, Boratyński A, Bugała W (in Polish) Sorus, Poznań, Kórnik

Burczyk J, Dzialuk A, Lewandowski A (2000) Genetic variability of Scots pine (Pinus sylvestris L.) on the seed orchard in Gniewkowo. (in Polish). Sylwan CXLIV (7): 65-74

Cheddadi R, Vendramin G, Litt T, François L, Kageyama M, Lorentz S, Laurent J, Beaulieu J, 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

Corander J, Tang J (2007) Bayesian analysis of population structure based on linked molecular information. Mathematical Biosciences 205 (1):19-31

Derory J, Scotti-Saintagne C, Bertocchi E, Le Dantec L, Graignic N, Jauffres A, Casasoli M, Chancerel E, Bodenes C, Alberto F, Kremer A (2010) Contrasting relationships between the diversity of candidate genes and variation of bud burst in natural and segregating populations of European oaks. Heredity 104 (5):438-448

Dvornyk V, Sirvio A, Mikkonen M, Savolainen O (2002) Low nucleotide diversity at the pal1 locus in the widely distributed Pinus sylvestris. Molecular Biology and Evolution 19 (2):179-188

Eckert AJ, Bower AD, Wegrzyn JL, Pande B, Jermstad KD, Krutovsky KV, St. Clair JB, Neale DB (2009) Association Genetics of Coastal Douglas Fir (Pseudotsuga menziesii var. menziesii, Pinaceae). I. Cold-Hardiness Related Traits. Genetics 182 (4):1289-1302

Excoffier L (2004) Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model. Molecular Ecology 13 (4):853-864

Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1:47-50

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources 10 (3):564-567

Garcia-Gil MR, Mikkonen M, Savolainen O (2003) Nucleotide diversity at two phytochrome loci along a latitudinal cline in Pinus sylvestris. Molecular Ecology 12 (5):1195-1206

Giertych M, Oleksyn J (1992) Studies on genetic variation in Scots pine (*Pinus sylvestris L.*) coordinated by IUFRO. . Silvae Genetica 41 (3):133 - 143

González-Martinez SC, Wheeler NC, Ersoz E, Nelson CD, Neale DB (2007) Association genetics in *Pinus taeda* L. I. Wood property traits. Genetics 175 (1):399-409

Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. Theoretical and Applied Genetics 38:226-231
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 (12):1247-1266
Hudson RR (2000) A New Statistic for Detecting Genetic Differentiation. Genetics 155 (4):2011-2014
Hudson RR, Boos DD, Kaplan NL (1992) A statistical test for detecting geographic subdivision. Mol Biol Evol 9 (1):138-151
Hudson RR, Kreitman M, Aguade M (1987) A Test of Neutral Molecular Evolution Based on Nucleotide Data. Genetics 116 (1):153-159
Hurme P, Repo T, Savolainen O, Paakkonen T (1997) Climatic adaptation of bud set and frost hardiness in Scots pine (Pinus sylvestris). Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere 27 (5):716-723
Ingvarsson PK, Garcia 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 (3):1845-1853
Karhu A, Hurme P, Karjalainen M, Karvonen P, Kärkkäinen K, Neale D, Savolainen O (1996) Do molecular markers reflect patterns of differentiation in adaptive traits of conifers? Theoretical and Applied Genetics 93 (1-2):215-221
Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. Ecology Letters 7 (12):1225-1241
Kinloch BB, Parks GK, Fowler CW (1970) White Pine Blister Rust: Simply Inherited Resistance in Sugar Pine. Science 167 (3915):193-195. doi:10.1126/science.167.3915.193
Krakau U-K, Liesebach M, Aronen T, Lelu-Walter M-A, Schneck V (2013) Scots Pine (Pinus sylvestris L.). In: Pâques LE (ed) Forest Tree Breeding in Europe, vol 25. Managing Forest Ecosystems. Springer Netherlands, pp 267-323
Kujala S, Savolainen O (2012) Sequence variation patterns along a latitudinal cline in Scots pine (Pinus sylvestris): signs of clinal adaptation? Tree Genetics & Genomes 8 (6):1451-1467
Latta RG (2004) Relating processes to patterns of genetic variation across landscapes. Forest Ecology and Management 197 (1-3):91-102
Le Corre V, Kremer A (2003) Genetic variability at neutral markers, quantitative trait loci and trait in a subdivided population under selection. Genetics 164 (3):1205-1219
Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25 (11):1451-1452
Lynch M, Crease TJ (1990) The analysis of population survey data on DNA sequence variation. Mol Biol Evol 7 (4):377-394
Naydenov K, Senneville S, Beaulieu J, Tremblay F, Bousquet J (2007) Glacial vicariance in Eurasia: mitochondrial DNA evidence from Scots pine for a complex heritage involving genetically distinct refugia at mid-northern latitudes and in Asia Minor. BMC Evolutionary Biology 7 (1):233
Neale DB, Ingvarsson PK (2008) Population, quantitative and comparative genomics of adaptation in forest trees. Curr Opin Plant Biol 11 (2):149-155. doi:10.1016/j.pbi.2007.12.004
Neale DB, Kremer A (2011) Forest tree genomics: growing resources and applications. Nat Rev Genet 12 (2):111-122
Nei M (1987) Molecular Evolutionary Genetics. New York, Columbia University Press
Oleksyn J, Tjoelker MG, Reich PB (1992) Growth and biomass partitioning of populations of European Pinus sylvestris L. under simulated 50° and 60° N daylengths: evidence for photoperiodic ecotypes. New Phytologist 120 (4):561-574

Parducci L, Jørgensen T, Tollefsrud MM, Elverland E, Alm T, Fontana SL, Bennett KD, Haile J, Matetovici I, Suyama Y, Edwards ME, Andersen K, Rasmussen M, Boessenkool S, Coissac E, Brochmann C, Taberlet P, Houmark-Nielsen M, Larsen NK, Orlando L, Gilbert MTP, Kjær KH, Alsos IG, Willerslev E (2012) Glacial Survival of Boreal Trees in Northern Scandinavia. Science 335 (6072):1083-1086

Petit RJ, Aguinagalde I, de Beaulieu J-L, Bittkau C, Brewer S, Cheddadi R, Ennos R, Fineschi S, Grivet D, Lascoux M, Mohanty A, Muller-Starck G, Demesure-Musch B, Palme A, Martin JP, Rendell S, Vendramin GG (2003) Glacial Refugia: Hotspots But Not Melting Pots of Genetic Diversity. Science 300 (5625):1563-1565

Prus-Glowacki W, Urbaniak L, Bujas E, Curtu AL (2012) Genetic variation of isolated and peripheral populations of *Pinus sylvestris* (L.) from glacial refugia. Flora 207 (2):150-158

Pyhäjärvi T, Garcia-Gil MR, Knürr T, Mikkonen M, Wachowiak W, Savolainen O (2007) Demographic History Has Influenced Nucleotide Diversity in European *Pinus sylvestris* Populations. Genetics 177 (3):1713-1724

Pyhäjärvi T, Salmela MJ, Savolainen O (2008) Colonization routes of *Pinus sylvestris* inferred from distribution of mitochondrial DNA variation. Tree Genetics & Genomes 4 (2):247-254

Ray N, Currat M, Excoffier L (2003) Intra-Deme Molecular Diversity in Spatially Expanding Populations. Molecular Biology and Evolution 20 (1):76-86. doi:10.1093/molbev/msg009

Sabor J (2006) Intraspecific variability of forest trees. Scots pine and Norway spruce. [In:] The elements of genetics and selective breeding of forest trees. Edi Sabor J (in polish) Wyd CILP, Warszawa

Salmela MJ, Cavers S, Cottrell JE, Iason GR, Ennos RA (2013) Spring phenology shows genetic variation among and within populations in seedlings of Scots pine (Pinus sylvestris L.) in the Scottish Highlands. Plant Ecology & Diversity 6 (3-4):523-536

Savolainen O, Pyhäjärvi T (2007) Genomic diversity in forest trees. Current Opinion in Plant Biology 10 (2):162-167

Savolainen O, Pyhäjärvi T, Knürr T (2007) Gene Flow and Local Adaptation in Trees. Annual Review of Ecology, Evolution, and Systematics 38 (1):595-619

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

Tajima F (1989) Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123 (3):585-595

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 Genetics & Genomes 5 (1):117-132

Wachowiak W, Boratyńska K, Cavers S (2013a) 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 (2):225-238

Wachowiak W, Iason GR, Cavers S (2013b) Among population differentiation at nuclear genes in native Scots pine (Pinus sylvestris L.) in Scotland. Flora - Morphology, Distribution, Functional Ecology of Plants 208 (2):79-86

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 (8):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 (1):29-40

Wachowiak W, Salmela MJ, Ennos RA, Iason G, Cavers S (2010) High genetic diversity at the extreme range edge: nucleotide variation at nuclear loci in Scots pine (*Pinus sylvestris* L.) in Scotland. Heredity 106:775-787

Wang XR, Szmidt AE, Lindgren D (1991) Allozyme differentiation among populations of *Pinus sylvestris* L. from Sweden and China. Hereditas 114 (3):219-226

Weir BS, Cockerham CC (1984) Estimating F-Statistics for the Analysis of Population Structure. Evolution 38 (6):1358-1370

Willis KJ, van Andel TH (2004) Trees or no trees? The environments of central and eastern Europe during the Last Glaciation. Quaternary Science Reviews 23 (23-24):2369-2387

Wright SI, Andolfatto P (2008) The Impact of Natural Selection on the Genome: Emerging Patterns in Drosophila and Arabidopsis. Annual Review of Ecology, Evolution, and Systematics 39 (1):193–213

Wright SI, Gaut BS (2005) Molecular population genetics and the search for adaptive evolution in plants. Molecular Biology and Evolution 22 (3):506-519

Zeng K, Shi S, Wu C-I (2007) Compound Tests for the Detection of Hitchhiking Under Positive Selection. Molecular Biology and Evolution 24 (8):1898-1908