10.14751/SZIE.2017.093

SZENT ISTVÁN UNIVERSITY

FACULTY OF HORTICULTURAL SCIENCE

DEPARTMENT OF BOTANY

PHYLOGEOGRAPHY AND ADAPTIVE GENETIC VARIATION OF SCOTS PINE (*PINUS SYLVESTRIS* L.) POPULATIONS FROM THE CARPATHIANS AND THE PANNONIAN BASIN

DOCTORAL (PH.D.) DISSERTATION

ENDRE GYÖRGY TÓTH

SUPERVISOR: DR. MÁRIA HÖHN

BUDAPEST

2017

10.14751/SZIE.2017.093

PH.D. SCHOOL

Name:	Doctoral School of Horticultural Science
Field:	Crop Sciences and Horticulture
Head of the Ph.D. school:	Prof. Dr. Éva Zámboriné Németh, Head of Department of Medicinal and Aromatic Plants SZENT ISTVÁN UNIVERSITY, Faculty of Horticultural Science
Supervisor:	Assoc. Prof. Dr. Mária Höhn, Head of Department of Botany and Botanical Garden of Soroksár SZENT ISTVÁN UNIVERSITY, Faculty of Horticultural Sciences

The applicant met the requirement of the Ph.D. regulations of the SZENT ISTVÁN UNIVERSITY and the thesis is accepted for the defense process.

....

Head of Ph.D. School Supervisor

10.14751/SZIE.2017.093

JURY MEMBERS

CHAIRMAN:

DR. TAMÁS PÓCS
(PROF. EMERITUS, MHAS)

SCIENTIFIC COMMITTEE:

DR. GYÖRGY DÉNES BISZTRAY (PH.D.)
DR. PÉTER ÓDOR (PH.D.)
DR. LÁSZLÓ NAGY (PH.D.)
DR. JÚLIA HALÁSZ (PH.D.)

OPPONENTS:

DR. SÁNDOR BORDÁCS (PH.D.)
DR. ZSUZSANNA BENYÓNÉ GYÖRGY (PH.D.)

SECRETARY:

DR. PAPP VIKTOR (PH.D.)

CONTENT

ABBREVIATIONS

A	Haplotypes per population
a.s.l.	Above sea level
ABC	Approximate Bayesian Computation
AMOVA	Analysis of molecular variance
A_R	Allelic richness
BF	Bayes factor
bp	Base pairs
BP	Before Present
B-SAP	Bacteria specific amplification polymorphism
C.I.	Confidence interval
cal.	Calibrated
cDNA	Complementary DNA
cpDNA	Chloroplast DNA
cpSSR	Chloroplast simple sequence repeats
D	Tajima's D test of non-neutral theory of molecular evolution (selection)
D_A	Nei's chord-distance
DIYABC	Do It Yourself ABC
DNA	Deoxyribonucleic acid
D_{sh}^2	Goldstein's mean genetic distance between individuals
EUFORGEN	European Forest Genetic Resources Programme
F	Fu and Li's F test of non-neutral theory of molecular evolution (selection)
FDR	False discovery rate
F_{IS}	Inbreeding coefficient
FLA	Fragment length analysis
F_{ST}	Fixation index
G_{ST}	Nei's coefficient of gene variation
H	Fu and Li's H test of non-neutral theory of molecular evolution (selection)
h	Haploid genetic diversity (for non-coding cpSSR loci)
Hd	Haplotype diversity
H_e	Expected heterozygosity
H_o	Observed heterozygosity
H_R	Haplotypic richness
HWE	Hardy-Weinberg equilibrium
IBD	Isolation by distance
ITS	Internal transcribed spacer
k	Average number of nucleotide differences
ka	Kilo ages/ kiloannus
kb	Kilo base pairs
LD	Linkage disequilibrium
LGM	Last Glacial Maximum (LGM: 21 ± 2 ka cal BP)
LPG	Late Pleniglacial (LPG: 26.5–15 ka cal BP)
MCMC	Markov chain Monte Carlo algorithm
mtDNA	Mitochondrial DNA
Mya	Million years ago
n	Number of samples analyzed (for candidate gene loci)
N_a	Number of different alleles
NCBI	National Center for Biotechnology Information
nDNA	Nuclear DNA

N_e	Effective number of haplotypes
N_{eff}	Number of effective alleles
Nh	Number of haplotypes (for candidate gene loci)
Nm	Absolute number of migrants
$n_{non\text{-}syn}$	Number of non-synonymous mutations
Non-syn.	Non-synonymous site
N_p	Number of private alleles
ns	Number of haploid sequences (for candidate gene loci)
nSSR	Nuclear simple sequence repeats
n_{syn}	Number of synonymous mutations
p/*P*	P-value, significance level (statistical)
PCA	Principal Component Analysis
PCR	Polymerase chain reaction
P_h	Number of private haplotypes
PhiPT (Φ_{PT})	Pairwise population genetic differentiation, an analogue of F_{ST}
QTL	Quantitative Trait Locus
r	Correlation coefficient
RFLP	Restriction fragment length polymorphism
r_{xy}	Association between two quantitative variables (matrices)
S	Number of polymorphic (segregating) sites
Sc	Scenario
Sing	Number of singleton mutations
SNP	Single-nucleotide polymorphism
SSR	Simple sequence repeats
Syn.	Synonymous site
TCS	Templeton-Crandall-Singh network analysis
TPM	Two-Phased Mutation model
UTR	Untranslated region
θ_π	Theta pi nucleotide diversity indicator
π	Average number of pairwise differences per site, nucleotide diversity

1. INTRODUCTION AND OBJECTIVES

Scots pine (*Pinus sylvestris* L.) is a long-lived coniferous tree species of the Pinaceae family which occupies a continuous range as the dominant tree species of the Eurasian taiga communities (Pravdin 1969). It is a key species of many forests types, various ecosystems like e.g. pine and pine-birch boreal forests, hemiboreal forests, mixed pine-birch or pine-pedunculate oak forests (Giertych and Mátyás 1991, Matías and Jump 2012, Pividori *et al.* 2016). On the southern and western edge of its wide distribution Scots pine has many disperse populations that are considered peripheral occupying ecologically different habitat types.

Peripheral populations have been widely studied across species' distribution range. Enzyme polymorphism revealed overall low structuring of populations, but elevated differentiation was reported between populations presumed to have derived from different glacial refugia (Müller-Starck *et al.* 1992). Studies performed on the southern European provenances showed that these are distinct from those occupying the northern European regions (Mejnartowicz 1979, Kieliszewska-Rokicka 1981).

Modern-day organelle and nuclear DNA marker studies (with non-coding microsatellite markers) presented high genetic diversity and differentiation at the European distribution periphery of Scots pine (Robledo-Arnuncio *et al.* 2005, Labra *et al.* 2006, Scalfi *et al.* 2009, Belletti *et al.* 2012). Populations from Central-Eastern Europe, particularly from the Carpathians and the Pannonian Basin, presented low level of differentiation among the populations and the impact of Holocene population fragmentation (Bernhardsson *et al.* 2016).

Former studies that aimed to elucidate adaptive genetic variation of *P. sylvestris* have found overall moderate level of diversity and differentiation in continental European populations. Additionally, signals of negative selection and effects of historical demography on nucleotide variation were detected (Pyhäyärvi *et al.* 2007, Wachowiak *et al.* 2009, 2011, Kujala and Savolainen 2012).

Along the the Carpathian Mountain range Scots pine is distributed in island-like isolated populations (Fekete and Blattny 1913), but there are also scattered natural populations sustained in mixed forest stands, with broad-leaved species in the western Pannonian Basin, at the foothills of the Alps (Pócs 1960, Fekete *et al.* 2014). The genetic structure of these peripheral populations of the Carpathian distribution was highly affected by the postglacial climate warming, forcing Scots pine to immigrate into edaphically specialized habitat types. Indeed, Scots pine natural populations are distributed in the Carpathians on a large elevation gradient, located in sites of divergent ecological conditions, including humid, cool peatbogs and sunny, dry, rocky outcrops. In addition, historical human-mediated activities further increased habitat fragmentation and considerably reduced population census sizes. In part, as a consequence of this, only isolated and island-like populations have been sustained (Giertych and Mátyás 1991). Moreover, forest activities and agricultural practices caused shifts in the species' distribution, Scots pine being even spread outside its natural range.

Drawing on macrofossil and pollen evidence, studies on the Quaternary vegetation history of *P. sylvestris* within this region conclude that Scots pine, along with other cold-tolerant and

drought-tolerant conifer taxa, inhabited the Carpathians and the Pannonian Basin in the full glacial and later in the beginning of the postglacial period (Rudner *et al.* 1995, Rudner and Sümegi 2001, Jankovská and Pokorný 2008). *In situ* findings also suggest that conifers and in particular boreal and cool temperate tree taxa like Scots pine in Central Europe and in the Carpathians survived the Last Glacial Maximum (LGM: 20,000–19,000 Before Present) in small, patchy and discontinuous glacial refugia (Willis *et al.* 1998, Magyari *et al.* 2014a, b). Species would have had sustained populations in isolated, so-called cryptic refugia, with favourable conditions both for Scots pine and for other boreal and temperate species (Rull 2009, 2010, Sommer and Zachos 2009). Altogether, Scots pine has a complex spatio-temporal history in Central-Eastern Europe during the Holocene, influenced mainly by oscillations in the climate and affected by human activities, as a consequence of which overall reduction of population size was experienced (Feurdean *et al.* 2007). Pollen records also indicate that populations on the lowlands of the Pannonian Basin dramatically declined during the Holocene (Willis *et al.* 1995, Magyari 2011). Furthermore, on mid-altitudinal to high-altitudinal sites in the Carpathians, species abundance varied greatly by location (Willis 1994, Birks and Ammann 2000).

Despite combined pollen, macrofossils and organelle DNA analysis that could detect glacial refugia in the Carpathian Basin along the Danube, previous molecular studies performed in the region reported lack of geographic structure both with mtDNA and cpDNA within the Carpathian Mountains (Cheddadi *et al.* 2006, Bernhardsson *et al.* 2016). Similarly, no variation and no phylogeographic structure in mitochondrial DNA was found in provenance trials conducted in the region by Čelepirović *et al.* (2009). Sequence variation studies, involving candidate gene loci, are also missing from this region, hence adaptive potential of these peripheral populations are yet unknown.

The overall objectives of this research were as follows:

- Highlight the current population structure of the selected peripheral Scots pine populations native to Central-Eastern Europe, most of which formerly were not included in molecular studies.
- Identify genetic relationships, degree of diversity and divergence and infer gene flow between the studied stands.
- Describe historical demographical processes (expansions-contraction) and circumscribe putative refugia within the studied region that might have existed in the time of the Pleistocene.
- Assess the nucleotide diversity, divergence at candidate gene loci to infer adaptive nucleotide variation of peripheral populations as signs of local adaptation.

2. LITERATURE REVIEW

2.1. Taxonomical position and botanical characteristics of Scots pine

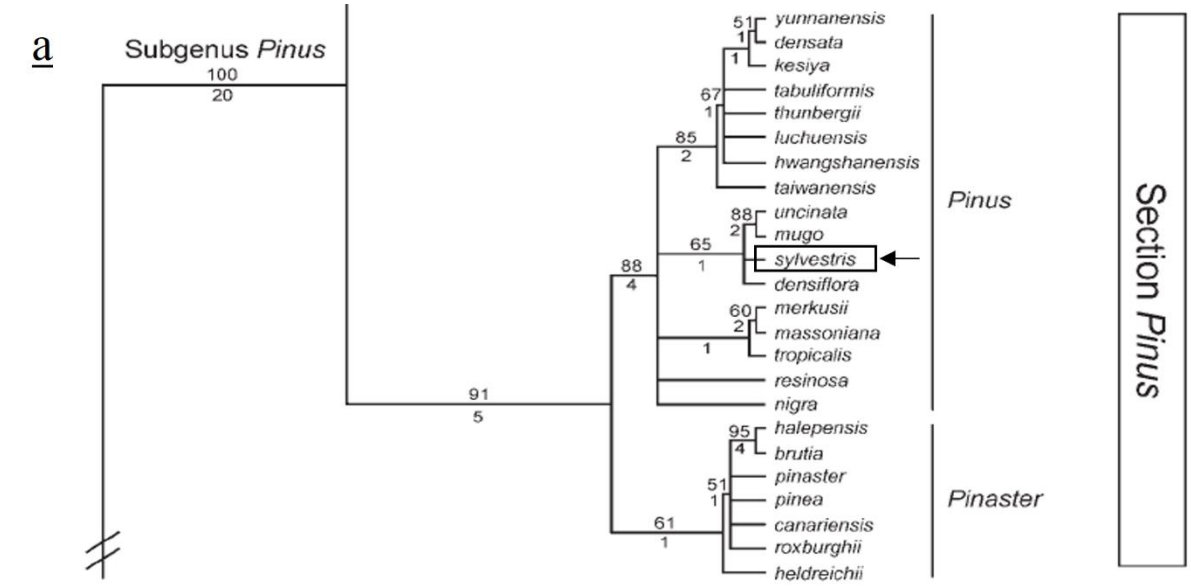
Scots pine (*Pinus sylvestris* L.) is a member of the *Pinus* genus, which is the largest extant genus among the conifers. According to Price *et al.* (1998), the genus comprises 111 species based on morphology, anatomy, cytology, crossability, secondary metabolites, proteins and DNA comparisons. Although, in 2001 Farjon recognized only 109 species belonging to the genus, in the recent study by Gernandt *et al.* (2005) 101 species were described. Their evaluation to infer phylogeny of the related taxa was based on chloroplast DNA sequence polymorphism (*matK, rbcL* and ITS regions). Additionally, within this study the classification was also evaluated with morphological and distributional traits to gain robustness for differentiation **(Fig. 1a, b)**.

According to these, Scots pine belongs to the subgenus *Pinus* (diploxylon or hard pines), section *Pinus* and subsection *Pinus* (Gernandt *et al.* 2005). The species chromosome number is 2n=24 (Mirov and Stanley 1959).

Due to the wide geographical distribution and morphological heterogeneity of *P. sylvestris* large number of subspecies and varieties have been described by different authors in the last century (144 by Carlisle 1958, 150 by Tutin *et al.* 1964 and 140 by Farjon 1998) although, only few of them are presently accepted (Farjon 1998, Price *et al.* 1998). Major variants of the *Pinus sylvestris* complex according to Price *et al.* (1998) are listed and considered as subspecies: subsp. *sylvestris* Linnaeus, subsp. *hamata* (Steven) Fomin, subsp. *kalundensis* Sukacev, subsp. *lapponica* Fries, subsp. *sibirica* Ledebour. According to The Plant List database (http://www.theplantlist.org, accessed: 01.01.2017) there are 95 records deposited, including forms, subspecies and varieties, but only *Pinus sylvestris* L. and two varieties are accepted (var. *hamata* Steven, var. *mongholica* Litv.).

Scots pine is a 30-40 meter high evergreen tree with a dark brown bark (on the lower part) and a pale ochre red and flaking trunk (on the upper part) (Skilling 1990, Tutin *et al.* 1964). The trunk can reach up to 1 m diameter in mature age. The crown of the tree is conical, but later becomes irregular **(Fig. 2a)**. The shape of the canopy is very versatile due to the wide geographical distributional range and diverse habitats. At young age the branches are in whorls around the trunk. Habit of pine trees in the boreal region are elongated, even at an elevated age they preserve the conical crown (Gencsi and Vancsura 1992). On its southern distribution Scots pine has an expanded crown with slightly stronger branches. The crown usually forming an irregular shape. The buds are acute, light brown and more or less resinous. The twigs are yellowish-green but later become grayish-brown. Leaves in paires, 30-70 mm long and about 2 mm thick. The needles are twisted and glaucous **(Fig. 2b)** and stay on the tree for three or four years. The color of the cones are dull yellowish brown. The shape of the cone is acute, deflexed and caducous. The apophysis is flat or shortly pyramidal on the back of the cone. The cones stay on the tree for 1 or 2 years. The seeds are winged and about 3-4 mm in size (with wing over 1 cm in length). The wing is about 3 times longer than the size of the seed (Skilling 1990, Gencsi and Vancsura 1992). The flowers are

blooming in May and are pollinated by wind, the seeds are dispersed by wind at the end of the next year (Tutin *et al.* 1964).



b

	leaf vascular bundles	fascicle sheaths	needles per fascicle	needle resin ducts	cone scale	mature cones	umbo prickle	seed wing	umbo position	distribution
Section *Pinus*										
subsect. *Pinus*	two	persistent	2 (3)	medial, external (septal)	thick	open	variable	articulate	dorsal	Eurasia (E NA)

Fig. 1: Consensus tree of combined *rbcL* and *matK* matrix (a) and synopsis of character variation in subsections of *Pinus* (b) according to Gernandt *et al.* (2005).

Fig. 2: Mature Scots pine individuals from the Eastern Carpathian, Poiana Stampei peat bog. (a): trunk and canopy characteristics, (b): twig, needle and cone characteristics.

2.2. Modern day distribution of Scots pine

Pinus sylvestris is the second most distributed conifer in the word and its area extends on the northern hemisphere from Western Europe to Central-Eastern Asia (Tutin *et al.* 1964, Debreczy and Rácz 2000, Hytteborn *et al.* 2005, Labra *et al.* 2006, Debreczy *et al.* 2011), and covers more

than 14 000 km from the Iberian Peninsula towards the Siberian plain, reaching the Sea of Okhotsk. This vast distribution of *Pinus sylvestris* is a consequence of its wide ecological tolerance (Hytteborn *et al.* 2005). Present distribution along the Eurasian range of Scots pine is discontinuous. The westernmost limit stretches from Portugal to western Scotland (Cipriano *et al.* 2013, Pavia *et al.* 2014) and even further to Ireland. A recent radiocarbon dated palynological evaluation confirms the existence of a native population in Rockforest in Ireland, where the species could have survived habitat loss caused by human exploitation (McGeever and Mitchell 2016). In the North (Scandinavia and Siberia), along the subarctic region Scots pine grows in a continuous distribution. However on the southern boundary, where during the Pleistocene Scots pine was more widespread, the following postglacial climate warming in the Holocene resulted in island-like and fragmented populations **(Fig. 3a, b)**. Scots pine survived in arid southern mountainous areas in Spain, Turkey and Crimea. However, in these regions, the gradual decline of populations and the expansion to higher latitudes has been prognostized (Hytteborn *et al.* 2005, Labra *et al.* 2006, Matías and Jump 2012, Cipriano *et al.* 2013). Due to its ability to grow on different altitudinal gradients, Scots pine populations are present at the sea level on the northern limits, and also even at 2000-2700 m altitude on the eastern coast of the Black Sea in Turkey (Turna and Güney 2008, Floran *et al.* 2011). The wide geographical range of the species and the various habitat types occupied by the populations leading to various growth forms and accumulation of a large gene stock provide good basis to study species' phylogeography (Turna 2003, Turna and Güney 2008, Laurentin 2009). Modern day distribution of the species is expected to change in the future due to the climate change and land use via anthropogenic events. Increased drought stress has resulted in slow growth rate, low recruitment, and in some cases resulted in massive mortality. Direct climatic effects are acting together with indirect effects due to altered biotic interactions including outbreaks of insects, pathogens, and parasites and increased herbivory linked to declining ecosystem productivity (Matías and Jump 2012).

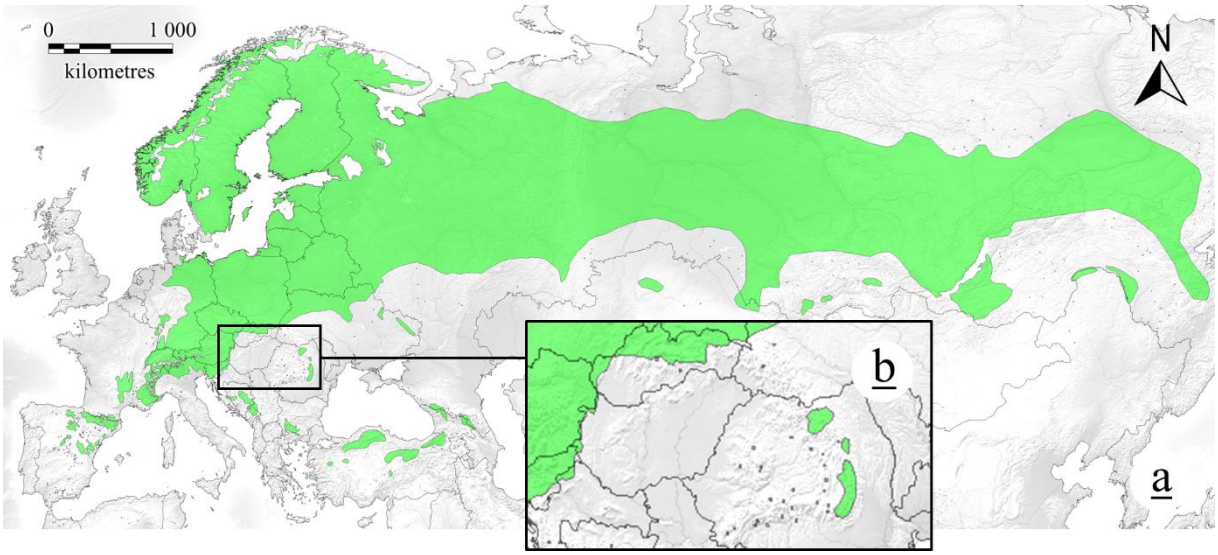


Fig. 3: Present day natural distribution of Scots pine. (a): across Eurasia, (b): in Central-Eastern Europe, along the Carpathian Mountains, and the Pannonian basin. The natural distribution of Scots pine is marked in green according to the EUFORGEN database, with modifications by the author.

2.3. Historical aspects of Scots pine

2.3.1. Quaternary history of Scots pine in Europe

Fossilized pollen and plant micro- and macrofossils show that in response to the Quaternary (ca. 2.8 Mya) climatic fluctuations the biosphere has experienced dramatic changes, including large-scale species' range shifts, population contractions, expansions and extinctions, as well as aggregation and disassociation of forest communities (Petit *et al.* 2008). Such a complex pattern of historical background can be seen also in case of *Pinus sylvestris*, most of the information deriving from pollen accumulation and plant fossils, both charcoal, macro- and megafossil occurrences. According to fossil data of closely related species, the ancestral gene pool of *Pinus* genus was located at the middle latitudes of North America and Western Europe (Mirov 1967, Millar 1993, Millar 1998). Beginning from the first known pine species *P. belgica,* dated to the early Cretaceous period, about 130 million years ago, pines were continuously found throughout the European continent from the lower Cretaceous (Mesozoic) to the Quaternary period. During this time the influence of the climate change has expanded and contracted the distribution range of the genus several times (Bennett *et al.* 1991, Bennett 1995, Richardson 1998, Kelly and Connolly 2000, Cohen *et al.* 2013). In the Quaternary period cyclic changes of the continental ice sheets highly affected the vegetation cover, and so the distribution pattern of the European conifers. Pines, including Scots pine, were present and population fluctuations were documented both in the glacial (full glaciation) and interglacial periods.

2.3.2. Pre-LGM and full-glacial history of Scots pine in Europe

High number of macrofossil evidences and sedimentary records show the presence of pines in the last full-glacial period between 100,000–18,000 BP (BP; Before Present) (Richardson 1998). Most of the fossils are charcoals of cold- or drought-tolerant coniferous taxa like *Pinus sylvestris*, *Pinus cembra, Larix, Picea* and *Juniperus* species originating from central and southern Moravia and from the territory of the Pannonian Basin (both from Austria and Hungary) (Jankovská and Pokorný 2008). One well-developed and another weakly developed charcoal layer was also found in the Carpathian Basin including molluscs and macro-charcoal remains dating back to between ca. 70,000 and 15,000 years BP. These fossils of woodland species mostly consists of *Picea* trees from the north and *Pinus sylvestris* from the south of the Carpathian basin (Rudner and Sümegi 2001). Several *Pinus sylvestris* charcoal macrofossils have been identified from all across Eastern Europe and especially from several sites of Hungary. These fossils were dated as early as 30,000–20,000 BP (Rudner *et al.* 1995, Richardson 1998). Another paleobotanical survey in the Hungarian plain at Hortobágy shows pollen fossils dating back to around 30,000–20,000 BP. This suggests that before the last glacial maximum the area was covered by mixed conifer-hardwood forests including *Pinus sylvestris,* and this species existed on both lower and higher floodplain zones (Magyari 2011). From the same region pollen diagram assemblages showed, not just the presence, but the dominance of Scots pine in woodlands between 35,000–30,000 BP from the Late Würm and later from the Holocene (Félegyházi and Tóth 2003; Sümegi *et al.* 2005). Similarly, pollen

and macrofossil evidences from the Northern Carpathians (Slovakia and Czech Republic) show, that Weichselian full-glacial montane forests were dominated by *Larix, Pinus cembra, Pinus sylvestris* and *Picea* between 50,000 and 16,000 BP (Jankovská and Pokorný 2008). This and many other *in situ* findings (Stieber 1967, Haessaert *et al.* 1996, Damblon 1997, Willis *et al.* 2000) prove that full-glacial forests existed in Southern and Central Europe, and survived because of special micro-environmental conditions. Such an example was evidenced from the Carpathian basin where Late Pleniglacial (LPG: 26.5–15 ka cal BP) pollen records showed that forest patches or scattered trees probably also sustained on north-facing hillslopes, and at moister sites of the loess plateaus. These patches were dominated by boreal and cool temperate tree species including coniferous taxa like *P. sylvestris, P. mugo, P. cembra* (Magyari *et al.* 2014a).

By the last glacial maximum (LGM: 21±2 ka cal BP), when the ice sheets were at their most recent maximum extension, pine species only survived in patchy small and discontinuous glacial refugia mostly on ice-free areas (Willis *et al.* 1998, Matías and Jump 2012). Refugial territories existed not just along the Southern European territories, but also in the northern latitudes like in Scandinavia, where fossils of Scots pine was also evidenced (Stewart *et al.* 2010, Matías and Jump 2012, Parducci *et al.* 2012). However, the largest refugial territories were localized in the Southern European regions, in Italy, in the Iberian Peninsula (Labra *et al.* 2006, Cheddadi *et al.* 2006) and in the Balkans. Furthermore there are evidences for refugial locations in the Alps and in East-Central Europe e.g. the Hungarian plain (Naydenov *et al.* 2005, Cheddadi *et al.* 2006, Feurdean *et al.* 2011). Morphological and molecular genetic studies also support that southern refugia were located in Turkey in Asia Minor (Pyhäjärvi *et al.* 2007, Pyhäjärvi *et al.* 2008, Jasińska *et al.* 2014). Northern refugia of boreal trees were found in Scandinavia (Parducci *et al.* 2012), and more recently for Scots pine near Moscow where it could possibly survived during the LGM (Buchovska *et al.* 2013). Major parts of Russia, except the westernmost parts and north-western coast, remained ice-free during the LGM (Svendsen *et al.* 2004). At these sites pines could have been also survived and after the glaciers had retreated, these source populations expanded into Western-Europe and Fennoscandia (Prus-Głowacki *et al.* 2011, Kyrkjeeide *et al.* 2014). There are evidences that these refugial zones were also shelters to many other vascular plants (Ehrich *et al.* 2008, Tollefsrud *et al.* 2008).

2.3.3. Post-glacial history of Scots pine in Europe

As a response to the climate warming and to the retreat of the continental ice sheets, the expansion of the refugial populations into the European continent started between 16,000 and 12,000 BP. (Richardson 1998, Pyhäjärvi *et al.* 2008). In the Eastern Carpathians the expansion of *Pinus* and *Betula* dominated forests started already around 16,300 BP (Magyari *et al.* 2014b), firstly Scots pine spread mainly on the low altitudinal sites (Pérez-Obiol and Juliá 1994). In Greece, increase of pine dominance has started 16,000 years BP (Bottema 1974) and in the Iberian and Italian Peninsula about 16,500 years BP (Pérez-Obiol and Juliá 1994). As Scots pine spread northward it has reached the western Alps and southern France about 14,000 years BP (Walker 1995, Cheddadi

et al. 2006.). According to Huttunen *et al.* (1992), increase of diploxylon pines like *Pinus sylvestris*, occurred much later in the high mountains, like in the Rhodope in Bulgaria from 12,700 BP.

Following the late-glacial period, at the beginning of the Holocene interglacial period, ca. 11,200 years ago, pine fossil records are highly complex, because both latitudinal changes and altitudinal migration interfered the distribution pattern. In the Rhodope, the Alps and in Central Europe, decrease of pine density was estimated (Huttunen *et al.* 1992, Lowe 1992), while in the Hungarian basin and on the Iberian Peninsula pines sustained their populations without disturbance and no change has been detected (Pérez-Obiol and Juliá 1994, Willis *et al.* 1998). In this period, also a noticeable altitudinal change has been revealed in the treeline elevation of Europe, such as in the Southern Carpathians, where treeline increase was estimated in response to the rapid climatic fluctuations (Magyari *et al.* 2012).

Following the retreat of the ice fields between 12,000–11,000 years ago in the north-western part of Europe only treeless tundra vegetation persisted. After the gradual warmup of the exposed tundra pines that were forced to migrate northward occupied the territories of the European lowlands (Huntley and Birks 1983). At the same time in Southern Europe in the Mediterranean region the previously steppe dominated vegetation started to be invaded by woodlands of deciduous species. Firstly, boreal populations such as *Betula* and *Salix* were dominating (Adams 1997), later the spread of *Corylus, Fraxinus, Quercus, Tilia* and *Ulmus* species caused the decline of *Pinus* species.

During the northward migration, pines reached most of the Scandinavian territory in 10,000 BP and began to increase their presence in Britain as well. It has become dominant in Finland about 9000 years BP, with uncertain origins of the populations (Pyhäjärvi *et al.* 2008, Richardson 1998). At the same time ca. 9900 years BP, Scots pine has become locally dominant in Scotland (Richardson 1998, Cheddadi *et al.* 2006, Matías and Jump 2012) and finally reached the maximum distribution area all-across Europe around 8000 years ago (Matías and Jump 2012, Kullman and Kjälgren 2006).

At the northern latitudes 7800 years ago dominance of pine's have further increased and reached northern edges of Fennoscandia, but later 6800 years ago a decline was estimated in the southern parts of this region. According to Kremenetsky *et al.* (1994) the distribution area extended from Fennoscandia eastward to the Siberian steppes at that time.

Because climate change enabled the northward spread of deciduous species also, like *Corylus* and *Quercus* (Richardson 1998), about 7000 years ago Scots pine started to retreat from several sites (Richardson 1998, Matías and Jump 2012). Beginning from this time, between 4800 and 4200 years BP the species declined also in the northern and western parts of the British Isles. Pollen diagrams from western Scotland and northern Ireland also reveal the decline of populations at about 4000 BP (Birks and Birks 1980, Birks and Williams 1983, Watts 1984 and 1988). In the Iberian Peninsula and in Italy between 5700 and 3200 years BP, populations from their maximum distribution have retreated to the present distribution area (Willis *et al.* 1998, Matías and Jump 2012). Meanwhile between 5000 and 3000 BP in northern Scandinavia, *P. sylvestris* gradually

suffered a withdrawal (Eronen 1979, Eronen and Hyvärinen 1982). These movements across Europe occurred due to the combination of changes in the climate and the ongoing anthropogenic activities. There are paleobotanical evidences of expanded forest fires during drought periods that destroyed pine forests and favoured the spread of warm tolerant, broad-leaf species (Richardson 1998). As a result Scots pine survived on marginal territories, in extreme conditions of poor edaphic soils, far from the refugia of deciduous species (Bennett 1984, Richardson 1998, Willis *et al.* 1998, Matías and Jump 2012).

Natural withdrawal of Scots pine has been continuing since 4000 BP till today, as a consequence of the current climate warming and anthropogenic influence via land use. The land use as the food producing economy increased, caused dramatical changes in the environment. Several factors like the character of soil, bedrock, hydrography and hydrological conditions became important for the growing human settlements and migrating groups (Sümegi *et al.* 2002). Anthropogenic activities such as grazing pressure, soil depletion by pasturing started to intensified in the Bronze and Iron Age (Willis *et al.* 1998, Sinclair *et al.* 1999, Cheddadi *et al.* 2006, Matías and Jump 2012). Moreover, in the mid Holocene, the burning of the *Pinus* woodlands, and also the timber production related to the mid to late Bronze Age human activity were probably highly influenced hydrological (climatic) changes, -likely increased aridity- which altered quick shrinkage and disappearance of populations (Tipping *et al.* 2007, Tipping *et al.* 2008).

In the recent history Scots pine populations are also affected by a secondary recolonization most likely caused by forest activities and agricultural practices. Roman occupation (ca. 2000 years ago) in Southern and Central Europe further reduced the size of animal populations and caused changes in farming practices so it has facilitated mountain pine forest development. As the result of the secondary colonization at these sites, *Pinus sylvestris* became an established woodland tree of the natural or human induced disturbed areas (Richardson 1998).

By concluding historical background of the species, Scots pine has an unexpectedly complex glacial and post-glacial history influenced mainly by oscillations of the climate, and recent and past anthropogenic activity.

2.4. Specificities of the inheritance of organelle and nuclear genomes in conifers

Plant DNA coexists in the cytoplasm in separate places: in the mitochondria, in plastids and in the nucleus. DNA of the three different organelles are inherited in different ways. The chloroplast and mitochondrial genomes are inherited uniparentally, while the nuclear genome is inherited biparentally (Mátyás 2002, Bock and Knoop 2012). The inheritance patterns of organelle and nuclear genes can be used to unravel the complexity of gene flow, as they are predicted to result in very different distribution of genetic diversity within and among populations (Petit *et al.* 1993, Petit *et al.* 2005, Petit and Vendramin 2007).

The haploid chloroplast DNA (cpDNA) of dicotyledons, where all angiosperm (broad-leaf) tree species belongs shows exclusively maternal inheritance. Pines from the gymnosperms, apart from a few exceptional cases have cpDNA of paternal inheritance, accordingly cpDNA is

distributed by pollen (Mátyás 2002). The chloroplast genome is highly conserved, it has lower mutation rate than the nuclear genome. In gymnosperms chloroplast-specific polymorphic assays are able to facilitate the analysis of population differentiation and gene flow (Powell *et al.* 1995).

The mitochondrial genome is also haploid. It is relatively volatile with multiple repetition of one or more prolonged sequences that can vary among species or even among populations. Accordingly, these sequences can be very useful for species identification. In most of the plants and animal species mitochondrial DNA (mtDNA) is inherited maternally (Birky 1995, Mátyás 2002, Forrest *et al.* 2000). Tree species that are exceptions belong to the gymnosperms like *Sequoia sp.*, *Calocedrus sp.* (Hipkins *et al.* 1994). Mitochondrial genome provides information about seed dispersal via maternally inheritance (Forrest *et al.* 2000).

Diploid nuclear genome shows the highest level of variability due to the recombination in the meiosis. Variability depends on the number of chromosomes, which is 2n = 24 in case of pines (Mátyás 2002). The analysis of the molecular organization revealed, that a significant proportion of the nuclear genome consists of repetitive sequences, that are not included in the transcription process (Plomion *et al.* 2011).The study of organelle and nuclear genomes allows not only a more thorough analysis of the inheritance, but permits the detection of ongoing genetic processes in populations, by understanding the interplay of evolutionary factors, demography and population structure.

2.5. Genetic diversity and differentiation at species distribution range periphery

Genetic characteristics of peripheral populations are strongly influenced by the interplay of genetic drift, gene flow and natural selection. These ongoing processes are also affected by demography and spatial distribution on the edge of species distribution (Eckert *et al.* 2008). Former studies evaluating ongoing processes at core and edge (peripheral) populations found differences in diversity and differentiation. A profound discussion of Safriel *et al.* (1994) explains these differences raised by selection and other processes affecting genetic variation.

Peripheral populations can be genetically degraded with low additive variance due to the reason that the required level of heterozygosity through gene flow rarely reach them (Carson 1959, Safriel *et al.* 1994) or have low genetic diversity due to genetic drift. According to Mayr (1966, 1970), even if peripheral populations are genetically depauperate, they can still experience high incoming gene flow which can suppress the local selection processes. Mayr also highlighted that isolated peripheral occurrences are less variable genetically and are unsuccessful in survival, but adaptation processes coupled with new environment will occasionally allow a population to occupy new niches. In this case population size, density and diversity can increase (Cook 1961).

From another aspect, peripheral populations can maintain high genetic variability as they can be adapted to the specific and fluctuating environment (Fisher 1930, 1950). If environmental conditions are unstable in the periphery populations either may evolve several adaptive genetic combinations (specialists), or genotypes with high phenotypic plasticity (generalists) are maintained. Peripheral populations rarely experience favorable environmental conditions,

therefore may undergo adversity selection (evolve to survive conditions) (Whittaker and Goodman 1979).

Historical demography has a substantial effect on the genetic diversity of populations. As species colonize new geographical gradient of environmental conditions, reproduction and survival will be the highest where ecological conditions are the most optimal. In these regions species abundance (population size and density) is expected to be the highest, while on the periphery species become less abundant. This hypothesis also predicts that natural populations at geographic range periphery will have lower genetic diversity compared to those located centrally. This phenomena is the "abundant center hypothesis" (Hengeveld and Haeck 1982, Sagarin and Gaines 2002, Sagarin *et al.* 2006).

Differences in genetic variation are not large, but often unequally partitioned within the species distribution ranges (Channell 2004, Eckert *et al.* 2008). Recently, 134 studies representing 115 species by Eckert *et al.* (2008) revealed that peripheral populations very often experience demographic processes, either historical, contemporary or both, that can lead to lower genetic diversity, but higher genetic differentiation, while other studies found that the manifestation of this hypothesis on the partition of genetic diversity occasionally is contrary **(Table 1)**.

Table 1: Partition of genetic diversity in literature that compares central (core) and peripheral populations of Gymnosperms as reviewed by Channell (2004).

Species name (common name)	Higher genetic diversity		Citation
	Core	Periphery	
Picea abies (L.) H. Karst. (Norway spruce)	-	Yes	Lagercrantz and Ryman 1990
Pinus contorta (Lodgepole pine)	Yes	-	Yeh and Layton 1979 Fazekas and Yeh 2001
Pinus edulis Engelm. (Pinyon pine)	-	Yes	Betancourt *et al.* 1991
Pinus jeffreyi Balf. (Jeffrey pine)	Yes	-	Furnier and Adams 1986
Pinus rigida Mill. (Pitch pine)	-	Yes	Guries and Ledig 1982
Pinus sylvestris L. (Scots pine)	-	Yes	Dvornyk 2001
Pseudotsuga menziesii (Mirb.) Franco (Douglas-fir)	Yes	-	Li and Adams 1989

Fragmentation likely occur along the species distribution periphery, where population sizes contemporary (periodically) change. Fragmentation process segregates species' continuous distribution into smaller spatially isolated habitats, which can highly affect genetic architecture (Young *et al.* 1996, Hampe and Petit 2005). This process can lead to an erosion of genetic variation which can cause loss of heterozygosity with reduced individual fitness in short term and limits species response to changing selection via reduced allelic richness (A_R) in long term (Frankel *et al.* 1995, Young *et al.* 1996). According to Gilpin (1991) and Raijmann *et al.* (1994) the major effects of fragmentation can increase random genetic drift, raise inbreeding in subpopulations and increase probability of local extinction. Fragmentation can give rise to bottlenecks and such populations continue to lose alleles by random genetic drift (Ellstrand and Elam 1993, Young *et al.* 1996). It is also evidenced that wind pollinated and seed dispersed tree species experience direct reduction of gene flow due to the increased interpopulation distances (Templeton *et al.* 1990).

Genetic divergence among fragmented populations are often increased, especially in those populations where the structure is unstable and the species sustained in dispersed occurrences (Fahrig and Merriam 1994, Hoffmann and Blows 1994, Hoffmann and Parsons 1997, Garciá-Ramos and Kirkpatrick 1997). Hampe and Petit (2005) highlighted that different evolutionary processes affect the species' northern ("Leading edge") and southern ("Rear edge") fragmented distribution range **(Fig. 4)**. Accordingly, processes can influence the diversity and the differentiation of the species. They emphasize that high levels of genetic differentiation are often observed among such isolated populations, leaving footprints also on diversity. Latter can greatly vary due to demographic events and local ecological factors.

Large scale historical demographic events such as extinction as well as colonization can also affect divergence of peripheral populations (Lande 1992, Whitlock 1992, Barton 2001, Lenormand 2002). Likewise, temporal and spatial variation in migration rates and effective population sizes may also influence diversity and divergence among subpopulation (Whitlock 1992, Young *et al.* 1996).

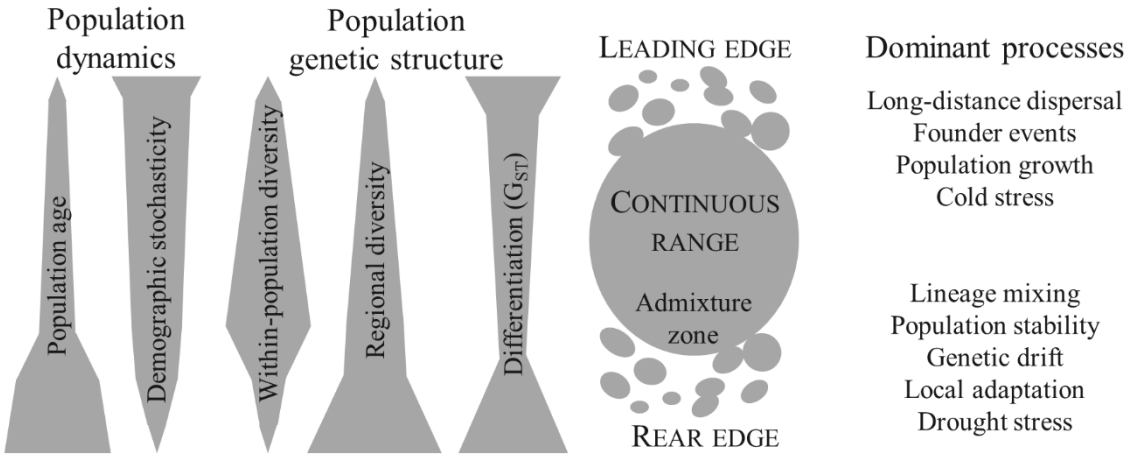


Fig. 4: Population features and genetic processes at the distribution edge of species ranges according to Hampe and Petit (2005).

2.6. Adaptation processes at species distribution range periphery

Adaptation to geographically peripheral (and ecologically marginal) habitats depends on complex interactions between dispersal, habitat quality, form and strength of selection, mating and genetic architecture of underlying traits. These genetic processes associated with adaptation are often acting simultaneously, hence are hardly traceable.

While core populations occupy the optimal niche that they belong to and thrive, populations at the range periphery are prone to local extinction and suffer from severe stress due to the temporal and spatial variation of the environment, demographic stochasticity and edge effects (Pulliam 2000, Kawecki 2008). Species peculiar intraspecific genetic variability that develops in such an environment is often expressed in adaptation to the local climate and other environmental factors. In this term, local adaptation is defined as the higher fitness (lifetime reproductive success) of local individuals compared with non-local individuals of the same species (Biere and Verhoeven 2008).

In the periphery, increased environmental stress with decreased genetic mixing due to isolation may lead to stronger local adaptation compared to the core. Local adaptation to the changing environmental conditions determine distribution ranges and likely affect species responses to climate change (Biere and Verhoeven 2008, Kreyling *et al.* 2014).

Habitats from core to the periphery become less suitable, therefore survival and reproduction declines, likewise population density and habitat occupancy. Conversely, a profound adaptation may result in range expansion over evolutionary time. This requires populations to be well adapted to their habitats, so that their abundance and persistence increase and higher number of offspring are produced (Kawecki 2008). These populations may evolve to colonization sources for a species, facilitating the occupation of outlying areas.

Dispersal has a profound influence on adaptation. Usually, gene flow is asymmetric, originating from the core to the periphery. Transmits alleles to the periphery and often suppress the effects of local adaptation (Holt and Gomulkiewicz 1997, Kawecki 2008). Gene flow from the central population is an important factor to replenish local genetic variation. It sustains diversity of marginal populations and increases the number of individuals (Kawecki and Holt 2002, Kawecki 2008). Although, gene flow suppresses ongoing local adaptation and new alleles/genotypes compete with the locals, the positive effects of dispersal may often be more important in promoting adaptation to marginal habitats (Kawecki 2003, 2008). Immigrant rare alleles/genotypes improve genetic fitness in the marginal habitats, hence the adaptive potential of the peripheral populations. Dispersal also acts backwards, in which a low proportion of gene flow from the periphery reaches the core population. Thus, some alleles/genotypes establish in the core and will transmit marginally favored alleles/genotypes back from the periphery.

Temporal environmental and ecological fluctuations often occur on the range periphery and can affect ongoing adaptation processes (Schmid 1985, Linhart and Grant 1996, Holt 2003, Leimu and Fischer 2008, Kawecki 2008). A good year temporary may increase population size and locally adaptive allele frequency and also makes local selection more effective, while a bad year may depress locally adapted population and suppress it with immigrant alleles (Callaghan *et al.* 1996, Ronce and Kirkpatrick 2001, Kawecki 2008). Temporal fluctuations along with spatial disturbances can have irreversible effects on adaptation, hence on evolution of specialization (Ronce and Kirkpatrick 2001). It can also result in migration-meltdown, which means that immigration brings locally maladapted alleles, decreases local density and increases immigration rate (Lenormand 2002). Competition of species within a habitat (both inter-individual and intra-individual level) cause temporally reduced density, more asymmetric gene flow, frequency-dependent selection in addition to drift. These trigger idiosyncratic effects on adaptation (Kawecki 2008).

Genetic drift removes rare advantageous recessive alleles more likely and selection on these recessive alleles is less effective against gene flow, which also mediates dominant alleles. Thus, non-recessive alleles are favorable for adaptation and less prone to the effect of gene-swamping e.g. loss of genetic variance at a locus under selection because gene flow is too high (Nagylaki 1975, Lenormand 2002, Kawecki 2008).

Ongoing selection affects adaptation linked loci and creates positive linkage-disequilibrium between favored alleles. This process increases genetic variation, therefore makes selection more effective (Lenormand 2002, Kawecki 2008). Inter-loci associations are only partially removed by recombination at the meiosis, but tends to reduce this disequilibrium and thus also the variance in fitness. Therefore, if migration load is high the efficacy of selection is low (Slatkin 1975, Lenormand and Raymond 1998, Lenormand 2002). Strong between-loci associations (high disequilibrium) are important when multiple allele combinations manifest in adaptation. If adaptation is mediated with a particular combination of alleles and if these alleles are rare they will occur together rarely, hence selection will act against them (Kawecki 2008).

Nevertheless, adaptation can occur if the direction of selection changes for an allele among habitats (core-periphery) and if the intensity of selection covaries negatively among habitats at several loci that are maintained as polymorphic by recurrent mutations (Kawecki *et al.* 1997, Lenormand 2002).

2.7. Phylogeography of the Carpathians

The Carpathian region was largely neglected in terms of phylogeographical studies compared to other European montaneous areas like the Alps or the Pyrenees. In the paper of Ronikier *et al.* (2011) the authors emphasize the need of a detailed literature review of the recent results, to assess intraspecific phylogeographical structure and differentiation of high-mountain plants within the Carpathians. All literature data confirm that the main phytogeographical division of the Carpathians is coherent with the geographical units defining the Western, Eastern and the Southern Carpathians (Georgescu and Doniţă 1965, Ronikier *et al.* 2011). Geographic segregation is clearly evidenced for some alpine herbaceous species like *Hypochaeris uniflora* (Mráz *et al.* 2007) and *Campanula alpina* (Ronikier *et al.* 2008), their intraspecific genetic structure differentiate the populations of the Western, Eastern and Southern Carpathians. Similarly, the ITS sequence variation in case of the montane-subalpine *Melampyrum sylvaticum* shows the separation of the Western and Eastern Carpathian populations (Tesitel *et al.* 2009).

Ronikier *et al.* (2011) highlighted that genetic differentiation identified between the Western and Eastern Carpathians might be due to the constitution of the Carpathian arch, where the lowering elevations from the Western towards the Eastern Carpathians and the intervening low depressions represent strong landscape barriers for alpine species (potentially persisted also during the LGM and earlier cold periods).

Genetic pattern identified for *Picea abies* shows congruent pattern with the perennial alpine species by delimiting the Western Carpathians from the Southern Carpathians. According to the authors, the Carpathians harbored two separate refugia for spruce, one being located in the Western Carpathians and another in the northern part of the Eastern Carpathians (Tollefsrud *et al.* 2008). Studies of *Abies alba* confirmed phylogeographyic structure within the Carpathians with a sharp boundary in the species distribution. Moreover a contact zone of divergent genetic lineages

between the Western and Eastern Carpathian regions was detected (Liepelt *et al.* 2009, Gömöry *et al.* 2012).

Pinus cembra, another subalpine conifer preserved high genetic diversity and remarkable spatial isolation between populations. This species shares common history with other above mentioned species. Although, significant genetic differentiation between the two parts of the natural range was low, Carpathian populations proved to be highly differentiated by cpSSR markers (Höhn *et al.* 2005, 2009, 2010). In the most recent study with nSSR markers only weak but detectable segregation was revealed between the Western and Eastern Carpathian populations. Results identified post-glacial contraction of the species' range with strong effects of genetic drift over historical gene flow (Lendvay *et al.* 2014).

Pinus mugo, the dwarf alpine conifer species presents high differentiation between sites located in the Western, North-Eastern (Ukrainian) Carpathians and the Alps (Dzialuk *et al.* 2012, Sannikov *et al.* 2011). Despite the fact that *P. mugo* has only scattered, island-like populations along the Carpathian arch (with an isolation probably longer than the Holocene) the species was able to survive the LGM at low elevational sites in the Western Carpathians, likewise in the regions of the Eastern Carpathians. It is also presumed that the North-Western parts of the Carpathians were colonized from a different refugia than the South-Eastern Carpathians (Tsaryk *et al.* 2006, Sannikov *et al.* 2011).

Differentiation of the Eastern Carpathian populations from the rest of the Carpathian regions (as for the above mentioned species), was also detected in case of the alpine dwarf woody *Salix herbacea* with AFLP data (Alsos *et al.* 2009). This species could have persisted in the Carpathians as well as in the Alps and the surroundings during the LGM.

Intraspecific differentiation along the Eastern and Southern Carpathians of different species is hard to be assessed, because spatial extent of the genetic lineages/structures does not match exactly (Ronikier *et al.* 2011). Topography of the massifs do not constitute barrier restricting gene flow for some alpine and sub-alpine species like *Campanula alpina* (Ronikier *et al.* 2008, 2011). Intraspecific genetic patterns of species suggest glacial survival in separate local refugia within the Eastern and Southern Carpathians (Ronikier *et al.* 2011).

The Central-Island Mountains (Apuseni) has a particular position in species' phylogeography. Coniferous species' populations share common pattern with those from the Southern Carpathians, hence often form one genetic cluster. This is the case for *A. alba*, *P. abies* and *Pinus sylvestris* (Liepelt *et al.* 2009, Tollefsrud *et al.* 2008, Tóth *et al.* 2017). Apuseni Mountains has low to moderate altitudinal sites, which harbor alpine species like *Arabis alpina* and *Hypochaeris uniflora*, but the origin of these populations is still uncertain. Most probably during the glacial periods Apuseni Mountains could have served as refugial area (Ronikier *et al.* 2011) and/or the present pattern might reflect lineages of colonization routes from the south, either from the Balkan or from the south-eastern parts of the Carpathians (Ehrich *et al.* 2007, Mráz *et al.* 2007).

2.8. Former studies of Scots pine phylogeography with non-coding molecular markers

2.8.1. Overview of mtDNA based molecular studies

Most mitochondrial DNA (mtDNA) markers show low level of intrapopulation and higher level of interpopulation variation in pines (Forrest *et al.* 2000). Although most of the two needle pines like Scots pine have winged seeds, genome dispersal depends on seed travelling (because mtDNA is maternally inherited) reaching shorter distances compared to the paternally inherited pollen, which spread through a larger area, resulting increased differentiation among the populations (Floran *et al.* 2011, Korpelainen 2004). A major difficulty in using mtDNA analysis is that the low level of variation in their exons and introns, although an indel has been discovered in the intron B/C of the mitochondrial *nad 1* gene (Naydenov *et al.* 2007, Soranzo *et al.* 2000). The mitochondrial *cox1* gene has also proved to be useful as an RFLP marker (Sinclair *et al.* 1999). RFLPs require complicated procedures is hardly applicable in case of large number of individuals (Forrest *et al.* 2000). Sinclair *et al.* (1998) studied 466 individuals from 20 natural pine populations in Scotland. A homologous probe was constructed for the *cox1* mitochondrial gene and used to detect mtDNA RFLP variation. They could distinguish two common (*A, B*) and one rare (*C*) variant. Mitotype *A* was present in all sites, but *B* was present only in three western populations. According to the geographical distribution of mitotypes, authors suggested that Scots pine derived not only from continental Europe via England, but also from a western refugia, probably existed in Ireland or western France. Later they have extended the research with further 18 populations from the continental Europe (Sinclair *et al.* 1999) and have detected three major mitotypes (*A, B* and *D*). The greatest mitotype diversity was found in the seven Spanish populations. Mitotype *D* was only present in the Sierra Nevada region. Although there were differences among the other European regions the rest showed little or no mtDNA diversity. Italian populations clearly showed mitotype *A*, while the Fennoscandians were fixed for *B*. Their results suggest that recent structure of Scots pine populations from Western Europe are supposedly deriving from three main sources (Sinclair *et al.* 1999). PCR based polymorphic marker system was also developed which is based on the variations of repeats numbers of SSR regions in the mitochondrial genome (Soranzo *et al.* 1999, Forrest *et al.* 2000). Soranzo *et al.* (2000) studied 747 individuals from 23 populations. They have found two distinct length variants (*A* and *B*), and by sequencing the 2.5 kb region six individuals from each haplotype (three *A* and three *B*). The two mitotypes differed by the insertion of a 31 bp fragment. Haplotype *A* was fixed in all the northern European populations, including Scotland, Poland, France, Lithuania and Czechia. The *A* and *B* variants were both present in the populations of the Iberian Peninsula. The *B* haplotype was dominant in the Pyrenees region and within some populations in Central Spain. Labra *et al.* (2006) studied eight populations including the Italian Alps and the Apennines. The analysis of the polymorphisms in the *nad 1* intron sequence confirmed, that the populations have the same mitotype (mitotype *A*) as the Central Europeans, which are characterized by the absence of the 31 bp deletion in the *nad 1* intron. Naydenov *et al.* (2007) reported a novel polymorphic mtDNA region in Scots pine, the intron 1 of the *nad7* gene, which is informative for genetically distinct maternal lineages. They have tested 54 populations from the Eurasian distribution of the species. Altogether four multi–locus haplotypes (mitotypes)

were observed for *nad1* and *nad7* introns. According to the geographic distribution, populations were highly structured. Haplotype *AA* (72,3 % of total) was largely distributed and found in most of the populations sampled. *CA* (5,8 %) was present mainly in Asia minor, in Turkey in the Pontide Mountains. Mitotype *AB* (4,6 %) was found in the high mountain regions of the Iberian Peninsula. Mitotype *BA* (17,3 %) was distributed in the lowland regions of middle to northern Eastern-Europe, and was also dominant in the Baltic region and Russia. Pyhäjärvi *et al.* (2008) also used the *nad1* and *nad7* introns proved to be previously polymorphic. They have sampled 37 populations from the western distribution of the species range from Europe. Similarly, to Naydenov *et al.* (2007) they have found four mitochondrial haplotypes (A, B, C and D). Haplotype *A* was present in all populations, considered to be an ancestral type. Haplotype *B* was fixed to the Iberian Peninsula. Haplotype *C* and *D* were completely new and not described earlier. *D* was restricted to Kalabak in Turkey, while *C* was found in central, northern (including Fennoscandia) and eastern part of Europe. Čelepirović *et al.* (2009) studied the *nad1* intron region on samples collected from two international provenance trials located in Croatia and Hungary. Provenances were mainly from Russia, Poland, Germany and some few from Central Europe and the Balkan Peninsula. All 42 provenances proved to belong to haplotype *A* with all examined 344 individuals.

Studies reporting the results of mitochondrial DNA markers are listed in **Table S1**, and major studies depicted on **Fig. 5**.

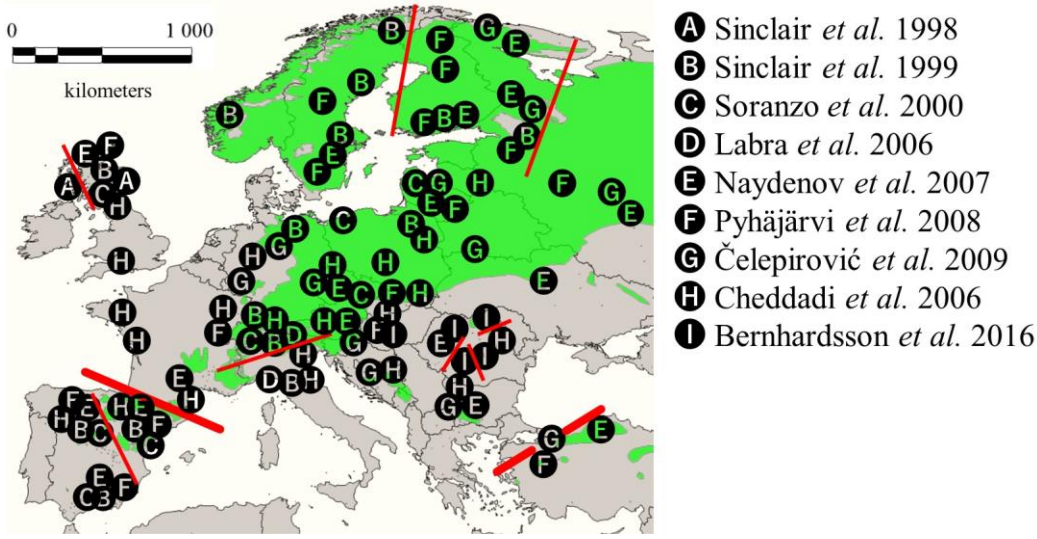


Fig. 5: Geographic map of mtDNA studies carried out on the natural distribution of Scots pine. Green area marks the current natural distribution of the species according to EUFORGEN database. Lines indicate structural delimitations identified by mtDNA markers (strong line: evidences supported by all studies, thin line: identified in regional studies.

2.8.2. Overview of cpDNA based molecular studies

After sequencing the complete chloroplast genome of *Pinus thunbergii* (Wakasugi *et al.* 1994) new potential marker assays became available (Vendramin *et al.* 1996). In 1995, cpDNA SSR (Simple Sequence Repeats) system was reported for pines, which is more effective than isozyme or RFLP markers (Floran *et al.* 2011, Powell *et al.* 1995). Provan *et al.* (1998) analyzed 15 populations from Scotland and Europe, with the result of higher diversity within the populations,

and small but significant genetic variation detected between populations. Diversity based on the haplotype frequency of Scottish populations was found to be higher than reported earlier in monoterpene and isozyme studies on the same populations by Kinloch *et al.* 1986. Kinloch also evidenced non-significant differences between Scottish and European populations. Naydenov *et al.* (2005) studied the structure of 12 peripheral Bulgarian populations with chloroplast SSRs and terpene analysis. According to the results of the cpSSRs, size variants per locus and number of haplotypes were similar to the results of Provan *et al.* (1998). Haplotype diversity of populations (*H)* was high in all studied mountains (Rila, Pirin, Rhodopes), which suggests that these regions might have been a refugia during the glacial depression. Similarly, high haplotypic diversity was found among populations of the Iberian Peninsula, Meseta region in Spain (Robledo-Arnuncio *et al.* 2005). The analysis of molecular variance showed that genetic variation among populations was low but significant. This was also supported by the results of slightly polymorphic izoenzyme loci by Prus-Głowacki *et al.* (2003). Paleobotanical informations suggest that there has been a recent fragmentation of a historically larger population on the Iberian Peninsula (Robledo-Arnuncio *et al.* 2005). Robledo-Arnuncio *et al.* (2004) analyzed 324 individuals with 6 cpSSR markers from 13 populations from the Iberian Peninsula and found very high haploid genetic variation, low differences among populations, and no clear geographic pattern in the distribution of genetic diversity. In their following study they have found high haploid genetic diversity among populations as previous studies indicated (Robledo-Arnuncio *et al.* 2005). According to AMOVA analysis, they revealed that low differentiation among populations at disjoint distribution could not account for the genetic divergence among the tested areas. The Iberian Peninsula along with Asia Minor (Turkey) was separated from other European populations also in the study of Wójkiewicz and Wachowiak (2016). Authors identified high divergence and substructuring of populations. Genetic clusters were unique and restricted to these two regions, which might also indicate that these regions did not contribute to postglacial recolonization of Europe. Since genetic diversity and differentiation could be accounted for substructuring in Iberian populations association studies were also carried out by using SSR sequences of the chloroplast genome, such as in the work of Soto *et al.* (2010). They aimed to study the association between neutral genetic diversity and species-specific climatic requirements in case of the Iberian pine species. Haplotypic diversity was substantially smaller in thermophilous species (*P. halepensis, P. pinea)* compared to mountain species, like Scots pine. Association between genetic diversity and summer precipitation showed positive correlation. They also found that summer-drought affected Mediterranean populations exhibiting a lower genetic variation, but *P. sylvestris* was an exception to this rule.

Combined microsatellite markers (cpSSR and nSSR) on the southern-western margin of the species distribution strengthened the hypothesis of relictary origin of the isolated native Portuguese populations. Moreover the pattern showed clear separation from other European populations. A local population from northern Portugal reflected gene flow with other populations especially with those from Spain. Furthermore, genetic differentiation presumably raised by geographic barriers, reduced gene flow or even fire propagation. Detected uniqueness of the studied two peripheral

populations also confirmed the glacial relict origin and supported the initial hypothesis of native population of Scots pine in Portugal (Pavia *et al.* 2014).

Cheddadi *et al.* (2006) studied 106 European populations by using cpSSR markers and mitochondrial *nad1* region to identify postglacial source of recolonization and to outline major genetic lineages of the species in Europe. Genetic results were combined with the paleobotanical data modelling ancient vegetation. They have identified two major refugia both from the Mediterranean region (Iberian Peninsula and Apennines), and also found that recolonization of Europe has happened most probably from another refugia located in the Eastern Alps.

Recently, Bernhardsson *et al.* (2016) applied similarly combined markers on Romanian Carpathian and Hungarian populations from East-Central Europe. Results revealed high level of genetic diversity and low level of differentiation among the peripheral populations, with the support of non-significant isolation by distance (IBD). Although, signs of inbreeding and genetic erosion was also reported as an estimated effect of recent population fragmentation and restricted gene flow, Tóth *et al.* (2017) confirmed that natural and relict populations are yet unaffected by the genetic consequences of isolation and fragmentation.

Studies reporting the results of chloroplast DNA markers are listed in **Table S2**, and major studies depicted on **Fig. 6**.

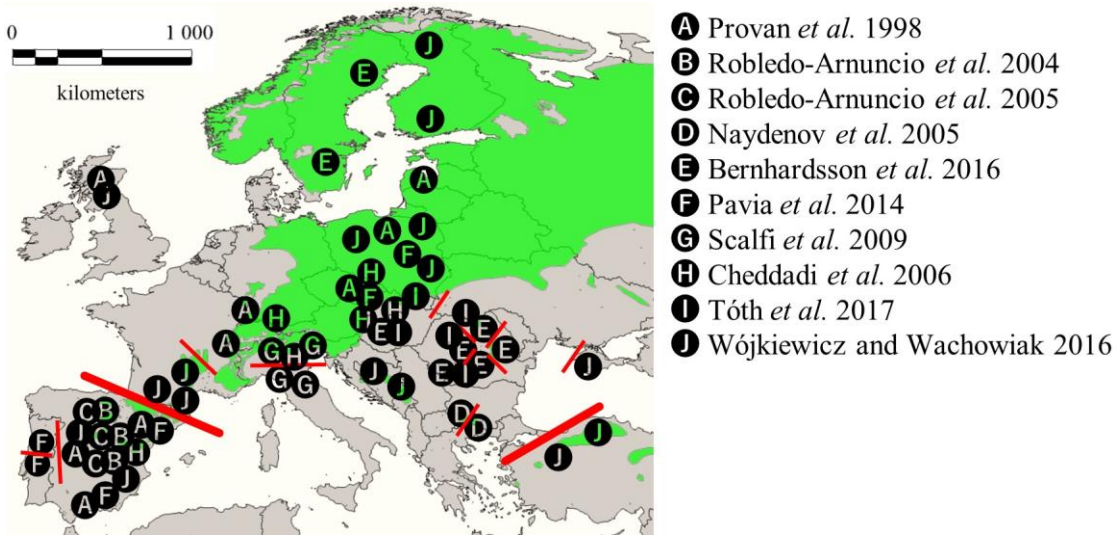


Fig. 6: Geographic map of cpDNA studies carried out on the natural distribution of Scots pine. Green area marks the current natural distribution of the species according to EUFORGEN database. Lines indicate structural delimitations identified by cpDNA markers (strong line: evidences supported by all studies, thin line: identified in regional studies.

2.8.3. Overview of nDNA based molecular studies

Kostia *et al.* (1995) developed and applied nuclear markers (SSRs) to detect polymorphism in *Pinus sylvestris*. Altogether eight microsatellites were identified from 6000 clones screened. Soranzo *et al.* (1998) successfully used an enrichment procedure (White and Powell 1997) to isolate SSRs and they also produced a set of nuclear SSR primer pairs for *P. sylvestris*. Sebastiani *et al.* (2012) developed fifty-five highly polymorphic microsatellite markers isolated from Scots

pine cDNA sequences and tested them for variability in three populations from Russia and Finland. According to their results, only ten of the tested markers displayed consistent and polymorphic patterns, which are useful for further fine-scale population studies.

Despite the fact that nuclear microsatellites are useful to study phylogeographic and gene flow patterns in conifers, such as in Scots pine, developing reliable new markers is difficult (González-Martínez *et al.* 2004, 2010). Due to this reason, cross-amplification of earlier designed SSR markers is crucial and cost-effective way to study highly polymorphic loci on several different species. González and Martínez *et al.* (2004) tested 19 nuclear microsatellite markers designed for *Pinus taeda* L. and 3 for *P. sylvestris* on seven Eurasian hard pine species. Their results show that transferability was slightly higher between the subsection *Pinus* (36–59%) than those to subsections *Pineae* and *Pinaster* (32–45%). Most recently two nSSR multiplex sets has been developed based on *P. taeda* and *P. sylvestris* primers for facilitating high-throughput genotyping. Moreover these sets were fully transferable to nine other *Pinus* species closely related to Scots pine (Ganea *et al.* 2015).

Relying on the existing and available molecular markers Scalfi *et al.* (2009) found significant among population differentiation in the studied populations from the Apennines by comparing them to those from the Alps. Their results also confirmed the genetic distinctiveness of Apennine populations and their possible derivation from different glacial refugia compared to those of the Alps. Belletti *et al.* (2012) used twelve nSSR markers selected from the published literature for studying 21 populations at the Italian Alpine region and the Apennines. Primers were mainly designed for *P. sylvestris* although five were for *Pinus taeda*, and these also proved to be useful (Elsik *et al.* 2000, Elsik and Williams 2001, Zhou *et al.* 2002). They have found high within population genetic diversity and clear separation between populations from the Alps and the Apennines. Bayesian clustering and PCA suggested an east-west subdivision of the populations. By using the previously designed *Pinus taeda* nSSR markers Naydenov *et al.* (2011) have assessed the genetic diversity and population structure of the Bulgarian populations with six markers. According to their results genetic evidence of bottlenecks and the existence of three distinct groups were presumed. Furthermore, nearly all studied populations presented a distinct genetic pattern, and the results confirmed the existence of genetic differentiation between populations from different mountain ranges and watersheds. Their results also supported the documented "mountain effect" influence on the genetic diversity: strong altitudinal variation in morphological and physiological parameters. Recently, Lučić *et al.* (2014) carried out analysis of the genetic variability of seven Scots pine populations in Serbia. According to the four SSRs, designed by Soranzo *et al.* (1998), genetic differentiation of population and spatial isolation by natural barriers was observed between the populations.

Studies discussing the results of nuclear DNA markers are listed in **Table S3**, and major studies depicted on **Fig. 7**.

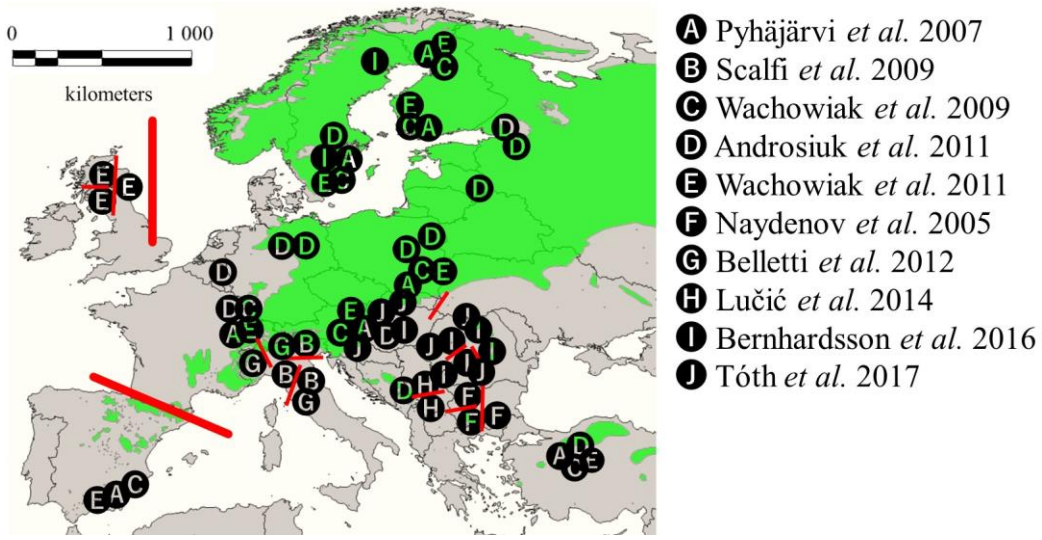


Fig. 7: Geographic map of nDNA studies carried out on the natural distribution of Scots pine. Green area marks the current natural distribution of the species according to EUFORGEN database. Lines indicate structural delimitations identified by nDNA markers (strong line: evidences supported by all studies, thin line: identified in regional studies.

2.9. Former studies of Scots pine phylogeography with candidate gene markers

2.9.1. Overview of SNP based molecular studies of Scots pine

Pyhäjärvi *et al.* (2007) analyzed sequence diversity of 16 sequenced nuclear loci (candidate genes), with variable putative function (*dhy-like, a3ip2, chcs, hprgp-like, gi, pal1, hlh1, lp2, nir, scl, phyo, phyn, phyp, co, adhc, laccase*) in eight *Pinus sylvestris* populations divided into four geographical groups. Among-group differences in the level of silent SNP (synonymous substitution) diversity were not detected and there was some evidence that linkage disequilibrium extended further in northern Europe than in Central Europe. They did not found strong imprints of selection at any specific loci and linkage disequilibrium indicated that selective sweeps have not had a strong impact on the nucleotide variation. Wachowiak *et al.* (2009) examined the nucleotide variation in known members of the dehydrin gene family (cold resistance candidate genes) in eight geographically distant natural stands across Europe. Patterns of nucleotide diversity were compared between northern and southern populations to reveal adaptation to cold. They have detected significant differentiation between populations in allelic frequency at *dhn1*, *dhn3*, and *abaH* loci. Haplotype clustering by geographical distribution was not evidenced at *dhn9*. However, moderate multi-locus nucleotide diversity was found compared to other conifer species. Tests of neutrality (Tajima's D, Fu and Li's H) identified excess of rare and high frequency derived variants by significantly negative multi-locus values (D=-0.72, H=-0.50). Androsiuk *et al.* (2011) applied a novel *KatG* gene based B-SAP marker to distinguish populations from distant geographic origin and to infer Quaternary refugia and postglacial migrations. Spatial variation of *KatG* gene entails two plausible routes of migrations directed from the South toward North-East and North-Central Europe. Putative origin of populations are either from the Balkan and/or from Hungarian refugia. Furthermore, unique alleles detected in Montenegro population (Balkan) unequivocally entails

restricted gene flow and diversification. *KatG* gene also proved to be a potential tool for adaptation studies via its possible linkage with QTLs controlling morphological features, as additionally suspected in this study. Wachowiak *et al.* (2011) studied the nucleotide polymorphism at 12 earlier tested candidate gene loci (*dhn1, dhn2, dhn3, dhn7, dhy2PP, abaR, a3ip2, ccoaomt, chcs, erd3, lp3-1, lp3-3*) across the environmental gradient in Scotland. Their results showed a recent bottleneck which affected the molecular pattern, and a slightly higher genetic variation in Scottish populations compared to the mainland populations, that pointed out their distinct origin. Deficit of rare nucleotide variants were detected in Scottish populations (multi-locus Tajima's D=0.316) as compared with mainland European populations (D =-0.379). The advanced neutrality test of Fay and Wu's H presented a high frequency derived single-nucleotide polymorphisms (H=-0.564 and -1.240, respectively). In the study of Savolainen *et al.* (2011) a set of candidate genes for cold tolerance, budset (*col1, gi, phyn, lp2, dhn1, cry1, ft4, prr1, ztl, myb*) and pollen dispersal records was combined with the earlier results of Pyhäjärvi *et al.* (2007) and with other phylogeographic studies to infer adaptive potential of northernmost Scots pine populations. The reviewed results showed little evidence of increased diversity and admixing colonization routes (genetic lineages) in Scandinavia. Phenotypic correlation with quantitative genetic predictions suggested that increased growth and survival counteracts ongoing selection and evolutionary responses at the northern periphery. Later, the same set of cold tolerance and budset candidate genes (only instead of *ft4*, *ftl2* was used) were used to test demography and adaptive clinal variation of Scots pine (Kujala and Savolainen 2012). Results indicated a recently occurred minor bottleneck, which revealed a reduction of population size about 1% of the current size. Low number of F_{ST} outliers (in *cry1* and *ztl*) were identified and were evidence for clinal selection (only *myb* indicated clinal selection and geographic differentiation) was reported. It is also noticeable that three genes from the set showed significant allele frequency clines along the latitudinal distribution (*dhn1, ftl2* and *prr1*).

The latest study of nucleotide diversity by Wachowiak *et al.* (2014) involved a total of 32 candidate genes related to cellular metabolism, transport, signal transduction and transcription regulation. 19 populations from Poland and 11 reference populations from across Europe were used. No significant differences in nucleotide diversity between populations were detected and 63-73% of detected SNPs were shared across the region. Neutrality test (Tajima's D) provided evidence for selection at six candidate gene loci (*Pr1_5*, *Pr1_46*, *PhytP*, *Pr4_12*, *Pr4_19*, and *Pr4_6*). A rapid decay of linkage disequilibrium was detected at a distance about 200bp (r^2= 0.2-0.4), likewise in the earlier study of Wachowiak *et al.* (2011) with 12 candidate gene loci in Scotland (distance: 400 bp, r^2= 0.2). 11 outlier SNP (across four gene) was detected via multi-locus distribution of F_{ST} values (in *Pr1_4*, *Pr11_4*, *Pr27_4*, *PhytP*). Most of the detected outlier SNPs (74%) were unique. Although, no significant differentiation was measured between Polish populations, principal coordinates analysis revealed similarities between Polish and Northern European populations.

Studies discussing the results of adaptive genetic diversity (based on sequence variation in candidate genes) are listed in **Table S4**.

2.10. Summary on former studies of Scots pine phylogeography

2.10.1. Current understanding and future perspectives of Scots pine phylogeography

According to the revealed phylogeographic patterns Scots pine was extremely widespread in Eurasia during the Quaternary being present even in the outskirts of the glaciated territories. As well, Scots pine survived in several southern latitude refugia across Europe in highly disperse isolated occurrences.

Putative refugia of the most widespread genetic lineage was detected from the Central Continental Europe. It is widely confirmed by all marker types and all authors that colonization presumably originated from the (Sub) Mediterranean areas like the Balkan Peninsula but also from around the Eastern Alps and the surroundings of the Danube plain (western Pannonian basin). Fennoscandia and northern European Baltic regions were most probably colonized from two main directions, from Western Europe and from the Russian Plain. Studies have reported that these migration processes took place between 14,000 BP and 8000 BP and presumably started on low elevational sites and later concomitantly with the rising temperature towards high altitudes. Other lineages seems to be less successful in the post-glacial recolonization process.

Molecular studies, macrofossil and pollen data analyses concluded that Scots pine in some regions was unable to colonize the vast areas of the European mainland. Lineages from these refugia, located in the Iberian Peninsula, Italian Apennines, Turkey, and Crimea and in some parts of Asia Minor did not contribute to the recolonization of Europe after the Last Glacial Maximum (LGM). Based on evidences from the compiled literature the highest regional heterogeneity (highest differentiation) revealed by all marker systems (mitochondrial, chloroplast and nuclear) can be located to the periphery, especially to those localities that were situated along the distribution edge of the species, like the Balkan Peninsula, Carpathians, Alps, Apennines and the Iberian Peninsula. Noticeably, all identified refugia have been sustained at the southern (Mediterranean) distribution periphery, and these extensions are at the present the most often studied and the best documented **(Fig. 8)**.

Less revisions were carried out on inland areas of Europe, like the Eastern European Plain (Sarmatic Plain) where genetic lineages are yet unidentified. Furthermore, genetic patterns at the north-western Scandinavian periphery in Norway or the northeast Black sea region (Turkey) and Caucasus Mountains (Russia) remained unravelled. Although there are unexploited distribution areas the newly discovered native western European habitats of Scots pine distribution can reveal additional information on phylogeographic history and may answer questions linked to colonization processes. Recently identified peripheral isolated Irish and Portuguese populations might be keys to reveal such historical events.

Newly applied sequence variation studies on candidate gene loci assumed that populations having important role in the colonization processes of the northern territories might be cold tolerant genotypes while the southern localities maintained genotypes sensitive to low temperature and were less moisture demanding. Postglacial climatic changes like drought or rising temperature seems to affect the expansion of the species. In late Holocene decrease in population sizes might be considered because of the ongoing climatic change and the recent anthropogenic activities that

have had also a major impact on the population structure of Scots pine and shaped its present day distribution.

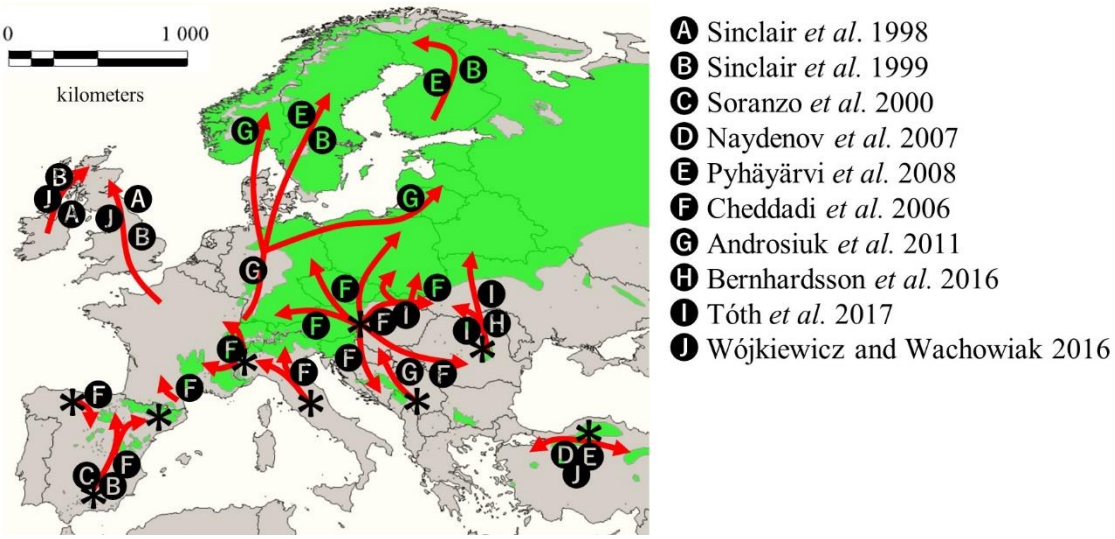


Fig. 8: Geographic map with studies carried out on the natural distribution of Scots pine. Green area marks the current natural distribution of the species according to EUFORGEN database. Red arrows present the detected post-glacial migration routes and glacial refugia (*) based on literature data.

3. MATERIAL AND METHODS

3.1. Plant material and the studied geographic range

Plant material originates from Central-Eastern Europe, from the area of the Carpathian Mountains and the Pannonian Basin. Additionally, a population was included from the northern part of the distribution range, from Estonia (core population) and another from the south, from the Balkan Peninsula (Bulgaria) **(Fig. 9a, b)**.

The Carpathian Mountain range extends approximately from 45ºN-50ºN to 18ºE-27ºE in Central-Eastern Europe and covers over 209,255 km^2 in an arc from Bratislava on the Austrian-Slovakian border to the southern Danube river at the Iron Gate. The range is about 1300-1500 km long and 100-350 km wide (Ronikier 2011, Quinn and Woodward 2014). It is not a continuous mountain massif, but a series of small isolated mountains and intervening basins (Quinn and Woodward 2014). Elevations are generally 750-2500 m a.s.l., the highest peak is the Gerlachovský Štit (2663 m a.s.l.) located in the Western Carpathians, namely in the High Tatra Mountains (Ronikier 2011). Several rivers and their tributaries have their headwaters within the region and about 90% of the drainage goes to the Black Sea. The north-western extension (Western Carpathians) of the range is a combination of Mesozoic rocks of the Beskid Mountains (in Czech Republic, Poland, Slovakia and Ukraine), hilly limestone plateaus at 395-485 m, also sandstone and shale sedimentary basins. The central part (High and Low Tatras) is delimited by a basin and it is built of Hercynian granite and lower Paleozoic sediments that have been strongly folded. Southern extensions are lower in height and are predominantly composed by Mesozoic limestone and dolomites. Karst formation is also prevalent within the region. The eastern extension (Eastern Carpathians) and the southern extension (Southern Carpathians) share several similarities. Both are characterized by mostly geological faulting events and encompass the Transylvanian Plateau basin. The eastern range is sandstone and shale, while the western is a Hercynian massif with sedimentary cover. Along the eastern range Tertiary volcanic cones and craters are scattered (Harghita, Caliman Mountains). In the Southern Carpathians a complex mix of Hercynian basement, limestone, andesitic volcanism thrust northward. The Transylvanian Central-Island Mountains (Apuseni Mountains) is a Hercynian massif with flat summits and deep valleys built by sedimentary limestone and volcanic rocks (Quinn and Woodward 2014).

The Pannonian Basin is encircled by the Carpathians (from north and east), by the Dinaric Alps (from south) and by the Southern and Eastern Alps (from west). The basement of the basin consists of a complex of igneous, metamorphic, and sedimentary rocks. The basin has developed after the contraction of the shallow Pannonian Sea (Quinn and Woodward 2014).

Natural Scots pine populations within the Carpathians usually grow at a wide range of altitudinal gradients, between 500-600 m a.s.l. to 1450-1600 m a.s.l., where Scots pine forms conifer-broadleaved mixed stands or pure coniferous stands (Bojnanský and Fargašová 2007).

In the Pannonian Basin Scots pine naturally grows in the foothill forest zone or at the border of the forest-steppe, approximately on 200-300 m a.s.l., but extends up to 500-600 m a.s.l. in conifer–broadleaved mixed stands (Giertych and Mátyás 1991, UNEP-DEWA 2007). Despite the

fact that Scots pine naturally occupies a wide distribution range, along the Carpathian Mountains and the Pannonian Basin the species' populations are scattered, island-like and mostly isolated from each-other. These stands are sustained mainly in mixed forest groups, with broad-leaved species in the western Pannonian Basin at the foothills of the Alps, at the so called Pre-Alpine region (Pócs 1960; Fekete *et al.* 2014). Carpathian habitats were highly affected by the postglacial climate warming, forcing Scots pine to immigrate into edaphically specialized habitat types. Indeed, Scots pine natural populations are distributed in the Carpathians on a large elevation gradient, located in sites of divergent ecological conditions, including humid, cool peat-bogs and sunny, dry, rocky outcrops. In addition, historical human mediated activities further increased habitat fragmentation and considerably reduced population census sizes. In part as a consequence of this, only isolated and island-like populations have been sustained (Pócs 1960, Giertych and Mátyás 1991, UNEP-DEWA 2007).

Altogether 20 natural and autochthonous populations were sampled from the highly fragmented distribution range of the species **(Fig. 9a, b)**. In total, 421 individuals were analysed with nuclear and chloroplast simple-sequence repeats (SSR) **(Table 2)**: Five populations from the Pannonian Basin in western Hungary (HFE, HKO, HOR, HZA, HVA), four from the Western Carpathians (Tatra Mountains), i.e. the Low Tatras (STU, SLI) and the westernmost ridges of High Tatras (SKV and SME), nine from Romania, four of which are found in the Transylvanian Central-Island Mountains (Apuseni) (RBI, RBE, RML, RMH), one in the western part of the Southern Carpathians, Latoriței (RPA), and four in the Eastern Carpathians (RFE, RCO, RMO, RPO). More distant populations from outside of the Carpathian region were also included, one from the Bulgarian Rila Mountains (BYU) and one from the Central Estonian Plain (ESE) as outgroups. Detailed population descriptions are included in the **Supplemental Materials**.

To test neutral and adaptive genetic variation, i.e. genetic variation under natural selection, 10 populations (96 individuals) were used. Four from the Hungarian Pannonian Basin (HKO, HVE, HZA, HFE), three populations from the Slovakian Western Carpathians (SME, STU, SKV) and another three from the Romanian Eastern Carpathians (RCO, RFE, RPO).

Most of the populations persist in ecologically extreme conditions, including dry, sunny outcrops with low soil availability, acidic peatbogs, and dry sandy substrates with low nutrient content. Needles were sampled from mature trees (8-30 from each population) with at least 30 m distance between individuals. However, in some of the populations samples were taken from shorter inter-individual distances due to relatively small population size (from 0.02 to >5.00 km^2) and limited sample availability. Plant material was stored on silica gel and frozen at -80°C until DNA extraction.

Table 2: List of the 20 island-like populations of *Pinus sylvestris* used in this study from the Central and Eastern European peripheral distribution of the species. RL: Rila Mountains, CE: Central Estonian Plain, PB: Pannonian Basin, WC: Western Carpathians, EC: Eastern Carpathians, SC: Southern Carpathians and CIM: Central-Island Mountains (Apuseni), respectively.

Code	Country and region		Site	Lat. (ºN)	Long. (ºE)	Altitude (m a.s.l.)	Area (~km^2)	Samp. size
SKV	SK	WC	Kvačianska valley	49.18	19.54	799	0.48	20
SME	SK	WC	Medzi Bormi	49.27	19.63	815	0.07	20
STU	SK	WC	Turková	49.02	19.91	1107	0.70	20
SLI	SK	WC	Liptovský Hrádok	49.04	19.74	729	0.02	10
RFE	RO	EC	Fântâna Brazilor	46.50	25.26	953	0.32	26
RPO	RO	EC	Poiana Stampei	47.30	25.12	878	1.43	20
RMO	RO	EC	Mohos	46.13	25.91	1052	0.58	29
RCO	RO	EC	Suhardul Mic	46.80	25.79	981	0.04	19
RPA	RO	SC	Lătoriței	45.38	23.91	753	3.42	23
RBI	RO	CIM	Roşia	46.84	22.37	393	0.13	8
RBE	RO	CIM	Scăriţa-Belioara	46.49	23.36	524	0.84	20
RML	RO	CIM	Mluha	46.33	23.34	925	0.10	10
RMH	RO	CIM	Călățele	46.73	23.02	913	0.58	11
HFE	HU	PB	Fenyőfő	47.35	17.77	252	4.49	24
HZA	HU	PB	Pethőhenye	46.87	16.92	306	0.04	19
HVE	HU	PB	Szalafő	46.87	16.30	231	0.08	20
HOR	HU	PB	Csörötnek	46.93	16.35	296	0.10	23
HKO	HU	PB	Kőszeg	47.34	16.45	468	0.24	20
BYU	BG	RL	Yundola	42.07	23.83	1561	>5.00	29
ESE	EE	CE	Estonia	58.88	26.19	102	>5.00	30

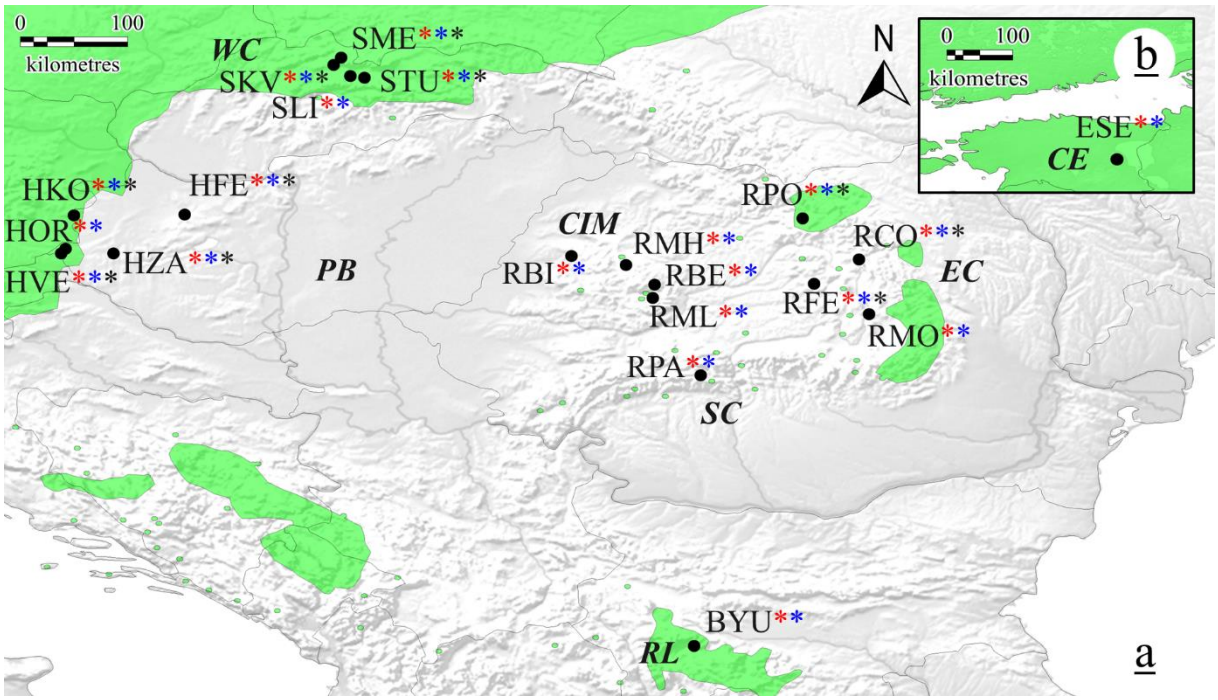


Fig. 9: Geographic location of the studied Scots pine populations. (a): Central-Eastern Europe, (b): Estonia. The acronyms stand for the population code in **Table 2**. Asterisks indicate types of marker (*; cpSSR, *; nSSR, *; SNP) applied to evaluate the population. The natural distribution of Scots pine is marked in green according to the EUFORGEN database, with modifications by the authors.

3.2. Laboratory method of DNA isolation, PCR, fragment size analysis and sequencing

Total DNA was extracted from 20-25 mg of plant material (one-year old needles) by using DNeasy Plant Mini Kit (QIAGEN, Valencia, CA, USA) according to the manufacturers' protocol.

For the non-coding microsatellite markers a preliminary test (PCR and Fragment Length Analysis (FLA)) was carried out to check polymorphism of the selected loci. At the first attempt, ten nuclear microsatellite markers (nSSR) were used (SPAG 7.14, SPAC 11.4, SPAC 11.6, SPAC 12.5 from Soranzo *et al.* (1998) and psyl16, psyl17, psyl19, psyl36, psyl42, psyl57 from Sebastiani *et al.* (2012)). Based on the results, two inconsistent loci (SPAC 11.6, SPAC 12.5) were excluded from further analysis. The remaining eight markers (SPAG 7.14, SPAC 11.4, psyl16, psyl17, psyl19, psyl36, psyl42, psyl57) showed reliable amplification and size pattern, therefore were scored and used for further analysis. These markers proved to be highly polymorphic in our populations. Genotyping of individuals at chloroplast SSR loci was performed with four markers, Pt-30204, Pt-15169, Pt-45002 and Pt-26081 originally developed for *Pinus leucodermis* (Vendramin *et al.* 1996). The four chloroplast microsatellite loci, used in this study, amplified successfully and proved to be highly polymorphic in the initial test.

A similar preliminary test (PCR and sequence analysis) was carried out to evaluate the selected candidate gene loci. Here, we examined the sequence variation by single-nucleotide polymorphism (SNP) at 21 candidate genes. Candidate loci were selected based on both polymorphism formerly detected in Scots pine or divergence from neutral expectations (sign of selection) and putative function of the gene potentially related to response to dehydrative stress.

First, *abaR*, *ccoaomt*, *chcs*, *dhn1*, *dhn2*, *dhn3*, *dhn7*, *dhy2PP*, *erd3*, *lp3-1* and *lp3-3* was evaluated based on the study of Wachowiak *et al.* (2011). Later, *lp5-like*, *sod-chl*, *ferritin*, *rd21A-like*, *sams-2*, *pal-1*, *cpk3*, *ppap12*, *pp2c*, *ug-2_498* (González-Martínez *et al.* 2006) was also included to the initial test. Finally, 11 candidate genes were selected (*abaR, ccoaomt, chcs, cpk3, dhn3, dhn7, dhy2PP, erd3, pal-1, ppap12, rd21A-like*), while the remaining loci were excluded due to weak and/or unsuccessful amplification, unreliable multiple banding pattern, sequencing error or unreliable sequence chromatograms. The studied loci include water dehydration stress response genes: known members of the dehydrin gene family (*dhn3, dhn7, dhy2PP*), abscisic acid responsive protein (*abaR*), caffeoyl CoA O-methyltransferase (*ccoaomt*), chalcone synthase (*chcs*), Calcium-dependent protein kinase (*cpk3*), early response to dehydration 3 protein (*erd3*), phenylalanine ammonia-lyase 1 (*pal1*), possible wall-associated protein kinase (*ppap12*), cysteine protease (Pseudotzain), similar to *rd21A* in Arabidopsis (*rd21A*) (González-Martínez *et al.* 2006, Wachowiak *et al.* 2011).

Details of the markers used for amplification of non-coding microsatellite (nSSR and cpSSR) regions and candidate gene loci are presented in **Table 3**.

Table 3: Applied microsatellite and candidate gene markers of the studied DNA regions and their attributes for PCR amplification.

cpSSR markers			*Primer sequence (5'-3')*	*Reference*
Loci	*Primer name*	T_a(°C)		
Pt-30204	F: Pt-30204-F R: Pt-30204-R	55	F: TCATAGCGGAAGATCCTCTTT R: CGGATTGATCCTAACCATACC	A
Pt-26081	F: Pt-26081-F R: Pt-26081-R	55	F: CCCGTATCCAGATATACTTCCA R: TGGTTTGATTCATTCGTTCAT	A
Pt-15169	F: Pt-15169-F R: Pt-15169-R	55	F: CTTGGATGGAATAGCAGCC R: GGAAGGGCATTAAGGTCATTA	A
Pt-45002	F: Pt-45002-F R: Pt-45002-R	55	F: AAGTTGGATTTTACCCAGGTG R: GAACAAGAGGATTTTTTCTCATACA	A
nSSR markers			*Primer sequence (5'-3')*	*Reference*
SPAG 7.14	F: SPAG 7.14-F R: SPAG 7.14-R	55	F: TTCGTAGGACTAAAAATGTGTG R: CAAAGTGGATTTTGACCG	B
SPAC 11.4	F: SPAC 11.4-F R: SPAC 11.4-R	65, 60*	F: TCACAAAACACGTGATTCACA R: GAAAATAGCCCTGTGTGAGACA	B
psyl16	F: psyl16-F R: psyl16-R	55	F: GCTCTGCCCATGCTATCACT R: TGATGCTACCCAATGAGGTG	C
psyl17	F: psyl17-F R: psyl17-R	55	F: TGGTCTGCAAATCAATCGAA R: GGGTAGGAATGCAAGTTAGGC	C
psyl19	F: psyl19-F R: psyl19-R	55	F: GGCTGTAATTGGCACAGGTT R: CGAGGTGGTACACAGCAACA	C
psyl36	F: psyl36-F R: psyl36-R	55	F: TATCATCGAGAGCCCCAAAA R: GAAAGGCGAAAGCAAAAGTG	C
psyl42	F: psyl42-F R: psyl42-R	55	F: CAACTTCAGCCTTGCAACAA R: CGACTTCATTTGGAACACCA	C
psyl57	F: psyl57-F R: psyl57-R	56	F: CCCCACATCTCTACAGTCCAA R: TGCTCTTGGATTTGTTGCTG	C
Candidate gene markers			*Primer sequence (5'-3')*	*Reference*
abaR	F: abaR-F R: abaR-R	60, 50*	F: AAGATTGCAATGATGAGTAAAGAAG R: CTTGGCCCATTTGTTGAAGCAGTGA	D
ccoaomt	F: ccoaomt-F R: ccoaomt-R	65, 55*	F: GCAGCAGAAGTGAAGGCTCAGA R: TCTTTCCATCATCGGGCAATG	D
chcs	F: chcs-F R: chcs-R	65, 55*	F: ACTCCCCCTAATGCGGTTGA R: CTTGGCTGCGGCTTCTTTC	D
cpk3	F: cpk3-F R: cpk3-R	65, 55*	F: TGGCCATCTATTTCTGGTGG R: CAATTGTCCCATCCCCATC	E
dhn3	F: dhn3-F R: dhn3-R	60, 50*	F: TACTCGTTATTAAGATGGCGCAACC R: CGATTGTACCCGAAGTCCCATTTAT	D
dhn7	F: dhn7-F R: dhn7-R	60, 50*	F: ATTAAGATGGCGGAAGAGCAACAGG R: TTGTACCCGAAGTCCCATTT	D
dhy2PP	F: dhy2PP-F R: dhy2PP-R	65, 55*	F: CTGCAGAGACTGTGCCTGAGC R: CCAGGGAGCTTTTCCTTGATCT	D
erd3	F: erd3-F R: erd3-R	60, 50*	F: GAACGGGTCCGTACATTTTCTG R: TGCCAGATTGATTGGCATAGAA	D
pal-1	F: pal-1-F R: pal-1-R	65, 55*	F: TTCCTGTGTTTGAAGCCGAGC R: TGATCTGCCACCCTTACATATTTCTG	E
ppap12	F: ppap12-F R: ppap12-R	65, 55*	F: ATTAGCAGGGCATCTGTCTG R: CACGCCTCTCATTTTCATC	E
rd21A-like	F: rd21A-like-F R: rd21A-like-R	65, 55*	F: TTGCTTCTCTTGGTCAATGC R: TCCTGTAAGCGTCACATGAT	E

T_a(°C): optimal annealing temperature of primers

*: Touch-down PCR (TD-PCR) protocol, modified by author. First value (T_a1): high initial annealing temperature, second value (T_a2): decreased annealing temperature.

Ref.: A: Vendramin *et al.* 1996, B: Soranzo *et al.* 1998, C: Sebastiani *et al.* 2012, D: Wachowiak *et al.* 2011, E: González-Martínez *et al.* 2006

Non-coding (nSSR and cpSSR) microsatellite primers were fluorescently labelled with 6-FAM (SPAG 7.14, SPAC 11.4, psyl17, psyl19, psyl42, and psyl57) and NED (psyl16, psyl36). Chloroplast primers were labelled with HEX (Pt-30204, Pt-26081) and 6-FAM (Pt-15169, Pt-45002).

Amplification of microsatellite loci was performed in a 25-µl reaction volume containing 20-80 ng DNA, 10X PCR reaction buffer, 2.5 mM $MgCl_2$, 0.2 mM dNTP mix, 2.5 µmol of each primer, 1 unit of Taq DNA polymerase (Waltham, MA, USA) and Milli-Q ultrapure water. PCR was performed in a thermocycler (Swift™ MaxPro Thermal Cycler, Esco Healthcare Pte, Singapore). Amplification protocols greatly varied according to the primer pairs used. A touch-down PCR protocol was carried out for SPAC 11.4 as follows: initial denaturation at 94°C for 3 min, followed by 5 cycles of 30 s at 94°C, 30 s at 65°C and 30 s at 72°C (decreasing the annealing temperature 1°C per cycle), followed by 25 cycles of 30 s at 94°C, 30 s at 60°C and 30 s at 72°C, and a final extension step of 5 min at 72°C. The protocol for SPAG 7.14 consisted of an initial step of 94 °C for 3 min, followed by 30 cycles of 94°C for 30 s, 55°C for 30 s, 72°C for 1 min, followed by 72°C for 10 min. The amplification profile for psyl16, psyl17, psyl19, psyl36 and psyl42 consisted of a first step at 94 °C for 4 min, and 35 cycles of 94°C for 30 s, 55°C for 30 s, 72°C for 40s, followed by a final 72°C for 8 min. A similar protocol was used for psyl57, but with an annealing temperature of 56°C. The PCR profile of the chloroplast SSRs consisted of an initial denaturation at 94°C for 5 min, followed by 30 cycles of 15 s at 94°C, 30 s at 55°C and 40 s at 72°C, and a final extension of 5 min at 72°C.

PCR of candidate gene loci was performed in a 15-µl reaction volume containing 20-80 ng DNA, 1.5 mM $MgCl_2$ (included in Promega 5x Green Buffer, Medison, WI, USA), 10 µM of each dNTP, 0.2 µM of each primer, 1 unit of Taq DNA polymerase (Promega Go-Taq, Medison, WI, USA) and Milli-Q ultrapure water. PCR reaction was performed in a GeneAmp 9700 thermal cycler (Foster City, CA, USA). A touch-down PCR protocol was carried out for all candidate loci, as follows: initial denaturation at 93°C for 3 min, followed by 10 cycles of 30 s at 93°C, 45 s at T_a1°C **(Table 3)** and 1:20 min at 72°C (decreasing the annealing temperature 1°C per cycle), followed by 30 cycles of 30 s at 93°C, 45 s at T_a2°C **(Table 3)** and 1:20 min at 72°C, and a final extension step of 10 min at 72°C. Electrophoresis to detect PCR products were carried out on a 1 % (w/v) ethidium bromide-stained agarose gel in 1xTBE buffer.

After amplification of non-coding loci, strong yields of PCR products were diluted 25 to 30 times for fragment analysis, which was performed on an automated sequencer ABI PRISM 3100 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). Band scoring was done using PeakScanner 1.0 (Applied Biosystems 2006), and all size scores were visually (manually) checked.

Following the successful amplification of candidate gene loci, samples were quality-checked on 2% agarose gel stained with GelRed (Biotium, Fremont, CA, USA) and then purified using Multiscreen filter plates (Merckmillipore, Billerica, MA, USA) and manifold in a vacuum in order to carry out the sequencing reaction. 4 µl (about 20 ng) of purified PCR product was mixed with (0.32 µl) Forward or Reverse Primer, 1.5X Sequencing Buffer and 0.5 µl of BigDye

Terminator v3.1 Ready Reaction Mix (ThermoFisher SCIENTIFIC, Waltham, MA, USA) giving a final volume of 6 µl. The sequencing thermal profile was characterized by a first step of denaturation at 94°C for 1 min followed by 25 cycle at 96°C for 10 sec, 50°C for 5 sec, 60°C for 3 min. Sequencing reactions were then purified by membrane filtration using MultiScreenHTS® HV, 0.45 µm Plates (Merckmillipore) and Sephadex G50 (GE Healthcare, Chicago, IL, USA) and analyzed on an Applied Biosystems 3500 capillary sequencer (Applied Biosystems, Foster City, CA, USA) with traditional Sanger DNA sequencing method.

BioEdit 7.2.5 (Hall 1999) sequence alignment editor software was used for editing sequence chromatograms and for visual inspection of all segregating sites detected. Alignments were done with CLUSTALW (Thompson *et al.* 1994) as implemented in BioEdit.

3.3. Statistical data analysis

3.3.1. Analysis of non-coding microsatellite dataset

Micro-Checker (van Oosterhout *et al.* 2004) was used to test all nSSR loci for null alleles and possible scoring errors derived from large allele dropout and the presence of microsatellite stutter bands. Then, standard population genetic diversity indices (number of different alleles (*N*a), the number of effective alleles (N_{eff}), the number of private alleles (N_p), expected heterozygosity (H_e), observed heterozygosity (H_o), and the inbreeding coefficient (F_{IS}) were calculated for each population using GenAlEx v.6.5 software (Peakall and Smouse 2006). For cpSSR loci, haploid genetic diversity (h) and N_a, N_{eff}, N_p were calculated also with GenAlEx v.6.5. For estimating haplotype diversity indices Haplotype Analysis software (Eliades and Eliades 2009) was used. Relying on the average squared sum of all allele size differences with the assumption of the stepwise mutation model (Morgante *et al.* 1998), we calculated mean genetic distance between individuals (D_{sh}^2; Goldstein *et al.* 1995). In addition, the number of haplotypes per population (*A*), the number of private haplotypes (P_h), the effective number of haplotypes (N_e), and haplotypic richness (H_R) were estimated using rarefaction (El Mousadik and Petit 1996).

Analysis of molecular variance (AMOVA) implemented in Arlequin v.3.5 software (Excoffier and Lischer 2010) was used for nSSR and cpSSR data to determine the partition of the genetic variation within and among populations. Significance tests were evaluated using a permutation approach with 999 replications. Similarly, Arlequin v.3.5 was used to detect deviation from expectations of Hardy-Weinberg equilibrium (HWE) at nuclear loci. The test of HWE employed a Markov-chain approximation (Guo and Thompson 1992). The number of steps after the burn-in period was set to 1,000,000 with a 100,000 dememorization.

To test for correlation between geographical (kilometres) and genetic (Pairwise Nei) distances at both nSSR and cpSSR markers, dissimilarity matrices were generated and tested for isolation by distance (IBD; Wright 1943) with Mantel test (Mantel 1967). The analysis was carried out in the online platform of Isolation-by-Distance Web Service (IBDWS) 3.16 (Jensen *et al.* 2005) and GenAlEx v.6.5, with 9999 and 1000 permutations, respectively.

Different approaches were employed to investigate the spatial genetic structure of the populations. A Bayesian clustering approach implemented in STRUCTURE 2.3.4 (Pritchard *et al.* 2000) was used to infer groups or subpopulations based on the nSSR dataset. We performed the analysis with an admixture model with correlated allele frequencies and a LOCPRIOR setup (according to preliminary test runs). The method uses sampling locations as prior information in the case of a relatively weak signal of structures (Hubisz *et al.* 2009). *K* value was set to 1-10 with a burn-in period of 10^5 steps followed by 10^6 repetitions of Markov Chain Monte Carlo (MCMC). Fifteen repetitions were set for each run. The web-based STRUCTURE HARVESTER (Earl 2012) was used to apply the Evanno method (Evanno *et al.* 2005) to detect the value of *K* (the number of genetic groups) that best fit the data. The 15 simulations were averaged using CLUMPP v1.1.2 (Jakobsson and Rosenberg 2007) and represented in the form of bar graphs using POPHELPER (Francis 2016). We used BAPS 6.0 (Corander and Mattinen 2006, Corander *et al.* 2008a, b) to conduct hierarchical clustering analyses of the chloroplast microsatellite dataset. BAPS was run with the maximal number of groups (*K*) set to 2–20 (equal or larger than the population number), and each run was replicated five times.

Genetic discontinuities corresponding to the change in genetic variation among populations were identified with Barrier 2.2 (Manni *et al.* 2004) on both nuclear and chloroplast SSR datasets. Spatial geographic coordinates were connected with Delaunay triangulation, and the corresponding Voronoi tessellations were projected. To identify the genetic barriers, Monmonier's maximum difference algorithm was applied, which traces a barrier along the Voronoi tessellation, starting from the edge for which the distance value is maximal. The boundary proceeds across adjacent edges until the forming boundary reaches the limit of the Delaunay triangulation or closes itself by forming a loop around a population. Within the analysis, barriers were set from 1 to 9. To test the significance of the detected barriers, 1000 resampled bootstrapped (population-pairwise) D_A genetic distance matrices (Nei's chord-distance) were calculated in MICROSATELLITE ANALYSER (MSA) software (Dieringer and Schlötterer 2003).

To detect any recent severe reduction in effective population size or possible expansion events in Scots pine populations, BOTTLENECK 1.2.02 was used on nSSR dataset (Cornuet and Luikart 1996). Bottlenecks cause low-frequency alleles to become transitorily less abundant (<0.1), while more intermediate-frequency alleles increase (Luikart *et al.* 1998). BOTTLENECK correlates expected heterozygosity (H_e) with observed heterozygosity (H_o) at mutation-drift equilibrium. The two-phased model (TPM) of mutation was applied as the most appropriate for microsatellite data (Di Rienzo *et al.* 1994, Piry *et al.* 1999). We used 5% of multistep changes and a variance among multiple (12) steps (Piry *et al.* 1999). For each population 2000 simulations were performed. Significance was assessed using the implemented Wilcoxon sign-rank test, which determines whether the average of standardized differences between H_o and H_e is significantly different from zero or not (Cornuet and Luikart 1996). Significant heterozygote excess relative to the number of alleles indicates a recent population bottleneck. Additionally, the "Mode-shift" qualitative descriptor of allele frequency distribution was applied to discriminate bottlenecked populations (Luikart and Cornuet 1998).

Analysis of the historical demographic changes was studied with Approximate Bayesian Computation implemented in DIYABC v2.0 software (Cornuet *et al.* 2014). We used the nuclear microsatellite data to estimate possible population divergence time, changes in the effective population sizes, and admixture events. ABC analysis allows to compare complex historical scenarios to infer past history of populations. To perform ABC analysis we defined three populations (from pre-selected 16 subpopulation) according to STRUCTURE results **(Fig. 16)**: Pop1 (Hungarian; HFE, HVE, HZA, HOR, HKO, Slovakian; SKV, SME, STU, SLI, and Romanian; RCO, RBI, RPA), Pop2 (Romanian; RFE, RPO, RMO), and the admixed Pop3 (Bulgarian; BYU). Estonian population was excluded from ABC analysis due to its outlier position and possible different origin. Population relationships design in the scenarios were mainly based on the result of STRUCTURE analysis and on calculated standard population genetic indices (e.g. H_e, G_{ST}). Several pilot runs were carried out to produce final competing scenarios. In all scenarios, Pop1 was specified as the population to be traced back to an ancestral population (as required for DIY ABC analysis), because this population had the highest heterozygosity (H_e). Six different scenarios **(Fig. 15a, b)** were examined following a two-step DIYABC procedure.

In the first step **(Fig. 15a)**, the three populations (Pop1, Pop2 and Pop3) were analyzed to infer historical divergences and times of admixture event, and to reveal history of recent population structure identified by STRUCTURE. The second analysis **(Fig. 15b)** focused on determining changes within population size by estimating effective sizes, and to refine potential ancient expansion-contraction in species history. In all tested scenarios, *t#* refers to timescale expressed as generation time, and *N#* to effective population size of the corresponding population.

The scenarios are as follows **(Fig. 15a)**:

- **Scenario 1** (hereafter: Sc1): (Hierarchical split model 1): Pop1 and Pop2 diverged at t2 from an ancestral population Na, and Pop3 diverged from Pop1 at t1.
- **Scenario 2** (Sc2): (Hierarchical split model 2): Pop1 and Pop2 diverged at t2 from ancestral population Na, and Pop3 diverged from Pop2 at t1.
- **Scenario 3** (Sc3): (Simple split model): Pop1 diverged from Na at t2, and Pop2 and Pop3 diverged from Pop1 at t1.
- **Scenario 4** (Sc4): (Isolation with admixture model 1): Pop1 and Pop2 diverged at t2 from ancestral population Na, and Pop3 was generated by admixture of Pop1 and Pop2 at t1.

The most possible scenario (Sc4) was tested for demographic changes assuming an expansion at both Pop1 and Pop2.

The two final scenarios are as follows **(Fig. 15b)**:

- **Scenario 5** (Sc5): (Isolation with admixture model 1): same as Scenario 4.
- **Scenario 6** (Sc6): (Isolation with admixture model 2): we assumed Pop1 and Pop2 expansion from the ancestral population. N1c and N2c represent a variation in the effective population sizes of Pop1 and Pop2, respectively, and are set bigger than Na and smaller than Pop1 and Pop2, respectively.

In all simulations, Generalized Stepwise Mutation model (GSM; Estoup *et al.* 2002) and a standard mutation rate of min. $1x10^{-3}$ and max. $1x10^{-4}$ were used **(Supplemental Materials Table S5 and S7)**. Mean number of alleles (N_A), genetic diversity (H_e) and size variance were calculated as summary statistics for single populations, and Mean number of alleles, genetic diversity, size variance and differentiation (F_{ST}) for population pairs. One million simulations were performed for each scenario, after the most likely scenario was evaluated by comparing posterior probabilities using logistic regression **(Supplemental Materials Table S8)**. Model checking option was used to test the goodness of fit of the scenarios with Principal Component Analysis (PCA) **(Supplemental Materials Table S6, Fig. S1)** to measure deviation between simulated and real data.

3.3.2. Analysis of coding candidate gene dataset

Prior to statistical estimations haplotype reconstruction of genotype data was carried out with PHASE algorithm (Stephens *et al.* 2001, Stephens and Donnelly 2003) implemented in DnaSP 5.10.1 (Librado and Rozas 2009), which uses a coalescent-based Bayesian method to infer haplotypes from sequence data. Coding and non-coding regions (introns, UTRs) were annotated based on homologous genes in the web-based NCBI (National Center for Biotechnology Information) database. Basic statistics including the number of SNPs, synonymous and non-synonymous mutations were calculated. The number of segregating sites (S) and π, the average number of pairwise differences per site (Nei 1987) to measure polymorphism was calculated with DnaSP. Nucleotide diversity indicator theta pi (θ_π) was determined with Arlequin v.3.5 software (Excoffier and Lischer 2010) for all populations and for all sites. This has been estimated from the infinite-site equilibrium relationship between the mean number of pairwise differences (π) and theta (θ).

Number of haplotypes (Nh) and haplotype diversity (Hd) were computed with DnaSP. Haplotype networks were constructed in the program PopART (Leigh and Bryant 2015) using TCS (Templeton-Crandall-Singh) network analysis (Clement *et al.* 2002) to evaluate genealogical relationships among sequences and to infer haplotypes proportion and relationship between the studied populations.

Analysis of molecular variance (AMOVA) implemented in Arlequin software was used for candidate gene sequence data to determine the partition of the genetic variation within and among populations. Significance tests were evaluated using a permutation approach with 999 replications. The proportion of variation was calculated for all genes. Pairwise population relationships of F_{ST} (matrix) was also calculated to perceive genetic affinities between populations.

To test that all mutations are selectively neutral, hence randomly evolving (Kimura 1983), and to identify deviations from this neutral expectation (loci under selection) Tajima's D (Tajima 1989) test was applied. Additionally, Fu and Li's D and F (1993) statistical tests were also evaluated.

We also tested if single SNPs are deviated from neutral expectations by detecting loci under selection with 3 distinct approaches. SNPs under selection either show uncommonly low or high differentiation compared to the average found in the gene or gene fragment. To identify potential non-neutral sites in the candidate genes calculations were performed using the FDIST2 approach developed by Beaumont and Nichols (1996) implemented in Lositan (Antao *et al.* 2008). SNPs with an unusually high F_{ST} are putatively under directional (positive) selection, while SNPs with low F_{ST} value are considered to be potentially under stabilizing (balancing) selection. 100,000 iterations and a 99.5% confidence interval was applied. Lositan implements a multi-test correction based on false discovery rates (FDR) to avoid high overestimation of the percentage of outliers. We set the highest possible criteria of the false discovery rate in the test (FDR≤0.01, with a threshold of 99%).

A similar method (FDIST approach) of Excoffier *et al.* (2009), implemented in Arlequin was also applied, where coalescent simulations are used to see if observed locus-specific F_{ST} values can be considered as outliers. We simulated a hierarchical island model with 50,000 iterations to generate the joint distribution of F_{ST} versus heterozygosity. SNP detected out of the 99% confidence intervals of the distribution are identified as outliers and being putatively under selection. Above the computed 1% quantile SNPs were determined to be under positive selection, while under the 1% quantile were considered to be candidates for balancing selection.

Additionally, a Bayesian method implemented in BayeScan 2.1 (Foll and Gaggiotti 2008) was used. The method estimates a posterior probability for each locus directly and use a reversible-jump MCMC approach to infer selection. BayeScan estimate population-specific F_{ST} coefficients, hence accommodates differences in demographic history and the extent of genetic drift between populations (Foll and Gaggiotti 2008). The test was carried out with 20 pilot runs (with a length of 5000 and with a burn-in of 50,000 repetitions), additional iterations were set to 50,000 with a thinning interval of ten. Prior odds were decreased to two, while SNPs positioned in candidate genes having higher possibility for being under selection than random SNPs, according to Csilléry *et al.* (2014). The Bayes factors (BF) i.e. the ratio of posterior probabilities were calculated to determine whether a neutral or selection model fits for each SNP. Then, SNPs were ranked according to their estimated posterior probability and loci over a posterior probability of 0.99 were retained as outliers. BayeScan utilizes Jeffreys' interpretation to determine the scale of selection, whereas the Log_{10} values of the posterior odds (PO) < 0.5 (BF <3) are considered as "substantial" evidence for selection (Jeffreys 1961, Foll 2012). Direction of selection is estimated based on the locus-specific component (alpha), if alpha values are positive the SNPs are under diversifying (i.e. directional) selection, while negative values indicate balancing or background selection (Foll and Gaggiotti 2008, Foll 2012).

The decay of linkage disequilibrium (LD) was measured for all loci as the correlation coefficient (r^2) using all informative sites (Hill and Robertson 1968) as implemented in DnaSP. The r^2 values for the linked loci were plotted against the genetic distance to observe the magnitude of the LD decay. Correlation of the linked loci and their significance were tested using Fisher's

exact test (Sokal and Rohlf 1981). Bonferroni correction was used to correct for multiple testing (Weir and Hill 1996).

Similarly to the microsatellite dataset, the Bayesian clustering approach implemented in STRUCTURE was used to infer groups or subpopulations in the candidate gene dataset. We performed the analysis prior with all polymorphic loci detected, later reduced to those sites/loci (38 SNPs) which were detected by the outlier loci analysis. An admixture model with correlated allele frequencies and a LOCPRIOR setup was also used. K value was set to 1-10 with a burn-in period of 10^5 steps followed by 10^6 repetitions of Markov chain Monte Carlo (MCMC). Fifteen repetitions were set for each run. STRUCTURE HARVESTER was used to apply the Evanno method to detect the value of K (the number of genetic groups) that best fit the data. The 15 simulations were averaged using CLUMPP and represented in the form of bar graphs using POPHELPER.

4. RESULTS

4.1. Analysis of non-coding microsatellite dataset

4.1.1. Chloroplast microsatellites

All four chloroplast microsatellite markers have been amplified successfully, and polymorphism was found at all loci. Haplotype analysis of the SSRs revealed 4 to 13 size variants per locus. A total of 36 size variants at the four loci were identified. These size variants combined into 141 haplotypes, 87 of which were private, having frequencies of <1.0%. The number of haplotypes detected in each population was 3 to 23. Haploid diversity (h) was balanced along the populations (without any outlier value) and ranged from 0.349 in RFE to 0.703 in RBI. Similarly, the number of alleles (N_a: 2.5-5.0) and the number of effective alleles (N_{eff}: 1.706-3.558) were balanced without any remarkable differences. Mean number of private alleles was the highest in SME (1). The overall mean of genetic distance between individuals (D_{sh}^2) was 5.792, but in SME (17.070), RBI (20.893), and RCO (11.986) this value was higher than the average **(Table 4)**.

Table 4: Carpathian populations of Scots pine (*Pinus sylvestris*) and details regarding standard indicators of genetic diversity for chloroplast simple-sequence repeat (SSR) datasets. Notations chloroplast (SSR) as indices follows: N_a = number of alleles, N_{eff} = number of effective alleles, N_p = mean number of private alleles, h = haploid genetic diversity, D_{sh}^2 = mean genetic distance between individuals, A = number of haplotypes detected in each population, P_h = number of private haplotypes, N_e = effective number of haplotypes, H_R = haplotypic richness.

	cpSSR								
Population	N_a	N_{eff}	N_p	h	D_{sh}^2	A	P_h	N_e	H_R
BYU	4.500	2.755	0.000	0.575	5.194	23	8	21.125	2.945
ESE	4.750	2.455	0.000	0.496	3.151	19	4	13.000	2.775
HFE	4.500	2.839	0.250	0.584	4.795	19	7	18.182	2.968
HVE	4.250	2.887	0.000	0.591	4.663	16	4	14.286	2.874
HZA	4.250	2.957	0.000	0.605	4.848	15	3	10.939	2.771
HOR	4.250	2.981	0.250	0.624	5.650	18	3	16.333	2.914
HKO	4.000	2.656	0.000	0.525	3.838	13	2	12.250	2.934
SKV	3.750	2.767	0.000	0.609	3.773	16	5	14.727	2.922
SME	5.000	2.648	1.000	0.552	17.070	10	5	5.898	2.374
STU	4.000	2.832	0.000	0.596	4.614	18	6	17.190	2.965
SLI	3.250	2.645	0.000	0.560	3.756	9	4	8.333	2.867
RFE	3.250	1.706	0.000	0.349	1.538	7	1	3.109	1.642
RPO	3.000	2.081	0.000	0.478	3.351	11	4	7.692	2.532
RMO	3.750	2.743	0.250	0.495	4.971	11	5	7.686	2.454
RPA	4.000	2.742	0.000	0.570	4.830	19	7	16.030	2.884
RBI	4.750	3.558	0.750	0.703	20.893	8	4	8.000	3.000
RCO	3.500	2.500	0.250	0.553	11.986	12	5	9.783	2.780
RBE	3.500	2.154	0.250	0.436	2.836	11	6	7.681	2.541
RML	2.500	1.773	0.000	0.412	0.792	3	1	2.667	2.000
RMH	3.250	2.649	0.000	0.607	3.289	9	3	8.333	2.867
Mean	3.900	2.616	0.150	0.546	5.792	13.350	4.350	11.162	2.700

The geographic distribution of the twenty highest-frequency haplotypes is reported in **Fig. 10**. There were evident divergences in haplotype frequencies between the Eastern Carpathian populations on the one hand and the Transylvanian Central-Island Mountains (Apuseni) and Southern Carpathian populations on the other. Moreover, Western Carpathian populations showed

differences in haplotype proportions and occurrences in haplotypes when the four populations were separated into two groups; High Tatra (SME, SKV) and Lower Tatra (SLI, STU). It is also noticeable that three of the Hungarian populations (HZA, HVE, HFE) harboured only 7-8 of the 20 most common haplotypes. The average occurrence along the range of the 20 common haplotypes was 4-5, with some exceptions.

Nei's genetic distance between populations varied from 0.063 to 0.664, with a mean of 0.181. AMOVA analysis including all populations resulted in Φ_{PT}=0.074 ($p< 0.001$), meaning that 7% of genetic variance resides among and 93% within populations. Within the Carpathians (excluding ESE and BYU), AMOVA presented Φ_{PT}=0.081 ($p< 0.001$) value. The average gene flow (Nm; calculation based on F_{ST}) was estimated at 6.272 migrants per generation. The Mantel test revealed significant negative correlation between geographic and genetic distances among populations (r_{xy}: -0.176, $p< 0.045$) **(Fig. 13a)**. The Bayesian approach of population structuring estimated with BAPS identified three main clusters (K=3; **Fig. 11a, b**), which formed geographically distinct groups. Two major groups were detected: Populations from Western Hungary clustered together with those of the Eastern Carpathians and the Bulgarian population (1), while populations of the Western Carpathians (Tatras) and the Transylvanian Central-Island Mountains (Apuseni) grouped together with the Southern Carpathian and Estonian populations (2). SME population from the Tatras was separated as a distinct group.

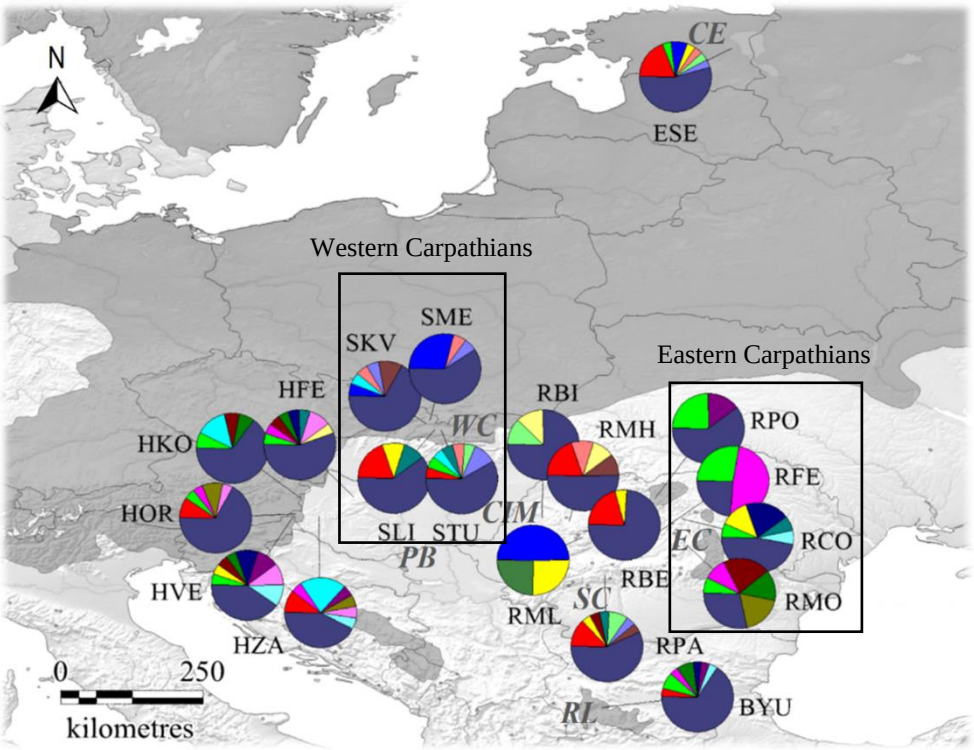


Fig. 10: Geographic location and distribution of the cpSSR haplotypes with the frequency of the 20 most common haplotypes among the populations sampled. The slate grey portion of the pie charts represents the remaining less frequent haplotypes. The acronyms stand for the population code in **Table 2**. The natural distribution of Scots pine is marked in grey according to the EUFORGEN database, with modifications by the author. RL: Rila Mountains, CE: Central Estonian Plain, PB: Pannonian Basin, WC: Western Carpathians, EC: Eastern Carpathians, SC: Southern Carpathians and CIM: Central-Island Mountains (Apuseni), respectively.

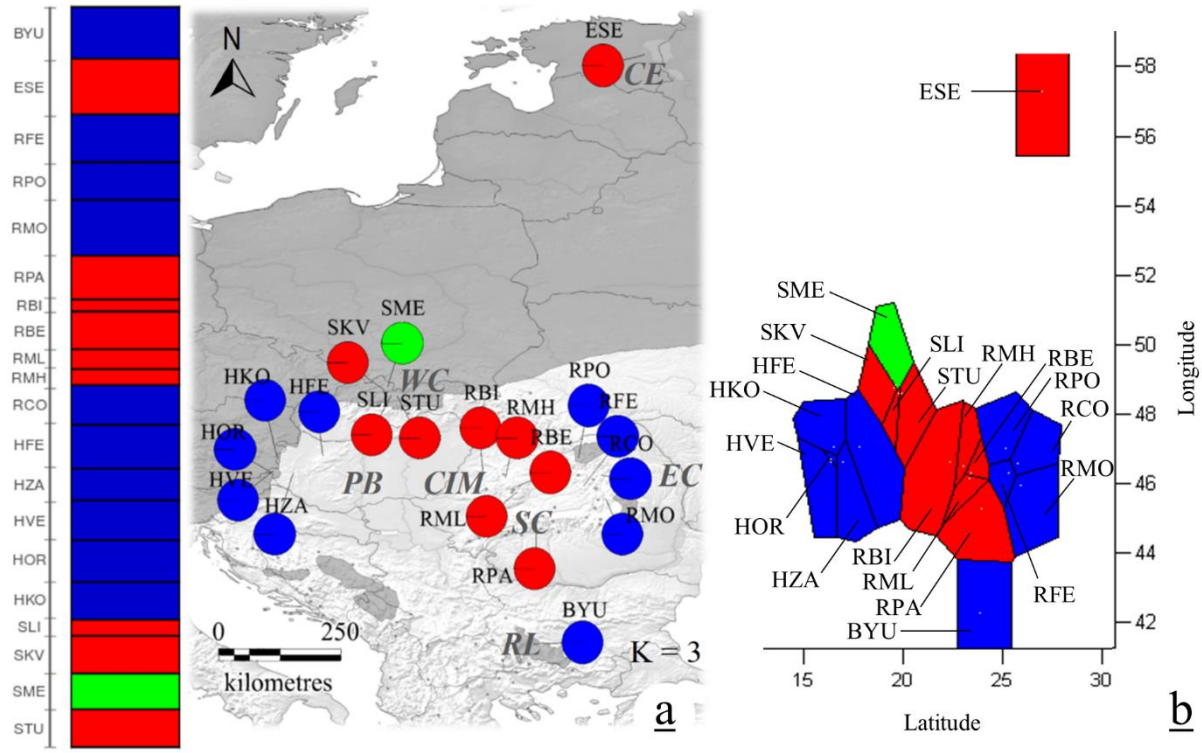


Fig. 11: Estimated spatial population structure K=3 (a and b) according to BAPS analysis (Corander and Mattinen 2006; Corander *et al.* 2008a, b). The acronyms stand for the population code in **Table 2**. The natural distribution of Scots pine is marked in grey according to the EUFORGEN database, with modifications by the author. RL: Rila Mountains, CE: Central Estonian Plain, PB: Pannonian Basin, WC: Western Carpathians, EC: Eastern Carpathians, SC: Southern Carpathians and CIM: Central-Island Mountains (Apuseni), respectively.

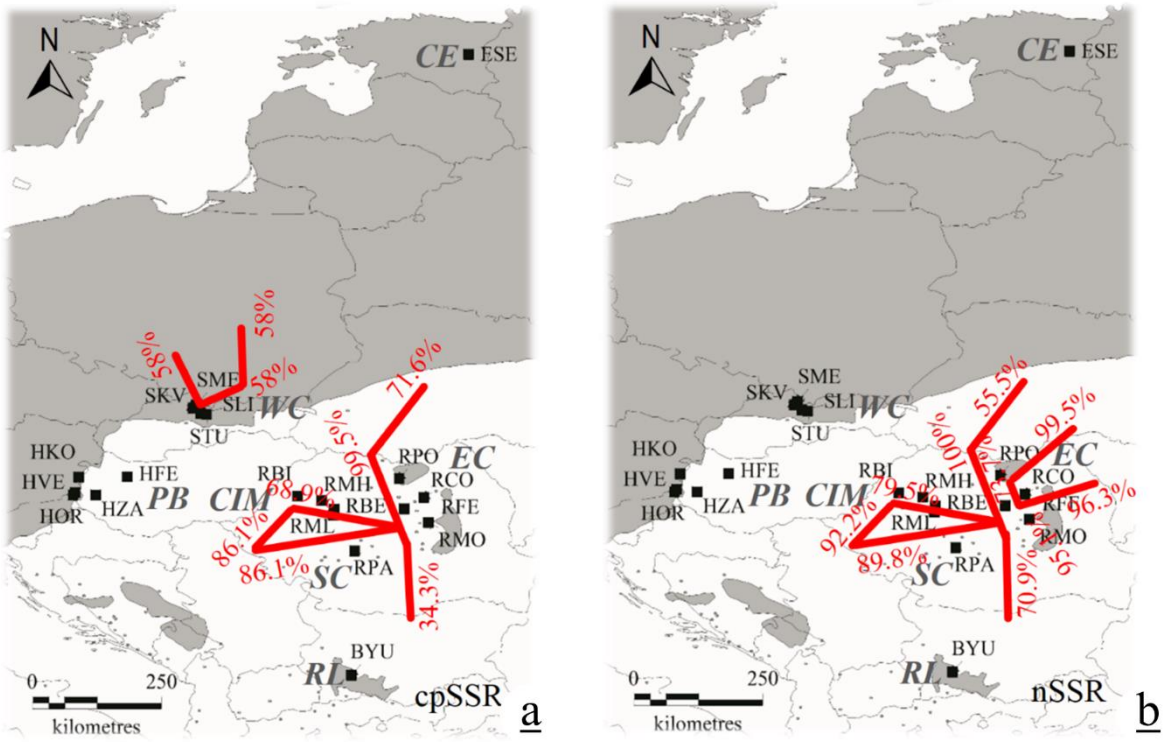


Fig. 12: Detected genetic discontinuities of the chloroplast (a) and nuclear (b) simple-sequence repeat (SSR) datasets with Barrier analysis (Manni *et al.* 2004). Acronyms stands for the population code in **Table 2**. Natural distribution of Scots pine is marked in gray according to EUFORGEN database and modified by the author. RL: Rila Mountains, CE: Central Estonian Plain, PB: Pannonian Basin, WC: Western Carpathians, EC: Eastern Carpathians, SC: Southern Carpathians and CIM: Central-Island Mountains (Apuseni), respectively.

The main barriers to gene flow as detected with BARRIER on the cpSSR dataset delimited the Eastern Carpathians populations with highest bootstrap support of 99.5% **(Fig. 12a)**. Between the Eastern and the Southern Carpathian populations, the barrier was only weak (34.3%). Another strong barrier (with up to 86.1% support) was drawn around the RML population in the Transylvanian Central-Island Mountains (Apuseni). In the High Tatra, SME population also showed genetic discontinuity with a 58% supported barrier.

4.1.2. Nuclear microsatellites

All loci amplified successfully and presented polymorphism in our studied region. Similarly, all loci largely conformed to the HWE and showed no significant deviations, although a few populations partially showed minor deviations, which cannot be associated with null alleles or non-neutral behaviour. Micro-Checker test confirmed the lack of null alleles.

A moderate level of intra-population variability was found. The mean of H_e was 0.586, which ranged from 0.493 in RMO to 0.648 in HKO. The H_o was 0.589, ranging from 0.488 in RMO to 0.652 in HKO. HKO population showed the highest value, both in the number of alleles (N_a: 8.250) and in the number of effective alleles (N_{eff}: 4.862). The mean number of private alleles was remarkably high in RPO (1.750), while in the rest of the populations the overall mean was 0.268. Inbreeding coefficient (F_{IS}) varied greatly, ranging from -0.2411 in RFE to 0.2830 in RPO, with a mean of -0.0329 **(Table 5)**.

Table 5: Carpathian populations of Scots pine (*Pinus sylvestris*) and details regarding standard indicators of genetic diversity for nuclear simple-sequence repeat (SSR) datasets. Notations for nuclear (SSR) as indices follows: N_a = number of alleles, N_{eff} = number of effective alleles, N_p = mean number of private alleles, H_o = observed heterozygosity, H_e = expected heterozygosity, F_{IS} = inbreeding coefficient.

	nSSR					
Population	N_a	N_{eff}	N_p	H_o	H_e	F_{IS}
BYU	7.875	4.198	0.250	0.593	0.584	-0.06329
ESE	7.625	4.671	0.250	0.529	0.561	-0.02113
HFE	7.625	4.675	0.625	0.576	0.604	-0.12653
HVE	7.250	4.299	0.125	0.580	0.622	0.18455
HZA	6.625	4.002	0.250	0.577	0.581	-0.04651
HOR	7.125	4.138	0.250	0.570	0.594	-0.08374
HKO	8.250	4.862	0.375	0.652	0.648	-0.09353
SKV	6.875	4.748	0.250	0.625	0.593	-0.02703
SME	6.625	4.579	0.000	0.552	0.641	0.19718
STU	6.625	4.231	0.500	0.644	0.636	-0.09195
SLI	5.750	3.930	0.125	0.664	0.647	-0.08621
RFE	4.750	2.375	0.000	0.608	0.524	-0.24113
RPO	5.750	3.096	1.750	0.521	0.515	0.28302
RMO	4.625	2.791	0.000	0.488	0.493	-0.08411
RPA	6.125	3.865	0.250	0.528	0.588	-0.08642
RBI	4.875	3.314	0.000	0.639	0.587	-0.02439
RCO	6.625	4.336	0.125	0.695	0.636	-0.03394
RBE	4.875	3.502	0.000	0.545	0.552	-0.06742
RML	3.500	2.815	0.000	0.538	0.517	-0.05882
RMH	6.000	4.216	0.250	0.669	0.598	-0.08696
Mean	6.268	3.932	0.268	0.589	0.586	-0.03292

AMOVA showed high molecular variance within individuals (90%) and relatively low molecular variance among individuals (3%), while among populations only a 7% variation was observed. Overall F_{ST} was 0.071 (p< 0.001), while within the Carpathians F_{ST} was 0.075 (p< 0.001). The number of migrants per generation (Nm) is estimated to be 3.247. Mantel test of IBD yielded no significance value (r_{xy}: 0.085, p< 0.233), and distribution of Nei's genetic distance over the geographic region did not show limited or restricted gene flow **(Fig. 13b)**.

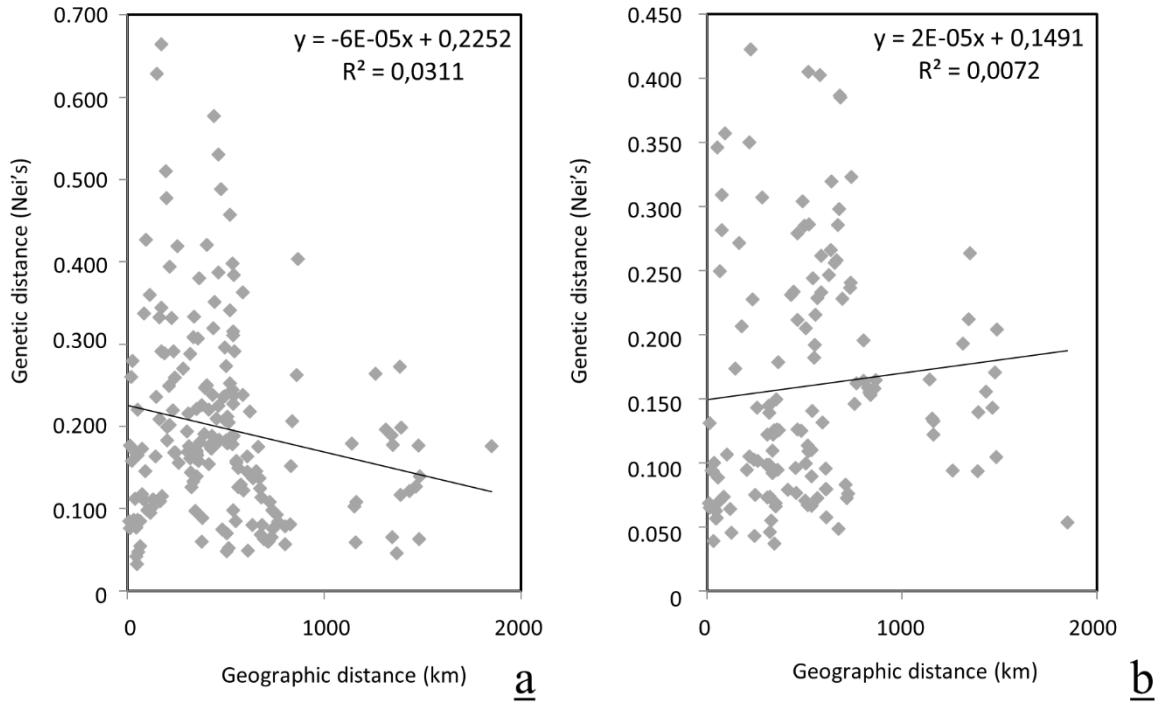


Fig. 13: Results of the Mantel test (Mantel 1967) for correlation between geographical (kilometres) and genetic (Pairwise Unbiased Nei) distances at cpSSR (a) loci (r_{xy}: -0.176, p< 0.045) and at nSSR (b) loci (r_{xy}: 0.085, p< 0.233).

By evaluating nuclear SSR variation with STRUCTURE when K=2, the two groups are **(Fig. 14)**: (1) Western Hungarian populations and the Southern Carpathian population, (2) Eastern Carpathians population with the Bulgarian population (BYU) and the northernmost population from Estonia (ESE). Populations from the Transylvanian Central-Island Mountains (Apuseni Mts.) are distributed in both clusters, RMH, RML, and RBE in Cluster 1 and RBI in Cluster 2. At K=3 two peat bog populations (RMO, RPO) from the Eastern Carpathians are delimited within Cluster 1. Although populations from Bulgaria and Estonia showed grouping with the Eastern Carpathian populations at K=2 and K=3, these populations were highly admixed **(Fig. 14)**.

To identify underlying subclusters, the two clusters (namely Cluster1 and Cluster2) were reanalysed separately (data not shown). Within Cluster 1, RFE was separated, while in Cluster 2, high admixture was detected without any clear substructuring among the Hungarian, Slovakian, and Romanian populations.

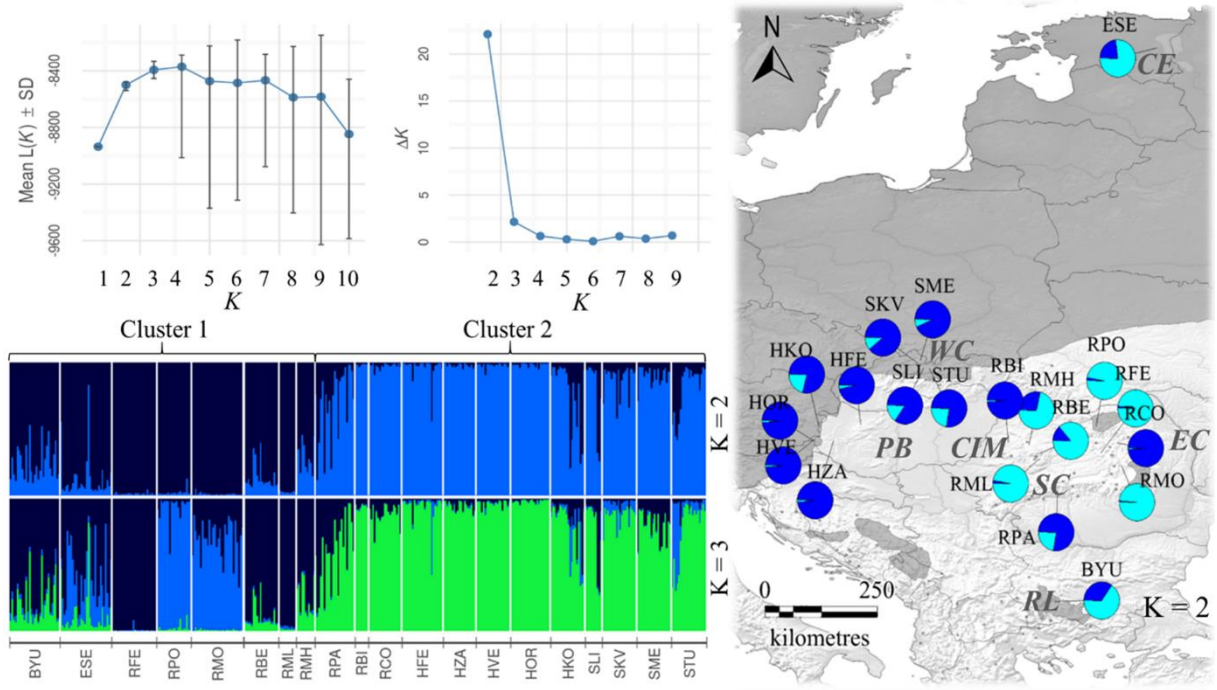


Fig. 14: Estimated population structure K=2 and K=3 of assignment analysis performed in STRUCTURE (Pritchard *et al.*, 2000). STRUCTURE HARVESTER (Earl 2012) resulted K=2 plotted on a geographic map. Most likely membership in a population is presented by the colour of the individual's thin line. The acronyms stand for the population code in **Table 2**. The natural distribution of Scots pine is marked in grey according to the EUFORGEN database, with modifications by the author. RL: Rila Mountains, CE: Central Estonian Plain, PB: Pannonian Basin, WC: Western Carpathians, EC: Eastern Carpathians, SC: Southern Carpathians and CIM: Central-Island Mountains (Apuseni), respectively.

BARRIER analysis identified major genetic discontinuities with high bootstrap support (from 70.9% to 100%) around the Eastern Carpathians, separating these populations from the rest **(Fig. 12b)**. Additionally, the single RCO population from rocky substrate was separated within the Eastern Carpathians with a 99.5% highest support. As in the case of the cpSSR dataset, around the RML population a strong barrier (up to 92.2% support) was detected. All the other barriers between the populations were weak and indicated a non-significant separation with <45% bootstrap support.

The BOTTLENECK analysis showed no evidence of significant excess or deficit of heterozygosity in most populations (19) under the TPM model. As an exception, a recent decline (heterozygote excess) was experienced in the RML population from the Central-Island Mountains (Apuseni). This was also supported by Wilcoxon sign-rank test. Additionally, a shift in the distribution of allele frequencies was detected in the RML population. In the case of non-bottlenecked Scots pine populations, the "Mode-shift" indicator test identified a normal L-shaped form of allele frequency distribution as expected in populations that are near to mutation-drift equilibrium. On the basis of the distribution of allele frequencies, we can assume that populations are randomly mating.

The Approximate Bayesian Computation (ABC) analysis, showed at the first evaluation **(Fig. 15a)**, that the most likely scenario is Sc 4, the admixture model, where posterior probability (0.7914, 95% C.I.: 0.7671–0.8156) was significantly higher than that for Sc 1 (0.1909, 95% C.I.: 0.1670–0.2148), Sc 2 (0.0145, 95% C.I.: 0.0083–0.0206) and Sc 3 (0.0033, 95% C.I.: 0.0000–0.0088). In Sc 4, the effective size of Na ancestral population was determined to be 479, and population sizes for N1 (Pop1), N2 (Pop2) and N3 (Pop3) were 16,000, 5540 and 5370, respectively. Estimated time for ancient divergence t2 (divergence of Pop1 and Pop2) was 3090 (95% C.I.: 867–8170) generations ago, and for recent admixture event and then divergence (Pop1-2-3) at t1 was 180 (95% C.I.: 23.6–823) generations ago. Assuming 50 years as generation time for extreme peripheral populations of Scots pine, t2 value scaled to 154.5 ka, and t1 to 9 ka. Considering a longer 60 years generation time, these increases to 185.4 ka for t2, and 10.8 ka for t1. Out of the 21 summary statistics, only one showed a significant difference between the observed and simulated data based on the posterior distributions, furthermore PCA analysis also showed that both simulated prior and posterior plots were clustered around the observed dataset **(Supplemental Materials Fig. S1)**, suggesting that scenario 4 generally fitted to the observed dataset. Estimated Type I error rate was 0.444 (Sc4) and the average Type II error rate was 0.0766.

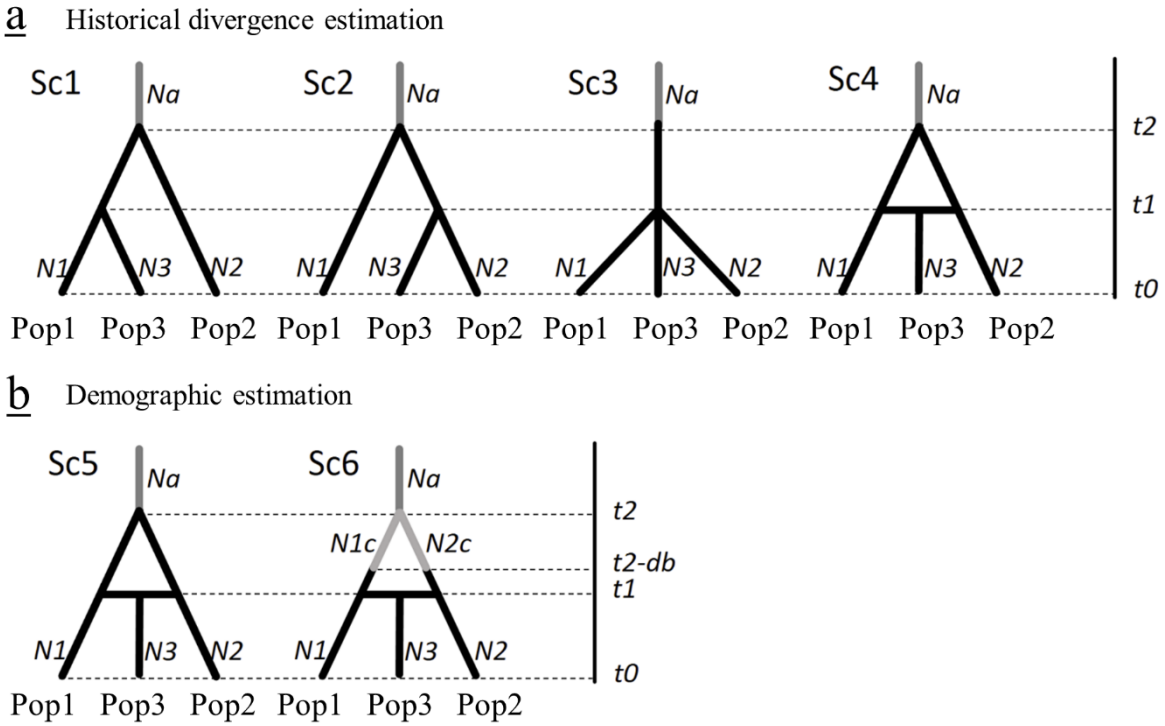


Fig. 15: The six demographic scenarios (a and b) tested on the Central-Eastern European Scots pine populations in DIYABC. In the tested scenarios t# represent the estimated time expressed in generation time, N# is the effective size of population Pop1, Pop2 and Pop3 respectively. N2c and N1c indicate expansion of ancient population expressed in effective population size, as like Na < N#c < N#.

The results of the second ABC analysis **(Fig. 15b)** indicated that Sc6 **(Fig. 16)** had the highest posterior probability (0.5670, 95% C.I.: 0.5307–0.6033), whereas Sc5 was less supported (0.4330, 95% C.I.: 0.3967–0.4693). The median values of the effective population sizes for the best scenario indicated that the ancestral population (Na, estimated to 430 (95% C.I.: 16.3–1790)) was 38.37, 14 and 13.58 times lower than Pop1, Pop2 and Pop3 actual populations. Accordingly, growth in effective population sizes showed that both Pop1 and Pop2 increased from Na to 5790 (95% C.I.: 1180–9770) and 2520 (95% C.I.: 423–7350) by t2-db time (915, 95% C.I.: 59.8–1920). Assessment of the most recent population sizes for N1, N2 and N3 resulted a further increase to 16,500, 6020 and 5840, respectively. Determined divergence times slightly differed from Sc4, and resulted t2 to 3560 (95% C.I.: 1040–8810) and t1 to 236 (95% C.I.: 29.4–876) in the final scenario.

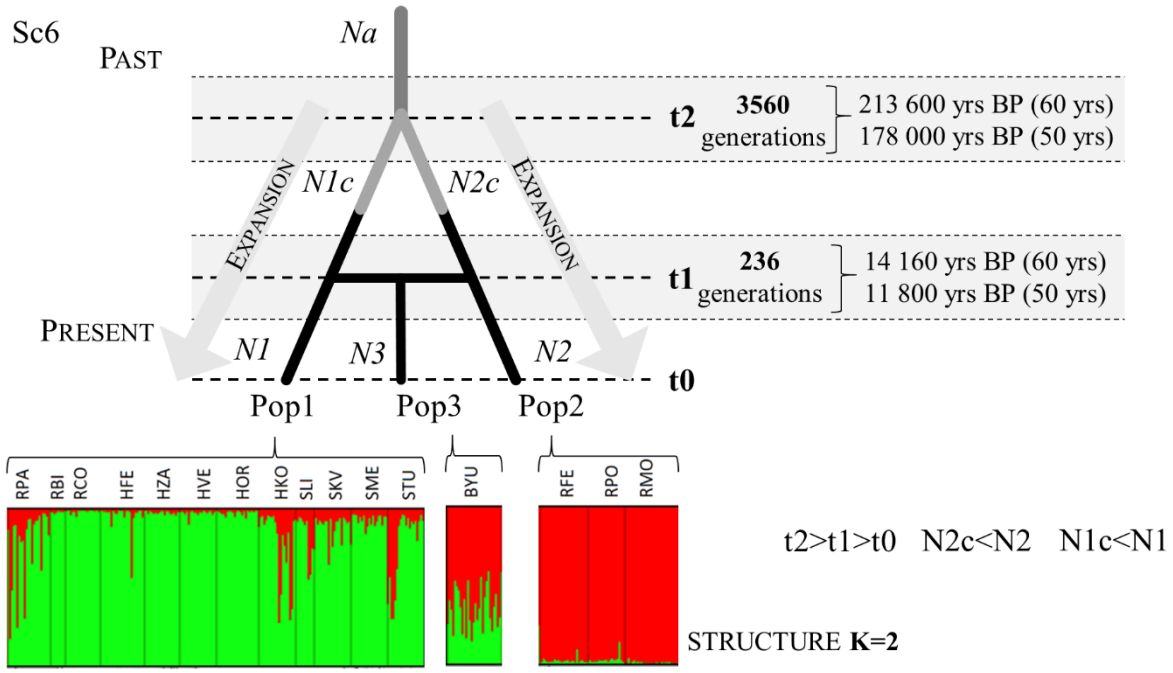


Fig. 16: The best supported demographic scenario with the highest posterior probability (0.5670, 95%CI: 0.5307–0.6033) resulted on the Central-Eastern European Scots pine populations in DIYABC. In the detected scenario t# represent the estimated time expressed in generation time, N# is the effective size of population Pop1, Pop2 and Pop3 respectively. N2c and N1c indicate expansion of ancient population expressed in effective population size, as like Na < N#c < N#.

4.2. Analysis of coding candidate gene dataset

11 candidate gene sequences related to draught stress response were evaluated in this study with a total length of 5,328 bp in 96 individuals. The minimum fragment size of 238 bp was amplified in *ppap12* gene, while the maximum of 1,005 bp was recorded in *rd21A*. The dataset provided a set of 232 polymorphic (segregating) sites consisting 158 SNPs and 74 singletons. On average, one mutation was found every 23 bp in the whole sample set. Altogether 87 non-synonymous, 55 synonymous and 90 non-coding substitutions (in UTRs and introns) were identified **(Table 6)**.

The highest number of substitution sites was found at *dhy2PP* and *rd21A* loci and in the SME and STU populations. The average population private substitutions per locus was between 0.500 and 1.900. Most of the population private substitution sites were detected in SME which was about twice as many as in other populations **(Supplemental Materials Table S9)**.

Table 6: The loci included in the nucleotide diversity analysis and the identified polymorphic sites.

			Base pairs screened				
Locus	**n**	***ns***	**Total**	**Coding region**	**Non-coding region**	***S***	**Sing**
abaR	95	190	378	321	57	17	5
ccoaomt	93	186	549	300	249	14	2
chcs2	95	190	341	231	110	16	2
cpk3	95	190	599	306	293	19	3
dhn3	61	122	268	216	52	11	4
dhn7	82	164	358	279	79	23	8
dhy2PP	96	192	454	357	97	34	13
erd3	95	190	679	474	205	14	9
pal1	96	192	459	276	183	15	4
ppap12	96	192	238	237	1	25	6
rd21A	80	160	1005	624	381	44	18
Total	984	1968	5328	3621	1707	232	74

n: number of samples analysed by locus
ns: number of haploid sequences
S: number of polymorphic (segregating) sites identified
Sing: total number of singleton mutations

The average nucleotide diversity (π) over all loci was π=0.0056 and Theta pi ($\theta\pi$) nucleotide diversity was $\theta\pi=2.15\times10^{-3}$. At all non-synonymous sites, the average π was 0.0028 and for synonymous sites the average π was 0.0091. All studied loci had higher values of π diversity at synonymous sites compared to the non-synonymous sites, except for *abaR* (π_{syn}=0.0014, $\pi_{non\text{-}syn}$=0.0049) and *dhn7* (π_{syn}=0.0024, $\pi_{non\text{-}syn}$=0.0053). Theta pi ($\theta\pi$) nucleotide diversity across the studied populations presented a balanced diversity ranging from 1.81×10^{-3} in RPO to 2.69×10^{-3} in SME. Highest values of $\theta\pi$ are detected at *dhy2PP* loci, which was higher than the average (4.24×10^{-3}-4.84×10^{-3}) in five populations (HVE, HFE, RFE, SME, SKV) from the studied ten **(Table 7, Supplemental Materials Table S10)**. An average haplotypic diversity (Hd) of 0.733 and an average number of 22.63 haplotypes (*Nh*) was present in the whole sample

set. The highest number of haplotypes was detected in *dhy2PP* and *rd21A*, while the lowest was in *ccoaomt* and *erd3*. TCS analysis across all studied loci resulted high number population specific haplotypes with low frequency and there were no major clade specific neither to a geographic region nor to a specific habitat type. Loci that presented the highest number of haplotypes (*dhy2PP* and *rd21A*), showed complex networks in which, all studied regions had private and low frequency haplotypes, except STU, RFE in *dhy2PP* and HFE in *rd21A*, respectively **(Supplemental Materials Fig. S2-12)**. The population SME showed the highest number of population specific haplotypes (33 in total) along with STU, HKO (15 and 16 in total) and SKV (15 in total) **(Supplemental Materials Table S11)**.

Table 7: Summary statistics of nucleotide and haplotype variation at the loci of the analysed candidate genes in Central-Eastern European Scots pine populations.

	Nucleotide diversity								**Haplotype diversity**	
	Total			Non-syn. sites		Syn. sites			Total	
Locus	π	θπ	k	$n_{non\text{-}syn}$	π	n_{syn}	π	$\pi_{non\text{-}syn}/\pi_{syn}$	*Nh*	Hd (s.d.)
abaR	0.0048	1.7440	1.8260	10	0.0049	5	0.0014	3.3020	23	0.824 (0.022)
ccoaomt	0.0021	0.9911	1.1800	2	0.0003	2	0.0017	0.1850	11	0.355 (0.045)
chcs2	0.0066	2.1278	2.2740	2	0.0001	7	0.0139	0.0080	17	0.760 (0.025)
cpk3	0.0045	2.4616	2.7380	0	0.0000	5	0.0168	0.0000	21	0.674 (0.028)
dhn3	0.0058	1.447	1.5740	7	0.0027	1	0.0099	0.2670	12	0.822 (0.018)
dhn7	0.0066	2.3081	2.3880	12	0.0053	3	0.0024	2.2230	25	0.729 (0.034)
dhy2PP	0.0093	3.9377	4.2420	10	0.0014	9	0.0131	0.1080	45	0.914 (0.013)
erd3	0.0010	0.7100	0.7280	6	0.0002	3	0.0003	0.5850	12	0.558 (0.028)
pal1	0.0045	2.0796	2.1070	7	0.0035	5	0.0193	0.1830	18	0.739 (0.022)
ppap12	0.0125	2.9161	2.9820	18	0.0119	7	0.0151	0.7880	24	0.819 (0.019)
rd21A	0.0031	3.0018	3.2100	13	0.0006	8	0.0057	0.1190	41	0.874 (0.016)
Mean	0.0056	2.1568	2.2954	-	0.0028	-	0.0091	0.7060	22.63	0.733

Nh: number of haplotypes
Hd: haplotype (gene) diversity
π: nucleotide diversity (Nei 1987)
θπ: Theta nucleotide diversity, $x10^{-3}$ (Tajima 1989)
k: average number of nucleotide differences
$n_{non\text{-}syn}$: number of non-synonymous mutations
n_{syn}: number of synonymous mutations

Decay of linkage disequilibrium (LD) was detected in only seven candidate gene loci, of which 4 loci (*abaR*, *dhn7*, *dhy2PP*, *pal1*) presented a slow decay (R^2=0.01-0.02) at a distance about 100-400 bp and 3 loci (*ccoaomt*, *erd3*, *rd21A*) presented a faster, but still relatively slow decay (R^2=0.04-0.09) at a distance about 100-400, -600 and -800 bp, respectively. Four candidate loci (*chcs*, *cpk3*, *dhn3*, *ppap12*) showed no intragenetic LD decay (R^2=<0.01) over the analyzed distance **(Supplemental Materials Fig. S13)**.

Test of Analysis of Molecular Variance (AMOVA) including all populations and by testing all candidate loci resulted an overall F_{ST}=0.063 (p<0.001), meaning that high molecular variance (93.63%) resides within populations, while among populations only a 6.36% variation was observed. High among population differentiation was detected at *ccoaomt* (13.8%), *cpk3* (11.4%) and *dhn3* (9.2%) loci, while all the other loci presented a low to moderate level of differentiation

ranging from 1.1% to 7.5% **(Supplemental Materials Table S12)**. F_{ST} value as measure of population differentiation was calculated between all population pairs. Maximum differentiation was found at RFE (at *ccoaomt, chcs, cpk3, rd21A* loci) and SME (at *dhy2PP, dhn7, ccoaomt* loci) populations (pairwise F_{ST}>0.20-0.50), while at the remaining populations a low differentiation (pairwise F_{ST}<0.05-0.10) was observed **(Supplemental Materials Fig. S14)**.

Bayesian clustering for genetic assignment evaluated with STRUCTURE indicated that the most likely number of groups of populations was K=4, using Evanno *et al.* (2005) ΔK criterion **(Fig. 17)**. The major group comprises most of the populations from across the studied range, similarly as detected at K=3, although at K=4 all populations within the group showed high level of admixture. Two distinct clusters were also detected including single populations: the Eastern Carpathian RFE and the Western Carpathian SME population.

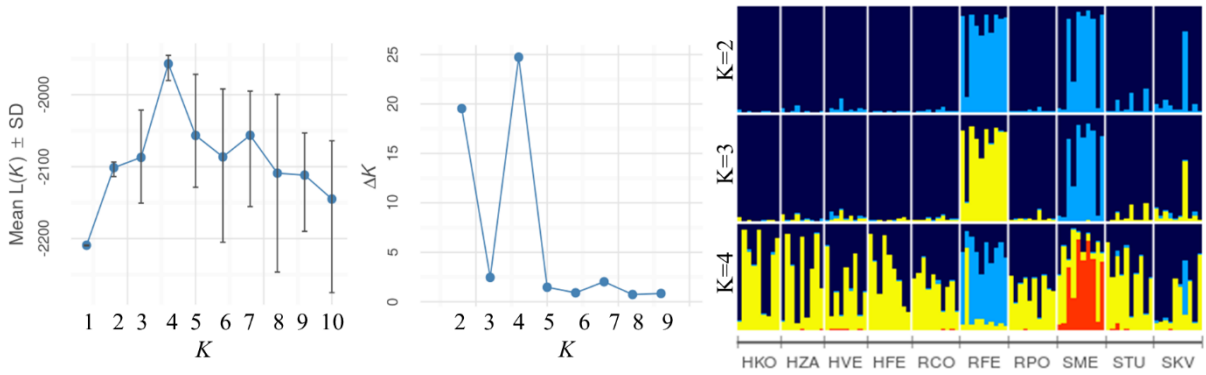


Fig. 17: Estimated population structure K=2, K=3 and K=4 of assignment analysis performed in STRUCTURE (Pritchard *et al.*, 2000) resulted on Central-Eastern European Scots pine populations. Most likely membership in a population is presented by the colour of the individual's thin line. The acronyms stand for the population code in **Table 2**.

Deviations from neutral expectations were studied with three different estimators, including Tajima's D, Fu and Li's D and F statistics, whereas Tajima's D values were negative for all tested loci. The D value was statistically significant for *erd3* (-1.7616; p<0.05) and *rd21A* (-1.7957; p<0.05), and it was about 2-fold higher than all remaining loci. Considering these loci significant deviation was also detected at the coding regions and at non-synonymous sites (*erd3*: -2.1015; p<0.05, *rd21A*: -19168; p<0.05). Additionally, *dhy2PP* presented a significantly negative value (-1.7949; p<0.05) for non-synonymous sites. In accordance with the Tajima's D statistics for *erd3* and *rd21A*, Fu and Li's D and F values were also negative and significant (-4.1613 and -3.9049; p<0.02, -2.7725 and -2.8331; p<0.05) for these loci. The value of *dhy2PP* was significant only for the Fu and Li's D test (-2.5375; p<0.05). It is noticeable that in case of Fu and Li's D and F value, with five exception (Fu and Li's D; *ccoaomt*, *chcs*, *cpk*, *pal1*, Fu and Li's F; *chcs*) all loci were negative **(Table 8)**.

Table 8: Summary statistics of tests of neutrality at analysed candidate genes in Central-Eastern European Scots pine populations.

	Tajima's neutrality					Fu and Li's neutrality	
Locus	Coding	Non-syn.	Syn.	Non-Syn/Syn	D	D	F
abaR	-1.2581	-0.7131	-1.6740	0.4259	-0.9874	-1.1578	-1.3112
ccoaomt	-1.2814	-1.0864	-0.8823	1.2313	-1.2943	0.2689	-0.3832
chcs2	-1.2565	-1.2919	-0.9368	1.3790	-0.4471	0.4443	0.1233
cpk3	0.4461	n.a.	0.4461	n.a.	-0.4301	0.1469	-0.0947
dhn3	-0.9448	-1.4620	1.4241	-1.0266	-0.5857	-1.3269	-1.2646
dhn7	-1.2910	-1.0728	-1.2492	0.8587	-1.1462	-1.7627	-1.8232
dhy2PP	-1.4635	-1.7949*	-0.7046	2.5471	-0.7844	-2.5375*	-2.1683
erd3	-2.1015*	-1.8783*	-1.4830	1.2665	-1.7616*	-4.1613**	-3.9049**
pal1	0.0093	-0.8089	1.0403	-0.7776	-0.4471	0.4443	-0.8493
ppap12	-0.8446	-0.7434	-0.7753	0.9589	-0.8446	-0.7411	-0.9419
rd21A	-1.9168*	-2.1725***	-0.9812	2.2141	-1.7957*	-2.7725*	-2.8331*

Statistical significance: *; p< 0.05, **; p< 0.02 ***; p< 0.01

We identified 16 SNP loci lying outside the 99.5% (C.I.=0.995) confidence interval, of which 9 were above the C.I., suggesting positive selection and 7 were below the C.I., suggesting balancing selection by applying the F_{ST} based FDIST2 approach implemented in Lositan.

Using the hierarchical method of Excoffier *et al.* (2009) as implemented in the Arlequin software, we identified 21 loci being under selection. 13 loci were candidate for positive selection, while 8 loci were candidate for balancing selection. By considering the highest criteria of 99% C.I. ($p \leq 0.01$) only 6 loci showed to be under selection, of which 4 loci fell below (candidates for balancing selection), while only two above the C.I. (candidates for positive selection).

The Bayesian method implemented in BayeScan detected 18 outliers with a Log_{10} Bayes factor between 0.12-1.50 and by applying a false discovery rate between 0.25-0.05. According to Jeffreys' interpretation, only 6 loci were identified as being a "substantial" evidence for selection (Log_{10} (PO); 0.5-1) and 3 for "strong" evidence for selection (Log_{10} (PO); 1-1.5), while the remaining was identified with less statistical support. The locus-specific component (alpha) was negative for 8 loci, suggesting balancing or purifying selection and positive for only 1 locus suggesting diversifying selection.

Altogether, we considered those loci to be true outliers which were detected with more than one approaches, while this way, SNP outliers are less prone to erroneous identification (Luikart *et al.* 2003, Tsumura *et al.* 2012). By this way, 13 loci out of the 38 detected was observed at different significance level **(Table 9, Supplemental Materials Fig. S15)**. These loci reside in 8 candidate gene and alter 6 non-coding, 1 synonymous and 6 non-synonymous nucleotide change. Seven loci were determined as potentially being under negative (balancing or purifying) selection and 6 loci as to be under positive (diversifying) selection. All loci detected for positive selection were in non-coding genomic regions, while loci under negative selection were located in coding regions and either altering non-synonymous or synonymous change.

Table 9: Overview of outlier SNPs detected being under selection across all analyzed candidate gene loci.

								Arlequin		**Lositan**			**Bayescan**		
Name	Locus	Position	SNPs	Subs. type[a]	Codon change	Product change	Direction of selection[b]	F_{ST}	*p*	F_{ST}	*P*	FDR	Log_{10}(PO)	alpha	FDR
SNP_1_60	*abaR*	60	C/G	NS	GCC/GGC	Ala/Gly	B	-0.0330	p≤0.01	-0.0326	<0.0001	FDR=0.01	1.1980	-1.3885	FDR=0.05
SNP_2_140	*ccoaomt*	140	G/C	NC	-	-	A	0.3639	p≤0.01	-	-	-	0.7782	1.1600	FDR=0.10
SNP_2_396	*ccoaomt*	396	G/C	NC	-	-	A	0.1532	p≤0.05	0.1578	1.0000	FDR=0.01	-	-	-
SNP_5_61	*dhn3*	61	A/C	NS	AAA/CAA	Lys/Gln	B	-0.0673	p≤0.01	-0.0485	0.0000	FDR=0.01	0.6799	-1.1520	FDR=0.15
SNP_6_297	*dhn7*	297	A/T	NC	-	-	A	0.2276	p≤0.05	-	-	-	0.2339	0.6727	FDR=0.20
SNP_6_292	*dhn7*	292	A/T	NC	-	-	A	0.1430	p≤0.05	0.1578	1.0000	FDR=0.01	-	-	-
SNP_7_357	*dhy2PP*	357	T/C	NC	-	-	A	0.2605	p≤0.05	0.2631	0.9985	FDR=0.10	0.2018	0.7102	FDR=0.25
SNP_9_240	*pal1*	240	A/C	NS	AAA/CAA	Lys/Gln	B	-0.0128	p≤0.05	-	-	-	0.6842	-1.1256	FDR=0.15
SNP_9_254	*pal1*	254	G/A	SY	GGG/GGA	Gly/Gly	B	-0.0128	p≤0.05	-	-	-	0.7278	-1.1370	FDR=0.10
SNP_10_112	*ppap12*	112	A/C	NS	ATT/CTT	Ile/Leu	B	-0.0301	p≤0.05	-0.0314	0.0000	FDR=0.01	0.2972	-0.8100	FDR=0.20
SNP_10_13	*ppap12*	13	C/G	NS	CAG/GAG	Gln/Glu	B	-0.0528	p≤0.01	-0.0530	0.0006	FDR=0.10	1.5036	-1.4919	FDR=0.05
SNP_10_220	*ppap12*	220	G/C	NS	GCT/CCT	Ala/Pro	B	-0.0538	p≤0.01	-0.0540	0.0006	FDR=0.10	1.5036	-1.5137	FDR=0.05
SNP_11_671	*rd21A*	671	T/C	NC	-	-	A	0.1426	p≤0.05	0.1578	1.0000	FDR=0.01	0.3608	0.9015	FDR=0.15

a: NC; Non-coding, NS; Non-synonymous, SY; Synonymous
b: Estimated direction of selection: A; positive (diversifying), B; negative (balancing or purifying)

5. DISCUSSION

5.1. Discussion of non-coding microsatellite results

Studying non-coding SSR aimed to assess the genetic diversity and differentiation, based on nuclear and chloroplast microsatellite marker analysis of *Pinus sylvestris* populations considered to be natural and inhabiting the species distribution range limits along the Carpathians and the Pannonian Basin. Both marker types showed overall high genetic variation along the range, suggesting that population fragmentation events might have taken place relatively recently.

Although the two marker types exhibit particular genetic patterns, there were also some congruencies in the detected spatial genetic structures of populations. BAPS and STRUCTURE analysis delimited Western Carpathian populations that proved to be different from those inhabiting the eastern range of the Carpathians **(Fig. 11 and 14)**. Furthermore, Eastern Carpathian populations were differentiated by cpSSR markers from the populations of the Transylvanian Central-Island Mountains (Apuseni) **(Fig. 11)**. The differentiation in the genetic pattern along the Carpathians has been recognized in earlier studies of other conifer species also. In the case of *Picea abies* or *Abies alba,* mitochondrial minisatellite regions and nSSRs delimited lineages of the Western and Eastern Carpathians and, accordingly, these patterns suggest different origins of populations from distinct glacial refugia (Tollefsrud *et al.* 2008, Liepelt *et al.* 2009, Gömöry *et al.* 2012). *Salix herbacea,* another arctic-alpine species from the Carpathian region, showed identical population structure highlighted by BAPS and STRUCTURE and confirmed the distinct origin of the Carpathian populations in question (Alsos *et al.* 2009). Moreover, the haplotype pattern of Scots pine revealed in our study is highly congruent with those observed by Höhn *et al.* (2009) for *Pinus cembra,* where the populations of the Western and the Eastern Carpathians were spatially separated on the basis of chloroplast SSR variation. However, this separation was not significant in the study by Lendvay *et al.* (2014), which was done using nuclear SSR markers. Based on a cluster analysis of cpSSR haplotype frequencies in *Picea abies,* Bucci and Vendramin (2000) revealed differentiated genetic pattern between Western and South-Eastern Carpathian populations. These findings correspond with the main geobotanical regions described earlier for the Carpathians (Georgescu and Donita 1965, Zemanek 1991, Ronikier 2011).

The Western Carpathian population, Medzi bormi (SME) from Slovakia, was the most outstanding, showing the overall highest number of private alleles (N_P= 1.000) and forming a distinct cluster in the BAPS. Moreover, Barrier analysis separated this population from those inhabiting the same range with a 58% bootstrap support **(Fig. 12)**. This might be explained by hybrid individuals of *Pinus rhaetica* (*P. sylvestris* x *P. mugo*) reported earlier from this peat bog (Staszkiewicz 1994, Kormut'ák *et al.* 2013). Bottleneck and restricted gene flow was detected by Barrier analysis in the Mluha population (MLA) from the Transylvanian Central-Island Mountains, which might be a consequence of the decreased population census size where closely related individuals are mating within an isolated stand **(Fig. 12)**. On the basis of a very early first description of the Mluha peatbog, in which Scots pine is not mentioned (Csató 1885), one plausible explanation would be the recent colonization of the peatbog by this species. Later, a low population

size was mentioned by Pacurar *et al.* (2010). Additionally, the Fantana Brazilor (RFE) population from the Eastern Carpathians presented as a conspicuously different group in the substructure analysis and also showed distinct haplotype proportion in the region. It is possible that this stand originates from a distinct refugium or might bear signs of earlier human influence.

On the basis of both chloroplast and nuclear microsatellite markers that revealed congruent structure with the previously mentioned conifer species, it is likely that the Carpathian populations of Scots pine harbour genetic material originating from at least two separate refugia, dating back to the Pleistocene. One refugium might have been situated around the Eastern Alps and the Hungarian Plain with the Danube region (Cheddadi *et al.* 2006, Tribsch and Schönswetter 2003), and the other might have existed in the Eastern Carpathians, where a high abundance of fossil pollen remains was reported (Feurdean *et al.* 2011). These two possible refugia were also reported for subalpine and alpine perennial plant species, such as *Hipochoeris uniflora* (Mráz *et al.* 2007) and *Campanula alpina* (Ronikier *et al.* 2008, Ronikier and Zalewska-Galosz 2014), which present similar delimitations in population structure and support both the North-eastern Alpine and East-Carpathian refugia. Eastern Carpathian populations might have also served as source populations in later Holocene colonization towards northern latitudes as described for other coniferous species (Latalowa and van der Knaap 2006, Feurdean *et al.* 2007, Tollefsrud *et al.* 2008). This is probably the case, since the Estonian population and the more southern Bulgarian population clustered together and exhibited a common structure.

The elevated genetic diversity revealed by our study compared to that found in the Romanian-Hungarian populations by Bernhardsson *et al.* (2016) may be attributed to the larger sampling area of our study, which included not just the Romanian Carpathians but also the Western Carpathians (the Tatras) and a higher number of Hungarian populations.

On a broader geographic scale, our results show, that genetic diversity values are generally the same as the values found in earlier studies of Scots pine inhabiting the European region, including the Balkan Peninsula. The mean detected expected heterozygosity (H_e: 0.586) of nuclear SSR shows similarities to the Bulgarian populations studied earlier by Naydenov *et al.* (2011), but much lower than even more southern peripheral populations from the Apennines and the southern Alps studied by Scalfi *et al.* (2009). Chloroplast haploid diversity (h: 0.546) was unexpectedly lower than the degree of diversity detected using similar markers (Vendramin *et al.* 1996, Soranzo *et al.* 1998) at the edge of the range populations in Italy and in Spain by Scalfi *et al.* (2009) and Robledo-Arnuncio *et al.* (2005). Despite the fact that genetic diversity estimated with nuclear markers was elevated in our region, chloroplast population diversity indices show detectable signs of segregation and fragmentation of these isolated populations, which can be the effect of restricted gene flow on a regional scale. Similarly, genetic discontinuity was also detected in both datasets with Barrier analysis.

As expected, a high number of cpSSR haplotypes was detected (36 size variants combined into 141 haplotypes) all over the range studied in the Carpathian populations. A high number of haplotypes was reported in earlier studies by Naydenov *et al.* (2005), Robledo-Arnuncio *et al.*

(2005) and Cheddadi *et al.* (2006), since these microsatellite regions have very high mutation rates (Vendramin and Ziegenhagen 1997, Vendramin *et al.* 1998, Provan *et al.* 1998).

We did not find signs of inbreeding in most of the populations studied, as in our study F_{IS} values were overall negative (-0.0329), except in the cases of HVE, SME, and RPO, for which F_{IS} values were found to be positive (0.184, 0.197 and 0.283). An earlier study by Bernhardsson *et al.* (2016) reported an overall positive F_{IS} value, potentially as a consequence of artificially maintained and human-restored populations. Alternatively, we consider that in HVE, SME, and RPO, in all likelihood the small population census size (0.02-4.49 km^2) and highly isolated habitat have increased the rate of selfing and might exhibit higher self-offspring viability (Savolainen *et al.* 1992), resulting in a slightly increased positive F_{IS} value. Although small and isolated populations are more vulnerable to inbreeding (Ellstrand and Elam 1993), our overall results regarding most of the population studied is in accordance with earlier statements according to which inbreeding takes generations to develop and/or even with a restricted gene flow, populations still maintain gene exchange. By estimating gene flow between populations, we detected a relatively high number of possible migrants per generation. Between populations, the value of the number of migrants per generation was elevated (Nm=6.272) for chloroplast and moderate (Nm=3.247) for nSSRs.

In accordance with no signs of inbreeding, BOTTLENECK analysis provided evidence that the Carpathian populations studied are not influenced by a recent genetic bottleneck. Long-lasting signs of bottlenecks require multiple generations to appear. Furthermore, the effects can vary not based only on the reduction size, but also depending on the duration period (Busch *et al.* 2007, Peery *et al.* 2012). It is most plausible that populations that today are isolated have undergone a recent fragmentation and isolation event. Macrofossil and pollen records indicate that conifer species like *Pinus sylvestris* with diploxylon pollen type have survived the LGM in the Carpathians and the Pannonian Basin (Rudner *et al.* 1995, Rudner and Sümegi 2001, Magyari 2011), and a strong withdrawal and population decline began only between 8000–10,000 years BP (or even later, depending on geographic location and elevation) in the Late Glacial/Holocene transition period to mid-Holocene (Tantau *et al.* 2003, 2006, Feurdean and Bennike 2004, Feurdean *et al.* 2007, Feurdean *et al.* 2012). Transition from coniferous stands to mixed forests has been detected by Mihai *et al.* (2007), and recently, within the last decades, increasing clear-cutting of the coniferous forests for pasturing has been reported (Motta *et al.* 2006).

Both marker types in our study presented a relatively high among-population differentiation, as in the cases of other peripheral study sites in the Italian Alps and the Apennines (Scalfi *et al.* 2009, Belletti *et al.* 2012). In our results, SSRs pairwise population differentiation was Φ_{PT}=0.071 in case of nuclear and Φ_{PT}=0.074 for chloroplast within the studied range. Moreover, there is presumably not an impervious barrier among regions, because neither gene flow nor inbreeding supports this. Our estimated differentiation in the Carpathian region might be related to the contact zone that has been established as a consequence of the migration of diverged lineages that survived glaciation in separated refugia and marked the geographical barrier detected.

Demographic history of Scots pine using molecular markers and involving macrofossil and pollen remains has been studied formerly across the European distribution (Cheddadi *et al.* 2006; Pyhäjärvi *et al.* 2007). However, the Quaternary history of the species in the region including the Carpathian Mountains and the Pannonian basin, has not been investigated yet by coalescent-based genetic analysis using Approximate Bayesian Computation approach. Fossil records from the Central-Eastern European region are available showing putative history of the species, like in the case of the Southern-Carpathians (Farcas *et al.* 1999, Magyari *et al.* 2012) or Northern- and Eastern-Carpathians (Magyari *et al.* 2014a, 2014b). Therefore, the results of our ABC analysis can provide a valuable information for linking fossil evidence to present day genetic pattern of Scots pine.

Our DIYABC analysis using 8 nSSR loci supported an admixture scenario, the SC 6, in which the two main detected gene pools (Pop1 and Pop2) separated at the same time, rather than the hierarchically split gradual divergence or simple split scenarios **(Fig. 16)**. Pop3, as a highly mixed population detected in our STRUCTURE analysis genetically infer with the two main populations Pop2 and Pop3 in distinct times presumably due to an admixture event. Potential admixture event has been confirmed along the run of demographic estimations, where gradual expansion of the populations were detected. This might have caused admixture of Pop1, Pop2 and Pop3, respectively. The estimated divergence times are strongly affected by the generation time of Scots pine, but it can greatly vary for conifers. Grotkopp *et al.* (2002) has estimated 5 years MGT (Minimum Generation Time) while Provan *et al.* (1999) up to 100 years as generation time for *Pinus sylvestris*. Based on our experiences we assumed a generation time to be approx. 50-60 years under extreme environmental conditions. If we assume this generation time, the first divergence time (t2), from the ancestral population, falls within 178 ka and 213.6 ka BP and the admixture event (t1) from 11.8 ka to 14.1 ka. Furthermore, we detected a population expansion taking place from t2 time, when diverged populations expanded and their effective population size increased from 430 individual up to 16,500 (Pop1), 6020 (Pop2) and 5840 (Pop3) by the time of admixture event.

Although, it is hard to make conclusions due to the lack of long time pollen records (going back to Pleistocene) from the Carpathians, there are strong evidences that *Pinus* (diploxylon) species were dominating from the mid-Pleistocene's transition to glacial to early Holocene interglacial period. Deep pollen cores from the Tenaghi Philippon peatland in Greece, showed an overall increase of *Pinus* pollen for the first time by 10% from 129 ka, which later steadily increased to 45% by 113 ka BP (ka; kilo ages/ kiloannus) (Milner *et al.* 2013). Moreover, sedimentological proxies from a recent study of Sadori *et al.* (2015) from lake Ohrid (western Balkan region, Albania) highlighted the high abundance of *Pinus* pollen, 10-87% between 245-189 ka, 14-83% between 161-121 ka and 9-77% between 70-12 ka. Accordingly, *Pinus* pollen concentration remained high during mid-Pleistocene to the LGM. These findings fit well to our detected population expansion causing extensive distribution of the species, which might have made possible the admixture event between the nowadays geographically distant stands. Since there is close relation between ice volumes, climate and forest expansion/contraction (Tzedakis *et*

al. 2006), it is certain that due to the favorable climate of the long lasting glacials coniferous species have maintained their population sizes in our studied region, despite the upper-Pleistocene's warmer interglacial and short dry-wet climate oscillations.

Our estimated admixture event for Scots pine, based on the ABC analysis, might have happened between 11.8 ka to 14.1 ka BP **(Sc6; Fig. 16)**, when Scots pine displayed a vast expansion, at the end of Younger Dryas and early Holocene, when *Pinus* (diploxylon) pollen percentages were at their maximum (Feurdean *et al.* 2011). Although, coniferous species like *Picea, Larix, Juniperus* including *Pinus sylvestris* survived the LGM in the Carpathians and in the Pannonian basin (Rudner *et al.* 1995, Rudner and Sümegi 2001, Magyari 2011), a strong reduction of conifers and expansion of deciduous species has started in the Late Glacial/Holocene transition period (Feurdean *et al.* 2012). This decline in *Pinus* pollen abundance has been detected at several sites along the Carpathian mountain range (Tantau *et al.* 2003, 2006, Feurdean and Bennike 2004, Feurdean *et al.* 2007), suggesting, that after the expansion and admixture, populations contracted causing fragmentation and reduction of population sizes. Furthermore, this late admixture event, and afterward the formation of recent population structure suggest a recently ongoing segregation and isolation of nowadays relictary populations.

5.2. Discussion of coding candidate gene results

Sequence diversity was evaluated at a set of drought stress related candidate gene loci to elaborate adaptive genetic variation of relict Scots pine populations at the species' Central-Eastern European periphery, the Carpathian Mountains and the Pannonian basin. Patterns of nucleotide divergence at all selected loci linked to dehydrative stress response have been formerly tested in various coniferous species by Pyhäjärvi *et al.* (2007), Grivet *et al.* (2009), Wachowiak *et al.* (2013). We aimed to identify signals of selection, local adaptation and population structuring at the extreme periphery of Scots pine.

Our data provided evidence for selection at the candidate gene loci. We have combined outlier detection approaches/algorithms, in order to gain robustness and to reduce statistical error rate, similarly to Tsumura *et al.* (2012), Csilléry *et al.* (2014) and Müller *et al.* (2015). Although, the different approaches resulted largely congruent results, we need to consider that errors are inevitable even when multiple methods are applied (Tsumura *et al.* 2012). Lositan and BayeScan are generally acknowledged for having the lowest Type II error rate (false negative) and have similarly low Type I error rate (false positive) for positive selection. Type I error for balancing selection is often high for Lositan and Arlequin (Narum and Hess 2011). Although BayeScan is more conservative in identifying outlier SNPs, it is considered to be the most efficient compared to other methods (Pérez-Figueroa *et al.* 2010, Henry and Russello 2013). Our results agreed that 13 significant outlier SNPs (true outliers) are under natural selection either located in coding or non-coding regions **(Table 9)**. Despite the fact that contrasting selection processes were estimated for these outliers (7 loci for negative; balancing or purifying, and 6 loci for positive; diversifying), results indicate a congruent ongoing process at the studied peripheral populations. Outliers suggest that there is a strong signal for selection acting to maintain genetic diversity and potentially to

counteract the effect of genetic erosion. Regardless the 6 loci detected to be under positive selection, while they are positioned in non-coding genomic regions, presumably do not influence the process of adaptation. Another possibility is that genetic structure has a substantial effect on the outlier analysis (the hierarchical island-model implemented in Arlequin considers geographic structure), likewise detected for *Pinus mugo*, *Pinus cembra* (Mosca *et al.* 2016) and for *Picea abies* (Scalfi *et al.* 2014). In addition, neutrality tests (Tajima's D, Fu and Li's D and F statistics) revealed that majority of the candidate genes presented highly negative values, which for *dhy2PP*, *erd3* and *rd21A* were statistically significant **(Table 8)**. This indicates that substitutions may be largely neutral or might also indicate selection against deleterious non-synonymous substitutions (indicate purifying selection). Our neutrality measures point to an excess of low-frequency polymorphism, hence indicate purifying selection, which likely acts against local adaptation in our peripheral populations. Moreover negative values can also refer to hitchhiking effects resulting from a selective sweep or are the result of historical demographic events (either spatial or demographic expansion) formerly revealed in case of forest tree species at drought stress response genes (Grivet *et al.* 2009, 2011, Homolka *et al.* 2013). Effects of historical demography (population size fluctuations effect on sequence variation) were identified by former studies on dehydrin genes and on other cold-related candidate genes for coniferous species, including *Pinus sylvesris* and *Pinus mugo* (Wachowiak *et al.* 2009, 2013, 2014). It should be considered that our sampled populations were proned to demographic fluctuations, as both cpSSR and nSSR markers revealed recent fragmentation in our former study (Tóth *et al.* 2017).

The decay of intragenic linkage disequilibrium (LD) was overall slow or negligible and shorter than ca. 100-400 bp in most genes. LD potentially is the result of complex interactions of several factors like population size, selection process, structure and history and depends on sequence length and nucleotide variation (Krutovsky and Neale 2005, Lalagüe *et al.* 2014). Relatively slow decay was estimated for Scottish populations of *P. sylvestris*, whereas LD decay was almost three times slower than that in mainland populations (Wachowiak *et al.* 2011). Slow decay was reported for *Pinus taeda* at drought stress-response candidate genes (Brown *et al.* 2004, González-Martínez *et al.* 2006). Our LD may present footprints of historical demographic expansion and/or selective sweep affecting the specific loci. Beside these, results could also suggest that randomly occurring unlinked substitutions accumulate to preserve genetic diversity. However for more accurate evaluation of LD further investigation is needed on longer sequences and at more loci.

Nucleotide diversity strongly depends on the investigated genes ranging from π=0.0010 to π=0.0125 (with a mean of π=0.0056) in this study, and thus, a comparison between different studies may be biased **(Table 7)**. Nevertheless, our mean value is much higher than detected in earlier studies for *P. sylvestris* (π=0.0014, Dvornyk *et al.* 2002), *P. radiata* (π=0.0018, Pot *et al.* 2005), *P. pinaster* (π=0.0024, Pot *et al.* 2005), *Picea abies* (π=0.0020, Heuertz *et al.* 2006), *Pseudotsuga menziesii* (π=0.0043, Eckert *et al.* 2009). In recent studies on *P. sylvestris*, including peripheral populations, values presented similar or even higher values, like in case of Scottish

populations (π=0.0078, Wachowiak *et al.* 2011), or populations in Central Europe and Spain (π=0.0056, Wachowiak *et al.* 2014).

Our measured haplotype diversity was generally higher (with a mean of Hd=0.733) than in studies by Wachowiak *et al.* (2009) (Hd=0.683) and Wachowiak *et al.* (2011) (Hd=0.650). However in the study of Wachowiak *et al.* (2014) (Hd=0.789) referring mainly to Scottish populations this value was almost the same. Our set of candidate genes for plant response to environmental stress preserved high variation, accordingly it is not affected by genetic erosion due to historical isolation, fragmentation. Presumably selection to local adaptation did not reduce genetic variation. The high genetic variation was also evidenced in our haplotype network analysis. A highly complex pattern was found but without geographic- or habitat-specific structuring. Additionally, extremely high number of population specific haplotypes were identified. The same issue was reported for *P. mugo*, where the high level of genetic variation was accounted for the complex haplotype pattern (Wachowiak *et al.* 2013).

There was no geographic structuring among the population according to the STRUCTURE result. However two populations were delimited. **(Fig. 17)**. The delta K method (ΔK) (Evanno *et al.* 2005) resulted an optimal value of K=4, separating SME and RFE populations from each other and from the remaining highly admixed populations. Our former study (Tóth *et al.* 2017) emphasized the distinct genetic fingerprint of SME populations, which was potentially assumed to be the consequence of hybrid individuals of *Pinus rhaetica* (*P. sylvestris* x *Pinus mugo*) and their specific genomic composition. The distinct clustering of RFE, both in our non-coding SSR and candidate gene study might be due to several factors, including introduced individuals (alien genetic material), effect of large-scale genetic drift or peculiar historical demographical events.

6. SUMMARY

During the Quaternary Scots pine was widespread in Eurasia, being present even in the outskirts of the glaciated territories. Scots pine survived in several southern latitude refugia across Europe in highly disperse isolated occurrences. Putative refugia of the most frequent genetic lineage was detected from Central Continental Europe. Presumably Scots pine originated from the (Sub) Mediterranean areas like the Balkan Peninsula but also from around the eastern Alps and the surroundings of the Danube plain (western Pannonian basin).

Molecular studies, macrofossil and pollen data analyses concluded that Scots pine in some regions was unable to colonize the vast areas of the European mainland. Lineages from these refugia in the Iberian Peninsula, Italian Apennines, Turkey, and Crimea and in some parts of Asia Minor did not contribute to the recolonization of Europe after the Last Glacial Maximum (LGM).

Based on the evidence from the compiled literature, the greatest regional heterogeneity (highest differentiation) revealed by all marker systems (mitochondrial, chloroplast and nuclear) can be located at the periphery of the species distribution, especially the Balkan Peninsula, the Carpathians, the Alps, the Apennines and the Iberian Peninsula. Notably, all identified refugia were sustained at the southern (Mediterranean) distribution periphery, and these are the most often studied and best documented at present.

Our microsatellite study involved natural peripheral Scots pine populations from Central-Eastern Europe most of which formerly were not studied by molecular markers. Results showed that the most differentiated region is the Romanian Carpathians, i.e. the Eastern Carpathians and the Transylvanian Central-Island Mountains (Apuseni Mts.). Our findings constitute evidence in support of the hypothesis that there were Pleistocene refugia in the region, also evidenced by some recent palynological records.

As natural Scots pine populations from the Carpathians still harbour high gene stocks (indicated by both nuclear and chloroplast diversity estimators) yet unaffected by isolation and genetic erosion, the studied populations represent valuable genetic resources for the species.

Our demographic analysis identified a major divergence from an ancient population with an effective population size growth followed by an admixture event that resulted in the recent population structure. Our results suggest two once existed, separate refugia for Scots pine in the studied region, one situated around the Eastern Alps and a second one in the Eastern Carpathians in accordance to recent fossil (charcoal, micro-, macro- and megafossil) evaluations. Demographic history highlighted that from the ancestral population two lineages have diverged and later due to the favourable climatic conditions of the mid-Pleistocene, populations underwent an expansion leading to an admixture event between 11.8–14.1 ka BP in the early Holocene.

The analysis of nucleotide variation at candidate loci revealed that peripheral populations have maintained high genetic diversity (large nucleotide polymorphism) despite being fragmented and isolated. The large variation is accounted for the weak structuring and high admixture detected in the structure analysis. Similarly, haplotype network evaluation showed no clade-specific structure neither linked to the habitat type nor to geographic position of the populations. The tests of neutrality and the outlier analysis presented selection at the studied draught stress related

candidate gene loci. Most of the detected outlier loci located in the coding regions are under negative (balancing or purifying) selection. This strong signal suggests that selection is actually acting to maintain genetic diversity and to counteract the effect of genetic erosion. Although, some loci were found to be under positive selection, these are located in non-coding regions (potentially not influence local adaptation).

Scots pine populations of the Carpathian Mountains and the Pannonian basin enhanced by their peripheral position and by their particular phylogeographic pattern, provide insights into the species' adaptive genetic variation, enriching our understanding of population genetic processes and allowing us to better assess the ongoing climate change and anthropogenic effects.

7. NEW SCIENTIFIC ACHIEVEMENTS

I. **Non-coding (chloroplast and nuclear) microsatellite markers revealed two distinct genetic lineages and overall geographic structuring of Scots pine populations from the Carpathian Mountains and the Pannonian basin.** Genetic segregation was revealed between the Western and the Eastern Carpathian populations. Results indicate that Scots pine populations harbour genetic material originating from at least two separate refugia. One refugium might have been situated around the Eastern Carpathians and another around the Eastern Alps, the Hungarian Plain with the Danube region.

II. **In accordance to former studies, despite fragmentation and isolation, high genetic diversity was preserved in the natural Scots pine populations of the Carpathians and the Pannonian basin.** This was evidenced by both non-coding (chloroplast and nuclear) microsatellite markers and coding candidate gene sequences. High genetic diversity was revealed by the high number of chloroplast haplotypes and the high expected heterozygosity at non-coding microsatellite loci. Candidate genes presented high nucleotide variation, including population specific substitutions both in coding and non-coding gene regions.

III. **Historical demographic expansion and recent fragmentation was detected for Scots pine populations inhabiting the Carpathians and the Pannonian basin, based on non-coding nuclear microsatellite markers.** A historical divergence from ancestral population was estimated in mid-Pleistocene between 178 ka and 213.6 ka BP with Approximate Bayesian Computation method. Later, the diverged populations have expanded, which led to an admixture event (between 11.8–14.1 ka BP). Further results evidenced a recent fragmentation in the early Holocene period.

IV. **Neutral genetic variation was evidenced at candidate gene loci indicating long-term maintenance of genetic diversity in peripheral Scots pine populations.** Signals of negative selection (balancing/purifying) were detected by neutrality tests. Outlier SNP analysis presented that the loci under selection are prone to maintain genetic diversity and counteract the effect of genetic erosion in the studied natural Scots pine populations.

8. ACKNOWLEDGEMENT

There are persons, institutions and funding sources who helped my work and whom I am greatly indebted.

First of all, I wish to thank the members of my dissertation committee for their constructive criticism and direction during the process of this dissertation.

Most of all, I would like to thank my supervisor Dr. Mária Höhn, for her expertise, assistance, guidance, and patience throughout the doctoral (Ph.D.) course. Without her help this thesis would not have been possible.

I am also indebted to the Department of Genetics and Plant Breeding staff, Szent István University, for their support and cooperation. Notably, I would like to mention and thank Dr. Pedryc Andrzej (†) for his professional support at the early stage of the research project. I will be forever grateful for his assistance.

I would like to thank the staff of the Institute of Biosciences and Bioresources (IBBR-CNR, Italy), especially Dr. Giovanni G. Vendramin and Dr. Francesca Bagnoli for their hospitality during my Erasmus grant in Italy. Furthermore, I'm grateful for financing some important steps of my research and for their valuable assistance and cooperation.

I would like to extend my sincere gratitude to the Forest Research Institute (ERTI) - NAIK, particularly like to acknowledge Dr. Klára Cseke and Dr. László Nagy for the the exhaustive laboratory work and professional support.

I would like to give special thanks to Dr. Tamás Pócs (Eszterházy Károly University), Dr. Ivan Iliev (University of Forestry, Bulgaria) and Dr. Tibor Baranec (Slovak University of Agriculture, Slovakia) for the assistance they offered with the collection of plant material and for the valuable discussions in the course of which they shared their insights.

I would also like to acknowledge the National Research, Development and Innovation Office (Hungary) for the grant of the Hungarian Scientific Research Fund [OTKA K101600].

Last but not least, I would like to thank my family, friends and everyone else who helped me to contribute in this project. I would especially like to thank my fiancée Nóra Pap. Thank you for accompanying me everyday.

9. REFERENCES

Adams, J. M. (1997). Global land environments since the last interglacial. Oak Ridge National Laboratory, TN, USA. Available at: <http://www.esd.ornl.gov/ern/qen/nerc.html> [Accessed: 10.04.2016].

Alsos, I. G., Alm, T., Normand, S., & Brochmann, C. (2009). Past and future range shifts and loss of diversity in dwarf willow (*Salix herbacea* L.) inferred from genetics, fossils and modelling. *Global Ecology and Biogeography*, 18(2), 223-239.

Androsiuk, P., Zielinski, R., & Polok, K. (2011). B-SAP markers derived from the bacterial KatG gene differentiate populations of *Pinus sylvestris* and provide new insights into their postglacial history. *Silva Fennica*, 45(1), 3-18.

Antao, T., Lopes, A., Lopes, R. J., Beja-Pereira, A., & Luikart, G. (2008). LOSITAN: a workbench to detect molecular adaptation based on a F_{ST}-outlier method. *BMC bioinformatics*, 9(1), 323.

Barton, N. H. (2001). Adaptation at the edge of a species' range. *Special Publication-British Ecological Society*, 14, 365-392.

Beaumont, M. A., & Nichols, R. A. (1996). Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London B: Biological Sciences*, 263(1377), 1619-1626.

Belletti, P., Ferrazzini, D., Piotti, A., Monteleone, I., & Ducci, F. (2012). Genetic variation and divergence in Scots pine (*Pinus sylvestris* L.) within its natural range in Italy. *European Journal of Forest Research*, 131(4), 1127-1138.

Bennett, K. D. (1984). The post-glacial history of *Pinus sylvestris* in the British Isles. *Quaternary Science Reviews*, 3(2-3), 133-155.

Bennett, K. D. (1995). Post-glacial dynamics of pine (*Pinus sylvestris* L.) and pinewoods in Scotland. *Our pinewood heritage*, 23-39.

Bennett, K. D., Tzedakis, P. C., & Willis, K. J. (1991). Quaternary refugia of north European trees. *Journal of Biogeography*, 103-115.

Bernhardsson, C., Floran, V., Ganea, S. L., & García-Gil, M. R. (2016). Present genetic structure is congruent with the common origin of distant Scots pine populations in its Romanian distribution. *Forest Ecology and Management*, 361, 131-143.

Betancourt, J. L., Schuster, W. S., Mitton, J. B., & Anderson, R. S. (1991). Fossil and genetic history of a pinyon pine (*Pinus edulis*) isolate. *Ecology*, 72(5), 1685-1697.

Biere, A., & Verhoeven, K. J. (2008). Local adaptation and the consequences of being dislocated from coevolved enemies. *New Phytologist*, 180(2), 265-268.

Birks, H. H., & Ammann, B. (2000). Two terrestrial records of rapid climatic change during the glacial–Holocene transition (14,000–9,000 calendar years BP) from Europe. *Proceedings of the National Academy of Sciences*, 97(4), 1390-1394.

Birks, H. J. B., & Birks, H. H. (1980). *Quaternary palaeoecology*. London: Edward Arnold. p. 289.

Birks, H. J. B., & Williams, W. (1983). Late-Quaternary vegetational history of the Inner Hebrides. *Proceedings of the Royal Society of Edinburgh. Section B. Biological Sciences*, 83, 269-292.

Birky, C. W. (1995). Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proceedings of the National Academy of Sciences*, 92(25), 11331-11338.

Bock, R., Knoop, V. (2012). *Genomics of chloroplasts and mitochondria (Vol. 35)*. Springer Science & Business Media. pp. 2-21.

Bojnanský, V., & Fargašová, A. (2007). *Atlas of seeds and fruits of Central and East-European flora: the Carpathian Mountains region.* Springer Science & Business Media. pp. 11-12.

Bottema, S. (1974). Late Quaternary vegetation history of northwestern Greece. Thesis, Groningen University, Netherlands. pp. 1-190.

Brown, G. R., Gill, G. P., Kuntz, R. J., Langley, C. H., & Neale, D. B. (2004). Nucleotide diversity and linkage disequilibrium in loblolly pine. *Proceedings of the National Academy of Sciences of the United States of America*, 101(42), 15255-15260.

Bucci, G., & Vendramin, G. G. (2000). Delineation of genetic zones in the European Norway spruce natural range: preliminary evidence. *Molecular Ecology*, 9(7), 923-934.

Buchovska, J., Danusevicius, D., Baniulis, D., Stanys, V. I., Siksnianiene, J. B., & Kavaliauskas, D. (2013). The location of the northern glacial refugium of Scots Pine based on mitochondrial DNA markers. *Baltic Forestry*, 19(1), 2-12.

Busch, J. D., Waser, P. M., & De Woody, J. A. (2007). Recent demographic bottlenecks are not accompanied by a genetic signature in banner- tailed kangaroo rats (*Dipodomys spectabilis*). *Molecular Ecology*, 16(12), 2450-2462.

Callaghan, T. V., Carlsson, B. Å., & Svensson, B. M. (1996). Some apparently paradoxical aspects of the life cycles, demography and population dynamics of plants from the subarctic Abisko area. *Ecological Bulletins*, 133-143.

Carlisle, A. (1958). A guide to the named variants of Scots pine (*Pinus silvestris* Linnaeus). *Forestry*, 31(2), 203-224.

Carson, H. L. (1959). Genetic conditions which promote or retard the formation of species. *Cold Spring Harbor Symposia on Quantitative Biology*, 24, 87-103.

Čelepirović, N., Ivanković, M., Gradečki-Poštenjak, M., Nagy, L., Borovics, A., Agbaba, S. N., & Littvay, T. (2009). Review of investigation of variability of nad1 gene intron B/C of mitochondrial genome in Scots pine (*Pinus sylvestris* L.). *Periodicum Biologorum*, 111(4), 453-457.

Channell, R. (2004). The conservation value of peripheral populations: the supporting science. *Species at Risk*, 1-17.

Cheddadi, R., Vendramin, G. G., Litt, T., François, 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(3), 271-282.

Cipriano, J., Carvalho, A., Fernandes, C., Gaspar, M. J., Pires, J., Bento, J., Roxo, L., Louzada, J., & Lima-Brito, J. (2013). Evaluation of genetic diversity of Portuguese *Pinus sylvestris* L. populations based on molecular data and inferences about the future use of this germplasm. *Journal of Genetics*, 92: 41–48.

Clement, M. J., Snell, Q., Walker, P., Posada, D., & Crandall, K. A. (2002). TCS: estimating gene genealogies. *Parallel and Distributed Processing Symposium, International Proceedings*, 2, 184.

Cohen, K. M., Finney, S. C., Gibbard, P. L., & Fan, J. X. (2013). The ICS international chronostratigraphic chart. *Episodes*, 36(3), 199-204.

Cook, L. M. (1961). The edge effect in population genetics. *The American Naturalist*, 95(884), 295-307.

Corander, J., & Marttinen, P. (2006). Bayesian identification of admixture events using multi-locus molecular markers. *Molecular Ecology*, 15(10), 2833-2843.

Corander, J., Marttinen, P., Sirén, J., & Tang, J. (2008a). Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics*, 9(1), 539.

Corander, J., Sirén, J., & Arjas, E. (2008b). Bayesian spatial modelling of genetic population structure. *Computational Statistics*, 23(1), 111-129.

Cornuet, J. M., & Luikart, G. (1996). Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, 144(4), 2001-2014.

Cornuet, J. M., Pudlo, P., Veyssier, J., Dehne-Garcia, A., Gautier, M., Leblois, R., Marin, J. M., & Estoup, A. (2014). DIYABC v2. 0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics*, 30(8), 1187-1189.

Csató, J., (1885). A Mluha nevű tó (Teu Mluhi) és viránya. [in English: Plants of lake Mluha (Teu Mluhi)] *Magyar Botanikai Lapok*, 9, 1-8.

Csilléry, K., Lalagüe, H., Vendramin, G. G., González- Martínez, S. C., Fady, B., & Oddou-Muratorio, S. (2014). Detecting short spatial scale local adaptation and epistatic selection in climate- related candidate genes in European beech (*Fagus sylvatica*) populations. *Molecular Ecology*, 23(19), 4696-4708.

Damblon, F. (1997). Palaeobotanical study of representative upper palaeolithic sites in the central European plain: a contribution to the SC-004 Project. *Préhistoire européenne*, 11, 245-254.

Debreczy, Zs., & Rácz, I. (2000). *Fenyők a föld körül*. Budapest, Dendrológiai alapítvány Kft. pp. 157-158.

Debreczy, Zs., Rácz, I., & Musia, K. (2011). *Conifers around the world*. Budapest, Dendropress. pp. 21-24.

Di Rienzo, A., Peterson, A. C., Garza, J. C., Valdes, A. M., Slatkin, M., & Freimer, N. B. (1994). Mutational processes of simple-sequence repeat loci in human populations. *Proceedings of the National Academy of Sciences*, 91(8), 3166-3170.

Dieringer, D., & Schlötterer, C. (2003). Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes*, 3(1), 167-169.

Dvornyk, V. (2001). Genetic variability and differentiation of geographically marginal Scots pine populations from Ukraine. *Silvae Genetica*, 50(2), 64-68.

Dvornyk, V., Sirviö, 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.

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

Earl, D. A. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation genetics resources*, 4(2), 359-361.

Eckert, A. J., Wegrzyn, J. L., Pande, B., Jermstad, K. D., Lee, J. M., Liechty, J. D., Tearse, B. R., Krutovsky, K. V., & Neale, D. B. (2009). Multilocus patterns of nucleotide diversity and divergence reveal positive selection at candidate genes related to cold hardiness in coastal Douglas fir (*Pseudotsuga menziesii var. menziesii*). *Genetics*, 183(1), 289-298.

Eckert, C. G., Samis, K. E., & Lougheed, S. C. (2008). Genetic variation across species' geographical ranges: the central–marginal hypothesis and beyond. *Molecular Ecology*, 17(5), 1170-1188.

Ehrich, D., Alsos, I. G., & Brochmann, C. (2008). Where did the northern peatland species survive the dry glacials: cloudberry (*Rubus chamaemorus*) as an example. *Journal of Biogeography*, 35(5), 801-814.

Ehrich, D., Gaudeul, M., Assefa, A., Koch, M. A., Mummenhoff, K., Nemomissa, S., Intrabiodiv Consortium, & Brochmann, C. (2007). Genetic consequences of Pleistocene range shifts: contrast between the Arctic, the Alps and the East African mountains. *Molecular Ecology*, 16(12), 2542-2559.

El Mousadik, A., & Petit, R. J. (1996). High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. *Theoretical and Applied Genetics*, 92(7), 832-839.

Eliades, N. G., Eliades, D.G. (2009). Haplotype Analysis: software for analysis of haplotype data. Georg-August University Göttingen, Germany. Available at: http://www.uni-goettingen.de/en/134935.html. [Accessed: 05.10.2016].

Ellstrand, N. C., & Elam, D. R. (1993). Population genetic consequences of small population size: implications for plant conservation. *Annual review of Ecology and Systematics*, 24(1), 217-242.

Elsik, C. G., & Williams, C. G. (2001). Low-copy microsatellite recovery from a conifer genome. *Theoretical and Applied Genetics*, 103(8), 1189-1195.

Elsik, C. G., Minihan, V. T., Hall, S. E., Scarpa, A. M., & Williams, C. G. (2000). Low-copy microsatellite markers for *Pinus taeda* L. *Genome*, 43(3), 550-555.

Eronen, M. (1979). The retreat of pine forest in Finnish Lapland since the Holocene climatic optimum: a general discussion with radiocarbon evidence from subfossil pines. *Fennia*, 157(2), 93-114.

Eronen, M., & Hyvärinen, H. (1982). Subfossil pine dates and pollen diagrams from northern Fennoscandia. *Geologiska Foreningens i Stockholm Forhandlingar*, 103(4), 437-445.

ESBN-EC (European Commission), (2005). *Soil Atlas of Europe. European Soil Bureau Network. European Commission.* Office for Official Publications of the European Communities, L-2995 Luxembourg 92-894-8120-X. Catalogue Number LB-37-01-744-EN-C. p. 128.

Estoup, A., Jarne, P., & Cornuet, J. M. (2002). Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology*, 11(9), 1591-1604.

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, 14(8), 2611-2620.

Excoffier, L., & Lischer, H. E. (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.

Excoffier, L., Hofer, T., & Foll, M. (2009). Detecting loci under selection in a hierarchically structured population. *Heredity*, 103(4), 285-298.

Fahrig, L., & Merriam, G. (1994). Conservation of fragmented populations. *Conservation Biology*, 8(1), 50-59.

Farcas, S., de Beaulieu, J. L., Reille, M., Coldea, G., Diaconeasa, B., Goeury, C., Goslar, T., & Jull, T. (1999). First 14 C 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(9), 799-807.

Farjon, A. (1998). *World checklist and bibliography of conifers*. Kew, Royal Botanic Gardens. pp. 1-289.

Farjon, A. (2001). *World checklist and bibliography of conifers*. (Second Edition), Kew, Royal Botanic Gardens. pp. 1-309.

Fazekas, A. J., & Yeh, F. C. (2001). Random amplified polymorphic DNA diversity of marginal and central populations in *Pinus contorta* subsp. *latifolia*. *Genome*, 44(1), 13-22.

Fekete, G., Molnár, Z., Magyari, E., Somodi, I., & Varga, Z. (2014). A new framework for understanding Pannonian vegetation patterns: regularities, deviations and uniqueness. *Community Ecology*, 15(1), 12-26.

Fekete, L., Blattny, T. (1913). Az erdészeti jelentőségú fák és cserjék elterjedése a Magyar Állam területén (The distribution of forest trees and shrubs in Hungary). *Erdészeti Lapok*, 52:743-758.

Félegyházi, E., & Tóth, C. (2003). A Halas-fenék lefűzodött medermaradvány üledékanyagának szedimentológiai, mikromineralógiai és palinológiai vizsgálata (Sedimentological and palynological study of the abandoned channel remnant deposits of the Halas-fenék). *Acta Geographica Debrecina*, 36, 21-30.

Feurdean, A., & Bennike, O. (2004). Late Quaternary palaeoecological and palaeoclimatological reconstruction in the Gutaiului Mountains, northwest Romania. *Journal of Quaternary Science*, 19(8), 809-827.

Feurdean, A., Björkman, L., & Wohlfarth, B. (2012). A paleoecological reconstruction of the Late Glacial and Holocene based on multidisciplinary studies at Steregoiu site (Gutai Mts., Romania). *Studia UBB Geologia*, 46(2), 125-140.

Feurdean, A., Tanţău, I., & Fărcaş, S. (2011). Holocene variability in the range distribution and abundance of *Pinus*, *Picea abies*, and *Quercus* in Romania; implications for their current status. *Quaternary Science Reviews*, 30(21), 3060-3075.

Feurdean, A., Wohlfarth, B., Björkman, L., Tantau, I., Bennike, O., Willis, K. J., Farcas, S., & Robertsson, A. M. (2007). The influence of refugial population on Lateglacial and early Holocene vegetational changes in Romania. *Review of Palaeobotany and Palynology*, 145(3), 305-320.

Fisher, R. A. (1930). *The genetical theory of natural selection: a complete variorum edition*. Oxford, Oxford University Press. pp. 1-121.

Fisher, R. A. (1950). Gene frequencies in a cline determined by selection and diffusion. *Biometrics*, 6(4), 353-361.

Floran, V., Sestras, R. E., & GIL, M. R. G. (2011). Organelle genetic diversity and phylogeography of Scots pine (*Pinus sylvestris* L.). *Notulae Botanicae Horti Agrobotanici Cluj-Napoca*, 39(1), 317.

Foll, M. (2012). Bayescan v2. 1 user manual. *Ecology*, 20, 1450-1462.

Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, 180(2), 977-993.

Forrest, I., Burg, K., & Klumpp, R. (2000). Genetic markers: tools for identifying and characterising Scots pine populations. *Forest Systems*, 9(1), 67-88.

Francis, R. M. (2016). Pophelper: an R package and web app to analyse and visualize population structure. *Molecular Ecology Resources*. DOI:10.1111/1755-0998.12509.

Frankel, O. H., Brown, A. H., & Burdon, J. J. (1995). *The conservation of plant biodiversity*. Cambridge, Cambridge University Press. pp. 126-147.

Fu, Y.-X., & W.-H. Li. (1993). Statistical tests of neutrality of mutations. *Genetics*, 133:693-709.

Furnier, G. R., & Adams, W. T. (1986). Geographic patterns of allozyme variation in Jeffrey pine. *American Journal of Botany*, 1009-1015.

Ganea, S., Ranade, S. S., Hall, D., Abrahamsson, S., & García-Gil, M. R. (2015). Development and transferability of two multiplexes nSSR in Scots pine (*Pinus sylvestris* L.). *Journal of Forestry Research*, 26(2), 361-368.

Garcia-Ramos, G., & Kirkpatrick, M. (1997). Genetic models of adaptation and gene flow in peripheral populations. *Evolution*, 21-28.

Gencsi, L., & Vancsura, R. (1992). *Dendrológia*. Budapest, Mezőgazda Kiadó. p. 728.

Georgescu, G. G., Donita, L., (1965). La division floristique de Roumanie. *I. Revue Roumaine de Biologie, Série Botanique*, 10, 357-369.

Gernandt, D. S., López, G. G., García, S. O., & Liston, A. (2005). Phylogeny and classification of *Pinus*. *Taxon*, 54(1), 29-42.

Giertych, M., & Mátyás, C. (1991). *Genetics of Scots pine*. Amsterdam, Elsevier. 3:19-76.

Gilpin, M. (1991). The genetic effective size of a metapopulation. *Biological Journal of the Linnean Society*, 42(1- 2), 165-175.

Goldstein, D. B., Linares, A. R., Cavalli-Sforza, L. L., & Feldman, M. W. (1995). An evaluation of genetic distances for use with microsatellite loci. *Genetics*, 139(1), 463-471.

Gömöry, D., Paule, L., Krajmerová, D., Romšáková, I., & Longauer, R. (2012). Admixture of genetic lineages of different glacial origin: a case study of *Abies alba* Mill. in the Carpathians. *Plant Systematics and Evolution*, 298(4), 703-712.

González-Martínez, S. C., Dubreuil, M., Riba, M., Vendramin, G. G., Sebastiani, F., & Mayol, M. (2010). Spatial genetic structure of *Taxus baccata* L. in the western Mediterranean Basin: past and present limits to gene movement over a broad geographic scale. *Molecular Phylogenetics and Evolution*, 55(3), 805-815.

González-Martínez, S. C., Ersoz, E., Brown, G. R., Wheeler, N. C., & Neale, D. B. (2006). DNA sequence variation and selection of tag single-nucleotide polymorphisms at candidate genes for drought-stress response in *Pinus taeda* L. *Genetics*, 172(3), 1915-1926.

González-Martínez, S. C., Robledo-Arnuncio, J. J., Collada, C., Díaz, A., Williams, C. G., Alía, R., & Cervera, M. T. (2004). Cross-amplification and sequence variation of microsatellite loci in Eurasian hard pines. *Theoretical and Applied Genetics*, 109(1), 103-111.

Grivet, D., Sebastiani, F., Alía, R., Bataillon, T., Torre, S., Zabal-Aguirre, M., Vendramin G. G., & González-Martínez, S. C. (2011). Molecular footprints of local adaptation in two Mediterranean conifers. *Molecular Biology and Evolution*, 28(1), 101-116.

Grivet, D., Sebastiani, F., González- Martínez, S. C., & Vendramin, G. G. (2009). Patterns of polymorphism resulting from long- range colonization in the Mediterranean conifer Aleppo pine. *New Phytologist*, 184(4), 1016-1028.

Grotkopp, E., Rejmánek, M., & Rost, T. L. (2002). Toward a causal explanation of plant invasiveness: seedling growth and life-history strategies of 29 pine (*Pinus*) species. *The American Naturalist*, 159(4), 396-419.

Guo, S. W., & Thompson, E. A. (1992). Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*, 48, 361-372.

Guries, R. P., & Ledig, F. T. (1982). Genetic diversity and population structure in pitch pine (*Pinus rigida* Mill.). *Evolution*, 36, 387-402.

Haesaerts, P., Damblon, F., Bachner, M., & Trnka, G. (1996). Revised stratigraphy and chronology of the Willendorf II sequence, Lower Austria. *Archaeologia Austriaca*, 80, 25-42.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, 41(41), 95-98).

Hampe, A., & Petit, R. J. (2005). Conserving biodiversity under climate change: the rear edge matters. *Ecology Letters*, 8(5), 461-467.

Hengeveld, R., & Haeck, J. (1982). The distribution of abundance. I. Measurements. *Journal of Biogeography*, 303-316.

Henry, P., & Russello, M. A. (2013). Adaptive divergence along environmental gradients in a climate- change- sensitive mammal. *Ecology and Evolution*, *3*(11), 3906-3917.

Heuertz, M., De Paoli, E., Källman, T., Larsson, H., Jurman, I., Morgante, M., Lascoux, M., & Gyllenstrand, N. (2006). Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway spruce [*Picea abies* (L.) Karst]. *Genetics*, 174(4), 2095-2105.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25(15), 1965-1978.

Hill, W. G., & Robertson, A. (1968). Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics*, 38(6), 226-231.

Hipkins, V. D., Krutovskii, K. V., Strauss, S. H., (1994). Organelle genomes in conifers: structure, evolution, and diversity. *Forest Genetics*, 1, 179-189.

Hoffmann, A. A. & Parsons, P. A. (1997). *Extreme Environmental Change and Evolution.* Cambridge, Cambridge University Press. pp. 1-272.

Hoffmann, A. A., & Blows, M. W. (1994). Species borders: ecological and evolutionary perspectives. *Trends in Ecology & Evolution*, 9(6), 223-227.

Höhn, M., Ábrán, P., & Vendramin, G. G. (2005). Genetic analysis of Swiss stone pine populations (*Pinus cembra* L. subsp. *cembra*) from the Carpathians using chloroplast microsatellites. *Acta Silvatica et Lignaria Hungarica*, 1, 39-47.

Höhn, M., Gugerli, F., Abran, P., Bisztray, G., Buonamici, A., Cseke, K., Hufnagel, L., Quintela-Sabarís, C., Sebastiani, F., & Vendramin, G. G. (2009). Variation in the chloroplast DNA of Swiss stone pine (*Pinus cembra* L.) reflects contrasting post-glacial history of populations from the Carpathians and the Alps. *Journal of Biogeography*, 36(9), 1798-1806.

Höhn, M., Hufnagel, L., Cseke, K., & Vendramin, G. (2010). Current range characteristics of Swiss stone pine (*Pinus cembra* L.) along the Carpathians revealed by chloroplast SSR markers. *Acta Biologica Hungarica*, 61(1), 61-67.

Holt, R. D. (2003). On the evolutionary ecology of species' ranges. *Evolutionary Ecology Research*, 5(2), 159-178.

Holt, R. D., & Gomulkiewicz, R. (1997). How does immigration influence local adaptation? A reexamination of a familiar paradigm. *The American Naturalist*, 149(3), 563-572.

Homolka, A., Schueler, S., Burg, K., Fluch, S., & Kremer, A. (2013). Insights into drought adaptation of two European oak species revealed by nucleotide diversity of candidate genes. *Tree Genetics & Genomes*, 9(5), 1179-1192.

Hubisz, M. J., Falush, D., Stephens, M., & Pritchard, J. K. (2009). Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, 9(5), 1322-1332.

Huntley B, Birks H. J. B. (1983). *An atlas of past and present pollen maps for Europe, 0-13,000 years ago.* Cambridge, Cambridge University Press. pp. 1-688.

Huttunen, A., Huttunen, R. L., Vasari, Y., Panovska, H., & Bozilova, E. (1992). Late-glacial and Holocene history of flora and vegetation in the Western Rhodopes Mountains, Bulgaria. *Acta Botanica Fennica*, 144, 63-80.

Hytteborn, H., Maslov, A. A., Nazimova, D. I., Rysin, L. P. (2005). Boreal forests of Eurasia. (*In*: Andersson, F. (Ed) *Ecosystems of the World 6. Coniferous Forests*.) Elsevier. Amsterdam. Netherland. pp. 23-99.

Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23(14), 1801-1806.

Jankovska, V., & Pokorný, P. (2008). Forest vegetation of the last full-glacial period in the Western Carpathians (Slovakia and Czech Republic). *Preslia*, 80(3), 307-324.

Jasińska, A. K., Boratyńska, K., Dering, M., Sobierajska, K. I., Ok, T., Romo, A., & Boratyński, A. (2014). Distance between south-European and south-west Asiatic refugial areas involved morphological differentiation: *Pinus sylvestris* case study. *Plant Systematics and Evolution*, 300(6), 1487-1502.

Jeffreys, H. (1961). *Theory of Probability*. Oxford, Oxford University Press. p. 432.

Jensen, J. L., Bohonak, A. J., & Kelley, S. T. (2005). Isolation by distance, web service. *BMC Genetics*, 6(1), 13.

Kawecki, T. J. (2003). Sex- biased dispersal and adaptation to marginal habitats. *The American Naturalist*, 162(4), 415-426.

Kawecki, T. J. (2008). Adaptation to marginal habitats. *Annual Review of Ecology, Evolution, and Systematics*, 321-342.

Kawecki, T. J., & Holt, R. D. (2002). Evolutionary consequences of asymmetric dispersal rates. *The American Naturalist*, 160(3), 333-347.

Kawecki, T. J., Barton, N. H., & Fry, J. D. (1997). Mutational collapse of fitness in marginal habitats and the evolution of ecological specialisation. *Journal of Evolutionary Biology*, 10(3), 407-429.

Kelly, D. L., & Connolly, A. (2000). A review of the plant communities associated with Scots pine (*Pinus sylvestris* L.) in Europe, and an evaluation of putative indicator/specialist species. *Forest Systems*, 9(1), 15-39.

Kieliszewska-Rokicka, B. (1981). Geographic and climatic differentiation of electrophoretic forms of esterase, glutamate dehydrogenase and peroxidase in Scots pine tissues. *Acta Societatis Botanicorum Poloniae*, 50(3), 481-492.

Kimura, M. (1983). *The neutral theory of molecular evolution*. Cambridge, Cambridge University Press. pp. 117-137.

Kinloch, B. B., Westfall, R. D., & Forrest, G. I. (1986). Caledonian Scots pine: origins and genetic structure. *New Phytologist*, 104(4), 703-729.

Klimko, M., & Bykowska, J. (2015). Pollen morphology of *Pinus uliginosa* Neum. and *Pinus uncinata* Ramond ex DC in an experimental culture. *Steciana*, 19(3).

Kormut'ák, A., Camek, V., Branná, M., Celková, M., Vooková, B., Manka, P., Galgóci, M., Bolecek, P., & Gömöry, D. (2013). Introgressive hybridization between Scots pine and mountain dwarf pine at two localities of northern Slovakia. *Folia Oecologica*, 40(2), 201.

Korpelainen, H. (2004). The evolutionary processes of mitochondrial and chloroplast genomes differ from those of nuclear genomes. *Naturwissenschaften*, 91(11), 505-518.

Kostia, S., Varvio, S. L., Vakkari, P., & Pulkkinen, P. (1995). Microsatellite sequences in a conifer, *Pinus sylvestris*. *Genome*, 38(6), 1244-1248.

Kremenetsky, C. V., Tarasov, P. E., & Cherkinski, A. E. (1994). A history of the Kazakhstan "island" pine forests in Holocene. *Botanicheskiy Zhurnal*, 79(3), 13-29.

Kreyling, J., Buhk, C., Backhaus, S., Hallinger, M., Huber, G., Huber, L., Jentsch, A., Konnert, M., Thiel, D., Wilmking, M., & Beierkuhnlein, C. (2014). Local adaptations to frost in marginal and central populations of the dominant forest tree *Fagus sylvatica* L. as affected by temperature and extreme drought in common garden experiments. *Ecology and Evolution*, 4(5), 594-605.

Krutovsky, K. V., & Neale, D. B. (2005). Nucleotide diversity and linkage disequilibrium in cold-hardiness-and wood quality-related candidate genes in Douglas fir. *Genetics*, 171(4), 2029-2041.

Kujala, S. T., & 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.

Kullman, L., & Kjällgren, L. (2006). Holocene pine tree- line evolution in the Swedish Scandes: Recent tree- line rise and climate change in a long- term perspective. *Boreas*, 35(1), 159-168.

Kyrkjeeide, M. O., Stenøien, H. K., Flatberg, K. I., & Hassel, K. (2014). Glacial refugia and post-glacial colonization patterns in European bryophytes. *Lindbergia*, 37, 47-59.

Labra, M., Grassi, F., Sgorbati, S., & Ferrari, C. (2006). Distribution of genetic variability in southern populations of Scots pine (*Pinus sylvestris* L.) from the Alps to the Apennines. *Flora-Morphology, Distribution, Functional Ecology of Plants*, 201(6), 468-476.

Lagercrantz, U., & Ryman, N. (1990). Genetic structure of Norway spruce (*Picea abies*): concordance of morphological and allozymic variation. *Evolution*, 38-53.

Lalagüe, H., Csilléry, K., Oddou-Muratorio, S., Safrana, J., de Quattro, C., Fady, B., González-Martínez, S. C., & Vendramin, G. G. (2014). Nucleotide diversity and linkage disequilibrium at 58 stress response and phenology candidate genes in a European beech (*Fagus sylvatica* L.) population from southeastern France. *Tree Genetics & Genomes*, 10(1), 15-26.

Lande, R. (1992). Neutral theory of quantitative genetic variance in an island model with local extinction and colonization. *Evolution*, 381-389.

Latałowa, M., & van der Knaap, W. O. (2006). Late Quaternary expansion of Norway spruce *Picea abies* (L.) Karst. in Europe according to pollen data. *Quaternary Science Reviews*, 25(21), 2780-2805.

Laurentin, H. (2009). Data analysis for molecular characterization of plant genetic resources. *Genetic Resources and Crop Evolution*, 56(2), 277-292.

Leigh, J. W., & Bryant, D. (2015). Popart: full- feature software for haplotype network construction. *Methods in Ecology and Evolution*, 6(9), 1110-1116.

Leimu, R., & Fischer, M. (2008). A meta-analysis of local adaptation in plants. *PloS One*, 3(12), e4010.

Lendvay, B., Höhn, M., Brodbeck, S., Mîndrescu, M., & Gugerli, F. (2014). Genetic structure in *Pinus cembra* from the Carpathian Mountains inferred from nuclear and chloroplast microsatellites confirms post-glacial range contraction and identifies introduced individuals. *Tree Genetics & Genomes*, 10(5), 1419-1433.

Lenormand, T. (2002). Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*, 17(4), 183-189.

Lenormand, T., & Raymond, M. (1998). Resistance management: the stable zone strategy. *Proceedings of the Royal Society of London B: Biological Sciences*, 265(1409), 1985-1990.

Li, P., & Adams, W. T. (1989). Range-wide patterns of allozyme variation in Douglas-fir (*Pseudotsuga menziesii*). *Canadian Journal of Forest Research*, 19(2), 149-161.

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

Liepelt, S., Cheddadi, R., de Beaulieu, J. L., Fady, B., Gömöry, D., Hussendörfer, E., Konnert, M., Litt, T., Longauer, R., Terhürne-Berson, R., & Ziegenhagen, B. (2009). Postglacial range expansion and its genetic imprints in *Abies alba* (Mill.)—a synthesis from palaeobotanic and genetic data. *Review of Palaeobotany and Palynology*, 153(1), 139-149.

Linhart, Y. B., & Grant, M. C. (1996). Evolutionary significance of local genetic differentiation in plants. *Annual Review of Ecology and Systematics*, 237-277.

Lowe, J. (1992). Late glacial and early Holocene lake sediments from the northern Apennines, Italy–Pollen stratigraphy and radiocarbon dating. *Boreas*, 21(3), 193-208.

Lučić, A., Popović, V., Nevenić, M., Ristić, D., Rakonjac, L., Ćirković-Mitrović, T., & Mladenović-Drinić, S. (2014). Genetic diversity of Scots pine (*Pinus sylvestris* L.) populations in Serbia revealed by SSR markers. *Archives of Biological Sciences*, 66(4), 1485-1492.

Luikart, G., & Cornuet, J. M. (1998). Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. *Conservation Biology*, 12(1), 228-237.

Luikart, G., Allendorf, F. W., Cornuet, J. M., & Sherwin, W. B. (1998). Distortion of allele frequency distributions provides a test for recent population bottlenecks. *Journal of Heredity*, 89(3), 238-247.

Luikart, G., England, P. R., Tallmon, D., Jordan, S., & Taberlet, P. (2003). The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, 4(12), 981-994.

Magyari, E. K. (2011). Late Quaternary vegetation history in the Hortobágy steppe and middle Tisza floodplain, NE Hungary. *Studia Botanica Hungarica*, 42, 185-203.

Magyari, E. K., Jakab, G., Bálint, M., Kern, Z., Buczkó, K., & Braun, M. (2012). Rapid vegetation response to Lateglacial and early Holocene climatic fluctuation in the South Carpathian Mountains (Romania). *Quaternary Science Reviews*, 35, 116-130.

Magyari, E. K., Kuneš, P., Jakab, G., Sümegi, P., Pelánková, B., Schäbitz, F., Braun, M., & Chytrý, M. (2014a). Late Pleniglacial vegetation in eastern-central Europe: are there modern analogues in Siberia? *Quaternary Science Reviews*, 95, 60-79.

Magyari, E. K., Veres, D., Wennrich, V., Wagner, B., Braun, M., Jakab, G., Karatson, D., Pal, Z., Ferenczy, G., St-Onge, G., Rethemeyer, J., Francois, J. P., von Reumont, F., Schabitz, F. (2014b). Vegetation and environmental responses to climate forcing during the Last Glacial Maximum and deglaciation in the East Carpathians: attenuated response to maximum cooling and increased biomass burning. *Quaternary Science Reviews*, 106: 278-298.

Manni, F., Guerard, E., & Heyer, E. (2004). Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by using Monmonier's algorithm. *Human Biology*, 76(2), 173-190.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, 27(2 Part 1), 209-220.

Matías, L., & Jump, A. S. (2012). Interactions between growth, demography and biotic interactions in determining species range limits in a warming world: the case of *Pinus sylvestris*. *Forest Ecology and Management*, 282, 10-22.

Mátyás, Cs. (2002). *Erdészeti - természetvédelmi genetika*. Budapest, Mezőgazdasági Kiadó. pp. 356-362.

Mayr, E. (1966). *Animal species and evolution*. Cambridge, Massachusetts, Harvard University Press. pp. 297-481.

Mayr, E. (1970). *Populations, species, and evolution: an abridgment of animal species and evolution*. Cambridge, Massachusetts, Harvard University Press. pp. 106-331.

McGeever, A. H., & Mitchell, F. J. (2016). Re- defining the natural range of Scots Pine (*Pinus sylvestris* L.): a newly discovered microrefugium in western Ireland. *Journal of Biogeography*, 43(11), 2199-2208.

Mejnartowicz, L. (1979). Genetic variation in some isoenzyme loci in Scots pine (*Pinus sylvestris* L.) populations. *Arboretum Kornickie*, 24(9), 1-1.

Mihai, B., Savulescu, I., & Sandric, I. (2007). Change detection analysis (1986–2002) of vegetation cover in Romania: a study of alpine, subalpine, and forest landscapes in the Iezer Mountains, Southern Carpathians. *Mountain Research and Development*, 27(3), 250-258.

Millar, C. I. (1993). Impact of the Eocene on the evolution of *Pinus* L. *Annals of the Missouri Botanical Garden*, 471-498.

Millar, C. I. (1998). Early evolution of pines. (*In*: Richardson, D. M. (Ed), *Ecology and Biogeography of Pinus*.) Cambridge, Cambridge University Press, pp. 69-94.

Milner, A. M., Müller, U. C., Roucoux, K. H., Collier, R. E., Pross, J., Kalaitzidis, S., Christanis, K., & Tzedakis, P. C. (2013). Environmental variability during the Last Interglacial: a new high-resolution pollen record from Tenaghi Philippon, Greece. *Journal of Quaternary Science*, 28(2), 113-117.

Mirov, N. T. (1967). *The genus Pinus*. New York, The Ronald Press Company. pp. 1-602.

Mirov, N. T., & Stanley, R. G. (1959). The pine tree. *Annual Review of Plant Physiology*, 10(1): 223-238.

Morgante, M., Felice, N., Vendramin, G. G. (1998). Analysis of hypervariable chloroplast microsatellites in *Pinus halepensis* reveals a dramatic genetic bottleneck. (*In*: Karp, A., Isaac, P. G., Ingram, D. S. (Eds) *Molecular tools for screening biodiversity*, Springer Netherlands. pp. 407-412.

Mosca, E., Gugerli, F., Eckert, A. J., & Neale, D. B. (2016). Signatures of natural selection on *Pinus cembra* and *P. mugo* along elevational gradients in the Alps. *Tree Genetics & Genomes*, 12(1), 9.

Motta, R., Morales, M., & Nola, P. (2006). Human land-use, forest dynamics and tree growth at the treeline in the Western Italian Alps. *Annals of Forest Science*, 63(7), 739-747.

Mráz, P., Gaudeul, M., Rioux, D., Gielly, L., Choler, P., & Taberlet, P. (2007). Genetic structure of *Hypochaeris uniflora* (Asteraceae) suggests vicariance in the Carpathians and rapid post-glacial colonization of the Alps from an eastern Alpine refugium. *Journal of Biogeography*, 34(12), 2100-2114.

Müller, M., Seifert, S., & Finkeldey, R. (2015). A candidate gene-based association study reveals SNPs significantly associated with bud burst in European beech (*Fagus sylvatica* L.). *Tree Genetics & Genomes*, 11(6), 116.

Müller-Starck, G., Baradat, P., & Bergmann, F. (1992). Genetic variation within European tree species. *New Forests*, 6(1-4), 23-47.

Nagylaki, T. (1975). Conditions for the existence of clines. *Genetics*, 80(3), 595-615.

Narum, S. R., & Hess, J. E. (2011). Comparison of FST outlier tests for SNP loci under selection. *Molecular Ecology Resources*, 11(s1), 184-194.

Naydenov, K. D., Naydenov, M. K., Tremblay, F., Alexandrov, A., & Aubin-Fournier, L. D. (2011). Patterns of genetic diversity that result from bottlenecks in Scots Pine and the implications for local genetic conservation and management practices in Bulgaria. *New Forests*, 42(2), 179-193.

Naydenov, K. D., 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.

Naydenov, K. D., Tremblay, F. M., Alexandrov, A., & Fenton, N. J. (2005). Structure of *Pinus sylvestris* L. populations in Bulgaria revealed by chloroplast microsatellites and terpenes analysis: provenance tests. *Biochemical Systematics and Ecology*, 33(12), 1226-1245.

Nei, M. (1987). *Molecular evolutionary genetics*. New York, Columbia University Press. pp. 39-352.

Pacurar, I., Clapa, D., Sotropa, A. (2010). Research on Mluha peatland, Apuseni Mts. *Agricultura Stiinta si Practica*, 3(4), 75-76.

Parducci, L., Jørgensen, T., Tollefsrud, M. M., Elverland, E., Alm, T., Fontana, S. L., Bennett, K. D., Haile, J., Matetovici, I., Suyama, Y., Edwards, M. E., Andersen, K., Rasmussen, M., Boessenkool, S., Coissac, E., Brochmann, C., Taberlet, P., Houmark-Nielsen, M., Larsen, N. K., Orlando, L., Gilbert, M. T. P., Kjær, K. H., Alsos, I. G., Willerslev, E. (2012). Glacial survival of boreal trees in northern Scandinavia. *Science*, 335(6072), 1083-1086.

Pavia, I., Mengl, M., Gaspar, M. J., Carvalho, A., Heinze, B., & Lima-Brito, J. (2014). Preliminary evidences of two potential native populations of *Pinus sylvestris* L. in Portugal based on nuclear and chloroplast SSR markers. *Austrian Journal of Forest Science*, 131(1), 1-22.

Peakall, R. O. D., & Smouse, P. E. (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, 6(1), 288-295.

Peery, M. Z., Kirby, R., Reid, B. N., Stoelting, R., Doucet-Bëer, E., Robinson, S., Vásquez-Carrillo, C., Pauli, N. J., & Palsbøll, P. J. (2012). Reliability of genetic bottleneck tests for detecting recent population declines. *Molecular Ecology*, 21(14), 3403-3418.

Pérez- Figueroa, A., García- Pereira, M. J., Saura, M., Rolán- Alvarez, E., & Caballero, A. (2010). Comparing three different methods to detect selective loci using dominant markers. *Journal of Evolutionary Biology*, 23(10), 2267-2276.

Pèrez-Obiol, R., & Julià, R. (1994). Climatic change on the Iberian Peninsula recorded in a 30,000-yr pollen record from Lake Banyoles. *Quaternary Research*, 41(1), 91-98.

Petit, R. J., & Vendramin, G. G. (2007). Plant phylogeography based on organelle genes: an introduction. (*In:* Weiss, S., Nuno, F. (Eds) *Phylogeography of southern European refugia*.) Dordrecht, Springer Netherlands. pp. 23-97.

Petit, R. J., Duminil, J., Fineschi, S., Hampe, A., Salvini, D., & Vendramin, G. G. (2005). Invited review: comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Molecular Ecology*, 14(3), 689-701.

Petit, R. J., Hu, F. S., & Dick, C. W. (2008). Forests of the past: a window to future changes. *Science*, 320(5882), 1450-1452.

Petit, R. J., Kremer, A., & Wagner, D. B. (1993). Finite island model for organelle and nuclear genes in plants. *Heredity*, 71(6), 630-641.

Piry, S., Luikart, G., & Cornuet, J. M. (1999). BOTTLENECK: A computer program for detecting recent reductions in the effective population size suing allele frequency data. *Journal of Heredity*, 90(4), 502-503. DOI: 10.1093/jhered/90.4. 502.

Pividori, M., Giannetti, F., Barbati, A., Chirici, G. (2016). European Forest Types: tree species matrix. In: San-Miguel-Ayanz, J., de Rigo, D., Caudullo, G., Houston Durrant, T., Mauri, A. (Eds.), *European Atlas of Forest Tree Species*. Publ. Off. EU, Luxembourg, pp. e01f162+ (https://w3id.org/mtv/FISE-Comm/v01/e01aa69)

Plomion, C., Bousquet, J., & Kole, C. (2011). *Genetics, Genomics and Breeding of Conifers*. New York, CRC Press. pp. 13-263.

Pócs, T. (1960). Die zonalen Waldgesellschaften Südwestungarns. (The zonal forest communities of southwestern Hungary) *Acta Botanica Hungarica*, 6, 75-105.

Pot, D., McMillan, L., Echt, C., Le Provost, G., Garnier- Géré, P., Cato, S., & Plomion, C. (2005). Nucleotide variation in genes involved in wood formation in two pine species. *New Phytologist*, 167(1), 101-112.

Powell, W., Morgante, M., McDevitt, R., Vendramin, G. G., & Rafalski, J. A. (1995). Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines. *Proceedings of the National Academy of Sciences*, 92(17), 7759-7763.

Pravdin, L. F. (1969). *Scots pine: variation, intraspecific taxonomy and selection*. London, Annarbor Humphrey Science Publishers Ltd., p. 208.

Price, R. A., Liston, A., & Strauss, S. H. (1998). Phylogeny and systematics of Pinus. (*In*: Richardson, D. M. (Ed) *Ecology and biogeography of Pinus*. Cambridge, Cambridge University Press. pp. 49-68.

Pritchard, J. K., Stephens, M., Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155, 945-959.

Provan, J., Soranzo, N., Wilson, N. J., Goldstein, D. B., & Powell, W. (1999). A low mutation rate for chloroplast microsatellites. *Genetics*, 153(2), 943-947.

Provan, J., Soranzo, N., Wilson, N. J., McNicol, J. W., Forrest, G. I., 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 of London B: Biological Sciences, 265(1407), 1697-1705.

Prus-Głowacki, W., Stephan, B. R., Bujas, E., Alia, R., & Marciniak, A. (2003). Genetic differentiation of autochthonous populations of *Pinus sylvestris* (Pinaceae) from the Iberian Peninsula. *Plant Systematics and Evolution*, 239(1-2), 55-66.

Prus-Głowacki, W., Urbaniak, L., Bujas, E., Curtu, L. A. (2011). Genetic variation of isolated and peripheral populations of *Pinus sylvestris* (L.) from glacial refugia. *Flora,* 207: 150-158.

Pulliam, H. R. (2000). On the relationship between niche and distribution. *Ecology Letters*, 3(4), 349-361.

Pyhäjärvi, T., Garcia-Gil, M. R., 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, M. J., & Savolainen, O. (2008). Colonization routes of *Pinus sylvestris* inferred from distribution of mitochondrial DNA variation. *Tree Genetics & Genomes*, 4(2), 247-254.

Quinn, A. J., & Woodward, L. S. (2014). Carpathian Mountains. (*In*: Presley, L. (Ed) *Earth's Landscape: An Encyclopedia of the World's Geographic Features.*) Santa Barbara, ABC-CLIO. pp. 137-138.

Raijmann, L. E., Van Leeuwen, N. C., Kersten, R., Oostermeijer, J. G. B., Den Nijs, H., & Menken, S. B. (1994). Genetic variation and outcrossing rate in relation to population size in *Gentiana pneumonanthe* L. *Conservation Biology*, 8(4), 1014-1026.

Richardson, D. M. (1998). *Ecology and Biogeography of Pinus*. Cambridge, Cambridge University Press. pp. 3-118.

Robledo-Arnuncio, J. J., Alia, R., & Gil, L. (2004). High levels of genetic diversity in a long-term European glacial refugium of *Pinus sylvestris* L. *Forest Genetics.* 11: 239-248.

Robledo- Arnuncio, J. J., Collada, C., Alia, R., & Gil, L. (2005). Genetic structure of montane isolates of *Pinus sylvestris* L. in a Mediterranean refugial area. *Journal of Biogeography*, 32(4), 595-605.

Ronce, O., & Kirkpatrick, M. (2001). When sources become sinks: migrational meltdown in heterogeneous habitats. *Evolution*, 55(8), 1520-1531.

Ronikier, M. (2011). Biogeography of high-mountain plants in the Carpathians: An emerging phylogeographical perspective. *Taxon,* 60(2), 373-389.

Ronikier, M., & Zalewska-Gałosz, J. (2014). Independent evolutionary history between the Balkan ranges and more northerly mountains in *Campanula alpina* sl (Campanulaceae): genetic divergence and morphological segregation of taxa. *Taxon*, 63(1), 116-131.

Ronikier, M., Cieślak, E., & Korbecka, G. (2008). High genetic differentiation in the alpine plant *Campanula alpina* Jacq. (Campanulaceae): evidence for glacial survival in several Carpathian regions and long-term isolation between the Carpathians and the Alps. *Molecular Ecology*, 17(7), 1763-1775.

Rudner, E., Sümegi, P., Tóth, I., Beszeda, I., Hertelendi, E. (1995). The vegetation of the upper-Weichselian in the central and southern part of the Great Hungarian Plain. *7th European Ecological Congress.* p. 108.

Rudner, Z. E., & Sümegi, P. (2001). Recurring Taiga forest-steppe habitats in the Carpathian Basin in the Upper Weichselian. *Quaternary international*, 76, 177-189.

Rull, V. (2009). Microrefugia. *Journal of Biogeography*, 36(3), 481-484.

Rull, V. (2010). On microrefugia and cryptic refugia. *Journal of Biogeography*, 37(8), 1623-1625.

Sadori, L., Koutsodendris, A., Masi, A., Bertini, A., Combourieu-Nebout, N., Francke, A., Kouli, K., Joannin, S., Mercuri, A. M., Panagiotopoulos, K., Peyron, O., Torri, P., Wagner, B., Zanchetta, G., & Donders, T. H. (2015). Pollen-based paleoenvironmental and paleoclimatic change at Lake Ohrid (SE Europe) during the past 500 ka. *Biogeosciences Discussions*, 12(18), 15461-15493.

Safriel, U. N., Volis, S., & Kark, S. (1994). Core and peripheral populations and global climate change. *Israeli Journal of Plant Sciences*, 42, 331-45.

Sagarin, R. D., & Gaines, S. D. (2002). The "abundant centre" distribution: to what extent is it a biogeographical rule? *Ecology Letters*, 5(1), 137-147.

Sagarin, R. D., Gaines, S. D., & Gaylord, B. (2006). Moving beyond assumptions to understand abundance distributions across the ranges of species. *Trends in Ecology & Evolution*, 21(9), 524-530.

Sannikov, S. N., Petrova, I. V., Schweingruber, F., Egorov, E. V., & Parpan, T. V. (2011). Genetic differentiation of *Pinus mugo* Turra and *P. sylvestris* L. populations in the Ukrainian Carpathians and the Swiss Alps. *Russian Journal of Ecology*, 42(4), 270-276.

Savolainen, O., Kärkkäinen, K., & Kuittinen, H. (1992). Estimating numbers of embryonic lethals in conifers. *Heredity*, 69, 308-314.

Savolainen, O., Kujala, S. T., Sokol, C., Pyhäjärvi, T., Avia, K., Knürr, T., Kärkkäinen K., Hicks, S. (2011). Adaptive potential of northernmost tree populations to climate change, with emphasis on Scots pine (*Pinus sylvestris* L.). *Journal of Heredity*, 102(5), 526-536.

Scalfi, M., Mosca, E., Di Pierro, E. A., Troggio, M., Vendramin, G. G., Sperisen, C., La Porta, N., & Neale, D. B. (2014). Micro-and macro-geographic scale effect on the molecular imprint of selection and adaptation in Norway spruce. *PloS one*, 9(12), e115499.

Scalfi, M., Piotti, A., Rossi, M., & Piovani, P. (2009). Genetic variability of Italian southern Scots pine (*Pinus sylvestris* L.) populations: the rear edge of the range. *European Journal of Forest Research*, 128(4), 377.

Schmid, B. (1985). Clonal growth in grassland perennials: III. Genetic variation and plasticity between and within populations of *Bellis perennis* and *Prunella vulgaris*. *The Journal of Ecology*, 819-830.

Sebastiani, F., Pinzauti, F., Kujala, S. T., González-Martínez, S. C., & Vendramin, G. G. (2012). Novel polymorphic nuclear microsatellite markers for *Pinus sylvestris* L. *Conservation Genetics Resources*, 4(2), 231-234.

Sinclair, W. T., Morman, J. D., & Ennos, R. A. (1998). Multiple origins for Scots pine (*Pinus sylvestris* L.) in Scotland: evidence from mitochondrial DNA variation. *Heredity*, 80(2), 233-240.

Sinclair, W. T., Morman, J. D., & Ennos, R. A. (1999). The postglacial history of Scots pine (*Pinus sylvestris* L.) in western Europe: evidence from mitochondrial DNA variation. *Molecular Ecology*, 8(1), 83-88.

Skilling, D. D. (1990). *Pinus sylvestris* L. (*In*: Burns, R. M., Honkala, B. H. (Eds) *Silvics of North America. Vol. 1. Conifers*.) US Department of Agriculture. pp. 1042-1055.

Slatkin, M. (1975). Gene flow and selection in a two-locus system. *Genetics*, 81(4), 787-802.

Sokal, R. R., Rohlf, F. J. (1981). *Biometry*. (Second Edition) New York, W. H. Freeman and Company. pp. 219

Sommer, R. S., & Zachos, F. E. (2009). Fossil evidence and phylogeography of temperate species "glacial refugia"and postglacial recolonization. *Journal of Biogeography*, 36, 2013-2020.

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.

Soranzo, N., Provan, J., & Powell, W. (1998). Characterization of microsatellite loci in *Pinus sylvestris* L. *Molecular Ecology*, 7(9), 1260-1261.

Soranzo, N., Provan, J., & Powell, W. (1999). An example of microsatellite length variation in the mitochondrial genome of conifers. *Genome*, 42(1), 158-161.

Soto, A., Robledo- Arnuncio, J. J., González- Martínez, S. C., Smouse, P. E., & Alia, R. (2010). Climatic niche and neutral genetic diversity of the six Iberian pine species: a retrospective and prospective view. *Molecular Ecology*, 19(7), 1396-1409.

Staszkiewicz, J. (1994). Differentiation of populations of *Pinus* × *rhaetica* (Pinaceae) from the "Mezi Bormi"Nature Reserve of Slovakia. *Fragmenta Floristica et Geobotanica, Series Polonica*, 1, 223-233.

Stephens, M., & Donnelly, P. (2003). A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *The American Journal of Human Genetics*, 73(5), 1162-1169.

Stephens, M., Smith, N. J., & Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. *The American Journal of Human Genetics*, 68(4), 978-989.

Stewart, J. R., Lister, A. M., Barnes, I., & Dalén, L. (2010). Refugia revisited: individualistic responses of species in space and time. *Proceedings of the Royal Society of London B: Biological Sciences*, 277(1682), 661-671.

Stieber, J. (1967). A magyarországi felsőpleisztocén vegetációtörténete az anthrakotómiai eredmények (1957-ig) tükrében. *Földtani Közlöny,* 97, 308-317.

Sümegi, P., Bodor, E., Törőcsik, T. (2005). The origins of alkalisation in the Hortobágy region in the light of the palaeoenvironmental studies at Zám-Halasfenék. (*In*: Gál, E., Juhász, I., Sümegi, P. (Eds) *Environmental archaeology in northeastern Hungary.* (Varia Archaeologica Hungarica 19.). Budapest, Archaeological Institute of the Hungarian Academy of Sciences. pp. 115-126.

Sümegi, P., Kertész, R., Hertelendi, E. (2002). Environmental Change and Human Adaptation in the Carpathian Basin at the lateglacial/postglacial transition. *British Archaeological Report*, 1043, 171.

Svendsen, J. I., Alexanderson, H., Astakhov, V. I., Demidov, I., Dowdeswell, J. A., Funder, S., Gataullin, V., Henriksen, M., Hjort, C., Houmark-Nielsen, M., Hubberten, H. W., Ingolfsson, O., Jacobsson, M., Kjaer, K., Larsen, E., Lokrantz, H., Lunkka, J. P., Lysa, A., Mangerud, J., Matioushkov, A., Murray, A., Möller, P., Niessen, F., Nikolskaya, O., Polyak, L., Saarnisto, M., Siegert, C., Siegert, M. J., Spielhagen, R. F. & Stein, R. (2004). Late Quaternary ice sheet history of northern Eurasia. *Quaternary Science Reviews*, 23(11), 1229-1271.

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

Tantau, I., Reille, M., de Beaulieu, J. L., & Farcas, S. (2006). Late Glacial and Holocene vegetation history in the southern part of Transylvania (Romania): pollen analysis of two sequences from Avrig. *Journal of Quaternary Science*, 21(1), 49-61.

Tantau, I., Reille, M., de Beaulieu, J. L., Farcas, S., Goslar, T., & Paterne, M. (2003). Vegetation history in the Eastern Romanian Carpathians: pollen analysis of two sequences from the Mohoş crater. *Vegetation History and Archaeobotany*, 12(2), 113-125.

Templeton, A. R., Shaw, K., Routman, E., & Davis, S. K. (1990). The genetic consequences of habitat fragmentation. *Annals of the Missouri Botanical Garden*, 13-27.

Tesitel, J., Malinova, T., Stech, M., & Herbstova, M. (2009). Variation in the *Melampyrum sylvaticum* group in the Carpathian and Hercynian region: two lineages with different evolutionary histories. *Preslia*, 81(1), 1-22.

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22(22), 4673-4680.

Tipping, R., Ashmore, P., Davies, A. L., Haggart, B. A., Moir, A., Newton, A., Sands, R., Skinner, T., & Tisdall, E. (2008). Prehistoric Pinus woodland dynamics in an upland landscape in northern Scotland: the roles of climate change and human impact. *Vegetation History and Archaeobotany*, 17(3), 251-267.

Tipping, R., Ashmore, P., Davies, A., Haggart, B. A., Moir, A., Newton, A., Sands, R., Skinner, T., & Tisdall, E. (2007). Peat, pine stumps and people: interactions behind climate, vegetation change and human activity in wetland archaeology at Loch Farlary, northern Scotland (*In*: Barber, J., Clark, C., Cressey, M., Crone, A., Hale, A., Henderson, J., Housley, R., Sands, R., Sheridan, A., & Scottish Wetland Archaeology Project (SWAP) (Eds.) *Archaeology from the Wetlands: Recent Perspectives: Proceedings of the 11th WARP Conference*, Edinburgh, Society of Antiquaries of Scotland. 11th Annual WARP conference. pp. 157-164.

Tollefsrud, M. M., Kissling, R., Gugerli, F., Johnsen, Ø., Skrøppa, T., Cheddadi, R., van der Knaap, W. O., Latałowa, M., Terhürne-Berson, R., Litt, T., Geburek, T., Brochmann, C. & Sperisen, C. (2008). Genetic consequences of glacial survival and postglacial colonization in Norway spruce: combined analysis of mitochondrial DNA and fossil pollen. *Molecular Ecology*, 17(18), 4134-4150.

Tóth, E. Gy., Vendramin, G. G., Bagnoli, F., Cseke, K., & Höhn, M. (2017). High genetic diversity and distinct origin of recently fragmented Scots pine (*Pinus sylvestris* L.) populations along the Carpathians and the Pannonian Basin. *Tree Genetics & Genomes*, 13(2), 47.

Tribsch, A., & Schönswetter, P. (2003). Patterns of endemism and comparative phylogeography confirm palaeo-environmental evidence for Pleistocene refugia in the Eastern Alps. *Taxon*, 52(3), 477-497.

Tsaryk, I., Didukh, Y. P., Tasenkevich, L., Waldon, B., & Boratynski, A. (2006). *Pinus mugo* Turra (Pinaceae) in the Ukrainian Carpathians. *Dendrobiology*, 55, 39-49.

Tsumura, Y., Uchiyama, K., Moriguchi, Y., Ueno, S., & Ihara-Ujino, T. (2012). Genome scanning for detecting adaptive genes along environmental gradients in the Japanese conifer, *Cryptomeria japonica*. *Heredity*, 109(6), 349-360.

Turna, I. (2003). Variation of some morphological and electrophoretic characters of 11 populations of Scots pine in Turkey. *Israel journal of plant sciences*, 51(3), 223-230.

Turna, I., & Güney, D. (2008). Altitudinal variation of some morphological characters of Scots pine (*Pinus sylvestris* L.) in Turkey. *African Journal of Biotechnology*, 8(2).

Tutin, T. G., Heywood, V. H., Burges, N. A., Valentine, D. H., Walters, S. M., Webb, D. A. (1964). *Flora Europea, volume I.* Cambridge, Cambridge University Press. pp. 34-35.

Tzedakis, P. C., Hooghiemstra, H., & Pälike, H. (2006). The last 1.35 million years at Tenaghi Philippon: revised chronostratigraphy and long-term vegetation trends. *Quaternary Science Reviews*, 25(23), 3416-3430.

UNEP-DEWA United Nations Environment Programme Division of Early Warning, & Assessment. (2007). Carpathians environment outlook 2007. United Nations Pubns. Publisher. pp. 90-99.

van Oosterhout, C., Hutchinson, W. F., Wills, D. P., & Shipley, P. (2004). MICRO- CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, 4(3), 535-538.

Vendramin, G. G., & Ziegenhagen, B. (1997). Characterisation and inheritance of polymorphic plastid microsatellites in Abies. *Genome*, 40(6), 857-864.

Vendramin, G. G., Anzidei, M., Madaghiele, A., & Bucci, G. (1998). Distribution of genetic diversity in *Pinus pinaster* Ait. as revealed by chloroplast microsatellites. *TAG Theoretical and Applied Genetics*, 97(3), 456-463.

Vendramin, G. G., Lelli, L., Rossi, P., & Morgante, M. (1996). A set of primers for the amplification of 20 chloroplast microsatellites in *Pinaceae*. *Molecular Ecology*, 5(4), 595-598.

Wachowiak, W., Balk, P. A., & 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.

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(2), 225-238.

Wachowiak, W., Salmela, M. J., Ennos, R. A., Iason, G., & Cavers, S. (2011). High genetic diversity at the extreme range edge: nucleotide variation at nuclear loci in Scots pine (*Pinus sylvestris* L.) in Scotland. *Heredity*, 106(5), 775-787.

Wachowiak, W., Wójkiewicz, B., Cavers, S., & Lewandowski, A. (2014). High genetic similarity between Polish and North European Scots pine (*Pinus sylvestris* L.) populations at nuclear gene loci. *Tree Genetics & Genomes*, 10(4), 1015-1025.

Wakasugi, T., Tsudzuki, J., Ito, S., Nakashima, K., Tsudzuki, T., & Sugiura, M. (1994). Loss of all ndh genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. *Proceedings of the National Academy of Sciences*, 91(21), 9794-9798.

Walker, M. J. C. (1995). Climatic changes in Europe during the last glacial/interglacial transition. *Quaternary International*, 28, 63-76.

Watts, W. A. (1984). The Holocene Vegetation of The Burren, western Ireland. (*In*: Haworth, E. Y., Lund, J. W. G. (Eds) *Lake Sediment and Environmental History*. Leicester, Leicester University Press. pp. 359-376.

Watts, W. A. (1988). Quaternary Vegetation Cycles. (*In*: Edwards, K. J., Warren, W. P. (Eds) *The Quaternary History of Ireland.*) London, Academic Press. pp. 155-184.

Weir, B. S., & Hill, W. G. (1996). Genetic data analysis II. *Genetical Research*, 68(2), 183-188.

White, G., & Powell, W. (1997). Isolation and characterization of microsatellite loci in *Swietenia humilis* (Meliaceae): an endangered tropical hardwood species. *Molecular Ecology*, 6(9), 851-860.

Whitlock, M. C. (1992). Temporal fluctuations in demographic parameters and the genetic variance among populations. *Evolution*, 608-615.

Whittaker, R. H., & Goodman, D. (1979). Classifying species according to their demographic strategy. I. Population fluctuations and environmental heterogeneity. *American Naturalist*, 185-200.

Willis, K. J. (1994). The vegetational history of the Balkans. *Quaternary Science Reviews*, 13(8), 769-788.

Willis, K. J., Bennett, K. D., Birks, H. J. B. (1998). The late Quaternary dynamics of pines in Europe. (*In*: Richardson D.M. (Ed), *Ecology and Biogeography of Pinus*.). Cambridge, Cambridge University Press. pp. 107-121.

Willis, K. J., Rudner, E., & Sümegi, P. (2000). The full-glacial forests of central and southeastern Europe. *Quaternary Research*, 53(2), 203-213.

Willis, K. J., Sümegi, P., Braun, M., & Tóth, A. (1995). The late Quaternary environmental history of Bátorliget, NE Hungary. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 118(1-2), 25-47.

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.

Wright, S. (1943). Isolation by distance. *Genetics*, 28(2), 114.

Yeh, F. C., & Layton, C. (1979). The organization of genetic variability in central and marginal populations of lodgepole pine *Pinus contorta* spp. *latifolia*. *Canadian Journal of Genetics and Cytology*, 21(4), 487-503.

Young, A., Boyle, T., & Brown, T. (1996). The population genetic consequences of habitat fragmentation for plants. *Trends in Ecology & Evolution*, 11(10), 413-418.

Zemanek, B. (1991). The phytogeographical boundary between the East and West Carpathians – past and present. *Thaiszia*, 1, 59-67.

Zhou, Y., Bui, T., Auckland, L. D., & Williams, C. G. (2002). Undermethylated DNA as a source of microsatellites from a conifer genome. *Genome*, 45(1), 91-99.

10. SUPPLEMENTAL MATERIALS

Methodology

Population descriptions summarized as follows:

Region	Code	Description
Western Carpathians		
1	**SKV**	The Kvačianska valley (Kvacsianszka-völgy) population is located in a narrow valley facing north–south direction at the western foothills of the High Tatras (Magas-Tátra). Scots pine grows on the top of the steep valley walls and along the river stream in the cracks of the bedrock. Soil types are eutric and dystric cambisols on Triassic dolomite and calcareous rock (ESBN-EC 2005). The average yearly mean temperature at the site is about 5.0°C (min.:-9.3°C, max.:19.9°C) and the average yearly precipitation is about 1050 mm (Hijmans *et al.* 2005). National nature reserve.
2	**SME**	The Medzi Bormi population is located at the north-west foothills of the High Tatras, where Scots pine grows in a peat-bog, which was established on the impenetrable flysch basement. Hybrid individuals of *P. sylvestris* and *P. mugo* (*P.* × *rhaetica*) also were identified within the peat-bog community (Kormut'ák *et al.* 2012, Klimko and Bykowska 2015). The dominant soil type in the region is dystric cambisol (ESBN-EC 2005). The average yearly mean temperature at the site is about 5.5°C (min.:-9.0°C, max.:20.6°C) and the average yearly precipitation is about 1012 mm (Hijmans *et al.* 2005). Natura 2000 protected site.
3	**STU**	The Turková population is located in the north-east Low Tatras (Alacsony-Tátra) region and is situated in the valley of the river Čierny Váh (Fekete-Vág). Trees are growing on the steep cliff side above the hydro-power plant placed at the mid-section of the river. The prevailing rock type is degraded limestone, where Scots pine grows at the edge of the mixed forest on the rocky substrate. The dominant soil types within the region are dystric cambisol, rendzic leptosol and endoskeletic umbrisol, respectively (ESBN-EC 2005). The average yearly mean temperature at the site is about 3.7°C (min.:-9.9°C, max.:17.9°C) and the average yearly precipitation is about 1088 mm (Hijmans *et al.* 2005). Natura 2000 protected site.
4	**SLI**	The Liptovský Hrádok (Liptóújvár) population is positioned in the northern Low Tatras region, surrounded by the historical city. Scattered individuals are growing along the limestone outcrops isolated by the urban areas. Soil types are dystric cambisol and calcaric fluvisol (ESBN-EC 2005). The average yearly mean temperature at the site is about 6.5°C (min.:-8.6°C, max.:22.2°C) and the average yearly precipitation is about 847 mm (Hijmans *et al.* 2005).
Eastern Carpathians		
5	**RFE**	The Fântâna Brazilor (Fenyőkút) populations is situated in the Eastern Carpathian Corund-Hargitha (Korond-Hargita) region. Trees are growing in the constantly drying peat-bog located on the flat volcanic plateau. The dominant soil types within the region are eutric and eutric-andic cambisols (ESBN-EC 2005). The average yearly mean temperature at the site is about 5.8°C (min.:-9.8°C, max.:20.9°C) and the average yearly precipitation is about 708 mm (Hijmans *et al.* 2005). Natura 2000 protected site.

6	**RPO**	The Poiana Stampei population is located in the northern Eastern Carpathian Dorna valley, at the northern foothills of the Călimani Mountains (Kelemen-havasok). The population sustained within the peat-bog, which is established in the depression of the Dorna valley on crystalline and Eocene grit stone bedrock. The dominant soil types within the region are eutric cambisol and episkeletic podzol (ESBN-EC 2005). The average yearly mean temperature at the site is about 5.8°C (min.:-9.7°C, max.:20.9°C) and the average yearly precipitation is about 739 mm (Hijmans *et al.* 2005). Ramsar and Natura 2000 protected site.
7	**RMO**	The Mohos population is located in the southern Eastern Carpathian Intra-Carpathian Volcanic Range in the Harghita Mountains (Hargita). The population inhabits the eastern twin volcanic crater of the Ciomadul (Csomád) mountain. Trees are growing in peat with the mosaic of some open-water patches. The dominant soil types are stagnic luvisol and dystric cambisol (ESBN-EC 2005) The average yearly mean temperature at the site is about 5°C (min.:-10°C, max.:20°C) and the average yearly precipitation is about 736 mm (Hijmans *et al.* 2005).
8	**RCO**	The Suhardul Mic (Kis-Cohárd) population positioned in the Eastern Carpathian Bicaz (Békás) valley (southern Giurgeu Mountains (Gyergyói-havasok)). The population is sustained on the top of the southern facing rock surfaces above the Lacul Roşu (Gyilkos-tó). The dominant soil types within the region are leptosol, episkeletic podzol and dystric cambisol (ESBN-EC 2005). The average yearly mean temperature at the site is about 5.4°C (min.:-9.9°C, max.:20.6°C) and the average yearly precipitation is about 688 mm (Hijmans *et al.* 2005).
Southern Carpathians		
9	**RPA**	The Lătoriței (Latorica) mountain population is located east from Parâng (Páreng) mountain massif in a narrow east–west orientated valley in the Southern Carpathians. Scots pine growing on the top of the steep valley walls and along the Lătoriței river in the cracks of the vertical rock walls. The dominant soil types within the region are episkeletic podzol and dystric cambisol (ESBN-EC 2005). The average yearly mean temperature at the site is about 7.1°C (min.:-7.8°C, max.:22.2°C) and the average yearly precipitation is about 747 mm (Hijmans *et al.* 2005).
Central-Island Mountains (Apuseni)		
10	**RBI**	The Roşia (Biharrósa) population is located in the Pădurea Craiului Mountains in the north-western Apuseni massif. The Scots pine population is growing in the north–south facing Cheile Albioarei valley (Roșia valley) along the narrow and steep valley walls in a mixed forest community. The dominant soil types within the region are eutric cambisol and chromic luvisol (ESBN-EC 2005). The average yearly mean temperature at the site is about 9.3°C (min.:-5.6°C, max.:25.4°C) and the average yearly precipitation is about 655 mm (Hijmans *et al.* 2005).
11	**RBE**	The Scărița-Belioara (Bélavár) population is situated in the southern Gilău Mountains (Gyalui-havasok) in the central Apuseni massif. The Scots pine population is growing along the vertical rock walls and sharp peaks of degraded limestone outcrops admixed with *Larix decidua.* Individuals are scattered both at the top edge of the rocky plateau and on the south-eastern facing vertical walls. The dominant soil type within the region is dystric cambisol (ESBN-EC 2005). The average yearly mean temperature at the site is about 5.4°C (min.:-7.5°C, max.:19.5°C) and the average yearly precipitation is about 846 mm (Hijmans *et al.* 2005).
12	**RML**	The Mluha population located in the northern part of the Metaliferi Mountains (Erdélyi-Érchegység) in the central-eastern Apuseni massif. Low number of trees are growing in the constantly drying raised bog community located on a flat plateau, which extends from north-east to south-west parallel to Arieş river (Aranyos-patak)(belongs to the Arieș valley water catchment area). The dominant soil type within the region is dystric cambisol (ESBN-EC 2005). The average yearly mean temperature at the site is about 5.2°C (min.:-7.5°C, max.:19.1°C) and the average yearly precipitation is about 872 mm (Hijmans *et al.* 2005).

13	**RMH**	The Călăţele population located in the eastern Gilăului Mountains (Gyalui-havasok) in the north-eastern Apuseni massif. The population inhabits an intensively exploited peat-bog with only few individuals isolated at the edge of the peatland. The peatland is covered with mixed vegetation and in a process of rehabilitation. The dominant soil types within the region are eutric and dystric cambisols (ESBN-EC 2005). The average yearly mean temperature at the site is about 6.8°C (min.:-6.9°C, max.:21.6°C) and the average yearly precipitation is about 765 mm (Hijmans *et al.* 2005).
Pannonian Basin		
14	**HFE**	The population of Fenyőfő is situated in the northern foothills of the Bakony mountains in the central Transdanubian Mountains range (Dunántúli-középhegység). This population exhibits pure stands of Scots pine growing on slightly calcareous sandy substrate with low nutrient content. The dominant soil types within the region are gleyic phaeozem and chromic luvisol (ESBN-EC 2005). The average yearly mean temperature at the site is about 9.6°C (min.:-4.9°C, max.:25.8°C) and the average yearly precipitation is about 574 mm (Hijmans *et al.* 2005). National nature reserve since 1954.
15	**HZA**	The Pethőhenye population is located in the western Hungarian Zala county in the hilly Zalaegerszeg micro region. Here, Scots pine occupy eroded rocky outcrops. Low number of trees sustained on the shallow soil layer consist hapli-lithic leptosol and eutric histosol soils, with the mosaic of alkaline sandstone outcrops (ESBN-EC 2005). The average yearly mean temperature at the site is about 9.1°C (min.:-5.3°C, max.:25°C) and the average yearly precipitation is about 797 mm (Hijmans *et al.* 2005).
16	**HVE**	The Szalafő (Őserdő) forest reserve population is located within the territory of the Őrségi Nemzeti Park. The population grows in a mixed forest stand at the hilly slopes of the Vas county. The population belongs to the foothills of the Alps (Pre-Alpine). Within the region the soils are acidic with high loess, clay and gravel content. The dominant soil type is gleyic luvisol (ESBN-EC 2005). The average yearly mean temperature at the site is about 10.1°C (min.:-4.4°C, max.:26.2°C) and the average yearly precipitation is about 672 mm (Hijmans *et al.* 2005). National nature reserve.
17	**HOR**	Csörötnek population is situated south from the Rába river (belongs to the Rába valley water catchment area) within the territory of Őrségi Nemzeti Park. The population belongs to the foothills of the Alps (Pre-Alpine). The habitat of Scots pine is near the loamy (sand and clay, often with gravel) stream bank terraces of Rába river. The dominant soil types are dystric fluvisol and gleyic luvisol (ESBN-EC 2005). The average yearly mean temperature at the site is about 9.2°C (min.:-5.3°C, max.:25.1°C) and the average yearly precipitation is about 768 mm (Hijmans *et al.* 2005).
18	**HKO**	The Kőszeg population is located in the Kőszeg Mountains, which belongs to the foothill region of the Alps (Pre-Alpine). The fragments of natural vegetation includes scattered population of Scots pine sustained within the western-central part of the mountain range on calcareous bedrock (ridges of Péterics hegy). The dominant soil type within the region is hapli-lithic leptosol (ESBN-EC 2005). The average yearly mean temperature at the site is about 7.2°C (min.:-6.6°C, max.:22.6°C) and the average yearly precipitation is about 789 mm (Hijmans *et al.* 2005).
Outgroup populations		
19	**BYU**	The population of Yundola region is situated in the Rila Mountains in central Bulgaria. Both extensive pure and mixed Scots pine dominated forests (with spruce and fir) naturally occurs within the region at the southern slopes of Rila Mountain. The dominant soil types are dystric cambisol and haplic umbrisol (ESBN-EC 2005). The average yearly mean temperature at the site is about 4.6°C (min.:-8.1°C, max.:19.3°C) and the average yearly precipitation is about 697 mm (Hijmans *et al.* 2005).
20	**ESE**	The Estonian population originates from the Endla Nature Reserve at the foot on the Pandivere Upland of the Central Estonian Plain. Here, a pure Scots pine forest is growing in the extensive fresh-water peat-bog system. The dominant soil type within the region is stagnic luvisol (ESBN-EC 2005). The average yearly mean temperature at the site is about 4.7°C (min.:-10.2°C, max.:21.5°C) and the average yearly precipitation is about 612 mm (Hijmans *et al.* 2005). Ramsar and Natura 2000 protected site.

Tables and Figures

Tables

Table S1: List of main results and conclusions of major publications related to mitochondrial markers and the phylogeographic studies on *Pinus sylvestris*.

Mitochondrial DNA studies			*Main conclusions*
Date	*References*	*Markers*	
1998	Sinclair *et al.*	mtDNA cox1	Two common and one rare mitotypes were detected. Scotish populations of *P. sylvestris* have been derived from several places, probably from a western refugium (Ireland) or western France.
1999	Sinclair *et al.*	mtDNA cox1	Three major mitotypes were detected. In Spanish populations all three major mitotypes were found, and gene diversity was high. European populations showed little or no mtDNA diversity within regions, but there were marked differences between the regions. Present-day populations of Europe have been derived from three different sources. Spanish populations have not took a part in recolonization of Europe.
1999	Soranzo *et al.*	mtDNA nad3 rps12	Identified mononucleotide microsatellites in an intergenic region between mitochondrial genes. First report on the occurrence of microsatellite polymorphism.
2000	Soranzo *et al.*	mtDNA nad1	Two distinct haplotype was identified in Europe. On the Iberian Peninsula both haplotypes were detected, in Continental Europe and the Scottish populations were all fixed for one haplotype. Spanish populations have not contributed in recolonization after glaciation.
2006	Labra *et al.*	mtDNA nad1 ISSR	Italian populations have the same mitotype as the Central Europeans. ISSR analysis showed that Alpine populations have a higher genetic variability than Apennine samples. Genetic isolation between the Alpine and Northern Apennine populations from the early Holocene.
2007	Naydenov *et al.*	mtDNA nad7	Four genetically distinct ancestral lineages were detected. Lineage limited to the Iberian Peninsula was confirmed. New lineage was found confined to Turkey (Asia Minor).
2008	Pyhäjärvi *et al.*	mtDNA nad7	Iberian Peninsula's populations have not contributed in the recolonization of Europe. Populations from Turkey did not have role in the colonization as it is isolated from the main distribution. There might be a glacial refugium in Asia Minor, and populations in this region may be more connected to Asian than to Europe.
2009	Čelepirović *et al.*	mtDNA nad1	42 provenances were tested from the broad range of Eurasia. No variability detected in mitochondrial nad1 region in all analyzed samples. All samples presented haplotype A.

Table S2: List of main results and conclusions of major publications related to chloroplast markers and the phylogeographic studies on *Pinus sylvestris*.

Chloroplast DNA studies			*Main conclusion*
Date	*References*	*Markers*	
1995	Powell *et al.*	cpSSR	By screening chloroplast genome for polymorphism cpSSRs were identified in three pine species (*Pinus contorta, Pinus sylvestris, Pinus thunbergii*), Analysis of 305 individuals from seven populations of *Pinus leucodermis* revealed the presence of four distinct size variants.
1998	Provan *et al.*	cpSSR	Small but significant differentiation has been reported between Scottish and the mainland European populations. Examination of the distribution of a mutation (duplications in the repeat region) suggests that Wester Ross region (Shieldaig, Coulin and Loch Maree) may be a refugia.
2004	Robledo-Arnuncio *et al.*	cpSSR	Higher haplotypic genetic diversity than in other pine species and big differences between isolated populations and unclear geographic pattern in the distribution of haplotypes was revealed. High genetic diversity within populations and with very little but significant variation among them was identified.
2005	Robledo-Arnuncio *et al.*	cpSSR	Higher values of haplotypic diversity than at thermophilic pine species in the Iberian Peninsula was detected. Disjoint montane distribution could not account for the genetic divergence among areas. Their results also suggests that valleys have served as corridors for historical gene exchange among presently distant populations.
2005	Naydenov *et al.*	cpSSR terpenes	High haplotype diversity was found in all studied mountains. Rila, Pirin, Rhodope regions might have been refugia during the glacial depression. Terpene analysis was able to distinguish two groups, and results confirmed the hypothesis that gene flow was more intensive vertically than horizontally.
2006	Cheddadi *et al.*	cpSSR mtDNA fossil	Two identified Mediterranean refugia was confirmed. Refugial area identified east of the Alps, in the Hungarian Plain and the Danube region might have played a major role during the entire recolonization process of Europe.
2010	Soto *et al.*	cpSSR	Significant positive correlation between population genetic diversity and summer precipitation was highlighted. Thermophilous species have smaller haplotypic diversity.
2014	Pavia *et al.*	cpSSR nSSR	Relict status of Portuguese populations are evidenced. Results also suggest that distinct characteristics of the two studied populations might be explained by different origins.
2016	Bernhardsson *et al.*	mtDNA cpSSR nSSR	Study reveals high genetic diversity and recently ongoing genetic erosion, which suggests strong impact on Holocene population fragmentation based on combined marker systems in the studied Romanian and Hungarian populations.
2016	Wójkiewicz and Wachowiak	cpSSR	Scots pine in Europe is genetically highly variable and does not possess a homogenous gene pool. Spanish and Turkish populations differed from other European populations. Isolation by distance was found between the southern populations.

Table S3: List of main results and conclusions of publications related to nuclear markers and the phylogeographic studies on *Pinus sylvestris*.

Nuclear DNA studies			*Main conclusion*
Date	*References*	*Markers*	
1995	Kostia *et al.*	nSSR	Scots pine (*Pinus sylvestris*) genomic libraries were constructed and eight microsatellites were identified.
1998	Soranzo *et al.*	nSSR	Designed seven polymorphic SSR primer pairs for *Pinus sylvestris.*
2009	Scalfi *et al.*	nSSR cpSSR	Glacial refugium in the Apennines was detected. Erosion in population genetic variability was identified. High level of genetic diversity maintained and significant among population differentiation was found.
2011	Naydenov *et al.*	nDNA	12 populations of *P. sylvestris* were tested, 3 large clusters were determined namely: "Mesta River watershed"; "marginal" and "central" Rhodopes. Bottlenecks and "Mountain effect" have been evidenced.
2012	Belletti *et al.*	nSSR	East-West subdivision of the Italian Alps was revealed. Apennines clearly separated from Alps. High within population diversity was detected. Genetic erosion undergone in the Apennines region.
2012	Sebastiani *et al.*	nSSR	Designed and characterized 10 polymorphic nuclear microsatellite markers for Scots pine.
2014	Lučić *et al.*	nSSR	7 populations were studied in Serbia, where geographical distance determined the clustering of the populations. Isolation by distance and geographical barriers were also estimated.
2017	Tóth *et al.*	nSSR cpSSR	High genetic diversity and distinct origin was revealed within the Carpathian region. There was no sign of bottleneck events and inbreeding. Barriers in geneflow were identified.

Table S4: List of main results and conclusions of publications related to adaptive genetic diversity of Scots pine.

Nuclear DNA studies			*Main conclusion*
Date	*References*	*Markers*	
2007	Pyhäyärvi *et al.*	nDNA SNP	There were no among-group differences in the level of silent nucleotide diversity in the studied four geographical group. Evidence that linkage disequilibrium extended mostly towards northern Europe than central.
2009	Wachowiak *et al.*	nDNA SNP	Moderate multi-locus nucleotide diversity at silent sites was found. Significant differentiation between populations in allelic frequency of haplotype structure was identified. Allelic dimorphism with no evidence of haplotype clustering along the geographical distribution was revealed. Significantly negative multi-locus tests of neutral theory of molecular evolution was present.
2011	Androsiuk *et al.*	nDNA B-SAP (*KatG*)	B-SAP marker system based on the *KatG* gene differentiated 19 populations across Europe. Results determined two groups, the North and the South according to their geographic origins. B-SAP record theorize two plausible routes of migrations directed from the South toward Northern Europe.
2011	Wachowiak *et al.*	nDNA SNP	Scotland, western populations showed slightly reduced nucleotide diversity. Relatively recent bottleneck reduced the population to about 2% of the present size. Higher genetic variation in Scottish populations compared to mainland populations was detected.
2011	Savolainen *et al.*	nDNA SNP	Combined studies (genetic architecture, pollen dispersal, phenology) of northern European populations highlighted that evolutionary processes on the periphery are greatly vary and species responses including phenotypic plasticity and adaptation are substantially not independent. Rapid responses of adaptation to climate change is still uncertain.
2012	Kujala and Savolainen	nDNA SNP	Little signs of local adaptation. Candidate genes subjected to clinal sequence variation across the latitudinal distribution of the species in Europe. Species-wide selection is suggested in some of the studied genes.
2014	Wachowiak *et al.*	nDNA SNP	High genetic similarity of Polish and Northern European populations indicate most likely a common postglacial history. Polish populations differ from the more isolated North-West and South European populations. Diversifying selection acting across environmental gradients due to latitudinal differences in photoperiodism and temperature.

Table S5: Parameter estimates for best demographic scenarios (Sc4 for (A) and Sc6 for (B)) based on DIYABC analysis (Cornuet *et al.* 2014). N1= Pop1 effective population size; N2=Pop2 effective population size; N3=Pop3 effective population size; Na= effective population size of the ancestral population; times are considered from present (0) backwards in time, t2=divergence of Pop1 from Pop2; t1=admixture of Pop1 and Pop2, and divergence of Pop1, Pop2 and Pop3; μmic=mean mutation rate of microsatellites; pmic=mean parameter of geometric distribution (GSM, Generalized Stepwise Mutation Model); snimic=individual locus SNI (Single Nucleotide Insertion/deletion) rate.

Run	Parameter	Mean	Median	Mode	Quantiles			
					2,5%	5%	95%	97,5%
(A)	N1	1.53e+004	1.60e+004	1.79e+004	7.42e+003	8.77e+003	1.96e+004	1.98e+004
	N2	5.66e+003	5.54e+003	4.87e+003	2.14e+003	2.54e+003	9.18e+003	9.56e+003
	N3	5.45e+003	5.37e+003	5.02e+003	1.38e+003	1.85e+003	9.26e+003	9.63e+003
	t1	2.44e+002	1.80e+002	8.38e+001	2.36e+001	3.42e+001	6.94e+002	8.23e+002
	ra	3.65e-001	3.51e-001	3.38e-001	3.10e-002	5.97e-002	7.33e-001	8.41e-001
	t2	3.49e+003	3.09e+003	2.20e+003	8.67e+002	1.12e+003	7.31e+003	8.17e+003
	Na	6.16e+002	4.79e+002	5.49e+001	1.92e+001	3.84e+001	1.67e+003	1.82e+003
	μmic	4.84e-004	4.55e-004	3.29e-004	1.86e-004	2.13e-004	8.57e-004	9.08e-004
	pmic	2.65e-001	2.76e-001	3.00e-001	1.64e-001	1.89e-001	3.00e-001	3.00e-001
	snimic	4.04e-006	3.44e-006	2.82e-008	3.57e-008	6.62e-008	9.77e-006	1.00e-005
(B)	N1	1.60e+004	1.65e+004	1.92e+004	9.15e+003	1.04e+004	1.96e+004	1.98e+004
	N2	6.09e+003	6.02e+003	5.51e+003	2.39e+003	2.82e+003	9.44e+003	9.70e+003
	N3	5.83e+003	5.84e+003	5.21e+003	1.62e+003	2.17e+003	9.41e+003	9.70e+003
	t1	3.01e+002	2.36e+002	1.14e+002	2.94e+001	4.49e+001	7.89e+002	8.76e+002
	ra	3.77e-001	3.69e-001	3.80e-001	4.67e-002	8.09e-002	7.08e-001	8.03e-001
	t2	3.92e+003	3.56e+003	2.74e+003	1.04e+003	1.29e+003	7.88e+003	8.81e+003
	db	9.41e+002	9.15e+002	7.73e+002	5.98e+001	1.03e+002	1.85e+003	1.92e+003
	N1c	5.73e+003	5.79e+003	5.52e+003	1.18e+003	1.68e+003	9.51e+003	9.77e+003
	N2c	2.91e+003	2.52e+003	2.08e+003	4.23e+002	6.05e+002	6.57e+003	7.35e+003
	Na	5.73e+002	4.30e+002	9.50e+000	1.63e+001	3.23e+001	1.63e+003	1.79e+003
	μmic	4.40e-004	4.03e-004	3.14e-004	1.76e-004	1.98e-004	8.03e-004	8.68e-004
	pmic	2.70e-001	2.81e-001	3.00e-001	1.75e-001	1.99e-001	3.00e-001	3.00e-001
	snimic	4.34e-006	3.95e-006	4.28e-008	4.92e-008	9.16e-008	9.83e-006	1.00e-005

Table S6: Model checking for the most likely demographic scenarios (Sc4 for (A) and Sc6 for (B)) based on DIYABC analysis (Cornuet *et al.* 2014). NAL=mean number of alleles, HET=mean gene diversity (Nei, 1987), FST=proportion of the total genetic variance contained in a subpopulation relative to the total genetic variance, N2P=mean number of alleles between populations, H2P=mean gene diversity between populations, 1=Pop1, 2=Pop2, 3=Pop3.

Summary statistics	(A)		(B)	
	Observed value	Proportion (simulated<observed)	Observed value	Proportion (simulated<observed)
NAL_1_1	14.1250	14.1250	0.7415	0.7715
NAL_1_2	8.3750	8.3750	0.5485	0.5645
NAL_1_3	7.8750	7.8750	0.4900	0.5065
HET_1_1	0.6444	0.6444	0.1490	0.1620
HET_1_2	0.5845	0.5845	0.1270	0.1345
HET_1_3	0.5967	0.5967	0.0930	0.1055
VAR_1_1	9.5368	9.5368	0.8240	0.8160
VAR_1_2	7.1135	7.1135	0.7520	0.7630
VAR_1_3	7.2150	7.2150	0.7290	0.7230
N2P_1_1&2	15.8750	15.8750	0.7965	0.8080
N2P_1_1&3	14.3750	14.3750	0.6920	0.7070
N2P_1_2&3	10.7500	10.7500	0.6510	0.6635
H2P_1_1&2	0.6492	0.6492	0.1420	0.1555
H2P_1_1&3	0.6467	0.6467	0.1445	0.1605
H2P_1_2&3	0.5975	0.5975	0.1120	0.1160
V2P_1_1&2	9.0947	9.0947	0.7950	0.7910
V2P_1_1&3	9.3237	9.3237	0.8130	0.8030
V2P_1_2&3	7.2280	7.2280	0.7450	0.7440
FST_1_1&2	0.0693	0.0693	0.6675	0.5305
FST_1_1&3	0.0540	0.0540	0.8665	0.7910
FST_1_2&3	0.0295	0.0295	0.6110	0.5075

Table S7: Estimation of prior parameters distribution in DIYABC analysis (Cornuet *et al.* 2014).

Run	Parameter	Minimum	Maximum
	Effective population size		
(A-B)	N1	4000	20000
(A-B)	N2	10	10000
(A-B)	N3	10	10000
(B)	Na	1	2000
(B)	N1c	10	10000
(B)	N2c	10	10000
	Time scale in generations		
(A-B)	t1	10	1000
(A-B)	t2	10	10000
	Duration of event		
(B)	db	10	2000
	Admixture		
(A-B)	ra	0,001	0,999
	Mutation model		
(A-B)	Mean mutation rate	1×10^{-4}	1×10^{-3}
(A-B)	Individual locus mutation rate	1×10^{-5}	1×10^{-2}
(A-B)	Mean coefficient P	1×10^{-1}	3×10^{-1}
(A-B)	Individual locus coefficient P	1×10^{-2}	9×10^{-1}
(A-B)	Mean SNI rate	1×10^{-8}	1×10^{-5}
(A-B)	Individual locus SNI rate	1×10^{-9}	1×10^{-4}

Table S8: Posterior probability of the final six tested demographic scenario and its 95% confidence interval based on the logistic estimate according to DIYABC (Cornuet *et al.* 2014).

Run	Scenario	Posterior probability	95% CI [lower - upper]
(A)	1	0.1909	[0.1670 - 0.2148]
(A)	2	0.0145	[0.0083 - 0.0206]
(A)	3	0.0033	[0.0000 - 0.0088]
(A)	4	0.7914	[0.7671 - 0.8156]
(B)	5	0.4330	[0.3967 - 0.4693]
(B)	6	0.5670	[0.5307 - 0.6033]

Table S9: Number of substitutions and number of private substitutions in parentheses in all studied population at analysed candidate gene loci.

	Population											
Locus	**HKO**	**HZA**	**HVE**	**HFE**	**RCO**	**RFE**	**RPO**	**SME**	**STU**	**SKV**	**Mean**	**s.d.**
abaR	8(0)	7(1)	9(1)	9(1)	8(0)	3(0)	7(0)	6(1)	9(0)	10(1)	7.600(0.500)	2.011(0.527)
ccoaomt	5(0)	7(1)	5(0)	0(0)	4(0)	8(1)	1(1)	4(3)	4(0)	6(0)	4.400(0.600)	2.459(0.966)
chcs2	7(0)	5(0)	10(2)	9(0)	5(0)	4(0)	5(0)	13(4)	7(0)	7(0)	7.200(0.600)	2.781(1.350)
cpk3	8(1)	8(1)	10(0)	9(1)	7(0)	5(0)	7(0)	8(0)	9(2)	6(0)	7.700(0.500)	1.494(0.707)
dhn3	4(0)	7(1)	4(0)	2(0)	6(1)	5(0)	3(0)	4(2)	6(1)	4(0)	4.500(0.500)	1.509(0.707)
dhn7	5(0)	5(0)	11(1)	6(1)	11(2)	5(1)	6(0)	13(3)	8(1)	6(1)	7.600(1.000)	2.989(0.943)
dhy2PP	13(2)	17(1)	16(1)	14(2)	9(0)	10(1)	13(1)	17(4)	14(0)	19(4)	14.200(1.600)	3.155(1.430)
erd3	4(2)	4(1)	2(1)	1(0)	2(0)	1(0)	5(2)	6(4)	2(0)	3(0)	3.000(1.000)	1.700(1.333)
pal1	4(0)	4(0)	6(1)	4(0)	4(0)	4(1)	7(0)	11(2)	6(0)	6(1)	5.600(0.500)	2.221(0.707)
ppap12	8(0)	9(1)	9(0)	10(1)	11(1)	11(0)	9(0)	13(6)	9(2)	12(0)	10.100(1.100)	1.595(1.853)
rd21A	20(7)	7(1)	10(1)	6(0)	11(0)	7(1)	5(1)	8(1)	27(6)	17(1)	11.800(1.900)	7.223(2.470)
Total	86(12)	80(8)	92(8)	70(6)	78(4)	63(5)	68(5)	103(30)	101(12)	96(8)	-	-

Table S10: Nucleotide diversity indicator Theta pi ($\theta\pi$ x10^{-3}) in all studied population at analysed candidate gene loci.

	Population											
Locus	**HKO**	**HZA**	**HVE**	**HFE**	**RCO**	**RFE**	**RPO**	**SME**	**STU**	**SKV**	**Mean**	**s.d.**
abaR	2.2679	1.4117	1.8954	1.4771	2.3976	0.8655	1.8830	1.5555	1.7309	1.9558	1.7440	0.4444
ccoaomt	0.9477	0.9833	0.5555	0.0000	0.7578	3.3315	0.2333	1.4473	0.4000	1.2549	0.9911	0.9373
chcs2	2.5098	1.7254	2.4967	2.3833	2.1578	0.8052	1.8631	3.6526	1.7157	1.9684	2.1278	0.7340
cpk3	1.9019	2.7843	2.4902	3.2549	2.9052	0.5000	2.6157	2.9157	2.5157	2.7320	2.4616	0.7744
dhn3	1.5303	2.1868	1.6071	1.1666	1.6483	1.3916	0.9477	1.4642	1.5714	0.9560	1.4470	0.3663
dhn7	1.1503	1.6416	3.3296	2.3250	2.4771	2.1105	1.5620	4.5384	2.1000	1.8461	2.3081	0.9822
dhy2PP	2.7189	3.9607	4.5359	4.5620	3.3947	4.2473	2.7736	4.8736	3.7210	4.5894	3.9377	0.7693
erd3	0.8039	0.8366	0.7320	0.2941	0.5263	0.5210	0.9263	0.9789	0.6052	0.8758	0.7100	0.2173
pal1	1.7320	1.9150	2.2483	1.8692	1.6157	1.7684	2.4000	3.2631	1.8789	2.1052	2.0796	0.4803
ppap12	2.1176	2.5098	3.0392	2.4313	2.9736	4.2263	2.5842	2.9157	2.9684	3.3947	2.9161	0.5882
rd21A	3.7385	2.1307	2.3464	2.5111	2.8833	2.7878	2.2000	2.0842	5.3105	4.0261	3.0018	1.0486
Mean	1.9471	2.0078	2.2978	2.0249	2.1580	2.0505	1.8172	2.6990	2.2289	2.3368	-	-

Table S11: Population specific haplotypes according to TCS (Templeton-Crandall-Singh) network analysis in all studied population at analysed candidate gene loci.

	Population										
Locus	**HKO**	**HZA**	**HVE**	**HFE**	**RCO**	**RFE**	**RPO**	**SME**	**STU**	**SKV**	**Mean**
abaR	-	-	1	1	2	-	-	1	1	1	1.1667
ccoaomt	-	-	-	-	-	1	1	3	1	-	1.5000
chcs2	1	-	2	1	-	-	-	3	-	-	1.7500
cpk3	1	3	1	2	2	-	-	-	1	1	1.5714
dhn3	-	1	-	-	1	-	-	1	1	-	1.0000
dhn7	1	-	-	1	2	1	-	4	3	2	2.0000
dhy2PP	4	3	1	4	2	-	3	7	-	5	3.6250
erd3	2	1	1	-	-	-	1	1	-	1	1.1667
pal1	-	-	1	-	-	1	1	5	1	1	1.6667
ppap12	-	2	-	2	1	-	1	5	2	-	2.1667
rd21A	6	1	2	-	2	3	2	3	6	4	3.2222
Total	15	11	9	11	12	6	9	33	16	15	-

Table S12: Summary of genetic differentiation and analysis of molecular variance (AMOVA) at analysed candidate gene loci.

		AMOVA	
Locus	F_{ST}	**Among pop.**	**Within pop.**
abaR	0.0411	4.1167	95.8832
ccoaomt	0.1384	13.8403	86.1596
chcs2	0.0752	7.5205	92.4794
cpk3	0.1140	11.4053	88.5946
dhn3	0.0922	9.2244	90.7756
dhn7	0.0662	6.6212	93.3788
dhy2PP	0.0788	7.8833	92.1167
erd3	0.0269	2.6976	97.3023
pal1	0.0111	1.1159	98.8840
ppap12	0.0181	1.8192	98.1807
rd21A	0.0382	3.8234	96.1765
Mean	0.063	6.3698	93.6301

Figures

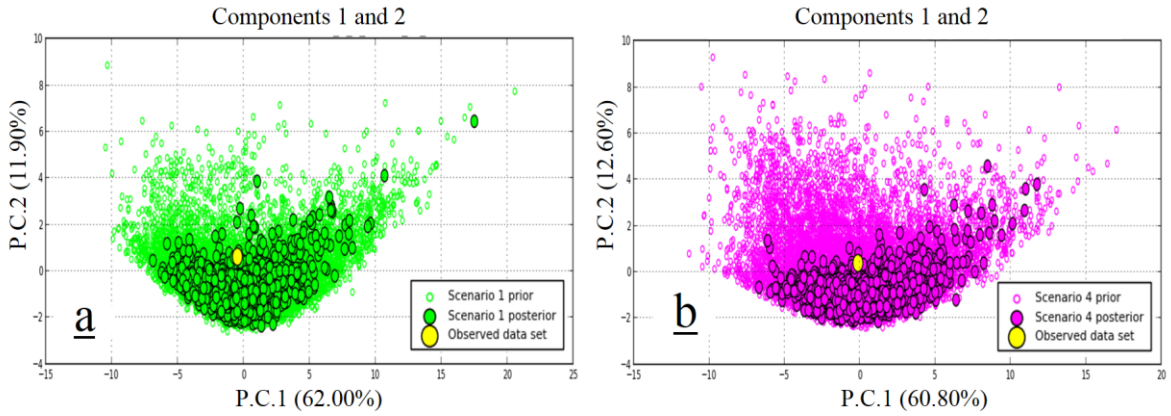


Fig. S1: Principal Component Analysis (PCoA) of Run A (a) and B (b) in DIYABC analysis (Cornuet *et al.* 2014).

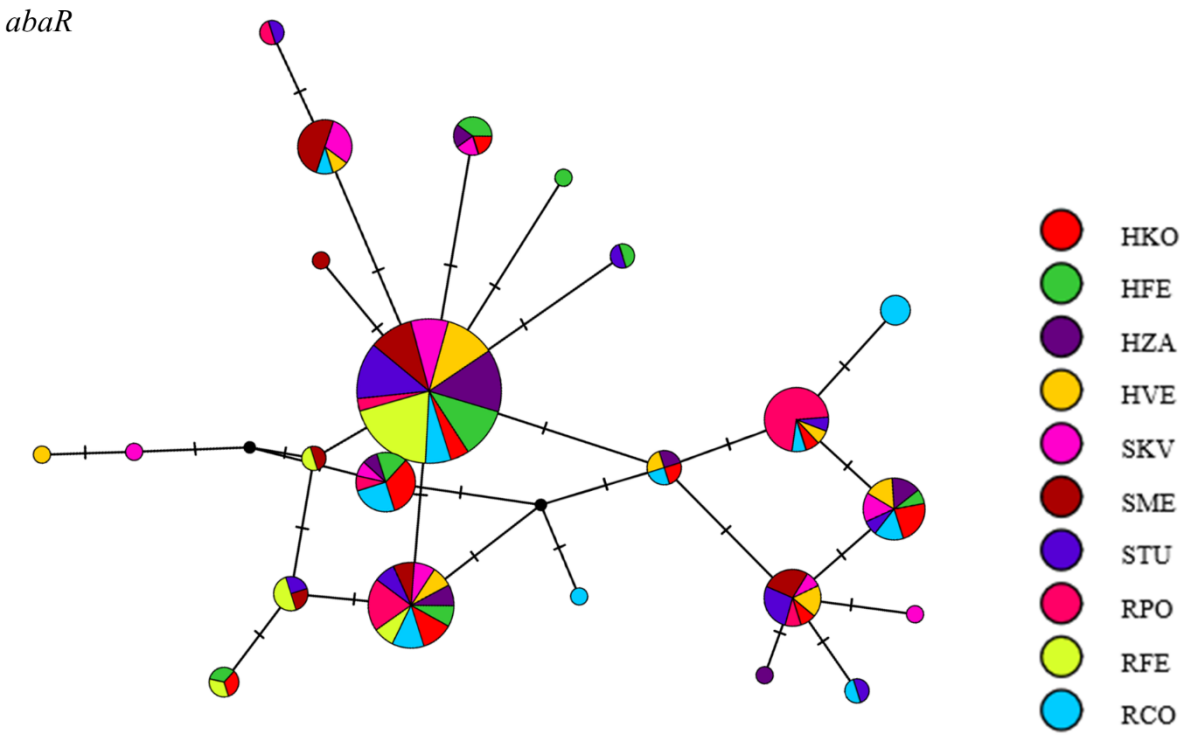


Fig. S2: Haplotype network obtained from the TCS analysis of *abaR* candidate loci sequences. The size of the circle represents the frequency of each haplotype. Different colors correspond to the different populations. Black dots indicate missing intermediate haplotypes that were unobserved in the analyzed sample set. Hash marks on branches represent mutation steps (number of base pair changes) between haplotypes.

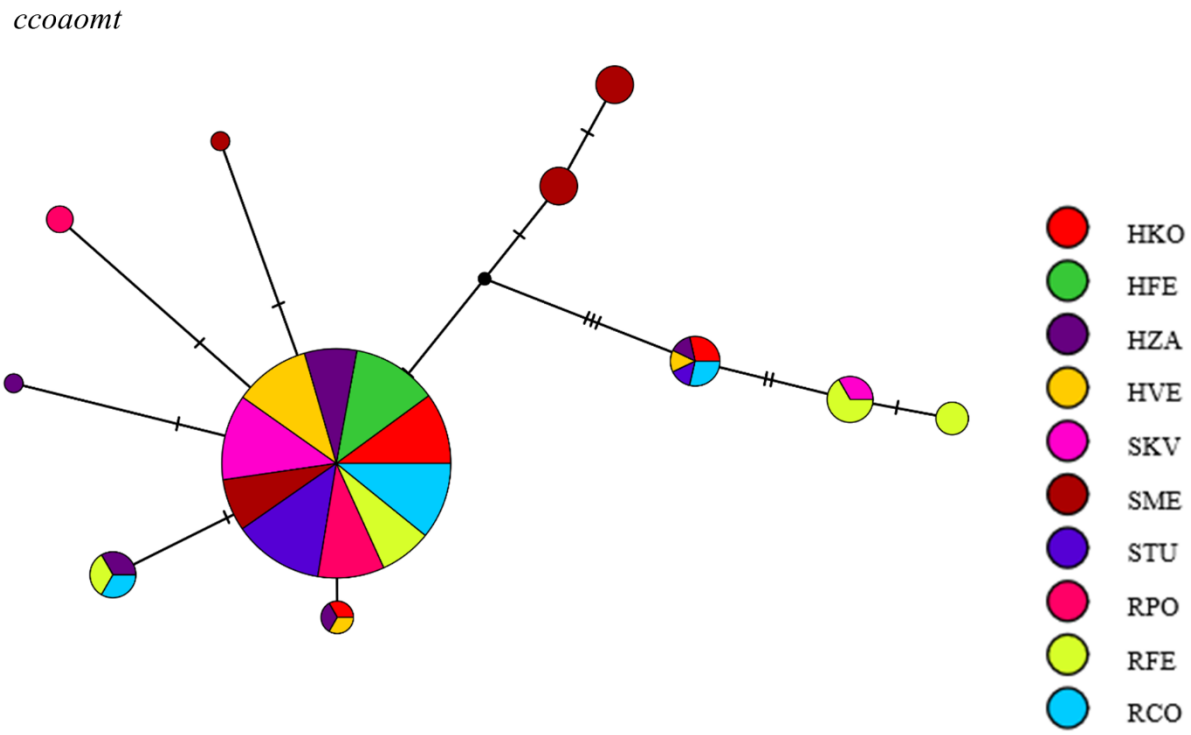


Fig. S3: Haplotype network obtained from the TCS analysis of *ccoaomt* candidate loci sequences. The size of the circle represents the frequency of each haplotype. Different colors correspond to the different populations. Black dots indicate missing intermediate haplotypes that were unobserved in the analyzed sample set. Hash marks on branches represent mutation steps (number of base pair changes) between haplotypes.

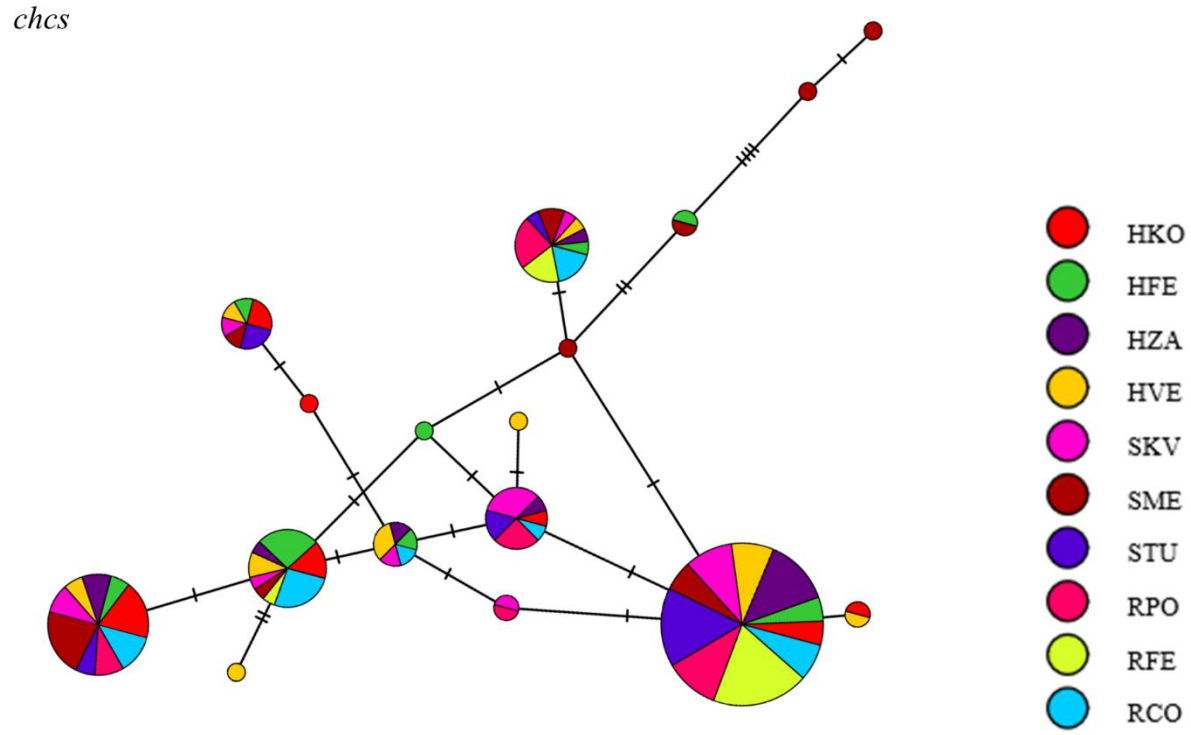


Fig. S4: Haplotype network obtained from the TCS analysis of *chcs* candidate loci sequences. The size of the circle represents the frequency of each haplotype. Different colors correspond to the different populations. Black dots indicate missing intermediate haplotypes that were unobserved in the analyzed sample set. Hash marks on branches represent mutation steps (number of base pair changes) between haplotypes.

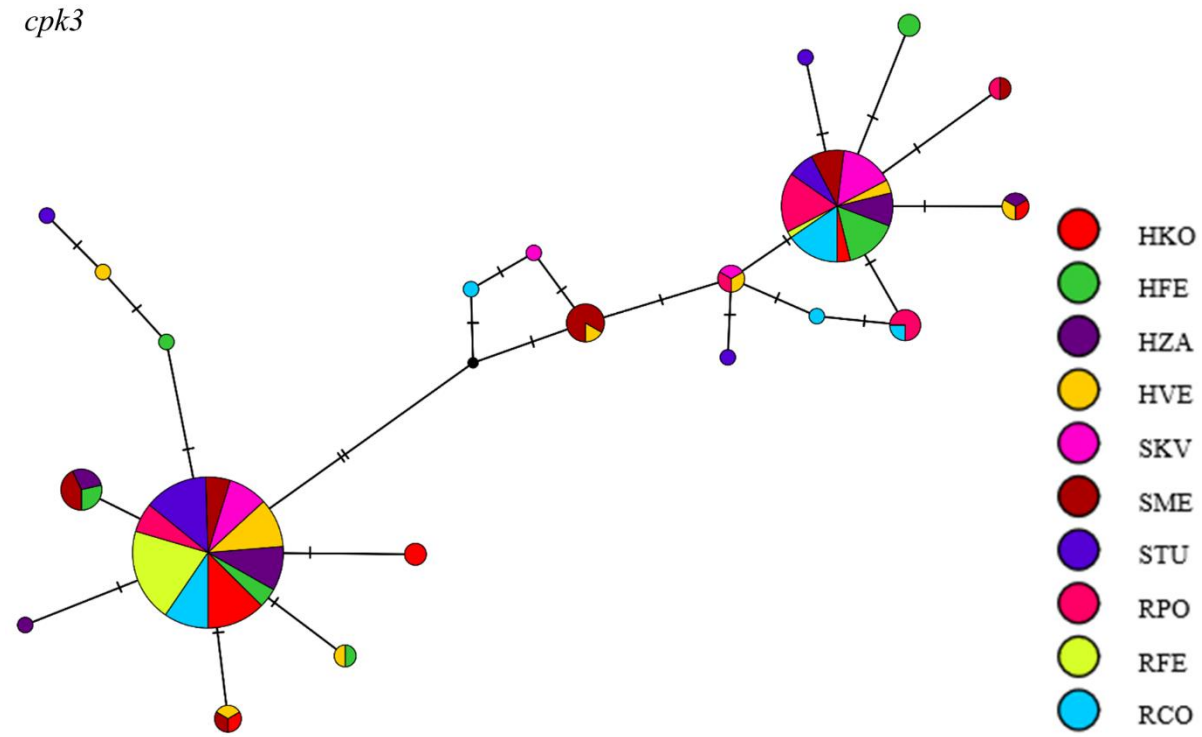


Fig. S5: Haplotype network obtained from the TCS analysis of *cpk3* candidate loci sequences. The size of the circle represents the frequency of each haplotype. Different colors correspond to the different populations. Black dots indicate missing intermediate haplotypes that were unobserved in the analyzed sample set. Hash marks on branches represent mutation steps (number of base pair changes) between haplotypes.

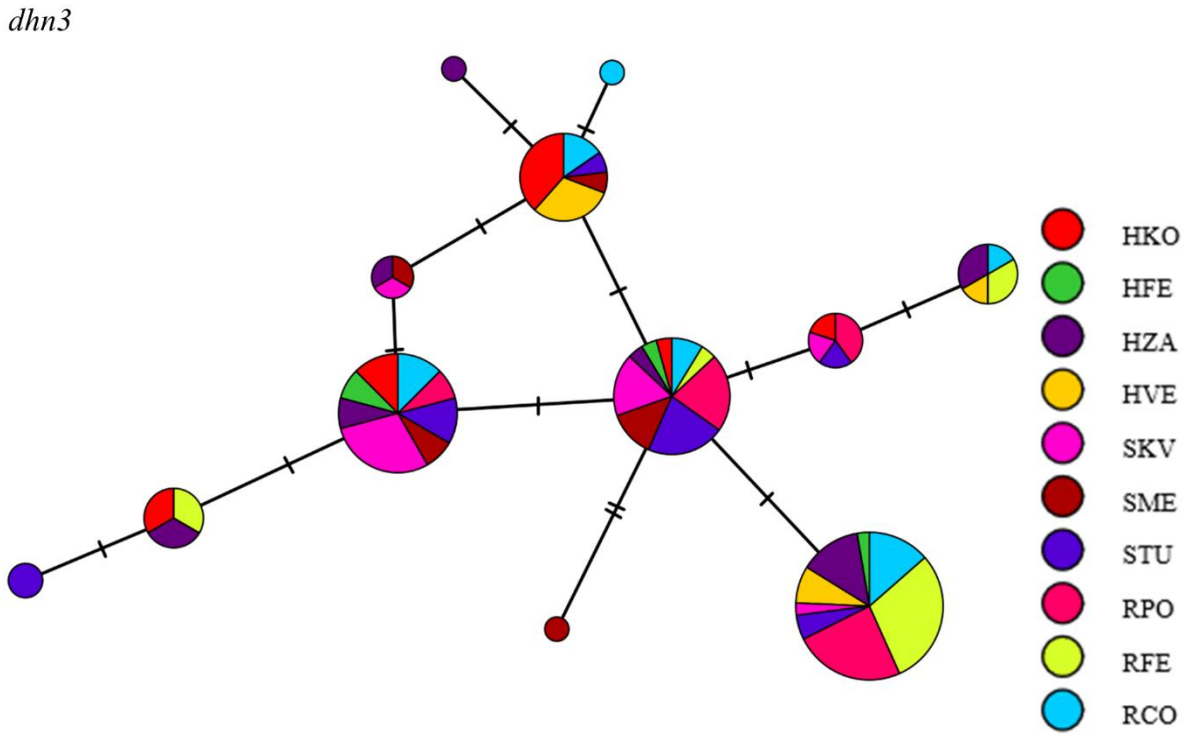


Fig. S6: Haplotype network obtained from the TCS analysis of *dhn3* candidate loci sequences. The size of the circle represents the frequency of each haplotype. Different colors correspond to the different populations. Black dots indicate missing intermediate haplotypes that were unobserved in the analyzed sample set. Hash marks on branches represent mutation steps (number of base pair changes) between haplotypes.

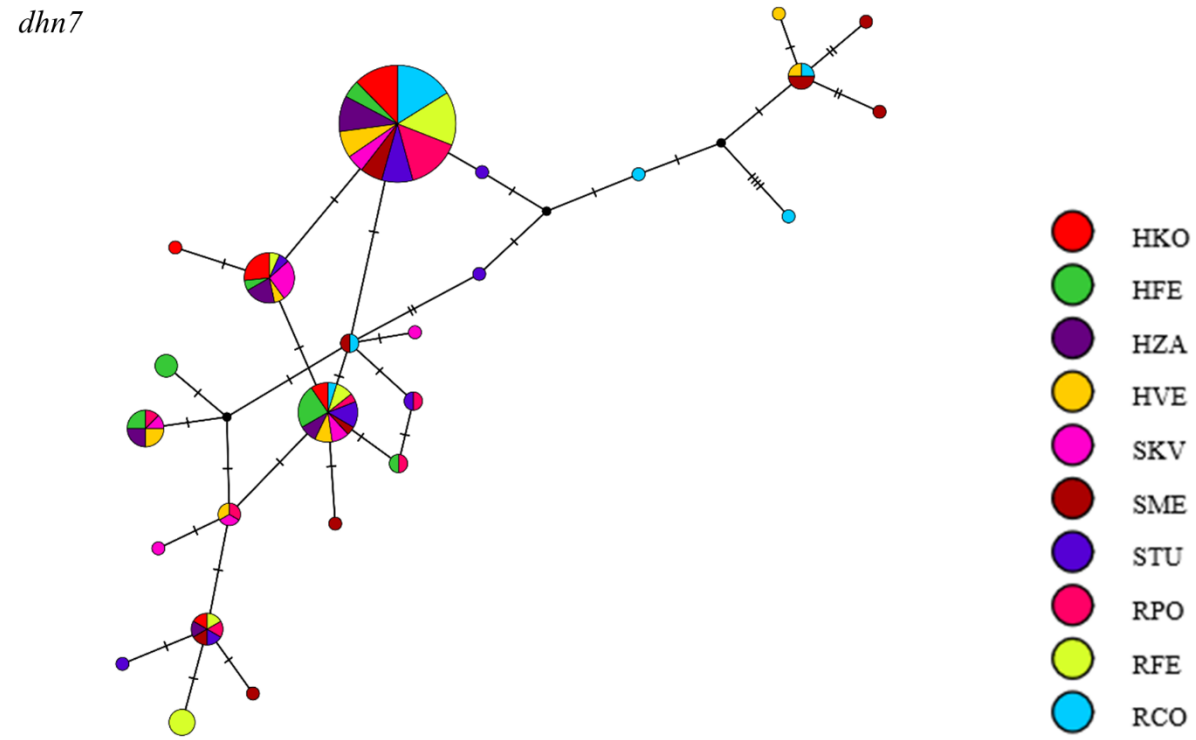


Fig. S7: Haplotype network obtained from the TCS analysis of *dhn7* candidate loci sequences. The size of the circle represents the frequency of each haplotype. Different colors correspond to the different populations. Black dots indicate missing intermediate haplotypes that were unobserved in the analyzed sample set. Hash marks on branches represent mutation steps (number of base pair changes) between haplotypes.

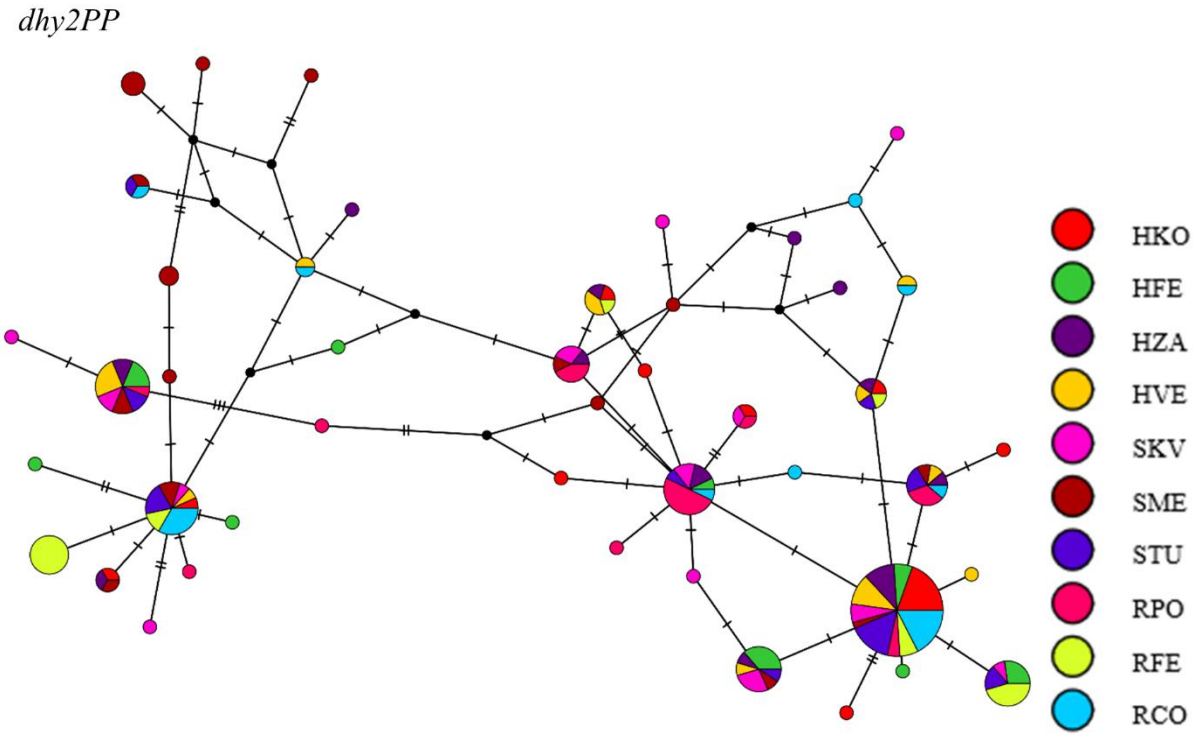


Fig. S8: Haplotype network obtained from the TCS analysis of *dhy2PP* candidate loci sequences. The size of the circle represents the frequency of each haplotype. Different colors correspond to the different populations. Black dots indicate missing intermediate haplotypes that were unobserved in the analyzed sample set. Hash marks on branches represent mutation steps (number of base pair changes) between haplotypes.

erd3

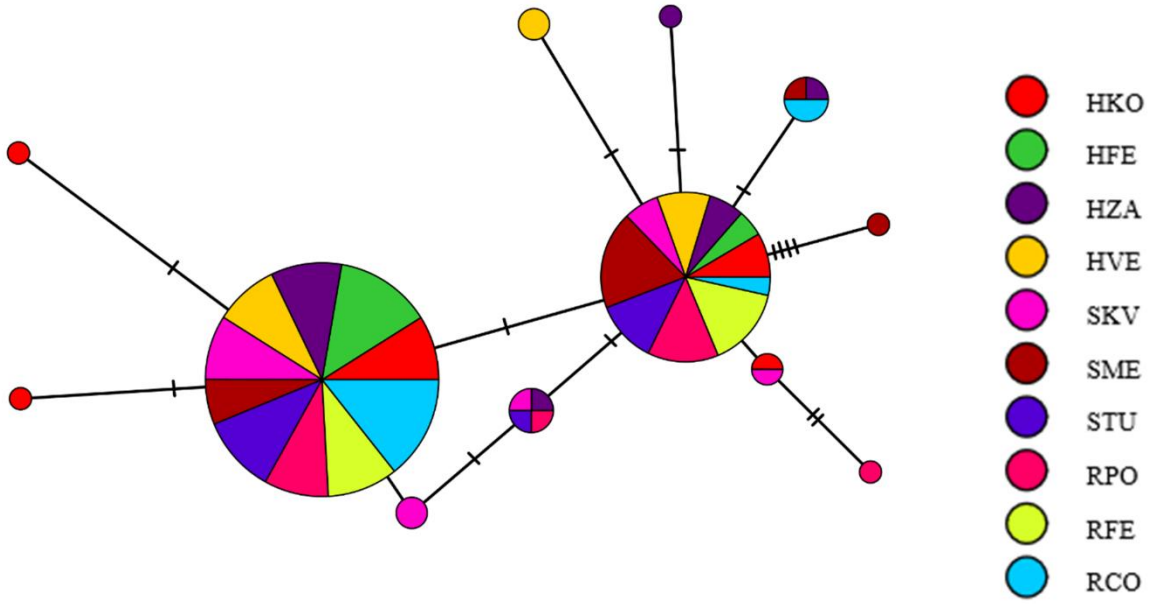


Fig. S9: Haplotype network obtained from the TCS analysis of *erd3* candidate loci sequences. The size of the circle represents the frequency of each haplotype. Different colors correspond to the different populations. Black dots indicate missing intermediate haplotypes that were unobserved in the analyzed sample set. Hash marks on branches represent mutation steps (number of base pair changes) between haplotypes.

pal1

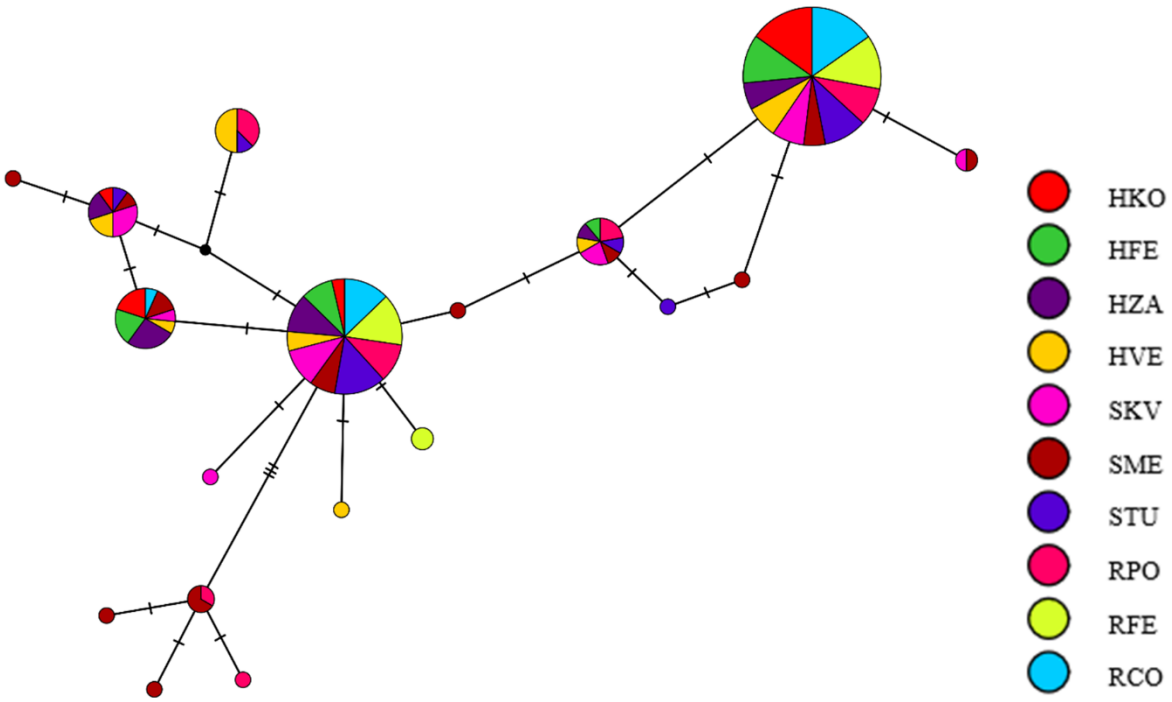


Fig. S10: Haplotype network obtained from the TCS analysis of *pal1* candidate loci sequences. The size of the circle represents the frequency of each haplotype. Different colors correspond to the different populations. Black dots indicate missing intermediate haplotypes that were unobserved in the analyzed sample set. Hash marks on branches represent mutation steps (number of base pair changes) between haplotypes.

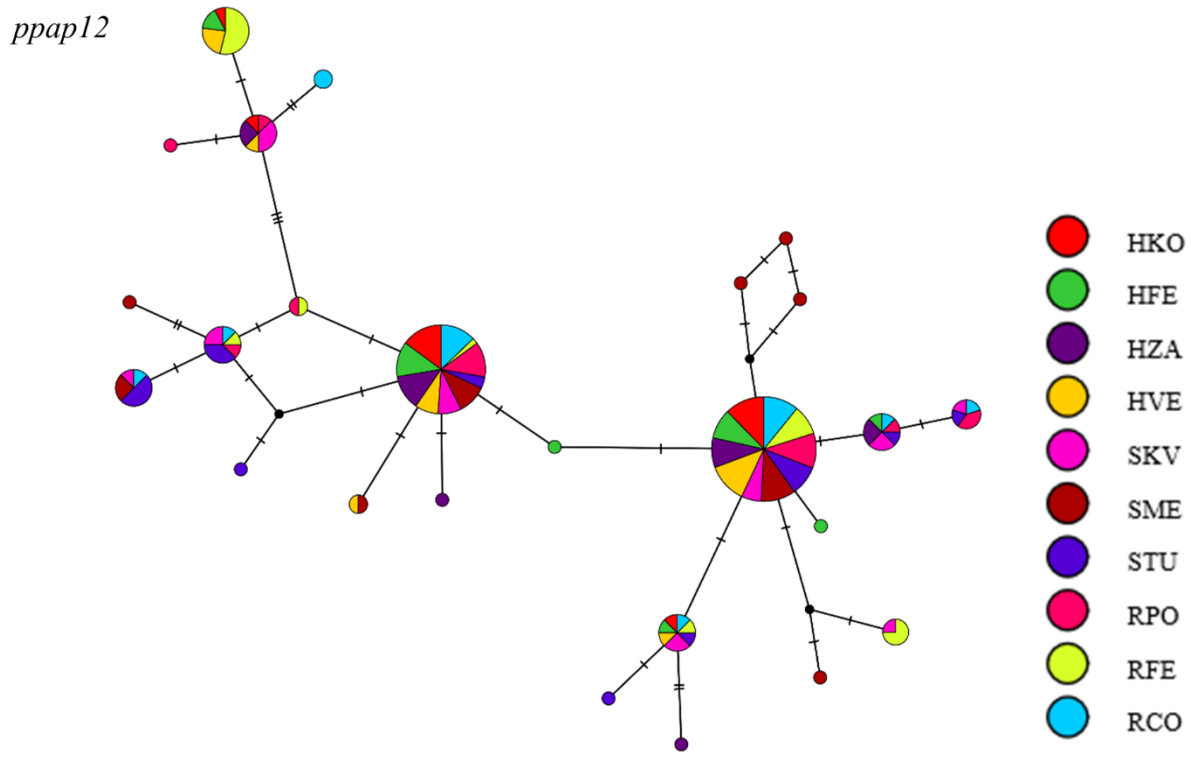


Fig. S11: Haplotype network obtained from the TCS analysis of *ppap12* candidate loci sequences. The size of the circle represents the frequency of each haplotype. Different colors correspond to the different populations. Black dots indicate missing intermediate haplotypes that were unobserved in the analyzed sample set. Hash marks on branches represent mutation steps (number of base pair changes) between haplotypes.

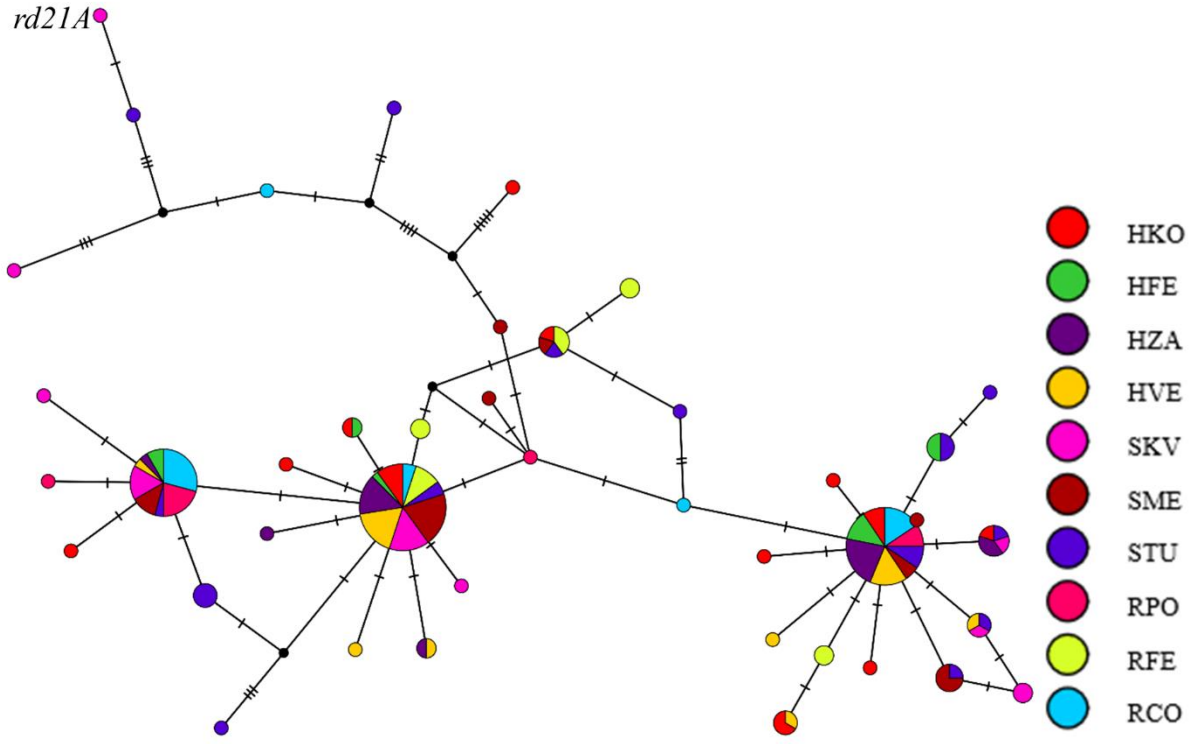


Fig. S12: Haplotype network obtained from the TCS analysis of *rd21A* candidate loci sequences. The size of the circle represents the frequency of each haplotype. Different colors correspond to the different populations. Black dots indicate missing intermediate haplotypes that were unobserved in the analyzed sample set. Hash marks on branches represent mutation steps (number of base pair changes) between haplotypes.

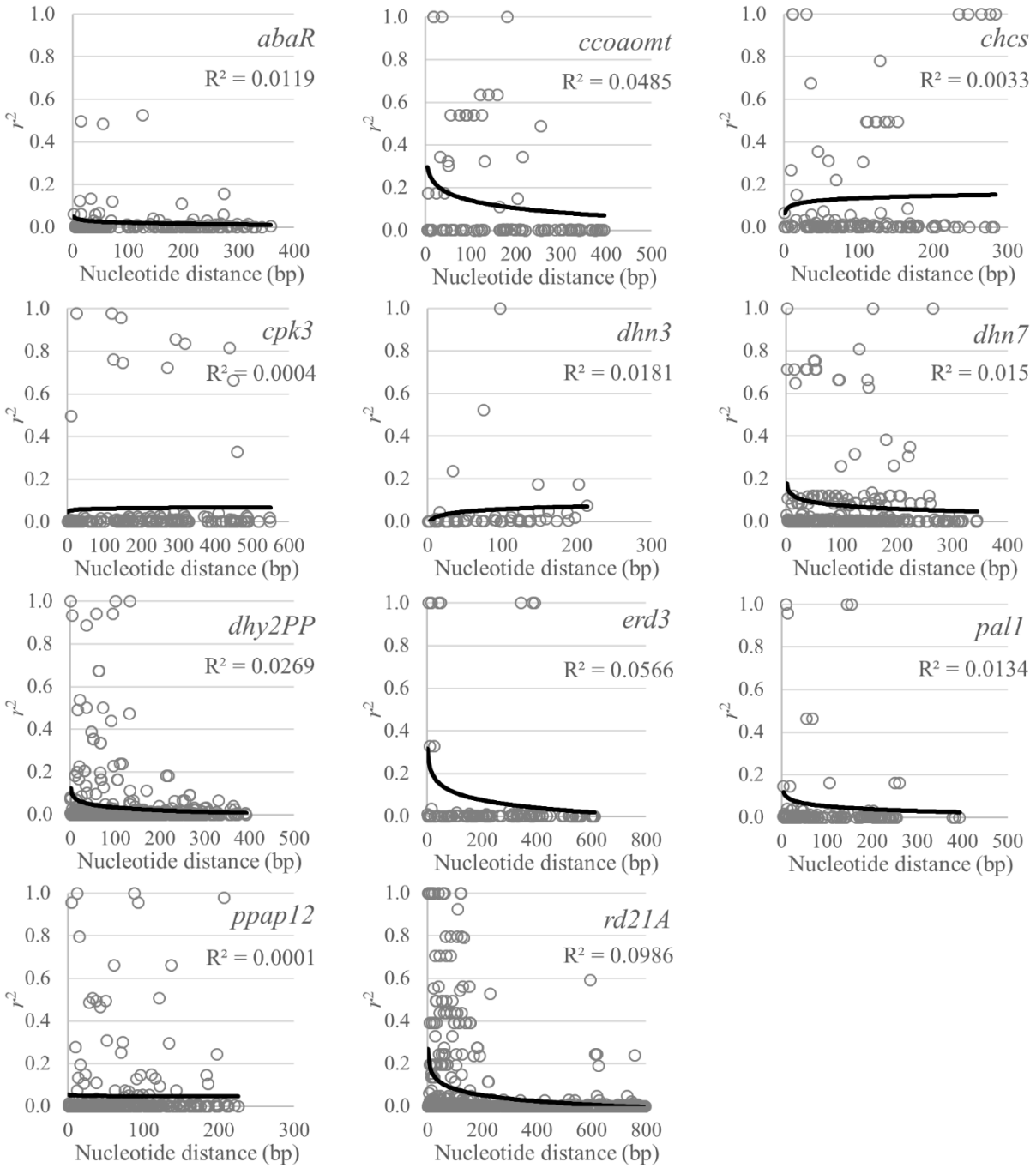


Fig. S13: Scatter plot of the squared correlation coefficient of allele frequencies (r^2) as a function of distance in base pairs between all polymorphic sites detected. Decay of linkage disequilibrium (LD) is shown by a non-linear fitting curve for each of the candidate loci separately.

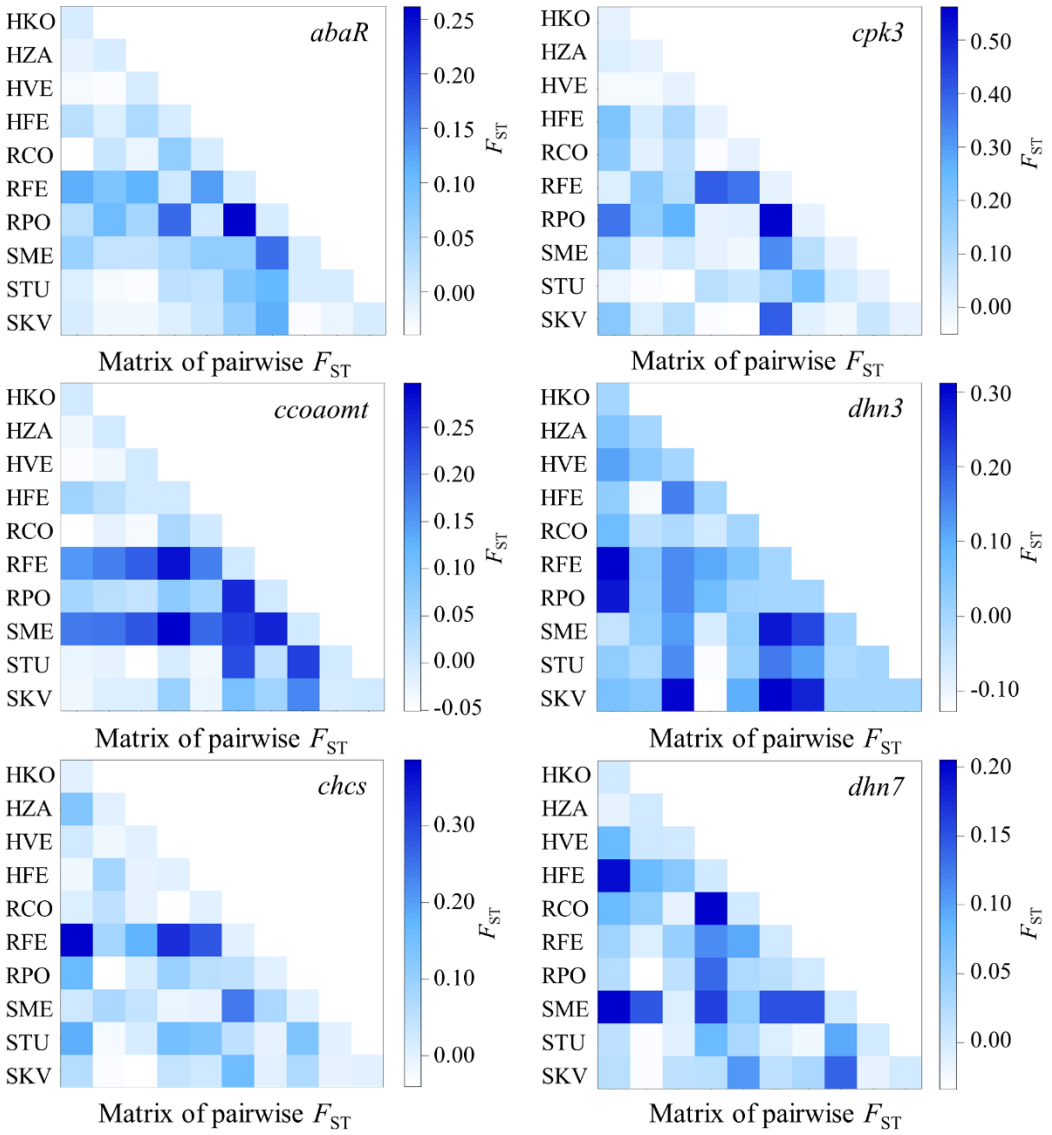


Fig. S14: Heatmap of population pairwise F_{ST} values estimated for candidate gene loci sequence data. Darker colors represent higher F_{ST} values between compared populations.

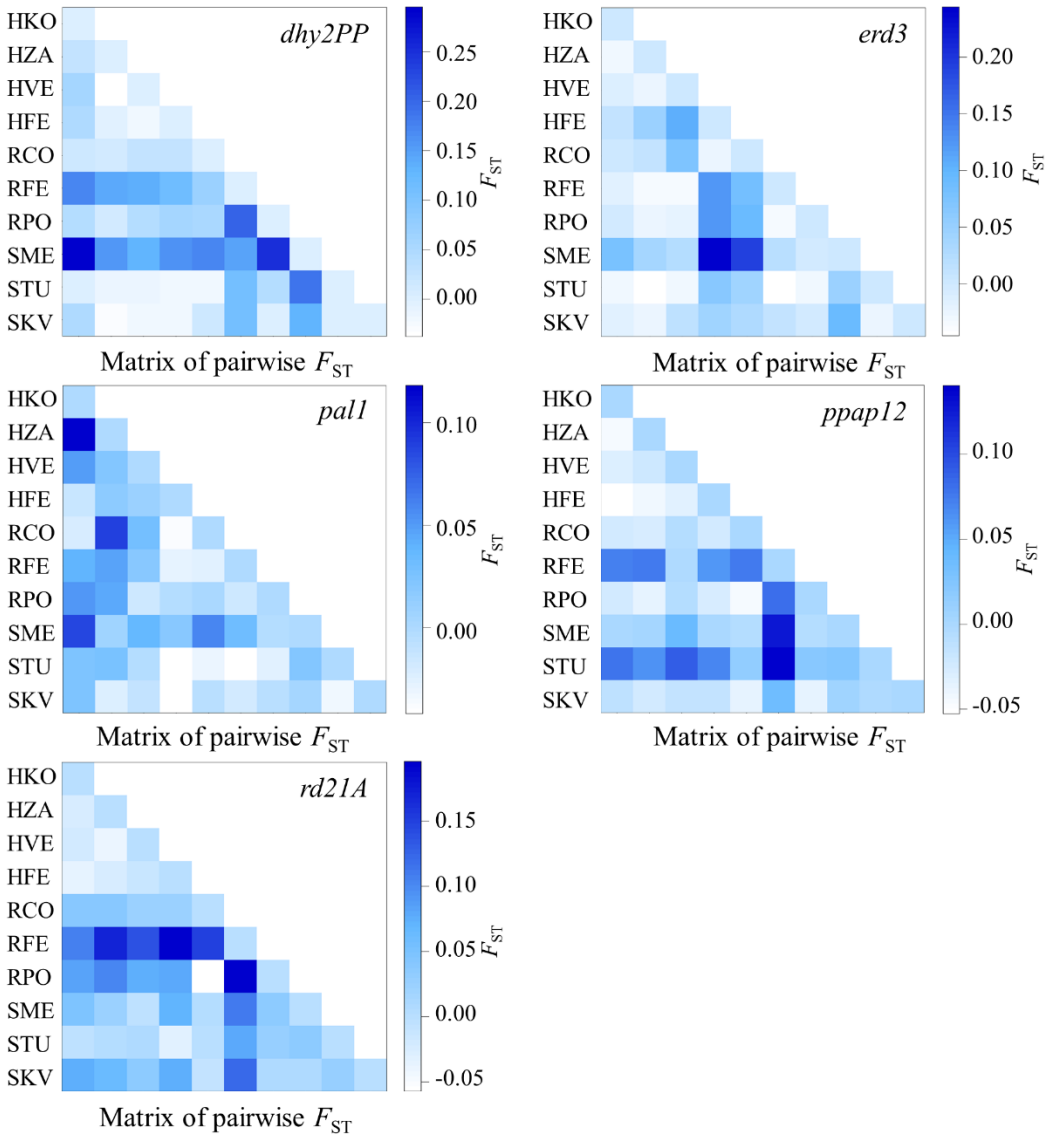


Fig. S14 (Continued): Heatmap of population pairwise F_{ST} values estimated for candidate gene loci sequence data. Darker colors represent higher F_{ST} values between compared populations.

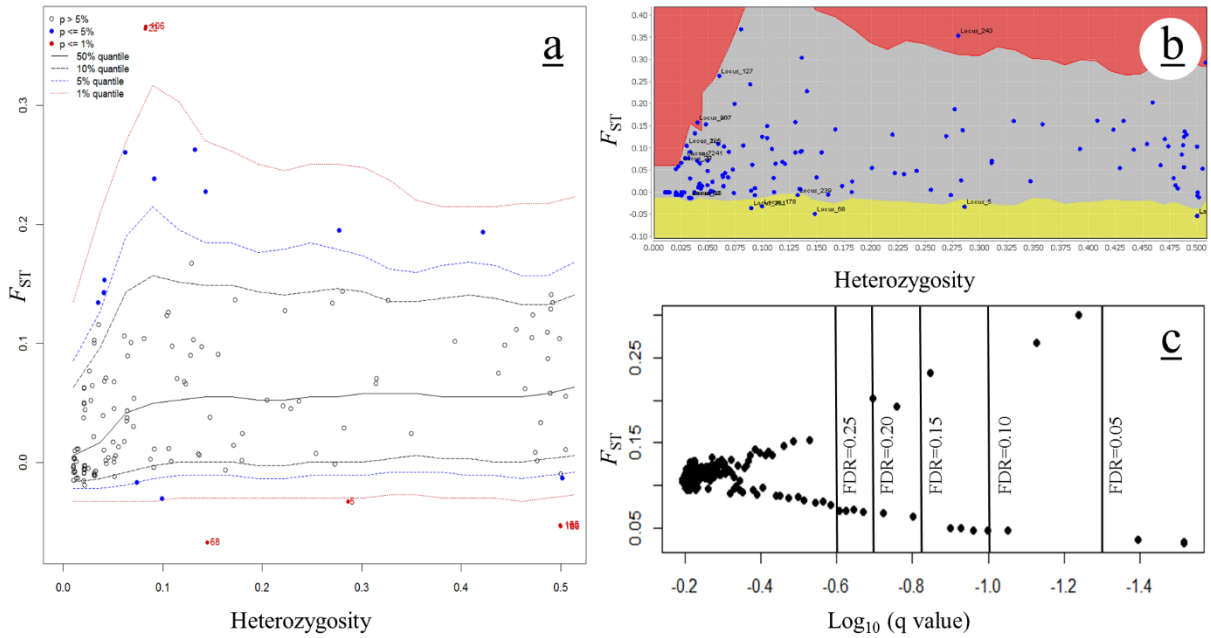


Fig. S15: Candidate loci under selection were identified using two F_{ST} (a; Arlequin by Excoffier and Lischer 2010 and b; Lositan by Antao *et al.* 2008) and one Bayesian (c; BayeScan by Foll and Gaggiotti 2008) based approaches.