Universität für Bodenkultur Wien

University of Natural Resources and Applied Life Sciences, Vienna, Austria
Department of Applied Plant Sciences and Plant Biotechnology

Institute of Applied Genetics and Cell Biology (IAGZ)
Head of Institute: O.Univ. Prof. Dr. Phil. Josef Glössl

GENETIC STRUCTURE OF NATURAL POPULATIONS OF SERBIAN SPRUCE [*PICEA OMORIKA* (PANČ.) PURK.]

Dissertation for obtaining a doctorate degree at the University of Natural Resources and Applied Life Sciences, Vienna, Austria

Under the supervision of
Ao.Univ.Prof. Mag.rer.nat. Dr.nat.techn. Herta Steinkellner

and co-supervision of
O.Univ.Prof. Dipl.-Ing. Dr.nat.techn. Gerhard Glatzel

Submitted by

Jelena M Aleksić

Vienna, July 2008

Acknowledgement

It seems that these days the clock is ticking faster and faster and we have to be quick and very nimble to keep up with the present because the very next moment we might realize that we are already in the past. This holds particularly for the science and especially in the science it seems that the majority of workers are replaceable.

But, I do not think so. There are people and there are moments which mark the science and which mark one's life forever. From the bottom of my heart, I truly respect those people and those moments. They are making our lives rich.

I am very pleased because I have a chance to express my deepest gratitude to all people which helped me in one way or another during the last three years and enabled me to come to this point when I am supposed to write this acknowledgement as the last, but not the least, part of my dissertation.

Firstly, I would like to thank my supervisor, Ao.Univ.Prof. Mag.rer.nat. Dr.nat.techn. Herta Steinkellner for a great support in a crucial moment, for endless understanding and kind words. Thanks to her, this dissertation has a form as it is now and it is finished. I truly respect that.

Than, I would like to thank O.Univ.Prof. Dipl.-Ing. Dr.nat.techn. Gerhard Glatzel for accepting to be an examiner on a very short notice, and for fruitful suggestions on dissertation improvement.

Now, I would like to express my deepest gratitude to Dr. Berthold Heinze. Berthold is not just a colleague who was endlessly helping me to find a solution for the lab problems, to interpret the results and to erase thousands 'the'-s and 'a'-s and to add new ones. He was

giving me support for the long three years and he is a good friend. There are no such kind words to express my feelings, so I will just say – thank you!

But now, when I want to express my gratitude to my other colleagues from the BFW, I realized that actually, every single person from the BFW helped me in one way or another to come to this point.

DI Michael Mengl was endlessly helping me in the lab, in the field and as a friend. Mag. Barbara Fussi is also my dear friend and we had not just nice times, but we did some interesting foldings and other analyses. She was always there in critical situations. Ing. Thomas Thalmayr was taking care of my wild PC and he is a dear friend too. Ing. Lambert Weissenbacher was of great help in the field and he and Ing. Wilfried Nebenführ are my dear colleagues. Dr. Heino Konrad, always calm, was always ready to help. Dr. Silvio Schueler was of great help in interpreting results and analyzing the data. DI Karl-Manfred Schweinzer was helping in making a great map. DI Dr. Georg Frank gave a final touch to the conclusion. Karin Robitchek was of great support all time long. My dear roommate, Ing. Gerald Golesh, already told me that he will miss me, as well as Ing. Christian Wurzer. Also, Renate Slunsky, DI Rudolf Litschauer, DI Karl Sieberer, DI Ilse Strohschneider were always friendly and helpful. Well, it would be really easier to provide a BFW address book and to write a million thanks there.

Ao.Univ.Prof. Dipl.-Ing. Dr.nat.techn. Christian Stauffer was of great support all time long and he helped me in finalizing the last analysis.

Many thanks to the Dean, DDr. Markus Gerhold and Petra Kranyak.

Katharina Engel and whole OEAD team were always friendly and supportive.

My dear colleagues from the National Park Tara – Ing. Duško Milekić and Ing. Marijana Lakić helped me to collect the samples and provided numerous data for this dissertation.

Special thanks to Ormon Sultangaziev who is a dear friend and who gave me his notebook when my wild PC, with life of his own, but with all my data, prematurely and suddenly died five days before submitting this dissertation.

Of course, a great help and support came from my best friends, Mag. Milka Brdar, Milena Brdar and DI Neven Jokanović. Without them, I'm not sure that I would make it.

The last, but not the least, I want to express my deepest gratitude to my another supervisor, Univ.Doz. Dipl.-Forstwirt DDr. Thomas Geburek. Thanks to Thomas, I am here today because three years ago he selected me as the best candidate out of 13 countries and awarded me with the first Bioversity International fellowship on forest genetic resources. Together, we decided to focus on *Picea omorika* and we were an excellent team all time long. I learned a lot from him. So, I am truly sorry that he is engaged with other obligations right now and he is not able to finalize this thesis with me.

All samples were collected with permission and support of the National Park 'Tara', Serbia, Ministry of Agriculture, Forestry and Water Management and Forest Enterprise 'Šume Republike Srpske', Republic of Srpska, Bosnia and Herzegovina.

This work is financially supported by the Republic of Austria and Bioversity International. I also want to express my gratitude to Dr. Jozef Turok, Dr. Jarkko Koskela and Dr. Michele Bozzano – a dear colleague who was literally always helping me.

This dissertation is dedicated to my bellowed father, mother, brother and his family.

Sincerely,

Jelena

Zusammenfassung

Picea omorika (Panč.) Purk. steht auf der „Roten Liste gefährdeter Arten“ der IUCN, aufgrund der geringen natürlichen Verbreitung im nördlichen Balkan (West – Bosnien, Herzegovina und Ost – Serbien). Obwohl man seit bereits mehr als 130 Jahre über diese Art bescheid weiß und mehr als 600 Artikel in diesem Zeitraum über *Picea omorika* verfasst wurden (hauptsächlich im ehemaligen Jugoslavien), sind die vorhandenen Daten sehr umstritten. Daher scheint es notwendig, grundsätzliche Fragestellungen wie: genetische Diversität, Reproduktion, Genfluss, evolutionsbiologische Beziehungen zu anderen Fichtenarten, etc. näher zu bearbeiten.

Die genetische Diversität in natürlichen Populationen von *P. omorika* wird mithilfe von fünf hoch polymorphen nuclearen EST-SSRs in 50 Bäumen von zwei Populationen und ein mitochondrialer Marker – das zweite Intron der NADH Dehydrogenase Untereinheit 1 Gen (nad1i477) in 200 Bäumen aus zehn Populationen. Eine überraschend hohe genetische Diversität vergleichbar mit anderen weit verstreuten Fichten trat am nuklearen Locus (H_e = 0.830) auf. Die Variabilität am mitochondrialen Lokus ergibt sich aufgrund der unterschiedlichen Anzahl der 34 bp Minisatelliten. Die genetische Diversität ist sehr gering - H_S = 0.075, H_T = 0.225, aufgrund der Predominanz eines einzelnen Haplotypen A (eine 34 bp Wiederholung). Die Differenzierung zwischen den Populationen an diesem Genort (G_{ST} = 0.668) ist vergleichbar mit jenen Werten von weiter verbreiteten Fichtenarten bei Verwendung eines oder mehrerer mtDNA Genorte.

Aufgrund der unterschiedlichen Muster der räumlichen Verteilung der genetischen Variabilität am nad1i477, werden zwei getrennte Genpools in *P. omorika* angenommen. Der erste Genpool repräsentiert acht Populationen mit einem fixierten Haplotyp A, während der zweite, hoch polymorphe Genpool eine einzige Population ‘Studenac’ (H_S =0.468) repräsentiert. Wird diese Population von den Analysen ausgeschlossen, dann zeigen sich alle populationsgenetischen Parameter einheitlich in dieser Art (H_S = 0.031, H_T = 0.033 and G_{ST} = 0.070). Deshalb muss die Wichtigkeit der Stichprobenentnahme hervorgehoben werden, vor allem für die Einschätzung genetischer Diversität bei Arten die in der Vergangenheit weitläufig verteilt waren und gegenwärtig auf kleinere Gebiete begrenzt sind.

Zwei mitochondriale Introns der Gruppe II – nad1i477 und nad5i230 wurden verwendet, um die evolutionäre Verwandtschaft zwischen *Picea omorika* und anderen Fichtenarten, speziell der Nordamerikanischen Fichten *P. mariana* und *P. rubens*, erneut zu evaluieren. Die Rekonstruktion der Sekundärstruktur von nad1i477 ermöglichte die Identifizierung von mikrostrukturellen Veränderungen, die zum Auftreten von zwei großen Indels (bis zu 1.300 bp) an diesem Genort führten, welche von unterschiedlicher Sequenzorganisation und - variabilität gekennzeichnet sind. Indel A ist aus mehreren Tandem-Wiederholungen von ca. 30bp zusammengesetzt (beinhaltet 34bp und 32bp Minisatelliten, über die schon früher berichtet wurde). Dieses Indel A ist nur in Fichtenarten der Gruppe A vorhanden - in *P.*

omorika, *P abies* und in einigen Asiatischen Fichten. Während innerhalb des Indels B nur gelegentliche Duplizierungen des Wiederholungsmotifs von maximal 30bp beobachtet wurden. Dieses Indel ist nur in Fichtenarten der Gruppe B vorhanden – in *P. smithiana*, *P. schrenkiana* und in allen Amerikanischen Fichtenarten einschließlich *P. mariana* und *P. rubens*.

Darüberhinaus führte die Verwendung dieses phylogenetischen Signals von Längenmutationen entweder mit einfacher oder komplexer Kodierung der Indels zu substantieller Verbesserung der Auflösung, vor allem bei nad1i477, wo die Längenmutationen eine bedeutende Rolle in der Sequenzevolution spielen. Aufgrund beider mtDNA Genorte, ist *P. omorika* näher mit *P. abies* und anderen Fichtenarten der Gruppe A verwandt, als mit *P. mariana* und *P rubens* (und andere Fichtenarten der Gruppe B). Basierend auf mehreren Grundlagen wie morphologischen Merkmalen, heutiger und früherer Verbreitung und zwei mitochondrialen Genorten, ist es höchst unwahrscheinlich, dass *P. omorika* nahe mit *P. mariana* and *P rubens* verwandt ist, trotz der erfolgreichen Kreuzung und nahen Verwandtschaft, die in zwei molekularen Phylogenien der Fichte berichtet wurden.

Schließlich zeigt auch *P. omorika* die komplexe Struktur der genetischen Diversität innerhalb des Balkans – den Hot-Spot für Biodiversität, der für andere Arten in dieser Region gefunden wurde. Sogar innerhalb einer Fläche von ca. 10.000 km^2 kann eine kleine Anzahl von natürlichen Populationen (ca. 30) zwei getrennten Genpools zugeordnet werden, die vor der letzten Eiszeit entstanden sind und durch unterschiedliche evolutionäre Geschichte und Migrationen in der Vergangenheit charakterisiert sind. Daher ist das Ziel, die verbliebenen genetischen Ressourcen von *P. omorika* zu erhalten, die bereits durch vergangene historische Ereignisse eingeengt wurden. Unter Berücksichtigung von extrem limitiertem Samenaustausch in dieser Fichtenart (*Nm* < 1) wird die Einführung von Erhaltungsmaßnahmen in Form von Mindesteingriff (minimum intervention) unterstützt und die Einrichtung einer *ex situ* Kollektion vorgeschlagen, die alle 419 Bäume der Population ‘Studenac’ beinhaltet – als Vertreterin eines zweiten Genpools.

Schlüsselwörter: *Picea omorika*, Mikrosatelliten, Genetische Diversität, Genfluss, Phylogeografie, Phylogenie, Mitochondrium, Intron der Gruppe II, Kodierung der Indels, Sekundärstruktur, Erhaltung der Biodiversität

Abstract

Picea omorika (Panč.) Purk. is IUCN red-listed species due to the limited natural range in northern Balkan (western Bosnia and Herzegovina and eastern Serbia). Although this species is known for more than 130 years and despite the fact than more than 600 articles have been published on this species during that period (however, mainly in former Yugoslavia), available data on this species are exceptionally controversial. Therefore, it seems necessary to shed light on some of the fundamental questions in this species, such as genetic diversity, reproductive strategy, gene dispersal, evolutionary relations to other spruce species, etc.

The genetic diversity in natural populations of *P. omorika* is assessed by utilizing five highly polymorphic nuclear EST-SSRs amplified in 50 trees in two natural populations and a mitochondrial marker – the second intron of the NADH dehydrogenase subunit 1 gene (nad1i477) amplified in 200 trees originating from ten natural populations. Surprisingly high genetic diversity comparable to that found in other widely distributed spruce species is found at nuclear loci (H_e = 0.830). The variability at mitochondrial locus is found to be due to the variable number of a 34 bp minisatellite and genetic diversity is found to be very low - H_S = 0.075, H_T = 0.225, due to the predominance of a single haplotype A (one 34 bp repeat). However, the among population differentiation at this locus (G_{ST} = 0.668) is found to be comparable to that reported in widely distributed *Picea* species using one or even more different mtDNA markers.

Due to the specific pattern of the spatial distribution of genetic variability at nad1i477, two separate gene pools in *P. omorika* are assumed. The first gene pool, represented by eight populations, is found to be fixed for the haplotype A, while the second, highly polymorphic gene pool is found to be represented by a single population ‘Studenac’ (H_S =0.468). When this population is omitted from the analyses, all population genetics parameters indicated genetic uniformity in this species (H_S = 0.031, H_T = 0.033 and G_{ST} = 0.070). Therefore, the importance of the sampling scheme is highlighted, especially in assessing genetic diversity in species widely distributed in the past which are currently confined to small areas.

Two mitochondrial group II introns – nad1i477 and nad5i230 are used for the re-evaluation of the evolutionary relations between *P. omorika* and other spruce species, especially North American spruces *P. mariana* and *P. rubens*. Reconstructions of the secondary structure of nad1i477 enabled the recognition of microstructural changes leading to the

occurrence of two large indels (up to 1.300 bp) at this locus, characterized by essentially different sequence organization and variability. Indel A, comprised of several tandem repeats of approx. 30 bp (including previously reported 34 and 32 bp minisatellites) is found only in group A spruce species – *P. omorika*, *P abies* and some Asian spruces, while within the indel B, only occasional duplications of a maximum 30 bp repeat motifs are observed. This indel is found only in group B spruce species – *P. smithiana*, *P. schrenkiana* and all American spruces, including *P. mariana* and *P. rubens*.

Furthermore, the utilization of the phylogenetic signal in length mutations by coding indels either by simple or complex indel coding substantially improved the resolution, especially at nad1i477 in which the length mutations are found to play a major role in sequence evolution. Based on both mtDNA loci, *P. omorika* is found to be closely related to *P. abies* and other group A spruce species rather than to *P. mariana* and *P rubens* (and other group B spruce species). Therefore, based on several lines of evidence, such as morphological features, current and past distributions and two mitochondrial loci, it is highly unlikely that *P. omorika* is closely related to *P. mariana* and *P rubens*, despite their successful crosses and close relations reported in two recently spruce molecular phylogenies.

Finally, the complex structuring of genetic diversity within the Balkan – the biodiversity hot-spot found in other species confined to this region is shown in *P. omorika* as well, because even within an area of approx. 10.000 km^2, a small number of natural populations (approx. 30) can be assigned to two separate gene pools of the pre-glacial origin, characterized by different evolutionary history and past migrations. Therefore, in aim to preserve the remaining genetic resources in *P. omorika*, already narrowed by past historical events and taking into account exceptionally limited seed flow in this species ($Nm < 1$), introduction of conservations measures through minimum intervention is supported and the establishment of an *ex situ* collection comprised of all 419 trees from the population ‘Studenac’ – the representative of a second gene pool, is recommended.

Key words: *Picea omorika*, microsatellites, genetic diversity, gene flow, phylogeography, phylogeny, mitochondrion, group II introns, indel coding, secondary structure, biodiversity conservation

Table of contents

Introduction

The Mediterranean Basin has high genetic variation in terms of number of species, subspecies divisions and allelic richness ('southern richness', Hewitt 2000). This region, and consequently the Balkan, is characterized as one of the biodiversity hotspots for conservation priorities, due to the exceptional concentration of endemic species exposed to the rapid habitat losses (Myers et al. 2000). It is well known that the current distribution of species is modeled not only by ecological features of the area, but also by geological history (Šilić 1990). Due to the pronounced orography and plasticity of the terrain within the Balkan and characteristic orientation of the mountain missives, Pleistocene glaciations were recorded at high mountains, while at the same time numerous species were confined to the refugial areas, usually deep canyons and other protected habitats (Šilić 1990). As a result, the Balkan is referred to as a refugial area during glaciations and the source for all eastern and for many western European species (Hewitt 2000, Petit et al. 2005). Wide spectrum of endemic and relict plant species present within the northern Balkan and within the area of the former Yugoslavia is making the flora of this region exceptionally rich (Šilić, 1990).

Despite such a high diversity, the Balkan is still referred to as a 'poorly known' region (cf. Hewitt 2000), due to two main reasons. Firstly, although substantial scientific work has been undertaken during the last century in the eastern European countries, this work was scarcely available to the worldwide scientific society. In the past period, scientific publications, at least those originating from the former Yugoslavia, were usually written in local languages or in German. Still, they were not broadly available because the exchange of scientific information was not as easy as today. Secondly, some areas of the former Yugoslavia were unexplored until relatively recently due to the inaccessibility of the terrain. Therefore, several new species endemic to this region were discovered during the last century or so. *Parietaria serbica* (Panč.), *Haplophyllum boissieranum* Vis. & Panč., *Eryngium serbicum* Panč., *Stachis anisochila* Vis. & Panč., *Ramonda nathaliae* Panč. & Petr., *Valeriana bertiscea* Panč., *Scabiosa fumarioides* Vis. & Panč., *Campanula secundiflora* Vis. & Panč. and *Reichardia macrophylla* Vis. & Panč. are only few plant species endemic to Serbia which were discovered by the Serbian botanist Josif Pančić at the end of the XIX century. But, maybe the most remarkable example was the discovery of a new conifer species – Serbian spruce [*Picea omorika* (Panč.) Purk.], which became a synonym for the Professor Pančić's life accomplishment.

Serbian Academy of Sciences and Arts (SANU) organized a conference dedicated to the Prof. Pančić's work and achievements in 1967 and a monography "Pačićev zbornik u spomen 150. godišnjice njegovog rođenja" was published on that occasion. There, Fukarek (1967) described how Prof. Pančić discovered Serbian spruce and in memory and honor to Prof. Pančić, a short overview from the above mentioned publication is presented.

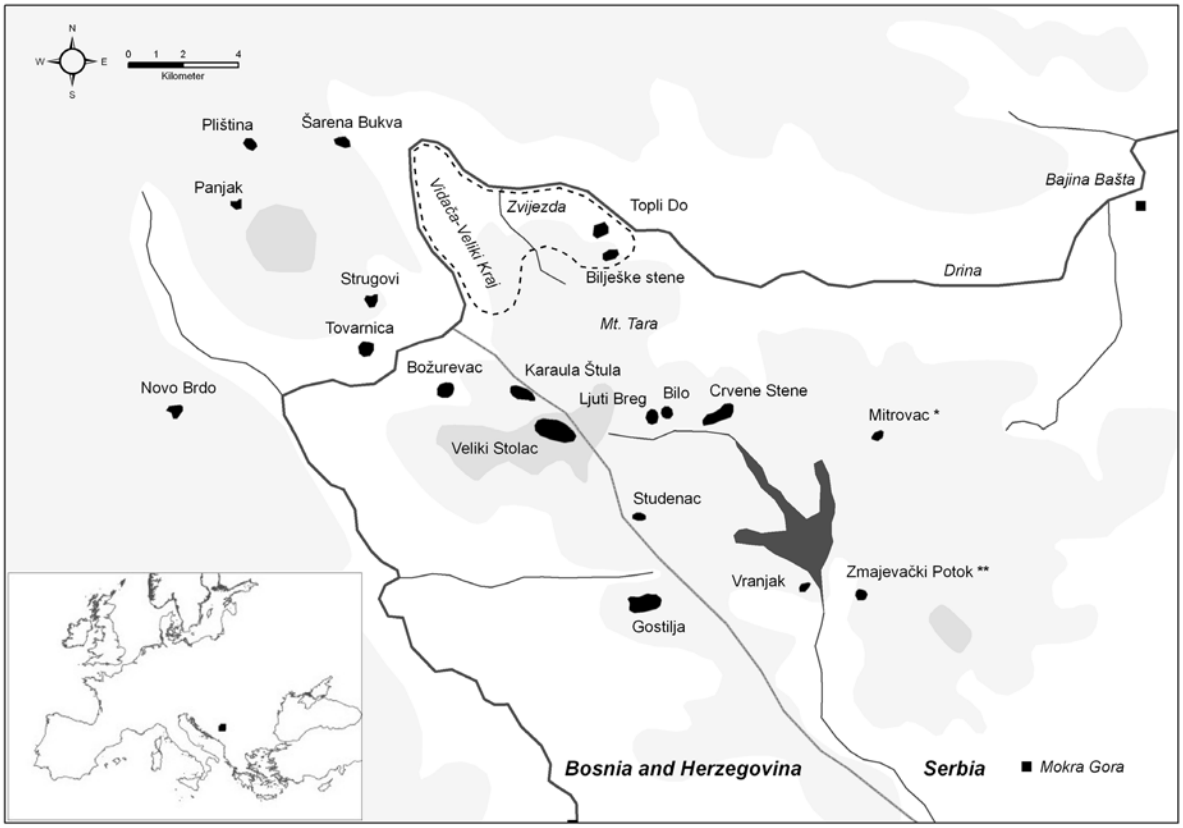
In 1855, during one of the training excursions for students at Mts. Tara and Zlatibor (Serbia), Prof. Pančić heard from the local people about 'omorika', a rare tree that grows in this area, but nobody was able to tell where to find it. Few years later, Prof. Pančić asked the authorities to provide conifer branches from this region for educational purposes. He thought that he might get a branch from this tree as well. Within a large sample he received, two branches - although without cones - were unknown to him and he realized that this was the tree he was looking for. He asked the authorities again to provide the branches, this time with cones, and to describe the exact location of the tree from which the branches were taken. Unfortunately, the location was not recorded. Later on, Prof. Pančić organized several training excursions within this region, but he did not found omorika, although according to the described routs, he was passing by several natural populations. After 20 years of searching, in 1875, Prof. Pančić went again to the Mt. Tara and finally, near the hamlet of Zaovine, he found omorika.

As it was tedious to discover Serbian spruce, it was also difficult to prove to the scientific society that this was a new conifer species. Prof. Pančić sent samples of omorika to several colleagues, among them to Prof. Grisebach, Göttingen, Germany and Prof. Purkyně, Vienna, Austria and all replies were similar - Prof. Pančić should refresh his botanical knowledge, because this was not a new species, but rather *P. obovata*, *P. orientalis*, *P*. *ayanensis* (syn. *P. jezoensis*) or *P. menziessi* (syn. *P. sitchensis*). Indeed, those species are similar to Serbian spruce, but neither of them was omorika. Therefore, Prof. Pančić collected samples from all above mentioned spruces, he described differences between them and Serbian spruce and he published a small monography 'Eine neue Conifere in den oestlichen Alpen' in Belgrade in 1876 (at that time, the northern Balkan was referred as southern Alps, J. Aleksić). Later on, Serbian spruce was acknowledged as a new conifer species (Bolle 1876, Purkyně 1877, Wettstein 1890).

Out of approximately 35 *Picea* species recognized by recent classifications and checklists (Schmidt 1998, Farjon 2001), only Serbian and Norway spruce [*Picea abies* (L.) Karst.] are confined to Europe. While latter species is widely distributed throughout Europe, Serbian spruce is confined to an extremely small area in western Bosnia and Herzegovina and

eastern Serbia (Figure 1). Due to such limited natural range, it has been legally protected in former Yugoslavia in 1964 and has been IUCN red-listed in 1998 (Conifer Specialist Group 1998). The remaining natural populations are confined to the region of approx. 10.000 km^2 where the River Drina meanders Mts. Zvijezda and Tara, e.g. between 43°21' and 44°08' north

Figure 1. The core area of *Picea omorika* natural distribution of approx. 10.000 km^2 in eastern Serbia and western Bosnia and Herzegovina



Within the complex 'Zvijezda-Vidača-Veliki Kraj' occupying approx. 2.000 ha, several smaller population and individual trees of *P. omorika* are present, but only populations 'Bilješke Stene' and 'Topli Do' are presented. Three populations are located south from the core area (not presented). All populations are found at limestone, except for 'Mitrovac*' – found on swampy terrain and 'Zmajevački Potok**' - found on serpentine.

and 18°37' and 19°45' east, and they form a very patchy meta-population. In Serbia, majority of *P. omorika* natural populations are found within a strictly protected area of the National Park

‘Tara’, while other natural populations and individual trees are in private property and are protected by the law. In National Park ‘Tara’, the main objective of conservation is biodiversity conservation with no active intervention and natural populations of Serbian spruce, as untouched forests, are left to free development. This level of protection corresponds to the category I of the Protected Area Management Categories of IUCN and to the category 1.1 of the Ministerial Conference on the Protection of Forests in Europe (MCPFE) classes of protected and protective forests.

In total, more than 30 natural populations comprised of several hundreds to several thousands of trees are found within the core area and only three populations are found south from this region. However, some authors reported much lower number of natural populations and the number of presumably remaining trees (e.g. Nasry et al. 2007). In a comprehensive study on effects of fire on reduction of *P. omorika* natural range conducted for more than ten years, Čolić (1966) described 102 sites on which either large populations, clusters of trees or even solitary trees are present. Some of those tree clusters and individual trees are believed to represent the remains from formerly widely distributed populations, while others are assumed to have arisen from wind dispersed seeds from neighboring populations (Čolić 1966, Gajić 1994). However, the cumulative area of all populations is certainly very small and probably does not exceed 60 ha (Burschel 1965).

In contrast to Norway spruce, Serbian spruce is confined to a warmer climate occupying altitudes ranging from 800 to 1600 m a.s.l. mostly at north–orientated, usually very steep slopes of the hills and the River Drina canyon (Figure 2). All populations are found at limestone, except for ‘Zmajevački Potok’ – found at serpentine, and ‘Mitrovac’ – found at swampy terrain. However, at latter site the natural regeneration is absent and only three remaining trees are found (J. Aleksić, M. Mengl and L. Weissenbacher, pers. observ.). Therefore, this population is currently experiencing extinction and it represents a cluster of trees remaining from a formerly wider population. Although Serbian spruce might form pure stands, it is usually mixed with other species such as Norway spruce, silver fir, black pine, aspen, beech and birch in varying proportions (Figure 3).

According to Vidaković (1991), Serbian spruce attains a height of 30 to 50 m. It has a narrow, pyramidal crown and slender trunk (Figure 4) and is characterized by slow growth rate. Crown density is moderate, lower branches are long, pendulous but with upward turning tips, middle ones are shorter and horizontal, while upper branches are the shortest and ascending. Bark is thin and reddish-brown, scaling-off and fissured (Figure 5). Young shoots are very thin,

Figure 2. *P. omorika* natural population ‘Studenac’

P. omorika trees are marked with yellow label

densely pubescent and grey-brown, while older shoots are glabrous. Buds are minute, 3 - 4 mm long, ovate-acute and not resinous. Needles are flattened, 1 - 2 cm long, up to 2 mm wide, acuminate to rounded, alternate. The upper surface is dark green, glossy and without stomata. The lower surface is marked with two white stomata bands, typical for this species. Needles on the lower side of the shoots are pectinate and leaf venation is parallel.

Serbian spruce is monoecious species. Male strobili are approx. 1 cm long, they are light red and located at the tips of the lower branches (Figure 6a). Female strobili are approx. 3 - 4 cm long, reddish-violet and confined to the upper branches at the very top of the crown (Figure 6b). They usually occur at the end of small twigs or they grow directly from the stem within the first 10 m of the crown tip (Jevtić 1960). Flowering starts at a rather early stage and occurs from April to June, depending on the habitat. Both, pollen and seed are dispersed by wind. The

Figure 3. *Picea omorika* natural population ‘Studenac’

pollen in Serbian spruce is approx. 75 to 87 µm long and approx. 52 to 60 µm wide (Jovančević 1962). Although it is smaller than of Norway spruce, the sedimentation velocity (approx. 5.2 cm s^{-1}) is nearly identical (Eisenhut 1961). Cones are 2 - 6 cm long, 2 - 3 cm wide, pendulous and ovate-oblong (Figure 7). They are violet when they are young while previous year cones are pale brown. The margins are entire to finely dentate. The number of fertile scales per cone is 66 to 90 (Figure 8) and cones are ripening in October and November. Seeds are 1.7 - 3.8 mm long and 1.1 - 2 mm wide, obovate, dark brown, winged. Wings are 5 - 8 mm long and 4 - 6 mm wide. According to Krstić (1950), the absolute weight of seeds with wings (1000 seeds) ranges from 1.50 to 4.95 g, and without wings from 1.10 to 4.10 g. A cone weighs between 1.24 to 9.21 g and 1 kg of cones containes on avarege 248 cones. One kg of larger cones (5 - 6 cm) yields 23.6 g of seeds with wings. Pintarić (1970) found very high and stable values for average germination rate (96%) and germination energy (78%) in seeds old up to three years, while those values decreased to 86.9% and 3.9%, respectively, in seeds stored for 6.5 years. Similar

pattern was observed by Gligorević-Danon (1969). Seedlings usually have 5 - 6 cotyledons which are 10 - 12 mm long.

Figure 4. The typical habitus of *Picea omorika* trees in natural population ‘Zmajevački potok

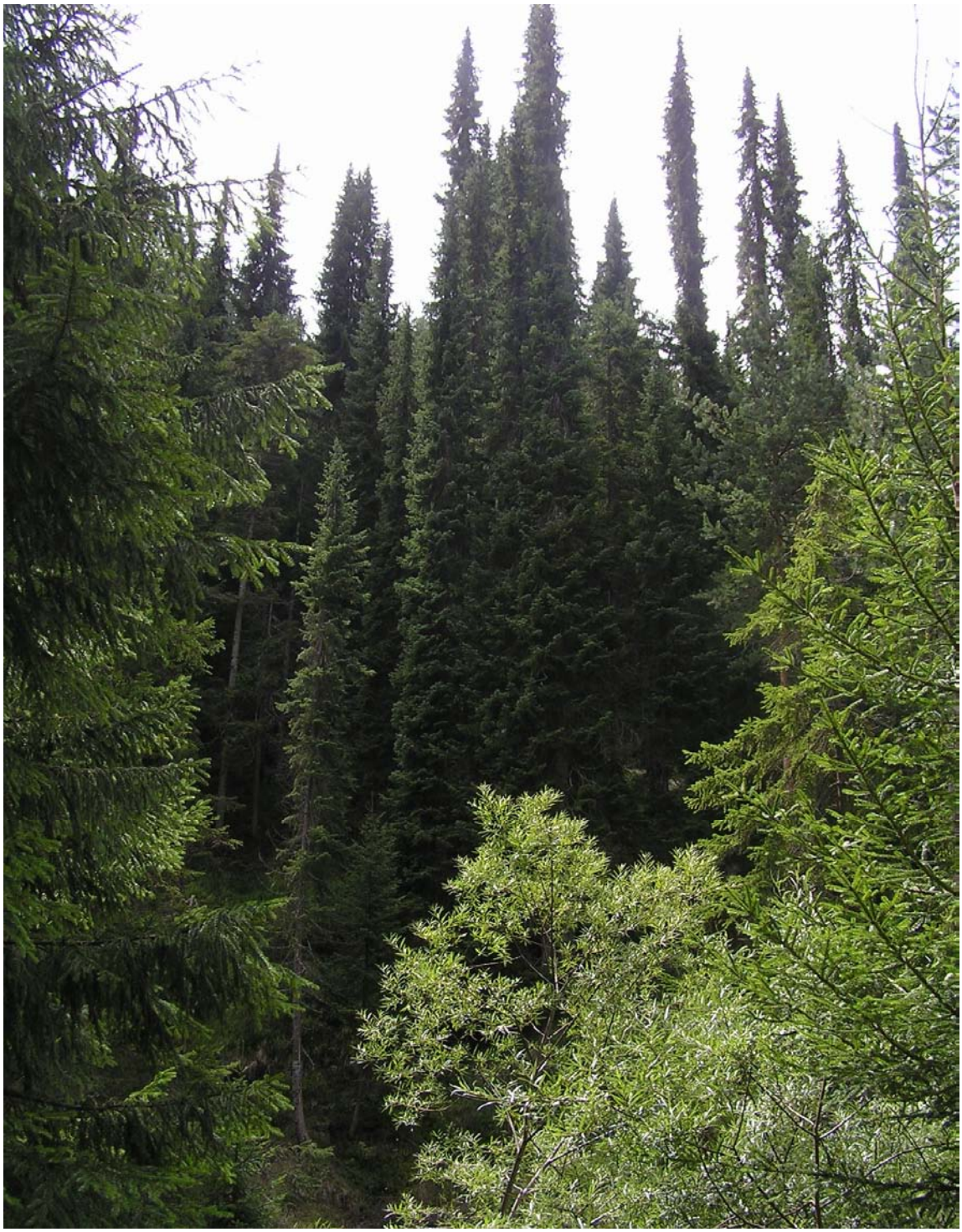


Figure 5. The bark in *Picea omorika*

In Serbian spruce, annual and individual oscillations in fructification (Figures 9a and 9b) are reported (Jevtić 1960, Čolić 1966, Gajić 1994). Although the masting years are observed in two-year intervals, abundant fructification might occur every year (Jevtić 1960). The author reported that fructification depends on the habitat, the age and the location of the tree within the population. For instance, a mature tree in a population 'Crvene Stene' can yield up to 6 kg of cones, while the cone yield in 'Zmajevački Potok' is 700 g per tree. Gajić (1994) reported that planted trees older than 100 years can yield 4 - 8 kg of cones. Investigating the abundance of fructification at 27 sites, Čolić (1966) reported full fructification at 20 sites, while the fructification was found to be absent at 1.5 sites (latter result has to be taken with cautions, because the population 'Mitrovac' was included in the observation and at some sites only few *P. omorika* trees are found, J. Aleksić). However, in plantations, Serbian spruce is characterized as a prolific seeder because almost annually trees of the dominant vegetation layer are shedding

seeds (Král 2002). Experimental data on effectiveness of Serbian spruce to distribute its genes through the wind-dispersed pollen or seeds in natural populations are not available.

Figure 6a. Male strobili in *Picea omorika* (photo M.Vidaković 1991)

Figure 6b. Female strobil in *Picea omorika* (photo M. Vidaković 1991)

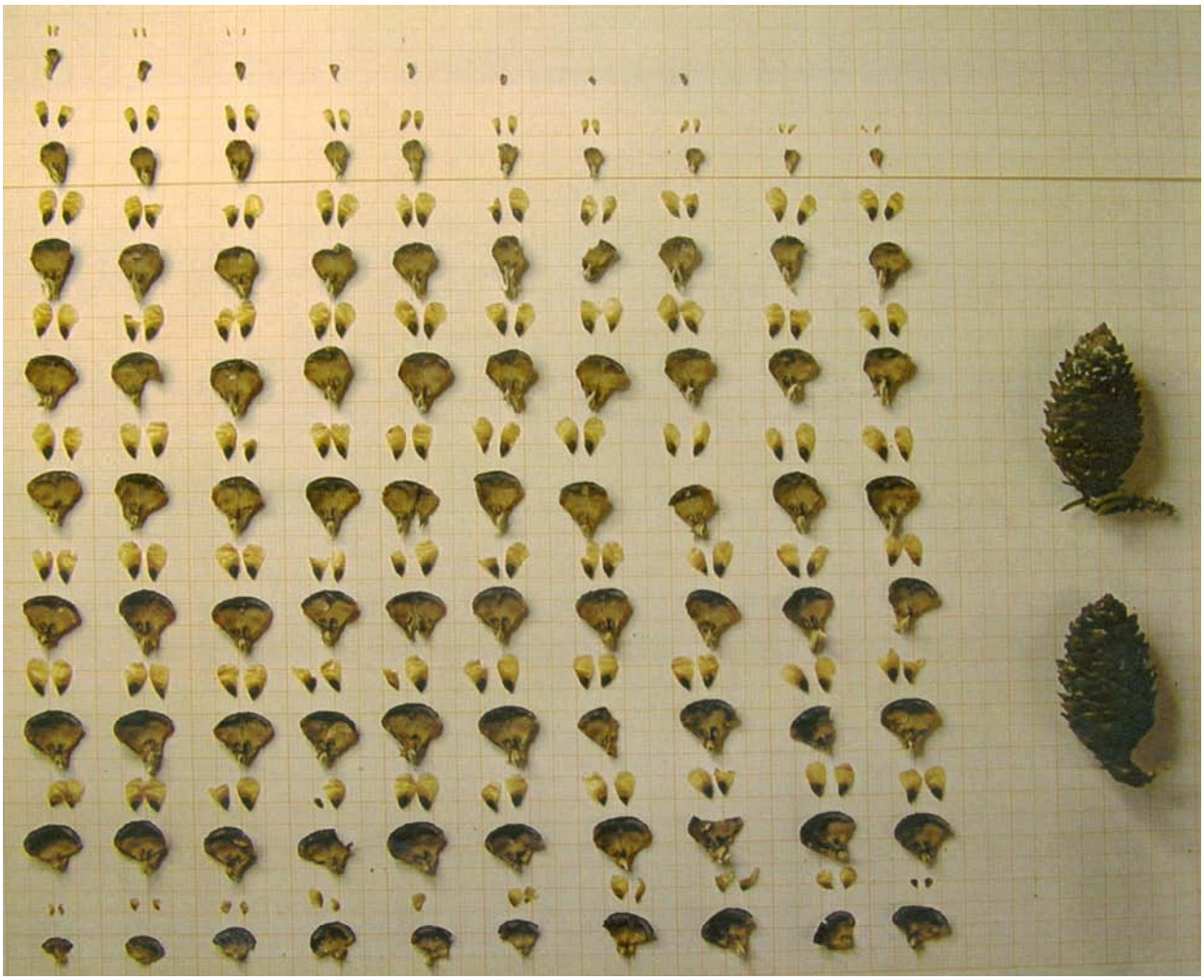
Figure 7. Two *Picea omorika* and a single *P. abies* cone

Čolić (1966) reported that openings after the fire are quickly inhabited by healthy seeds from the surviving trees and by seeds from the neighboring populations dispersed by the wind. Although the serotiny has been reported in Serbian spruce (Rushfort 1987 in Sigurgeirsson and Szmidt 1993), it has never been observed in nature, because the current years cones are shedding seeds during the same year (e.g. in October) and only the empty cones remain at trees during 3 or 4 years.

Serbian spruce reproduces sexually, although vegetative propagation has been reported (Čolić 1966, Vidaković 1991). Čolić (1966) found predominant sexual reproduction at 18 out of 27 investigated sites, while vegetative propagation was recorded at 1.6 sites. Although the number of investigated sites has to be taken with cautions, the author found the absence of natural regeneration at 7 sites.

During Tertiary and later interglacial periods *P. omoricoides* Weber, a close fossil relative of *P. omorika,* was probably widely distributed in Europe, although such early fossil findings are still unclear (Ravazzi 2002). Some macrofossils can be traced back to the Pliocene in former Czechoslovakia (Búžek et al. 1985). Also during the early Pleistocene, a wide distribution of this species is assumed in Europe together with Norway spruce due to the findings of fossil pollen

Figure 8. Cones, scales and seeds in *Picea omorika*



and needles at so distant places as the northern Alpine foreland and the Aegean coast (Ravazzi 2002). With the onset of the last glacial maximum, *P. omorika* was presumably restricted to a smaller region than *P. abies* and outlasted in the Balkan (Ravazzi 2002). After the retreat of the glaciers *P. omorika* could not – contrarily to more competitive species - enlarge its population size probably due to the lack of the migration corridors and/or stepping-stones in close vicinity of its refugium (cf. Ballian et al. 2006). However, fossil findings of *P. omorika*-like pollen in Serbia, e.g. in areas south-east relative to its current range indicate a wider distribution in the Balkan during the last 12 ky (Gigov, 1956). The reduction in natural range in this species is, at least partially, due to the occasional accidental and/or deliberate fire (Čolić 1966).

Figure 9a. The fructification in *Picea omorika* tree from the natural population ‘Bilješke Stene’. Current year cones are violet, previous year cones are pale brown.

Figure 9b. The fructification in *Picea omorika* (trees left and right) and *P. abies* (tree in the middle)

Typical *P. omorika* sites are characterized by high humidity, evenly distributed annual precipitation, heavy snow cover and low winter temperatures (Ostojić 2005). According to Wardle (1956), slow growth and intolerance of competition indicate that this species is not well adapted to its present environment, while Burschel (1965) claims that short branches and slender crowns are making this species well adapted to such conditions. However, the reaction norm of the species was not tested in common garden experiments and therefore its adaptability is still unknown. Contrary to the narrow ecological niche, allochthonous populations can survive even under extremely harsh conditions (Kuittinen et al. 1991) and Serbian spruce was found to be tolerant to early and late frosts (cf. Vidaković 1991) and industrial air pollution (Dallimore and Jackson 1967). Serbian spruce is one of the most commonly cultivated spruce

species in European parks and gardens (Schmidt 1998) and the range of cultivation and ecological amplitude certainly reflect the ability of this species to adapt to different sites.

As a pioneer species, Serbian spruce quickly occupies openings after fire and other catastrophes (Čolić 1966). However, in mixture with more competitive species, such as Norway spruce, it retreats to ravines and other areas less inhabitable by its competitors. Additive genetic variance for shade tolerance is rather low (Tucić and Stojković 2001) and thus natural regeneration of this opportunistic species strongly depends on available favorable habitats and competition by other tree species (Ostojić 2005). Recently, the biggest natural population of Serbian spruce in Serbia – 'Crvene Stene' confined to an area of 50 ha, has been investigated in aim to decipher ecological factors of natural maintenance and restoration (Ostojić 2005). The author reported that in dense forests where Serbian spruce is the main stand forming species, the natural regeneration of this species can be improved by occasional removal of its main competitors, such as Norway spruce, fir and beech. Therefore, the transition from the IUCN category I and MCPFE category 1.1 to categories II and 1.2, respectively, are recommended. However, at sites with open forest where *P. omorika* is not a main stand forming species, the removal of its competitors would actually enhance their natural regeneration rather than natural regeneration of Serbian spruce.

Serbian spruce is a self-fertile, but outcrossing species (Kuitinen and Savolainen 1992) carrying a relative high number of embryonic lethals (Savolainen et al. 1992). This species is assumed to have means other than early acting inbreeding depression to prevent the occurrence of selfed off-spring in mature seed, like spatial separation of male and female strobili, protogyny and earlier maturation of female vs. male strobili (Kuittinen and Savolainen 1992).

Selection against inbreeds usually appears early in the life of conifers, but the timing of the onset of inbreeding depression in Serbian spruce is weak in early development of the zygote, while in latter growth it is as strong as in other conifers (Kuittinen and Savolainen 1992). In an inbreeding study, Geburek (1986) found a surprisingly high level of inbreeding depression in 24-year old Serbian spruce: the growth retardation of selfs compared with that of outcrosses was of the same magnitude as in many other conifers. Langner (1959) did not find differences in full seed percentage of germination between selfs and outcrosses, but noticed a difference in growth in 1- and 2- year old seedlings. Both authors found significant differences in the amount of inbreeding depression between selfed progenies of individual trees. However, in an artificial hybridization study, Tucović and Isajev (1982) found the highest percentage of full seeds and

germination energy in selfed seeds of Serbian spruce and they recommended the utilization of inbreeding in establishing the seed orchard in this species. The only one generative seed orchard in *P. omorika* – 'Godovik', was established based of those findings (Tucović and Isajev 1986), although Langner (1959) suggested either the establishment of seed orchard in *P. omorika* with as many as 100 clones in order to avoid deleterious effects of selfing in progeny, or the establishment of a generative seed orchard in which the attention should be devoted to the origin of parent trees.

Among other factors affecting the levels and spatial distribution of genetic diversity within the species, geographical range appears to have large impact and usually, endemic species tend to become depauperate (Hamrick and Godt 1996). Novák (1927), Langner (1959) and Stern and Roche (1974) have stated that Serbian spruce is morphologically uniform, although Langner (1959) found variation in phenology. Geburek and Krusche (1985) found that Serbian spruce families were less variable in growth characters than *P. sitchensis* (Bong.) Carr. or their hybrids. However, the variability in several morphological features has been reported (Gajić 1994).

Schwerin (1929 in Vidaković 1991) described two varieties – var. serbica – with short branches and narrow-pyramidal crown, and var. borealis – with broad crown, occurring in Germany as a cultivar. Both varieties, as well as var. semidichotomy – with spontaneous dichotomy without visible biotic and abiotic stress, were observed in nature (J. Aleksić, M. Mengl, L. Weiszenbacher, pers. observ.). Although Isajev (1987) and Milovanović (2007) described those and four additional *P. omorika* phenogroups, some of them have been listed by Vidaković (1991) as horticultural forms.

Genetic markers have been used for the assessment of genetic diversity in Serbian spruce in only few studies and the results are surprisingly contradictory - Kuittinen et al. (1991) described a relatively high genetic diversity at allozyme loci in Serbian spruce, while fifteen years later, Ballian et al. (2006) reported reduced allozyme diversity. Nasry et al. (2007) clamed that this species is genetically depauperate based on chloroplast microsatellites.

Apart from the contradictory reports on morphological and genetic diversity in Serbian spruce, it seems that the systematic position of this species is also unresolved. Soon after the discovery of Serbian spruce, Willkomm (1887) originally divided the genus *Picea* into two sections, *Eupicea* and *Omorika* based on the leaf shape. Using the features of the seed cone, the cross-sectional shape of the leaf and the location of the stomata on the leaf, Mayr (1890) recognized 17 species and subdivided the genus into three sections, *Morinda*, *Casicta* and

Omorika. Since than, a number of classification schemes was proposed and all were based on either Willkomm's two-section (with *Picea* and *Omorika* as subsections within the same section) or Mayr's three-section system (Schmidt-Vogt 1977, LePage 2001). However, Serbian spruce was classified into the section/subsection *Omorika* in all classification schemes.

Crossability studies, also used as a measure of phylogenetic relations among spruces, revealed that Serbian spruce easily hybridizes with different members of the genus *Picea* (Wright 1955, Mikkola 1972, Fowler 1980, 1983 and references therein, Ledig 2004). *P. omorika* and its European congener - *P. abies* were found to have different genome sizes (Siljak-Yakovlev et al. 2002) and to hybridize with difficulties (Wright 1955, Fowler 1969 in Nkongolo 1999), corroborating the fact that natural hybrids between those species were never observed in nature. But, based on successful crosses with majority of North American spruces, Serbian spruce was assumed to provide a connection between the two North American spruce complexes, e.g. *P. mariana-rubens* complex and the *P. glauca-engelmannii-sitchensis-mexicana* complex (Ledig et al. 2004) and present, trans-continental distribution of those species is explained by vicariance.

Two recently published spruce phylogenies based on chloroplast RFLP analysis (Sugurgeirsson and Szmidt 1993) and combined chloroplast and mitochondrial markers (Ran et al. 2006) supported the close relationship between Serbian spruce and *P. mariana* and *P. rubens*, although Serbian spruce differs morphologically from those species classified into the section/subsection *Picea*.

Therefore, although Serbian spruce is known for more than 130 years and despite the fact that its bibliography extends over more than 400 papers published in 1878-1965 period (Fukarek 1967) and 178 papers published in 1965-1996 period (Šijak and Dinić 1996), it seems that some fundamental questions, such as morphological and genetic variation, reproductive strategy, gene dispersal, adaptability, systematic position and relations to other spruce species are still unknown.

On the other hand, recent improvements in genetic analysis and genotyping methods have resulted in a rapid expansion of the power of molecular markers to address all of the above mentioned questions (reviewed in Avise 1994, Sunnucks 2000, Zhang and Hewitt 2003, Schlötterer 2004). Nuclear microsatellites (SSRs) have emerged as one of the most popular molecular markers due to their high mutation rate (10^{-4} to 10^{-2} per locus per generation), codominance and high reproducibility, although potential homoplasy is usually highlighted as a

main source causing bias in inferences based on microsatellites (Goldstein and Pollock 1997, Schlötterer 2004, Varshney et al. 2005, Selkoe and Toonen 2006).

SSRs are short segments of DNA in which a specific motif of 1 to 6 bases is repeated up to a usual maximum of 60 repeats (Goldstein and Pollock 1997), although hipervariable loci with 180 repeats have been reported (Primmer et al. 1998). Morgante et al. (2002) found that SSRs are preferentially associated with non-repetitive DNA, Li et al. (2002) reported that dinucleotide repeats account for the majority of microsatellites for many species, while contrary to the most abundant GT/AC motif in mammals, the AA/TT and AT/TA motifs were found to be predominant in plants (Schöltterer 1998). In species with high rates of inbreeding, low population sizes and frequent or severe bottlenecks, typically low average polymorphism and heterozygosity, as well as shorter microsatellites have been recorded (DeWoody and Avise 2000, Neff and Gross 2001).

Levinson and Gutman (1987) suggested slipped-strand mispairing as a major mechanism generating length mutations. The mutational mechanisms underlying the evolution of microsatellites have been studied intensively (reviewed in Ellegren 2004). Traditionally, the infinite allele model (IAM) has been the model of choice for population genetics analysis because it is the simplest and the most general model (cf. Selkoe and Toonen 2006). According to IAM, every mutation event creates a new allele whose size is independent from the progenitor allele. The stepwise mutation model (SMM), originally applied by Ohta and Kimura (1973) is a model specific for microsatellites in which an allele is assumed to mutates up or down by a small number of repeats (Schlötterer and Tautz 1992). Di Renzo et al. (1994) provided evidence that a strict (single-step) stepwise mutation model may not be sufficient to account for allele frequency distribution at microsatellite loci and the authors proposed two-phase mutation model, in which most mutations are single step changes, but infrequent large jumps in repeat numbers are possible.

Traditionally, SSRs have been found by hybridizing repeat-enriched molecular probes in genomic libraries (e.g. van de Ven and McNicol 1996, Pfeiffer et al. 1997, Rajora et al. 2000, Hodgetts et al. 2001; Scotti et al. 2002). But, these methods usually yield relatively few SSRs which give complex multilocus amplification products (Echt et al. 1996, Soranzo et al. 1998) due to the large, repetitive genome in conifers (Pfeiffer et al. 1997; Bérubé et al. 2003). To overcome this obstacle, low-copy libraries (Elsik and Williams 2001) or unique sequences (expressed sequence tags i.e. ESTs) were used to develop robust, low copy SSRs (e.g. Rungis *et al.* 2004, Bérubé *et al.* 2007). Although EST-SSRs are believed to be less variable in comparison to

genomic SSRs, they show less null alleles and, as a major advantage, high level of transferability to related species (Rungis et al. 2004, Varshney et al. 2005). The AT repeat motif was also found to be the most abundant one in EST-SSRs in conifers (Rungis et al. 2004, Bérubé et al. 2007). EST-SSRs have been developed for different species from the genus *Picea* (Scotti et al. 2000, Schubert et al. 2001, Besnard et al. 2003, Pelgas et al. 2004, A'Hara and Cottrell 2007, Rungis et al. 2004), but not for *P. omorika*, although some of them have been used in a single trees for comparative purposes (e.g. Besnard et al. 2003, Rungis et al. 2004).

Plant mitochondrial genome is rather large (e.g. 1.000 kb in conifers, Kumar et al. 1995) and has slow rate of sequence evolution contrary to the frequent structural rearrangements (Palmer et al. 2000, Knoop 2004, Kubo and Mikami 2007). Such structural changes have been proved as useful in reconstructing phylogenetic histories (e.g. Gugerli et al. 2001a), although they make the development of universal mitochondrial primers with wide cross-species transferability difficult (cf. Jaramillo-Correa et al. 2003b). On the other hand, slow rate of evolution at sequence level, small effective population size for maternally inherited genomes (between one half and a quarter that of diploid data) and limited gene flow especially for maternally inherited genomes result in a strong differentiation among populations detected by mitochondrial markers in conifers (Laroche et al. 1997; Jaramillo-Correa et al. 2003b, 2004). Like in most conifers, mtDNA is maternally inherited in Serbian spruce (David and Keathley 1996).

However, the utilization of mtDNA in population genetic studies is possible only if mitochondrial loci exhibit an adequate level of intraspecific variability, which typically is not the case (Schaal et al. 1998; Lunt et al. 1998; Jaramillo-Correa et al. 2003b). An exceptional example is the second intron of the NADH dehydrogenase subunit1 gene (*nad*1 intron b/c). This intron is labelled as nad1i477 according to the nomenclature that numbers organelle introns according to the preceding homologous nucleotide in the continuous reading frame of *Marchantia polymorpha* (Dombrovska and Qiu 2004). Intraspecific variability at this locus has been reported in several *Picea* species, namely *P. abies* (Grivet et al. 1999; Sperisen et al. 2001), *P. mariana, P. glauca*, *P. pungens* (Jaramillo-Correa et al. 2003b, 2004) and *P. crassifolia* (Meng et al. 2007). Most of the variability was due to insertions/deletions (indels) and/or duplications with scarce DNA substitutions. However, in some species with limited natural range this locus was found to be monomorphic, e.g. in *P. rubens* (Jaramillo-Correa and Bousquet 2003a) and *P. chihuahuana* (Jaramillo-Correa et al. 2006). This mtDNA marker has

been recently amplified in several varieties of Serbian spruce (Milovanović et al. 2007), but it has never been tested in trees originating from natural populations.

Observed genetic structuring in plant populations is strongly affected by both, current pattern of microevolutionary forces, such as gene flow and selection, and by the phylogenetic history of populations and species (Templeton 1998, Schaal et al. 1998, Petit et al. 2005). The interaction of these two forces is particularly confounding in plants due to the reticulate nature of gene exchange between diverging lineages (Schaal et al. 2003). Classical models for describing genetic structuring in populations (e.g. Wright's *F* statistics) do not distinguish historical effects from recurrent processes. Estimates of gene flow (*Nm*) derived from those models assume that current population structure reflects an equilibrium between genetic drift and gene flow and they are applicable only for plant species with a stable population structure (Schaal et al. 1998). However, geographic patterns of genealogical structure across the range of species e.g. phylogeographic methods (Avise 2000) can potentially distinguish biogeographic patterns of genetic variation caused by gene flow from those caused by common ancestry (Templeton 1998, Schaal et al. 1998).

But, widespread application of this approach is hampered by a lack of appropriate molecular variation (Schaal et al. 1998). It has been shown that in trees with long life spans, present day genetic structure may still reflect the imprint of the pattern and speed of postglacial migration (Newton et al. 1999) and that mtDNA, due to limited seed dispersal, small effective population size and slow rate of sequence evolution enables identification of migration routes and glacial refugia in some species (e.g. Sperisen et al. 2001, Gugerli et al. 2001b). Therefore, mtDNA is referred to as one of the most powerful tools in phylogeographic studies in conifers (Cruzan and Templeton 2000). Phylogeographic inferences can be enhanced by integrating pallinological data (Cruzan and Templeton 2000) as well as ecological features of the species in question, such as breeding system (Schaal et al. 1998) and dispersal abilities including existence of current and past barriers to dispersal (Templeton 2004). As each species has potentially unique population structure and history that shapes its evolution, designing optimal sampling scheme is crucial in those studies (Templeton 1998).

On the other hand, described specific features of the plant mitochondrial genome caused its relatively scarce utilization in phylogenetic studies (e.g. Wissinger et al. 1991, Bakker et al. 2000, Freudenstein and Chase 2001). Molecular phylogenies at lower taxonomic levels are usually based on more variable non-coding regions, such as the internal transcribed spacers (ITS) of the nuclear ribosomal DNA (nrDNA), which are the most popular non-plastid regions for

species-level phylogenetic studies in plants (Feliner and Rosselo 2007), as well as highly variable chloroplast non-coding regions (Wang et al. 1999, Ran et al. 2006, Shaw et al. 2007).

However, the growing evidence exists proving that length mutations in both, cpDNA and mtDNA group II introns are reliable phylogenetic characters with often low levels of homoplasy (e.g. Wissinger et al. 1991, Bakker et al. 2000, Freudenstein and Chase 2001). Recent advances in understanding group II introns structure and function revealed that they are structurally constrained with a mosaic of highly conserved elements altering with sequence stretches that might be more or less freely evolving (Kelchner 2000, 2002, Löhne and Borsch 2005).

Classification of introns into three groups is based on their conserved RNA folding patterns according to the nomenclature of Michel and Dujon (1983) and Michel et al. (1989). Group II introns are restricted to plants, fungi and certain procaryotes. Primary function of the group II introns is to self-direct their extrication from gene transcripts prior to translation of the mRNA into the protein. Such autocatalytic function is enabled by their specific structure. All group II introns can be folded into the same secondary structure with six main stem-loop domains radiating from a central wheel (Michel et al. 1989). It has been suggested that some domains are more involved in the stability of the secondary structure because very different substitutions rates have been found among different domains (Learn et al. 1992). Thus, owing to the importance of accurate splicing, mutations that disrupt secondary and tertiary interactions are likely to be eliminated by strong selective pressures (Michel and Ferat 1995).

Variable regions in introns (and other non-coding segments) are characterized by relatively high number of length mutations in addition to base substitutions (Clegg et al. 1994, Golenberg et al. 1993, Gielly and Taberlet 1994, Laroche et al. 1997). Those regions are usually confined to the loops of the secondary structure (Kelchner 2000, Löhne and Borsch 2005) and the loop of the domain IV is frequently characterized by pronounced length variability (Löhne and Borsch 2005). Furthermore, mutational hot-spots have been described within this loop (e.g. Laroche and Bousquet 1999, Löhne and Borsch 2005). As discussed by Kelchner (2000), in such regions with pronounced length variability, the homology assessment is very difficult or even impossible.

Laroche et al. (1997) found positive correlation between the number of substitutions and the number of insertions/deletions (indels) per site in six mitochondrial introns and implied their potential phylogenetic use. Incorporation of indel characters in phylogenetic analysis is becoming more important because not coding such length mutations is equivalent to discarding

data (Simmons and Ochoterena 2000, Kelchner 2000). Simmons and Ochoterena (2000) provided the most elaborate treatment of coding indels for phylogenetic utilization. They suggested two alternative ways of indel coding, e.g. simple indel coding (SIC) and complex indel coding (CIC). In SIC, all gaps, regardless of length, are scored as separate presence/absence characters. Sequences with gaps that extend beyond both the 5' and 3' -termini of the gap being coded, as well sequences with gaps that extend beyond one terminus and to the other terminus, are scored as missing data for that character. Gapped positions are then treated as inapplicable for the nucleotide characters. However, this approach does not incorporate all available information and it is less informative than CIC. For CIC, gaps that share a common 5'- or 3'-terminus with another gap, are coded as separate mulistep characters. If two gaps in different sequences are entirely subsumed within a longer gap in a third sequence but do not overlap with one another, a symmetrical step matrix is applied to require two steps between the sequences with the non-overlapping gaps. Several alignment algorithms and software implementations for gap coding are currently available (see Müller 2006 and Simmons et al. 2007 and references therein) and the performance of different indel coding methods has been recently tested in simulations studies (Simmons et al. 2007, Ogden and Rosenberg 2007).

However, phylogenetic utility of mtDNA introns and, consequently, the use of the length mutations is not straight-forward and the risk of inaccurate homology assessment is still present (see Kelchner 2000). Gu and Li (1995) pointed out that mutational mechanisms leading to microstructural changes have to be considered for the correct primary homology assessment. Kelchner (2000, 2002) reviewed mutational mechanisms directing molecular evolution in non-coding regions in chloroplasts (and mitochondria) group II introns and highlighted that awareness and identification of those mechanisms are essential for improving both (i) the alignment, as the hypothesis of primary homology (cf. Simmons and Ochoterena, 2000) and (ii) the phylogenetic analysis itself, in which the hypothesis of primary homology is tested by congruence (cf. Simmons and Ochoterena, 2000).

Slipped-strand mispairing is suggested as the mechanism which generates length mutations (Levinson and Gutman 1987) resulting in short indels (e.g. Laroche et al. 1997, Kelchner 2000) and tandem minisatellite regions (e.g. Sperisen et al. 2001, Gugerli et al. 2001b). Latter regions, as well as minute inversions associated with hairpins are the most homoplasious characters in molecular phylogenetics (S. Kelchner, pers. comm.). Additional obstacles in accurate homology assessment are indels of surprising size that contain sequence content not readily identifiable in origin (Kelchner 2000). Such indels are resulting from frequent

intramolecular recombination within a discrete non-coding region or on a genomic scale (e.g. Laroche and Bousquet 1999, Graham et al. 2000, Kelchner 2000, Löhne and Borsch 2005) and they apparently do not annihilate possible secondary structure formation, same as smaller-scale deletions of loops, bulges and even parts of the stems (Kelchner 2000).

References

A'Hara SW, Cottrell JE (2007). Characterization of a suite of 40 EST-derived microsatellite markers for use in Sitka spruce (*Picea sitchensis* (Bong.) Carr.). *Silvae Genet* **56**: 3-4.

Avise JC (1994). *Molecular markers, natural history and evolution*. Chapman and Hall, New York, USA.

Avise JC (2000). *Phylogeography, the history and formation of species.* Harvard University Press, Cambridge, Massachusetts, London, England.

Bakker FT, Culham A, Pankhurst CE, Gibby M (2000). Mitochondrial and chloroplast DNA-based phylogeny of *Pelargonium* (Geraniaceae). *Am J Bot* **87**: 727-734.

Ballian D, Longauer R, Mikić T, Paule L, Kajba D, Gömöry D (2006). Genetic structure of a rare European conifer, Serbian spruce (*Picea omorika* (Pančić) Purk.). *Pl Syst Evol* **260**: 53-63.

Bérubé Y, Ritland C, Ritland K (2003). Isolation, characterization, and cross-species utility of microsatellites in yellow cedar (*Chamaecyparis nootkatensis*). *Genome* **46**: 353-361.

Bérubé Y, Zhuang J, Rungis D, Ralph S, Bohlmann J, Ritland K (2007). Characterization of EST-SSRs in loblolly pine and spruce. *Tree Genet Genomes* 3: 251-259.

Besnard G, Acheré V, Faivre Rampant P, Favres JM, Jeandroz S (2003). A set of markers developed from DNA sequence databanks in *Picea* (Pinaceae). *Mol Ecol Notes* **3**: 380-383.

Bolle C (1876). Eine neue Fichte in Serbien von Prof. Pančić entdeckt. *Sitzungsberichte d. Botanisch. Vereines d. Provinz Brandenburg*, Berlin.

Burschel P (1965). Die Omorikafichte. *Forstarchiv* **36**: 113-131.

Bůžek Č, Kvaček Z, Holý F (1985). Late Pliocene palaeo-environment and correlation of the Vildštejn floristic complex within central Europe. *Rozpravy Československé Akademie Věd* **95**: 1-72.

Clegg MT, Gaut BS, Learn Jr GH, Morton BR (1994). Rates and patterns of chloroplast DNA evolution. *Proc Natl Acad Sci USA* 91: 6795-6801.

Conifer Specialist Group (1998). *Picea omorika*. In: IUCN 2007. *2007 IUCN Red List of Threatened Species* (http://www.iucnredlist.org/search/details.php/30313/summ).

Cruzan MB, Templeton AR (2000). Paleoecology and coalescence: phylogeographic analysis of hypotheses from the fossil record. *Trends Ecol Evol* **15**: 491-495.

Čolić DB (1966). Požar kao ekološki faktor u sukcesiji zajednica Pančićeve omorike i redukovanju njenog areala. *Zaštita prirode* **33**, Beograd.

Dallimore W, Jackson AB, Harrison SG (1967). *A handbook of Coniferae and Ginkgoaceae*. 4th Ed. St. Martin's Press, New York. p 729.

David A., Keathley D. (1996) Inheritance of mitochondrial DNA in interspecific crosses of *Picea glauca* and *Picea omorika*. *Can. J. For. Res*. 26: 428-432.

DeWoody JA, Avise JC (2000) Microsatellite variation in marine, freshwater and anadromous fishes compared with other animals. *J Fish Biol* **56**: 461-473.

Di Renzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB (1994). Mutational processes of simple-sequence repeat loci in human populations. *Proc Natl Acad Sci USA* **91**: 3166-3170.

Dombrovska O, Qiu YL (2004). Distribution of introns in the mitochondrial gene nad1 in land plants: phylogenetic and evolutionary implications. *Mol Phylogen Evol* **32**: 246-263.

Echt CS, May-Marquardt P, Hseih M, Zahorchak R (1996). Characterization of microsatellite markers in eastern white pine. *Genome* **39**: 1102-1108.

Eisenhut G (1961). Untersuchungen über die Morphologie und Ökologie der Pollenkörner heimischer und fremdländischer Waldbäume. *Forstwiss Forsch* **15**:1-68.

Ellegren H (2004). Microsatellites: simple sequences with complex evolution. *Nat Rev Genet* **5**: 435-445.

Elsik CG, Williams CG (2001). Low-copy microsatellite recovery from a conifer genome. *Theor Appl Genet* **103**: 1189-1195.

Farjon A (2001). *World checklist and bibliography of conifers*. 2nd edition. Royal Bot Gard, Kew, London, UK.

Feliner GN, Rosselló JA (2007). Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. *Mol Phylogen Evol* **44**:911-919.

Fowler DP (1980). Hybridization of black spruce and Serbian spruce. *Can For Serv Marit For Cent Inf Rep* M-X-112.

Fowler DP (1983). The hybrid black x Sitka spruce, implications to phylogeny of the genus *Picea*. *Can J For Res* **13**: 108-115.

Freudenstein JV, Chase MW (2001). Analysis of mitochondrial nad1b-c intron sequences in *Orchidaceae*: utility and coding of length-change characters. *Syst Bot* **26**: 643-657.

Fukarek P (1967). Pančićevo otkriće omorike i njeno dalje proučavanje. In (ed) Josifović M, *Pančićev zbornik u spomen 150-godišnjice njegovog rođenja*. Srpska Akademija Nauka i Umetnosti, Odeljenje prirodno-matematičkih nauka, Beograd. pp 27-67.

Gajić M, Vilotić D, Karadžić D, Mihajlović Lj, Isajev V (1994). *Omorika – Picea omorika (Pančić) Purkyně na području Nacionalnog parka Tara (monografska studija)*. Nacionalni park Tara – Bajina Bašta and Šumarski fakultet, Beograd.

Geburek T, Krusche D (1985). Wachstum von Hybriden zwischen *Picea omorika* und *P. sitchensis* in Vegrleich zu den Elternarten. *Allg Forst-Jagdztg* **156**: 47-54.

Geburek T (1986). Some results of inbreeding depression in Serbian spruce (*Picea omorika* (Panč.) Purk.). *Silvae Genet* **35**: 169-172.

Gielly L, Taberlet P (1994). The use of chloroplast DNA to resolve plant phylogenies: noncoding versus rbcL sequences. *Mol Biol Evol* **11**: 769-777.

Gigov A (1956). Dosadašnji nalazi o postglacijalnoj istoriji šuma Srbije. *Institut za ekologiju i biogeografiju, Zbornik radova* **7/3**: 1-25.

Gligorević-Danon Z (1969). Klijanje semena *Picea omorika* Pančić and *Picea excelsa* L. zavisno od temperature i starosti semena. *God Biol inst Univ u Sarajevu* **22**: 5-20.

Goldestein DB, Pollock DD (1997). Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. *J Hered* **88**: 335-342.

Golenberg EM, Clegg MT, Durbin ML, Doebley J, Ma DP (1993). Evolution of a noncoding region of the chloroplast genome. *Mol Phylogenet Evol* **2**: 52-64.

Graham SW, Reeves PA, Burns ACE, Olmstead RG (2000). Microstructural changes in noncoding choroplast DNA: interpretation, evolution, and utility of indels and inversions in basal angiosperm phylogenetic inference. *Int J Plant Sci* **161**: S83-S96.

Grivet D, Jeandroz S, Favre JM (1999). Nad1 b/c intron polymorhism reveals maternal inheritance of the mitochondrial genome in *Picea abies*. *Theor Appl Genet* **99**: 346-349.

Gu X, Li WH (1995). The size distribution of insertions and deletions in human and rodent pseudogenes suggests the logarithmic gap penalty for sequence alignment. *J Mol Evol* 40: 464-473.

Gugerli F, Sperisen C, Büchler U, Brunner I, Brodbeck S, Palmer JD, Qiu Y-L (2001a). The evolutionary split pf Pinaceae from other conifers: evidence from an intron loss and a multigene phylogeny. *Mol Phylogenet Evol* **21**: 167-175.

Gugerli F, Sperisen C, Büchler U, Magni F, Geburek Th, Jeandroz S, Senn J (2001b). Haplotype variation in a mitochondrial tandem repeat of Norway spruce (*Picea abies*) populations suggests a serious founder effect during postglacial re-colonization of the western Alps. *Mol Ecol* **10**: 1255-1263.

Hamrick JL, Godt MJW (1996). Conservation genetics of endemic plant species. In: Avise JC, Hamrick JL (eds) *Conservation genetics: case histories from nature.* Chapman Hall: New York. pp 281-304.

Hewitt G (2000). The gentic legacy of the Quaternary ice ages. *Nature* **405**: 907-913.

Hodgetts RB, Aleksiuk MA, Brown A, Clarke C, Macdonald E, Nadeem S, Khasa D (2001). Development of microsatellite markers for white spruce (*Picea glauca*) and related species. *Theor Appl Genet* **102**: 1252-1258.

Isajev V (1987). Oplemenjivanje omorike (*Picea omorica*/Panč./Purk.) na genetsko selekcionim osnovama. PhD thesis, Faculty of Forestry, Belgrade.

Jaramillo-Correa JP, Bousquet J (2003a). New evidence from mitochondrial DNA of a progenitor-derivative species relationship between black spruce and red spruce (Pinaceae). *Am J Bot* **90**: 1801-1806.

Jaramillo-Correa JP, Bousquet J, Beaulieu J, Isabel N, Perron M, Bouillé M (2003b). Cross-species amplification of mitochondrial DNA sequence-tagged-site markers in conifers: the nature of polymorphism and variation within and among species in *Picea*. *Theor Appl Genet* **106**: 1353-1367.

Jaramillo-Correa JP, Beaulieu J, Bousquet J (2004). Variation in mitochondrial DNA reveals multiple distant glacial refugia in black spruce (*Picea mariana*), a transcontinental North American spruce. *Mol Ecol* **13**: 2735-2747.

Jaramillo-Correa JP, Beaulieu J, Ledig FT, Bousquet J (2006). Decoupled mitochondrial and chloroplast DNA population structure reveals Holocene collapse and population isolation in a threatened Mexican-endemic conifer. *Mol Ecol* **15**:2787-2800.

Jevtić J (1960). Neka zapažanja o urodu semena Pančićeve omorike na Tari. *Šumartsvo* **13**: 79-87.

Jovančević M (1962). Određivanje klijavosti polena šumskog drveća prema veličini, obliku i boji polenovih zrnaca. *Narodni šumar* 16: 493-502.

Kelchner SA (2000). The evolution of noncoding chloroplast DNA and its application in plant systematics. *Ann Miss Bot Gard* 87: 482–498.

Kelchner SA (2002). Group II introns as phylogenetic tools: structure, function, and evolutionary constraints. *Am J Bot* **89**: 1651-1669.

Knoop V (2004). The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. *Curr Genet* 46: 123-139.

Král D (2002). Assessing the growth of *Picea omorika* [Panč.] Purkyně in the Masaryk Forest Training Forest Enterprise at Křtiny. *J For Sci* **48**: 388-398.

Krstić M (1950). Morfološke i biometrijske pojedinosti fruktifikacije *Picea omorika* Panč. - Kvalitet semena. Institut za naučna istraživanja u šumarstvu NRS, Beograd, 1-23.

Kubo T, Mikami T (2007). Organization and variation of angiosperm mitochondrial genome. *Physiol Plantarum* **129**: 6-13.

Kuittinen H, Muona O, Kärkkäinen K, Borzan Ž (1991). Serbian spruce, a narrow endemic, contains much genetic variation. *Can J For Res* **21**:363-367.

Kuitinnen H, Savolainen O (1992). *Picea omorika* is a self-fertile but outcrossing conifer. *Heredity* **68**: 183-187.

Kumar R, Lelu M-A, Small I (1995). Purification of mitochondria and mitochondrial nucleic acids from embryogenic suspension cultures of a gymnospem, *Larix* x *leptoeuropaea*. *Plant Cell Rep* 14: 534-538.

Langner W (1959). Selbstfertilität und Inzucht bei *Picea omorica* (Pančić) Purkyne. *Silvae Genet* **8**: 84-93.

Laroche J, Bousquet J (1999). Evolution of the mitochondrial *rps3* intron in perennial and annual angiosperms and homology to *nad5* intron 1. *Mol Biol Evol* **16**: 441-452.

Laroche J, Li P, Maggia L, Bousquet J (1997). Molecular evolution of angiosperm mitochondrial introns and exons. *Proc Natl Acad Sci USA* **94**: 5722-5727.

Learn GH, Shore JS, Furnier GR, Zurawski G, Clegg MT (1992). Constraints on the evolution of plastid introns: the group II intron in the gene encoding tRNA-Val (UAC). *Mol Biol Evol* **9**: 856-871.

Ledig FT, Hodgskiss PD, Krutovskii KV, Neale DB, Eguiluz-Piedra T (2004). Relationships among the spruces (*Picea*, pinaceae) of southwestern North America. *Syst Bot* **29**: 275-295.

LePage BA (2001). New species of *Picea* A. Dietrich (Pinaceae) from the middle Eocene of Axel Heiberg Island, Arctic Canada. *Bot J Linn Soc* **135**: 137-167.

Levinson G, Gutman G (1987). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol* **4**: 203-221.

Li YC, Korol AB, Fahima T, Belies A, Nevo E (2002). Microsatellites: genomic distribution, putative functions, and mutational mechanisms: a review. *Mol Ecol* **11**: 2453-2465.

Löhne C, Borsch T (2005). Molecular evolution and phylogenetic utility of the *petD* group II intron: a case study in basal angiosperms. *Mol Biol Evol* **22**: 317-332.

Lunt DH, Whipple LE, Hyman BC (1998) Mitochondrial DNA variable number tandem repeats (VNTRs): utility and problems in molecular ecology. *Mol Ecol* **7**: 1441-1455.

Mayr H (1890). *Monographie der Abietineen des Japanischen Reiches*. M. Rieger'sche Universitäts-Buchhandlung, München.

Morgante M, Hanafey M, Powell W (2002). Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nature Genet* **30**: 194-200.

Michel F, Dujon B (1983). Conservation of RNA secondary structures in two intron families including mitochondrial-, chloroplast- and nuclear encoded members. *EMBO J* **2**: 33-38.

Michel F, Ferat J-L (1995). Structure and activities of group II introns. *Annual Rev Biochemistry* **64**: 435-461.

Michel F, Umesono K, Ozeki H (1989). Comparative and functional anatomy of group II catalytic introns - a review. *Gene* **82**: 5–30.

Mikkola L (1972). Crossability between *Picea omorika* (Pancic) Purkyne and *P. glauca* (Moench) Voss. *Ann Bot Fenn* **9**: 33-36.

Milovanović J, Isajev V, Krajmerova D, Paule L (2007). Allele polymorphism of nad1 gene of the Serbian spruce mitochondrial genome. *Genetika* **39**: 79-91.

Meng LA, Yang R, Abbott RJ, Miehe G, Hu T, Liu J (2006). Mitochondrial and chloroplast phylogeography of *Picea crassifolia* Kom. (Pinaceae) in the Qinghai-Tibetan Plateau and adjacent highlands. *Mol Ecol* **16**: 4128-4137.

Müller K (2006). Incorporating information from length-mutational events into phylogenetic analysis. *Mol Phylogenet Evol* **38**: 667–676.

Myers N, Mittermeier RA, Mittermeier CG, Fonseca GAB da, Kent J (2000). Biodiversity hotspots for conservation priorities. *Nature* **403**: 853-858.

Nasri N, Bojović S, Vendramin GG, Fady B (2007). Population genetic structure of the relict Serbian spruce, *Picea omorika*, inferred from plastid DNA. *Pl Syst Evol* (early online).

Neff BD, Gross MR (2001). Microsatellite evolution in vertebrates: inferences from AC dinucleotide repeats. *Evolution* **55**: 1717-1733.

Newton AC, Allnutt TR, Gillies ACM, Lowe AJ, Ennoss RA (1999). Molecular phylogeography, intraspecific variation and the conservation of tree species. *Trens Ecol Evol* **14**: 140-145.

Nkongolo KK (1999). RAPD and cytological analyses of Picea spp. From different provenances: genomic relationships among taxa. *Hereditas* **130**: 137-144.

Novák FR (1927). Zur fünfzigjährigen Entdeckung der *Picea omorica*. *Mitt Dtsch Dendrol Ges* **38**: 47-56.

Ogden TH, Rosenberg MS (2007). How should gaps be treated in parsimony? A comparison of approaches using simulation. *Mol Phylogenet Evol* 42: 817- 826.

Ohta T, Kimura M (1973). The model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a genetic population. *Genet Res* **22**: 201-204.

Ostojić D (2005). Ekološki činioci prirodnog održavanja i obnove cenopopulacija Pančićeve omorike u NP Tara. PhD thesis, Faculty of Forestry, Belgrade.

Palmer JD, Adams KL, Cho Y, Parkinson CL, Qiu Y-L, Song K (2000). Dynamic evolution of plant mitochondrial genomes: Mobile genes and introns and highly variable mutation rates. *Proc Natl Acad Sci USA* **97**: 6960-6966.

Pančić J (1877). *Eine neue Conifere in den oestlischen Alpen*. Serbischen Staatsdrukerei, Belgrade.

Pelgas B, Isabel N, Bousquet J (2004). Efficient screening for expressed sequence tag polymorphisms (ESTPs) by DNA pool sequencing and denaturing gradient gel electrophoresis (DGGE) in spruces. *Mol Breeding* **13**: 263-279.

Petit RJ, Hampe A, Cheddadi R (2005). Climate changes and tree phylogeography in the Mediterranean. *Taxon* **54**: 877-885.

Pintarić K (1970). Konzerviranje sjemena Pančićeve omorike (*Picea omorika* Panč.) u hermetički yatvorenim posidama i uticaj starosti na klijavost. *Šumarstvo* **24**: 3-12.

Primmer CR, Saino N, Moller AP, Ellegren H (1998). Unrevealing the processes of microsatellite evolution through analysis of germ line mutations in Barn Swallows *Hirundo rustica*. *Mol Biol Evol* **15**: 1047-1054.

Pfeiffer A, Olivieri AM, Morgante M (1997). Identification and characterization of microsatellites in Norway spruce (*Picea abies* K.). *Genome* **40**: 411-419.

Purkyně E (1877). Eine asiatische Konifere in den Balkanländern. *Österreichische Monatsschrift f Forstwesen Wien* **27**: 446-449.

Rajora OP, Rahman MH, Dayanandan S, Mosseler A (2000). Isolation, characterization, inheritance and linkage of microsatellite DNA markers in white spruce (*Picea glauca*) and their usefulnes in other spruce species. *Mol Gen Genet* **264**: 871-882.

Ran J-H, Wei X-X, Wang X-Q (2006). Molecular phylogeny and biogeography of *Picea* (Pinaceae): implications for phylogeographical studies using cytoplasmic haplotypes. *Mol Phylogenet Evol* **41**: 405-419.

Ravazzi C (2002). Late Quaternary history of spruce in southern Europe. *Rev Paleobot Palinol* **120**: 131-177.

Rungis D, Bérubé Y, Zhang J, Ralph S, Ritland CE, Ellis BE, Douglas C, Bohlmann J, Ritland K (2004). Robust simple sequence repeat markers for spruce (*Picea* spp.) from expressed sequence tags. *Theor Appl Genet* **109**: 1283-1294.

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

Schaal BA, Gaskin JF, Caicedo AL (2003). Phylogeography, haplotype trees, and invasive plant species. *J Hered* **94**: 197-204.

Schaal BA, Hayworth DA, Olsen KM, Rauscher JT, Smith WA (1998). Phylogenetic studies in plants: problems and prospects. *Mol Ecol* **7**: 465-474.

Schlötterer C (1998).Microsatellites. In Hoelzel AR (ed), *Molecular Genetic Analysis of Populations*. IRL Press at Oxford University Press, Oxford.

Schlötterer C (2004). The evolution of molecular markers - just a matter of fashion? *Nat Rev Genet* **5**: 63-69.

Schlötterer C, Tautz D (1992). Slippage synthesis of simple sequence DNA. *Nucl Acid Res* **20**: 211-215.

Schmidt PA (1989). Beitrag zur Systematic and Evolution der Gattung *Picea* A. Dietr. *Flora* **182**: 435-461.

Schmidt PA (1998). *Picea* A. Dietr. P. 1-14. In (eds) Schütt P, Schuck HJ, Lang U, Roloff A, Enzyklopädie der Holzgwächse. 14. Erg. Lfg., 12/95. Ecomed-Verlag, Landsberg.

Schmidt-Vogt H (1977). *Die Fichte, Ein Handbuh in zwei Bänden*. Verlag Paul Parey, Hamburg und Berlin.

Scotti I, Magni F, Fink R, Powell W, Binelli G, Hedley PE (2000). Microsatellite repeats are not randomly distributed within Norway spruce (*Picea abies* K.) expressed sequences. *Genome* **43**: 41-46.

Scotti I, Magni F, Paglis GP, Morgante M (2002). Trinucleotide microsatellites in Norway spruce (*Picea abies*): their features and the development of molecular markers. *Theor Appl Genet* **106**: 40-50.

Selkoe KA, Toonen RJ (2006). Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecol Lett* **9**: 615-629.

Sigurgeirsson A, Szmidt A (1993). Phylogenetic and biogeographic implications of choroplast DNA variation in *Picea*. *Nord J Bot* **13**: 233-246.

Shaw J, Lickey EB, Schilling EE, Small RL (2007). Comparison of whole choroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *Am J Bot* **94**: 275-288.

Shubert R, Mueller-Starck G, Riegel R (2001). Development of EST-PCR markers and monitoring their intrapopulational genetic variation in *Picea abies* (L.) Karst. *Theor Appl Genet* **103**: 1223-1231.

Siljak-Yakovlev S, Cerbah M, Coulaud J, Stoian V, Brown SC, Zoldos V, Jelenic S, Papes D (2002). Nuclear DNA content, base composition, heterochromatin and rDNA in *Picea omorika* and *Picea abies*. *Theor Appl Genet* **104**: 505-512.

Simmons MP, Müller K, Norton A (2007). The relative performance of indel-coding methods in simulations. *Mol Phylogenet Evol* **44**: 724-740.

Simmons MP, Ochoterena H (2000). Gaps as characters in sequence based phylogenetic analyses. *Syst Biol* **49**: 369–381.

Soranzo N, Provan J, Powell W (1998). Characterization of microsatellite loci in *Pinus sylvestris* L. *Mol Ecol* **7**: 1260-1261.

Sperisen C, Büchler U, Gugerli F, Mátyás G, Geburek Th, Vendramin GG (2001). Tandem repeats in plant mitochondrial genomes: application to the analysis of population differentiation in the conifer Norway spruce. *Mol Ecol* **10**: 257-263.

Stern N, Roche L (1974). *Genetics of forest ecosystems*. Springer-Verlag, Berlin.

Sunnucks P (2000). Efficient genetic markers for population biology. *Trends Ecol Evol* **15**: 199-203.

Šijak M, Dinić A (1996). Dodatak bibliografiji radova o Pančićrvoj omorici (*Picea omorika* Pančić). *Ekologija* **31**: 165-178.

Šilić Č (1990). *Endemične biljke*. 3rd edition. IP 'Svijetlost', Zavod za udžbenike i nastavna sredstva, Sarajevo, Zavod za udžbenike i nastavna sredstva, Beograd. p 228.

Templeton AR (1998). Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. *Mol Ecol* **7**: 381-397.

Templeton AR (2004). Statistical phylogeography: methods of evaluating and minimizing inference errors. *Mol Ecol* **13**:789-809.

Tucić B, Stojković B (2001). Shade avoidance syndrome in *Picea omorika* seedlings: a growth-room experiment. *J Evol Biol* **14**: 444-455.

Tucović A, Isajev V (1982). Uticaj različitih tipova oprašivanja na neka svojstva šišarica i semena omorike. *Glasnik Šumarskog Fakulteta u Beogradu, Ser. C, Pejzažna arhitektura* **59**: 59-65.

Tucović A, Isajev V (1986). Generativna semenska plantaža omorike. Izvodjački projekat. OOUR Institut za Šumarstvo, Šumarski fakultet, Belgrade.

van de Ven WTG, McNicol RJ (1996). Microsatellites as DNA markers in Sitka spruce. *Theor Appl Genet* **93**: 613-617.

Varshney RK, Graner A, Sorrells ME (2005). Genic mocrosatellite markers in plants: features and applications. *Trends Biotechn* **23**: 48-55.

Vidaković M (1991). *Conifers, morphology and variation*. 2nd edition. Grafički zavod Hrvatske, Zagreb.

Wang X-R, Tsumura Y, Yoshimaru H, Nagasaka K, Szmidt AE (1999). Phylogenetic relationships of Eurasian pines (*Pinus*, Pinaceae) based on chloroplast *rbcL*, *MAT*K, *RPL20-RPS18* spacer, and *TRNV* intron sequences. *Am J Bot 86*: 1742-1753.

Wardle P (1956). *Picea omorika* in its natural habitat. *Forestry* **29**: 91-117.

Wettstein R (1890). Die Omoricafichte, *Picea omorica* (Pančić) Purkyne. - Eine monographische Studie. Sitzungsberichte der Akademie d. Wissenschaften Wien. Matematisch-naturwissenschaft. Klasse 49 (1): 503-557.

Willkomm M (1887). *Forstliche Flora von Deutchland und Oesterrich*. CF. Winter'sche Berlagshandlung, Leipzig.

Wissinger B, Schuster W, Brennicke A (1991). *Trans*-splicing in Oenothera mitochondria: *nad1* mRNAs are edited in exon and *trans*-splicing group II intron sequences. *Cell* **65**: 473–482.

Wright JW (1955). Species crossability in spruce in relation to distribution and taxonomy. *For Sci* **1**: 319-349.

Zhang DX, Hewitt GM (2003). Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Mol Ecol* **12**: 563-584.

1 The aim of the dissertation

The aim of this dissertation is to address some of the fundamental questions concerning *Picea omorika* (Panč.) Purky., such as genetic diversity in natural populations, gene dispersal, survival during the last glacial maximum in the Balkan and evolutionary relations to other members of the genus *Picea* A. Dietr., especially Northern American species *P. mariana* and *P. rubens*. Highly polymorphic nuclear EST-SSRs and two mitochondrial loci (nad1i477 and nad5i230) are employed. The reconstruction of the secondary structure of the former group II intron is utilized for the improvement of the homology assessment and for more accurate phylogenetic inferences, while incorporation of the phylogenetic signal in length mutations, especially at nad1i477, is utilized for increasing the resolution in phylogenetic reconstructions. This work also contributes towards better understanding the highly structured genetic diversity in species confined to the Balkan - the biodiversity hot-spot, and towards conservation of the remaining genetic resources of *P. omorika*, an endemic and a Tertiary relict species.

2 Overview of all publications

As an outcome of this dissertation, several publications are either submitted or in preparation. Overview of all publications is given below.

2.1 Aleksić MJ, Schueler S, Mengl M, Geburek T (submitted): EST-SSRs developed for other *Picea* species reveal high genetic variation in *Picea omorika*

Summary: *Picea omorika* (Panč.) Purk. is a relict and an endemic species found exclusively in the Balkan within an area of approx. 10.000 km^2. Marker-based genetic knowledge in this conifer is very limited and partially contradictory. Therefore, 12 nuclear markers (10 EST-SSRs and 2 genomic SSRs) have been cross-species amplified using 50 trees originating from 2 natural populations. Five EST-SSRs amplified successfully and very high number of alleles per locus was found (7 to 18 alleles per locus) with a total of 61 alleles. Furthermore, a high number of private alleles was detected - 13 and 14 per population, respectively. Markers were selectively neutral, no linkage disequilibrium was detected and the genotype frequencies fitted Hardy–Weinberg proportions. Expected heterozygosity per locus ranged from 0.64 to 0.91 in both populations with an overall mean of 0.83. Considering the small remnant population sizes of *P. omorika* these values are unexpectedly high and comparable to values in *P. glauca* (Moench) Voss., *P. sitchensis* (Bong.) Carr. and *P. mariana* (Mill.) B.S.P. detected with an identical set of markers in a similar sample sizes.

2.2 Aleksić MJ, Geburek T (submitted): Was there more than one refugial population of an endemic Serbian spruce in the Balkans during the last glaciation?

Summary: The genetic variation in an endemic and a relict conifer species - *Picea omorika* (Panč.) Purk. was assessed by amplifying the partial sequence of the second intron of the mitochondrial NADH dehydrogenase subunit1 gene in 200 trees originating from 10 natural populations. Five haplotypes were detected and size variation was due to the variable number of a 34 bp minisatellite. In overall sample, haplotypic richness was 3.007 and all population genetics parameters were similar to those reported in widely distributed *Picea* species - H_S = 0.075, H_T = 0.225 and G_{ST} = 0.668. However, the distribution of haplotypes was not random, as predominant and presumably the ancestral haplotype A (one 34 bp repeat) was fixed in eight populations, while it was absent in a single, small population comprised of approximately 400 trees in which the highest gene diversity (H_S = 0.468) was observed. Current structuring of genetic variation in *P. omorika* supported by pallinological data is interpreted as a result of the secondary contact of two gene pools established long before the last glacial maximum and characterized by different evolutionary histories and past migrations. Our study supports the views that (i) a complex structuring can be found even within a small areas within the Balkan, and that (ii) the sampling scheme is essential in assessing genetic variability in species *per se*, especially in cases of formerly widely distributed species currently confined to small areas - when a single population from the second, highly polymorphic gene pool was omitted from the analyses, all population genetics parameters substantially decreased, misleadingly implying genetic uniformity in this species - H_S = 0.031, H_T = 0.033 and G_{ST} = 0.070. Almost absent seed flow in overall sample (Nm < 1) implied that isolation-by-distance seems to be very effective in *P. omorika* and, in dependence of the paternal gene flow, this has to be taken into consideration for delineation of conservation units of this species.

2.3 Aleksić MJ, Geburek T, Fussi B, Heinze B (submitted): Length mutations in two mitochondrial group II introns in *Picea omorika*: new insights into spruce phylogeny

Summary: Two mitochondrial group II introns - nad1i477 and nad5i230, were used to test a hypothesis of a close relationship between an European narrow endemic and a Tertiary relict - *P. omorika*, and two North American spruces - *P. mariana* and *P. rubens*, revealed by crossability studies and supported by two recently published molecular phylogenies. Our results

support the view that the recognition of microstructural changes leading to the occurrence of length mutations can improve the assessment of primary homology (i.e. alignment) and consequently, the phylogenetic analysis itself, and demonstrates that incorporation of length mutations in genealogical and phylogenetic reconstructions can substantially improve the resolution, especially in cases when length mutations play a major role in sequence evolution, as found at nad1i477. The domain IV in this intron is found to harbour two hypervariable large indels (up to 1300 bp) characterized by essentially different organization of variability and secondary structure. Indel A, present only in group A spruce species (*P. omorika*, *P. abies* and the majority of Asian species) is found to form a long hairpin suggesting that the loss of such a structure might result in the loss of the whole indel A, as found in group B spruce species. Similarly, group B spruce species (*P. schrenkiana* and *P. smithiana* and all American species, including *P. mariana* and *P. rubens*) are found to harbour indel B, organized in a stem-loop structure, implying that the loss of this structure would lead to the loss the whole indel B, as found in group A spruce species. Alternatively, indel B may have been gained by recombination. The concordance of such deep division among spruce species found at nad1i477 corresponds well to the five haplotypes described at the second, less variable mitochondrial intron nad5i230, and to geographical distribution of all species used in this study. The only species which did not follow the detected pattern was *P. breweriana,* characterized as an isolated, relict species, similarly as in case with *P. omorika*. Both species are found to be distantly related to *P. mariana* and *P. rubens*. A hypothesis of the eastern Asian origin of spruces is supported and based on our results, Wright's 'the most primitive species with generalized morphological features' - *P. koyamai,* still might be the basal spruce.

2.4 Aleksić MJ et al. (in preparation): The origin of hipervariability in mitochondrial group II intron nad1i477 in conifers: pine's nuclear ribosomal internal transcribed spacer1 in spruces?

Summary: Mitochondrial genome in plants is well known to acquire foreign sequences from both, nucleus and chloroplasts, and this phenomenon is assumed to occur frequently during the last 10.000 years or so. Mitochondrial group II intron nadi477 has been characterized by pronounced intra- and interspecific variability in spruces and pines and several minisatellites

have been described in both genera. The similar organization of the hipervariable domain IV in those genera implies their common origin and we demonstrate that the origin of high variability and the occurrence of described minisatellites are caused by the intracellular transfer of the nuclear ribosomal DNA internal transcribed spacer1 (nrDNA ITS1) fragment into the mitochondria. The fragment in question is found to harbour short subrepeats (SSRs) and since (i) pines SSRs 2 and 3 with typical conserved motifs were found to be easily aligned to tandem repeats found in spruces, and (ii) those SSRs have been reported to pair forming a hairpin, the intracellular transfer of such hairpin was assumed to occur in the common ancestor of those two lineages presumably 200 MYA, at the time of global environmental changes which probably caused increased receptivity of the mtDNA to the foreign sequences. Such transfer(s) might trigger the exchange of sequences between genomes in plants resulting in specific features and promiscuity of the plant mitochondrial genome.

2.5 Aleksić MJ et al. (in preparation): High genetic diversity in natural populations of an endemic and relict spruce *Picea omorika*

Summary: The genetic variability in *Picea omorika* (Panč.) Purk. was assessed by five nuclear EST-SSRs and the second intron of the mitochondrial NADH dehydrogenase subunit1 gene (nad1i477). Markers were amplified in 500 trees originationg from 10 natural populations. Nuclear EST-SSRs revealed high number of alleles per locus (10 to 52) and private alleles were found in five populations. Interstingly, in population 'Studenac,' 14 private nuclear SSRs were found and 13 of them were confined to a single locus. They were characterized by extremely high number of repeats, ranging from 69 to 89. In overal sample, H_e ranged from 0.668 to 0.845 with overal mean of 0.773. F_{IS} ranged from 0.049 to -0.042 and averaged to 0.008. Pairwise population F_{ST} ranged from 0.031 to 0.186 with surprisingly low overal mean of 0.102. AMOVA revealed that high amount of genetic variation is found within populations (89.96%), while only 10.04% can be assigned to differences between populations. At mitochondrial locus, six haplotypes were detected and variability was due to the variable copy number of a 34 bp minisatellite. However, the distribution of haplotypes was not random, as haplotype A was found to be fixed in eight populations and it was predominant in population 'Veliki Stolac'. Relatively rare haplotypes found in this population were predominant (B) or abundant (C and D) in

population 'Studenac', which did not harbour the most common type A. Private haplotypes were observed in both populations - J in 'Veliki Stolac' and E in 'Studenac', each found in a single tree. Our study supported the view of high genetic diversity in natural populations of *Picea omorika* and the finding of two different gene pools in this species, confined to an extremely small area in the Balkan.

2.6 Aleksić MJ et al. (in preparation): Recent admixture of three gene pools in a small natural population of an endemic and relict *Picea omorika*

Summary: Microspatial genetic structure in a single natural population of *P. omorika* comprised of 419 trees and confined to an extremely small area of 0.35 ha was analyzed by five nuclear EST-SSRs and a second intron of the NADH dehydrogenase subunit 1 gene (*nad*1 intron 2). Although all nuclear markers used were previously reported as neutral in two *P. omorika* natural populations and linkage disequilibrium was not significant, the neutrality of markers was not obtained and linkage disequilibrium was found between all marker pairs. Typing errors were eliminated as the possible cause for the described deviations and since global F_{IS} was found to be close to zero (0.013), Wahlund effect is ruled out as well. Therefore, it was assumed that this population was exposed either to inbreeding or admixture. Model based clustering methods (STRUCTURE and InStruct) revealed the presence of three different gene pools and inbreeding coefficients and selfing rates close to zero, unequivocally suggesting that the high genetic diversity within this population is a result of an recent admixture of three gene pools. The comparison of spatial locations of genotypes strongly assigned (>90%) to each of the three gene pools detected at nuclear loci and four haplotypes detected at mtDNA locus revealed the existence of the resident population characterized by haplotype B, first immigrant population characterized by haplotype C (immigration by seed dispersal) and second immigrant population in which the correlation between nuclear and mtDNA markers was not detected, implying immigration by pollen. Rare haplotypes D and E could not be assigned to either of three gene pools and they were assumed to originate by mutations within the second immigrant population. H_o and H_e were estimated to 0.644 and 0.653, respectively, and H_S, H_T and G_{ST} were found to be comparable to other widely distributed spruce species (0.094, 0.217 and 0.569, respectively). Non-random distribution of genotypes based on both, nuclear and mtDNA markers was

detected using Loiselle's kinship coefficient and within the first distance class positive correlation was found to be ten times stronger in latter case (0.015 vs. 0.150). Effective seed flow was found to be were limited (20.45 m), while *Sp* value was comparable to other wind pollinated and seed dispersed species (0.0091). However, due to the admixture of three gene pools found in this population and consequently, deviations from assumed neutrality and linkage equilibrium between markers, latter results must be interpreted with caution.

2.7 Aleksić MJ et al. (in preparation): Nuclear EST-SSRs in a relict *Picea omorika*: new insights into the evolution of microsatellites

Summary: The amplification of five nuclear EST-SSRs in 900 trees of a relict species *Picea omorika* revealed the pattern of allele frequencies which implied the existence of two classes of SSRs. Within the first class, the increase in length is probably possible to the certain maximum threshold which is locus specific (as well as the lower threshold), while point mutations were found as a mechanism regulating the length within the second class. Within this class, the occurrence of compound and imperfect SSRs is likely. Possible reasons for the existence of two SSR classes are discussed and the hypothesis of the emergence, existence and disappearance of SSRs is suggested, implying that within the genome, multinucleotide tandem repeats of low copy numbers corresponding to the lower threshold in first class as well as low copy motifs upstream of the active minisatellite in second class, might represent the remains of formerly active microsatellites. Recently reported microsatellites in another relict species - *Ginkgo biloba*, supported our hypothesis.

3 Selected publications

The following publications are presented as a part of the dissertation.

3.1 EST-SSRs developed for other *Picea* species reveal high genetic variation in *Picea omorika*

Aleksić MJ, Schueler S, Mengl M and Geburek T

ABSTRACT**. –** *Picea omorika* (Panč.) Purk. is a relict and an endemic species found exclusively in the Balkan within an area of approx. 10.000 km^2. Marker-based genetic knowledge in this conifer is very limited and partially contradictory. Therefore, 12 nuclear markers (10 EST-SSRs and 2 genomic SSRs) have been cross-species amplified using 50 trees originating from 2 natural populations. Five EST-SSRs amplified successfully and very high number of alleles per locus was found (7 to 18 alleles per locus) with a total of 61 alleles. Furthermore, a high number of private alleles was detected - 13 and 14 per population, respectively. Markers were selectively neutral, no linkage disequilibrium was detected and the genotype frequencies fitted Hardy–Weinberg proportions. Expected heterozygosity per locus ranged from 0.64 to 0.91 in both populations with an overall mean of 0.83. Considering the small remnant population sizes of *P. omorika* these values are unexpectedly high and comparable to values in *P. glauca* (Moench) Voss., *P. sitchensis* (Bong.) Carr. and *P. mariana* (Mill.) B.S.P. detected with an identical set of markers in a similar sample sizes.

KEY WORDS. - *Picea omorika*, EST-SSRs, genomic SSRs, genetic diversity.

INTRODUCTION

Serbian spruce [*Picea omorika* (Panč.) Purk.] is one of only two representatives of the genus *Picea* A. Dietr. in Europe. This conifer is a Tertiary relict and an endemic species with an extremely small and scattered natural distribution in south-eastern Europe i.e. in the Balkan. Apart from solitary trees and tree groups found in a larger area, more than 30 populations occupy an area of approx. 10.000 km2, although the cumulative area probably does not exceed 60 ha (BURSCHEL 1965). Serbian spruce has been legally protected in former Yugoslavia in 1964 and is listed in the 'IUCN Red List of Threatened Species' due to its limited natural range (CONIFER SPECIALISTS GROUP 1998).

To date, only three studies analyzed the genetic variability in this species by means of genetic markers, but results are partially contradictory. Recently, BALLIAN *et al.* (2006) and NASRI *et al.* (2007) using allozymes and chloroplast markers, respectively, reported very limited genetic diversity, contrasting the earlier results of KUITTINEN *et al.* (1991). For a better understanding of the remnant genetic structure in this species and to improve the knowledge of natural evolutionary processes such as mating and genetic drift, markers with a higher resolution, as e.g. nuclear simple sequence repeats (SSRs) are required.

Up to now SSRs have not been specifically developed for *P. omorika* although some markers were employed in single trees for comparative purposes (e.g. RUNGIS *et al.* 2004). In conifers, traditional development of SSRs from genomic libraries was only partially successful due to their large, repetitive genomes resulting in complex multilocus amplification products (ELSIK & WILLIAMS 2001, BÉRUBÉ *et al.* 2007). To overcome this obstacle, low-copy libraries (ELSIK & WILLIAMS 2001) or unique sequences (expressed sequence tags i.e. ESTs) were used to develop robust, low copy SSRs (RUNGIS *et al.* 2004, BÉRUBÉ *et al.* 2007). Although EST-SSRs are believed to be less variable in comparison to genomic SSRs, they have less null alleles and, as a major advantage, high level of transferability to related species (RUNGIS *et al*. 2004, SELKOE & TOONEN 2006, BÉRUBÉ *et al.* 2007). The most abundant class of EST-SSRs found in conifers are those with dinucleotide repeats and predominant AT repeat motif especially at the 3' ends (RUNGIS *et al*. 2004, BÉRUBÉ *et al.* 2007).

Our goals were to (1) identify a set of highly reliable and informative EST-SSRs for *P. omorika* out of 25 markers previously developed for different *Picea* species (2) to test 2 genomic SSRs developed for *P. abies* and (3) to provide a first insight into the genetic structure of two natural populations.

MATERIAL AND METHODS

Two populations, 'Topli Do' (TD) from Serbia and 'Veliki Stolac' (VS) from the Republic of Srpska, Bosnia and Herzegovina were selected for the study. The latter one is identical to the one analyzed previously by allozymes (KUITTINEN *et al.* 1991, BALLIAN *et al.* 2006). Populations were large (one to several thousands of trees) and geographically isolated by approximately 8 kilometers. Young twigs were collected from 25 trees from each population. Material was stored at low temperature during transportation and later kept at -60 °C until DNA extraction. Needles from each tree were cut into 2-3 mm long pieces and tissue (approx. 100 mg) was placed separately into 2 ml Eppendorf tubes. Initial drying was performed for one hour with methanol (500 µl per tube). Tissue was fully dried in a vacuum concentrator (Eppendorf) for 1.5 h. Two glass beads per tube were added and liquid nitrogen shock frozen tissue was grounded into fine powder using TissueLyzer (QIAGEN). Total DNA was extracted using Sigma GenElute Plant Genomic DNA Miniprep Kit.

RUNGIS *et al*. (2004) developed 25 EST-SSRs for *Picea glauca* (Moench) Voss, the interspecific hybrid *Picea glauca* (Moench) Voss *x Picea engelmannii* Parry ex Engelm. and *Picea sitchensis* (Bong.) Carr.. In their study, the authors further tested all EST-SSRs and 18 genomic SSRs developed by different authors (see RUNGIS *et al*. 2004) in *P. sitchensis*, *P. glauca* and *Picea mariana* (Mill.) B.S.P. in sample sizes of 5 to 20 trees. Finally, they tested 19 above mentioned EST-SSRs (e.g. all EST-SSRs with dinucleotide repeat motif and one with trinucleotide repeat motif) on a panel of 23 *Picea* species (one tree per species). Using the same sample, they tested all 18 genomic SSRs as well, including 3 markers developed by SCOTTI *et al.* (2000). In aim to identify polymorphic set of EST-SSRs in *P. omorika*, we selected informative ones out of 19 EST-SSRs tested in a large sample sizes in *P. sitchensis, P. glauca* and P. *mariana*. We

excluded 2 markers due to the fact that the best-matched protein was found in the chloroplast and mitochondrial genomes, respectively (RUNGIS *et al*. 2004), 2 markers with less than 3 alleles per locus found in *P. sitchensis*, 3 markers which are monomorphic in approximately 70% of the 23 *Picea* species tested (including a single *P. omorika* tree) and the only one marker with a trinucleotide repeat motif due to the low number of alleles found in all species. We included genomic SSRs developed by SCOTTI *et al.* (2000) and tested by RUNGIS *et al*. (2004), except for PAAC 23, due to its rare repeat motif (GT). A subset of 12 SSRs was obtained (Table 1) and initially screened on 20 *P. omorika* trees using agarose gels. Markers indicating polymorphism in this screening were further used. A final subset of 6 markers (EST-SSRs only) was identified and further employed in the overall sample using capillary electrophoresis.

< Table 1 >

PCRs were performed in total volume of 10 µl containing 20 ng template DNA, 10x PCR buffer, 1.5 mM $MgCl_2$, 200 µM dNTPs, 0.2 µM of each forward and reverse primers and 0.025 U/µl of Platinum® *Taq* DNA polymerase (Invitrogen). Forward primers were labeled with fluorescent dyes (WellRED Oligos, Proligo). PCRs were carried out in a PTC-100 thermal cycler (MJ Research) with the following protocol: 3 min initial denaturation at 94 °C, 30 cycles of 20 s denaturation at 94 °C, 20 s annealing at 53 °C and 40 s extension at 72 °C and 10 min final extension at 72 °C. PCR products were separated using the CEQ™ 8000 Genetic Analysis System (Beckman & Coulter) according to the manufacturer's instructions. In order to minimize scoring errors, additional PCRs and fragment separations were repeated in cases when large allele dropout was suspected. As a further control, separate DNA extractions, PCR amplifications and product separations were performed using 2 randomly re-selected trees from each population. Two fragments per locus were sequenced in both directions using an ABI 3770 automated sequencer (Applied Biosystems), corresponding sequences were aligned using CLC Free Workbench 3 (http:&&www.clcbio.com&index.php?id=354) and used for fragment length corrections and calculations of the flanking region lengths and repetition scores.

Selective neutrality of markers was assessed by the Ewens-Watterson neutrality test (EWENS 1972, WATTERSON 1978) implemented in ARLEQUIN 3.11 (EXCOFFIER *et al*. 2005). Null alleles and typing errors due to stuttering and large allele dropout were tested using MICRO-CHECKER (VAN OOSTERHOUT *et al*. 2004). Linkage disequilibrium

(LD), inbreeding coefficients (F_{IS}) and deviations from Hardy-Weinberg-expectations (HWE) were calculated using GENEPOP 4.0 (RAYMOND & ROUSSET 1995). Multiple comparisons of HWE were Bonferroni corrected (RICE 1989). Number of alleles (*A*), effective number of alleles ($A_e=1/1-H_e$), observed (H_o) and expected heterozygosity (H_e) were calculated.

RESULTS AND DISCUSSION

In conifers, employment of traditionally developed genomic SSRs is limited due their large, repetitive genome (ELSIK & WILLIAMS 2001, BÉRUBÉ *et al.* 2007). An advantage of EST-SSRs over traditional genomic SSRs was demonstrated in other conifers (see BÉRUBÉ *et al.* 2007) and in *P. omorika* as well. Genomic SSRs yielded multiple amplification products in the initial screening and preliminary separation by capillary electrophoresis (data not shown). Sequence data of the remaining six markers revealed no mutations in flanking regions and confirmed presence of the anticipated repeat motifs. Amplification of homologous sequences was expected due to the high transferability of EST-SSRs to related species (RUNGIS *et al*. 2004, BÉRUBÉ *et al.* 2007). Sequence alignment revealed that WS0016.O09 is identical to WS0073.H08, except that primers for the latter marker were designed closer to the repeat region. Therefore, WS0016.O09 was excluded.

The identified set of 5 markers enabled unambiguous genotyping of all 50 trees of 2 natural populations. Typing errors due to stuttering and large allele dropouts were not found (*P>0.05*). Null alleles were not detected (*P>0.05*), as expected in EST-SSRs (RUNGIS *et al*. 2004, SELKOE & TOONEN 2006, BÉRUBÉ *et al.* 2007). Since EST-SSRs were developed from the expressed portion of the genome, there is a higher probability that these markers are not neutral compared to genomic SSRs. However, Ewens-Watterson test revealed neutrality of EST-SSRs used in this study (*P>0.05*). Probably, this is because these markers are from the 3' untranscribed regions (BÉRUBÉ *et al.* 2007) which are assumed to be under less selective constraints. Also the absence of any LD supports neutrality. Furthermore, locus-specific selection would result in differential deviations from HWE in multilocus studies (VITALIS *et al.* 2001). Such pattern was not

detected, as deviations from HWE were not significant after sequential Bonferroni corrections. In conclusion, the identified EST-SSRs are suitable molecular tools in *P. omorika* especially for mating system studies and other population genetic studies for which neutrality is a prerequisite.

Number of alleles (*A*), effective number of alleles (A_e), observed (H_o) and expected heterozygosity (H_e) are shown in Table 2. The number of alleles per locus was similar in both populations with a total of 61 alleles. Private alleles were found, 13 in TD and 14 in VS. The effective number of alleles varied across loci averaging to 7.33 in the total sample. Two SSRs revealed a small excess of heterozygotes in both populations, but 3 SSRs were indifferent and showed either an excess or deficit of heterozygotes. The pooled data resulted in a mean F_{IS} value close to zero (0.012) indicating random mating. Expected heterozygosity per locus ranged from 0.64 to 0.91 with an overall mean of 0.83. Values for observed heterozygosity were similar. Estimates of heterozygosities are close to those in *P. glauca, P. sitchensis* and *P. mariana* detected with an identical set of markers in similar sample sizes (RUNGIS *et al*. 2004). This result was highly unexpected, considering i) the small population sizes of *P. omorika* and ii) findings of an extreme demographic bottleneck and random genetic drift by chloroplast markers (NASRI *et al.* 2007), as well as considering iii) the very limited remnant natural range of this conifer compared to the wide distribution ranges of the *P. glauca* and *P. mariana*, as well as distribution range of *P. sitchensis* (BURNS & HONKALA 1990).

< Table 2 >

Population VS used in this study was previously analyzed by allozymes (KUITTINEN *et al.* 1991, BALLIAN *et al.* 2006) with partially contradictory results. The results of BALLIAN *et al.* (2006) indicate a low genetic diversity (H_e=0.05) presumably due to genetic drift while KUITTINEN *et al.* (1991) found surprisingly high levels of genetic diversity (H_e=0.13) which is in the range of other spruce species. However, when identical isozyme loci in the two cited studies are considered, genetic differences between those studies are negligible (data not shown), corroborating our results and the view that *P. omorika* contains high genetic variation although being a narrow endemic. This is also supported by pronounced outcrossing rates (KUITTINEN & SAVOLAINEN 1992), high number of embryonic lethals (SAVOLAINEN e*t al.* 1992) and high inbreeding depression (GEBUREK 1986).

ACKNOWLEDGEMENTS

This work - financially supported by Bioversity International and the Republic of Austria - comprises a portion of the PhD dissertation of J.A. The authors thank L. Weißenbacher, M. Lakić and D. Milekić for their help in the field and B. Heinze for fruitful discussions. All samples were collected with permission and support of the National Park 'Tara', (Serbia), Ministry of Agriculture, Forestry and Water Management and Forest Enterprise 'Šume Republike Srpske' (Republic of Srpska, Bosnia and Herzegovina).

REFERENCES

BALLIAN D., LONGAUER R., MIKIĆ T., PAULE L., KAJBA D. & GÖMÖRY D., 2006. - Genetic structure of a rare European conifer, Serbian spruce (*Picea omorika* (Pančić) Purk.). *Pl. Syst. Evol.* **260**: 53-63.

BÉRUBÉ Y., ZHUANG J., RUNGIS D., RALPH S., BOHLMANN J. & RITLAND K., 2007. - Characterization of EST-SSRs in loblolly pine and spruce. *Tree. Genet. Genomes* **3**: 251-259.

BURNS, R.M. & HONKALA, B.H. (techn. coord.), 1990. - Silvics of North America, Vol. 1, Conifers. US Department of Agriculture, Washington DC.

BURSCHEL P., 1965. - Die Omorikafichte. *Forstarchiv* **36**: 113-131.

CONIFER SPECIALIST GROUP, 1998. - *Picea omorika*. *In*: IUCN 2007. *2007 IUCN Red List of Threatened Species* (http://www.iucnredlist.org/search/details.php/30313/summ).

ELSIK C.G. & WILLIAMS C.G., 2001. - Low-copy microsatellite recovery from a conifer genome. *Theor. Appl. Genet*. **103**: 1189-1195.

EXCOFFIER L., LAVAL G. & SCHNEIDER S., 2005. - Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1:** 47-50.

EWENS W.J., 1972. - The sampling theory of selectively neutral alleles. *Theor. Popul. Biol.* **3**: 87-112.

GEBUREK TH. 1986. - Some results of inbreeding depression in Serbian spruce (*Picea omorika* (Panč.) Purk.). *Silvae Genet.* **35**: 169-172.

KUITTINEN H. & SAVOLAINEN O., 1992. - *Picea omorika* is a self-fertile but outcrossing conifer. *Heredity* **68**: 183-187.

KUITTINEN H., MUONA O., KÄRKKÄINEN K. & BORZAN Ž., 1991. - Serbian spruce, a narrow endemic, contains much genetic variation. *Can. J. For. Res.* **21**: 363-367.

NASRI N, BOJOVIC S, VENDRAMIN G.G. & FADY B., 2007.- Population genetic structure of the relict Serbian spruce, *Picea amorika*, infreered from plastid DNA. *Pl. Syst. Evol.* (online early).

RAYMOND M. & ROUSSET F., 1995. - Genepop (version-1.2) - population genetics software for exact tests and ecumenicism. *J. Hered.* **86**: 248-249.

RICE W.R., 1989. - Analyzing tables of statistical tests. *Evolution* **43**: 223-225.

RUNGIS D., BÉRUBÉ Y., ZHANG J., RALPH S., RITLAND C.E., ELLIS B.E., DOUGLAS C., BOHLMANN J. & RITLAND K., 2004. - Robust simple sequence repeat markers for spruce (*Picea* spp.) from expressed sequence tags. *Theor. Appl. Genet.* **109**: 1283-1294.

SAVOLAINEN O., KÄRKKÄINNEN K. & KUITTINEN H., 1992. - Estimating numbers of embryonic lethals in conifers. *Heredity* **69**: 308-314.

SELKOE K.A. & TOONEN R.J., 2006. - Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecol. Lett*. **9**: 615-629.

SCOTTI I., MAGNI F., FINK R., POWELL W., BINELLI G. & HEDLEY P.E., 2000. - Microsatellite repeats are not randomly distributed within Norway spruce (*Picea abies* K.) expressed sequences. *Genome* **43**: 41-46.

VAN OOSTERHOUT C., WILLIAM F., HUTCHINSON D., WILLS P.M. & SHIPLEY P., 2004. - MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes* **4**: 535–538.

VITALIS R., DAWSON K. & BOURSOT P., 2001. - Interpretation of variation across marker loci as evidence of selection. *Genetics* **158**: 1811-1823.

WATTERSON G., 1978. - The homozygosity test of neutrality. *Genetics* **88**: 405-417.

Table 1. List of SSRs tested in *P. omorika.*

Locus	Repeat motif	NCBI accession number
WS0016.O09[1]	$(AT)_9$	CN480894
WS0019.F22	$(AT)_{13}$	CN480896
WS0022.B15	$(AG)_{12}$	CN480899
WS0023.B03	$(AT)_{10}$	CN480900
WS0053.K16	$(AT)_{13}$	CN480898
WS0071.J15[2]	$(AT)_{22}$	CN480902
WS0073.H08	$(AT)_{14}$	CN480903
WS0082.E23[2]	$(TA)_{11}$	CN480909
WS0082.O23[2]	$(TA)_{15}$	CN480910
WS0092.A19[2]	$(AC)_9$	CN480888
PAAC 17[2,3]	$(AC)_{36}$	AJ131107
PAAC 19[2,3]	$(CT)_{23}CAA(TG)_{12}$	AJ131108

[1] excluded from the study since identical to the WS0073.H08; [2] excluded after initial screening; [3] developed by SCOTTI *et al.* (2000), all other markers were developed by RUNGIS *et al.* (2004).

Table 2. Parameters of genetic variation in *P. omorika* at 5 loci compared to *P. glauca, P. sitchensis* and *P. mariana.*

	Locus	**WS00 22.B15**	**WS00 53.K16**	**WS00 73.H08**	**WS00 19.F22**	**WS00 23.B03**	**Total**
P. omorika	Size (bp)	*208-221*[1]	*225-243*	*236-242*	*404-412*	*216-234*	
	FR	160	200	193	349	156	
Pop. TD	Size (bp)	168-206	228-250	219-245	371-431	180-232	
	Size (rep)	4-23	14-25	13-26	11-41	12-38	**4-41**
	PA	-	2	3	2	6	**13**
	A	7	9	5	13	13	**47**
	A_e	5.80	7.04	2.82	9.50	9.35	**6.90**
	H_o&H_e	0.84&0.83	0.92&0.86	0.52&0.64	0.92&0.89	0.92&0.89	**0.82&0.82**
	F_{IS}	-0.015	-0.074	0.197	-0.029	-0.032	**0.000**
Pop. VS	Size (bp)	168-206	230-250	219-243	399-431	180-228	
	Size (rep)	4-23	15-25	13-25	25-41	12-36	**4-41**
	PA	-	3	3	3	5	**14**
	A	7	10	5	14	12	**48**
	A_e	5.26	7.16	3.78	11.04	11.56	**7.76**
	H_o&H_e	0.76&0.81	0.72&0.86	0.80&0.74	0.92&0.91	0.92&0.91	**0.82&0.84**
	F_{IS}	0.059	0.166	-0.090	-0.012	-0.007	**0.025**
Mean	***A***	**7**	**12**	**8**	**16**	**18**	61
(TD+VS)	$\boldsymbol{A_e}$	**5.53**	**7.10**	**3.30**	**10.27**	**10.45**	7.33
	$\boldsymbol{H_o}$&$\boldsymbol{H_e}$	**0.80&0.82**	**0.82&0.86**	**0.66&0.69**	**0.92&0.90**	**0.92&0.90**	0.82&0.83
	$\boldsymbol{F_{IS}}$	**0.021**	**0.046**	**0.045**	**-0.020**	**-0.019**	0.012
P. glauca	H_o&H_e	*0.92&0.85*[2]	*0.50&0.35*	*0.84&0.75*	*0.97&0.90*	*0.89&0.94*	**0.82&0.76**
P. sitchensis	H_o&H_e	*0.57&0.47*	*0.66&0.68*	*0.74&0.61*	*0.70&0.58*	*0.79&0.69*	**0.69&0.61**
P. mariana	H_o&H_e	*0.71&0.50*	*0.78&0.65*	*0.45&0.45*	*0.96&0.96*	*0.74&0.81*	**0.73&0.67**

Size (bp) – range of the fragment lengths in base pairs; FR – length of the flanking regions (in base pairs); Size (rep) - range of the number of repeats; PA - number of private alleles; *A* - number of alleles; A_e - effective number of alleles; H_o - observed heterozygosity; H_e - expected heterozygosity; F_{IS} - inbreeding coefficient.

[1] Data in italics were taken from RUNGIS *et al.* (2004) based on a single tree.

[2] Data in italics were taken from RUNGIS *et al.* (2004) based on sample sizes ranging from 16 to 20 trees.

3.2 Was there more than one refugial population of an endemic Serbian spruce in the Balkans during the last glaciation?

Aleksić MJ and Geburek T

Abstract

The genetic variation in an endemic and a relict conifer species - *Picea omorika* (Pančić) Purk. was assessed by amplifying the partial sequence of the second intron of the mitochondrial NADH dehydrogenase subunit1 gene in 200 trees originating from 10 natural populations. Five haplotypes were detected and size variation was due to the variable number of a 34 bp minisatellite. In overall sample, haplotypic richness was 3.007 and all population genetics parameters were similar to those reported in widely distributed *Picea* species - H_S = 0.075, H_T = 0.225 and G_{ST} = 0.668. However, the distribution of haplotypes was not random, as predominant and presumably the ancestral haplotype A (one 34 bp repeat) was fixed in eight populations, while it was absent in a single, small population comprised of approximately 400 trees in which the highest gene diversity (H_S = 0.468) was observed. Current structuring of genetic variation in *P. omorika* supported by pallinological data is interpreted as a result of the secondary contact of two gene pools established long before the last glacial maximum and characterized by different evolutionary histories and past migrations. Our study supports the views that (i) a complex structuring can be found even within a small areas within the Balkan, and that (ii) the sampling scheme is essential in assessing genetic variability in species *per se*, especially in cases of formerly widely distributed species currently confined to small areas - when a single population from the second, highly polymorphic gene pool was omitted from the analyses, all population genetics parameters substantially decreased, misleadingly implying genetic uniformity in this species - H_S = 0.031, H_T = 0.033 and G_{ST} = 0.070. Almost absent seed flow in overall sample (Nm < 1) implied that isolation-by-distance seems to be very effective in *P. omorika* and, in dependence of the paternal gene flow, this has to be taken into consideration for delineation of conservation units of this species.

Key words: *Picea omorika*, mitochondrial DNA, genetic diversity, phylogeography, Balkan's refugia, postglacial migration

Introduction

Global climatic changes during the past 2.4 million years associated with ice ages occurring regularly every 100 kilo-years (ky) and followed by 20 ky warm periods (e.g. Milankovitch cycles, Bennett 1990) had major impact on the present genetic structure of populations, species and communities (Hewitt 2000). For the long-term persistence of a species, survival within refugia was important during both, cool and warm periods, because disappearance of a species from such regions could lead to the complete extinction of that taxon during the next cycle (Taberlet et al. 1998).

The Balkan (and whole Mediterranean Basin) is characterized by high genetic variation in terms of number of species, subspecies divisions and allelic richness ('southern richness', Hewitt 2000) and is designated biodiversity hotspot for conservation priorities (Myers et al. 2000). It has been shown that all eastern and many western European species were confined to the Balkan's refugia during the ice ages (Hewitt 2000). Due to the pronounced orography and plasticity of the terrain (Čolić 1965), populations within this region were not exposed to large geographical displacements during the Holocene and they could descend and ascend mountains and rivers to track a suitable environment in response to climatic changes (Hewitt 2001, Petit et al. 2005). As a corollary, long-standing isolated populations appear to be the common phenomena in the Balkan. Recently, Petit et al. (2005) postulated that those populations should be considered relict rather than refugial. The authors highlighted such relict tree populations as an 'evolutionary heritage of disproportionate significance for the conservation of European plant diversity'.

It has been shown for several species that speciation as well as geographically distinct lineages within the species could be of Pliocene origin and that only few mutations distinguish post-glacial lineages (Hewitt 2001). Furthermore, as shown by Petit et al. (2005), heterogeneous rates of molecular evolution across lineages are inversely related with their stability during the ice ages. Among other factors affecting the

levels and spatial distribution of genetic diversity within the species, geographical range appears to have large impact (Hamrick and Godt 1996). Generally, geographically widespread species tend to posses more genetic polymorphisms, while endemic species tend to become depauperate. Low genetic diversity has been found in many populations of relict and/or endemic tree species (e.g. El Mousadik and Petit 1996, Huang et al. 2001, Ge et al. 2005 and references therein).

Out of more than 30 *Picea* A. Dietr. species recognized by recent clasifications and checklists (Farjon 2001), only two are confined to Europe. While Norway spruce [*Picea abies* (L.) Karst.] is widely distributed throughout Europe, Serbian spruce [*Picea omorika* (Pančˇ.) Purk.] is endemic to an extremely small area in the northern Balkans (western Bosnia and Herzegovina and eastern Serbia). This species was discovered in 1875 near the hamlet of Zaovine on the Tara Mountain in Serbia and is characterized as a relict of the Arcto-Tertiary flora (Fukarek 1967). However, little is known on genetic diversity of this species and the data are exceptionally controversial (see references in Discussion).

During Tertiary and later interglacial periods *P. omoricoides* Weber, a close fossil relative of *P. omorika,* was probably widely distributed in Europe, although such early fossil findings are still unclear (Ravazzi 2002). Some macrofossils can be traced back to the Pliocene in former Czechoslovakia (Búžek et al. 1985). Also during the early Pleistocene, a wide distribution of this species is assumed in Europe together with Norway spruce due to the findings of fossil pollen and needles at so distant places as the northern Alpine foreland and the Aegean coast (Ravazzi 2002). With the onset of the last glacial maximum (LGM), *P. omorika* was presumably restricted to a smaller region than *P. abies* and outlasted in the Balkans (Ravazzi 2002). After the retreat of the glaciers *P. omorika* could not – contrarily to more competitive species - enlarge its population size probably due to the lack of the migration corridors and/or stepping-stones in close vicinity of its refugium (Ballian et al. 2006). However, fossil findings of *P. omorika*-like pollen in Serbia, e.g. in areas south-east relative to its current range indicate a wider distribution in the Balkans during the last 12 ky (Gigov, 1956).

Like in most conifers, the mitochondrial genome (mtDNA) is maternally inherited in Serbian spruce (David and Keathley, 1996). Limited seed dispersal, small effective population size for maternally inherited genomes and long life span of trees have probably maintained a geographic genetic pattern primarily shaped by post-glacial

history (Newton et al. 1999). Such pattern of neutral genetic diversity is altered by drift and recurrent gene flow. Phylogeography, e.g. geographic patterns of genealogical structure across the range of species (Avise 2000) can potentially discriminate between those evolutionary forces shaping the current genetic structuring of a species (Templeton et al. 1995, Schaal et al. 1998). Therefore, mtDNA is one of the most powerful tools in studies of phylogeography in conifers, especially when enhanced by pallinological data (Cruzan and Templeton 2000) and considering species specific characteristics.

We amplified the partial second intron of the mitochondrial NADH dehydrogenase subunit1 gene (*nad*1 intron b/c) in 200 individuals of *P. omorika* originating from 10 natural populations to (i) assess potential variability at this locus, (ii) analyze geographical pattern of genealogical structuring and to (iii) infer migration routes of the species in relation to the last glacial maximum.

Materials and methods

Study species

Serbian spruce is characterized by the narrow ecological niche and, in comparison to *P. abies*, is confined to a warmer climate occupying altitudes ranging from 800 to 1600 m a.s.l. predominantly at north–orientated, usually very steep slopes of the hills and the River Drina canyon (Ostojić 2005). Such sites are characterized by high humidity, evenly distributed annual precipitation, heavy snow cover and low winter temperatures. As a pioneer species, Serbian spruce quickly occupies openings after fire and other catastrophes (Čolić 1966). However, in mixture with more competitive species, such as Norway spruce, it retreats to ravines and other areas less inhabitable by its competitors. Additive genetic variance for shade tolerance is rather low (Tucić and Stojković 2001) and thus natural regeneration of this opportunistic species strongly depends on available favorable habitats and competition by other tree species (Ostojić 2005).

Serbian spruce is a monoecious species and reproduces sexually, while vegetative regeneration has been reported only occasionally (Čolić 1966). It is a self-fertile, but outcrossing species (Kuittinen and Savolainen 1992) carrying a relative high number of embryonic lethals (Savolainen et al. 1992). Although it can artificially easily hybridize with different members of the genus *Picea* (Ledig et al. 2004), there is no introgression from *P. abies* in nature. Its pollen is smaller than of Norway spruce, but the sedimentation velocity (approximately 5.2 cm s^{-1}) is nearly identical (Eisenhut 1961). In plantations, Serbian spruce is characterized as a prolific seeder because almost annually trees of the dominant vegetation layer are shedding seeds (Král 2002). However, annual and individual oscillations are reported in natural populations (Čolić 1966, Gajić et al. 1994). Experimental data on effectiveness of Serbian spruce to distribute its genes through the wind-dispersed pollen or seeds in natural populations are not available.

Study sites

Majority of Serbian spruce natural populations are confined to the region where the river Drina meanders Mts. Zvijezda and Tara - between 43°21' and 44°08' north and 18°37' and 19°45' east. They represent the core area of the species forming a very patchy meta-population and only three populations are found south from this region. Within the core area, small patches of trees and even isolated single individuals are frequent. Some of them appear to be remains from formerly larger populations, while others are assumed to have arisen from neighboring populations by sporadic wind-dispersed seeds (Gajić et al. 1994). Reports that less than 1000 trees of this IUCN red-listed species are remnant (Conifer Specialist Group 1998) are probably underestimating the total population size found today. More than 30 populations, each consisting of several tens to several thousands of trees are scattered within an area of approximately 10.000 km^2 (Ostojić 2005). However, the cumulative area is certainly very small and probably does not exceed 60 ha (Burschel 1965).

The list of natural populations used in this study is presented in Table 1 and their geographical locations are presented in Figure 1. All populations are comprised of several small, high density patches of Serbian spruce adjacent to sites where this

confers is mixed with other species (i.e. Norway spruce, silver fir, black pine, aspen, beech and birch) in varying proportions. Pure stands of Serbian spruce have rarely been found. In Serbia, where River Drina meanders Mts. Zvijezda and Tara, several populations and patches of Serbian spruce trees are found within the complex 'Zvijezda-Vidača-Veliki kraj' (Gajić et al. 1994) which can be separated into 'Zvijezda' (approx. 700 ha) and 'Vidača-Veliki kraj' (approx. 1.300 ha). Due to inaccessibility, we could only collect material from two populations located at 'Zvijezda'. Sampling at Mt. Tara was exhaustive, including a site 'Mitrovac - MI', which is the atypical population because it is found at swampy terrain. Unfortunately, we could only find three surviving trees at the latter site and we treated MI not as a population, but as a patch of trees, remains from a former population. Two populations from the Republic of Srpska, Bosnia and Herzegovina were sampled as well, as neighboring populations at the southern edge of the species range. Two and five of the populations used in this study were also used by Ballian et al. (2006) and Nasri et al. (2007), respectively, but samples were taken independently.

>Table 1 approximately here

Samples were taken from 20 trees per population. The age of the sampled trees was not recorded, but we sampled exclusively mature trees indicated by a diameter at breast height exceeding 25 cm. All sampled trees within each population were evenly distributed across the space. Young twigs were kept at low temperature during transportation and were stored at -60 °C prior to DNA extraction.

>Figure 1 approximately here

DNA extraction, PCR amplification and fragment separation

Fresh needles from each tree were cut into 2-3 mm long pieces and tissue (approx. 100 mg) was placed separately into 2 ml Eppendorf tubes. Initial drying was performed for one hour with methanol (500 µl per tube). Tissue was fully dried in a vacuum concentrator (Eppendorf) for 1.5 h. Two glass beads per tube were added and liquid nitrogen shock frozen tissue was grounded into fine powder using TissueLyzer System (QIAGEN). Total DNA was extracted using Sigma GenElute Plant Genomic DNA Miniprep Kit. The *nad*1 intron b/c was amplified using the forward and reverse primers

developed by Sperisen et al. (2001). PCRs were performed in a total volume of 10 µl containing 20 ng of the template DNA, 10x PCR buffer, 1.5 mM $MgCl_2$, 200 µM dNTPs, 0.1 µM of each forward and reverse primer and 0.025 U/µl of Platinum® *Taq* DNA polymerase (Invitrogen). Both primers were labeled with different fluorescent dyes (WellRED Oligos, Proligo). PCRs were carried out in a PTC-100 thermal cycler (MJ Research) with the following touchdown protocol: 3 min initial denaturation at 94 °C, 9 cycles of 30 s denaturation at 94 °C, 30 s annealing at 67 °C, 1 min extension at 72 °C with the progressive decrease of the temperature by 1 °C per cycle, followed by 29 cycles of 30 s denaturation at 94 °C, 30 s of annealing at 57 °C, 1 min extension at 72 °C and final extension of 10 min at 72 °C. All fragments were digested with the blunt-end restriction enzyme *Eco*RV (Applied Biosystems) and were separated using CEQ 8000 Genetic Analysis System (Beckman and Coulter) according to the manufacturer's instructions. Full lengths of the mitochondrial fragments (haplotypes) were obtained by summing the lengths of the corresponding 5' and 3' fragments. Two fragments per each full size variant (except for the rare haplotypes which were found in a single tree each) were sequenced in both directions using ABI 3770 automated sequencer (Applied Biosystems). Sequences were deposited in Gene Bank (accessions EU649707-EU649710 and EU649712).

Genetic analysis

Observed number of haplotypes (*nh*), estimates of unbiased average gene diversity H_S [equivalent to expected heterozygosity for diploid data (Weir, 1996)] and haplotypic richness (*R*) were calculated per population and for overall sample. Observed number of haplotypes and average gene diversity, as well as H_T - total gene diversity, G_{ST} and N_{ST} were calculated in overall sample and subsequently in overall sample excluding population ST. G_{ST} is the proportion of the total diversity due to differences between populations based on haplotype composition (Nei, 1987), while N_{ST} additionally includes the genetic distance among haplotypes. To obtain statistical significance between G_{ST} and N_{ST}, 1000 permutations were run. As described by Dumolin-Lapègue et al. (1997), N_{ST} will be higher than G_{ST} if, on average, pairs of different haplotypes from the same population have more similar sequences than pairs of different haplotypes from separate

populations. All calculations were performed using PERMUT 2.0 (Pons and Petit 1996). Haplotypic richness (*R*) was calculated using HAPLODIV (Pons and Petit 1995).

An analysis of molecular variance (AMOVA) (Excoffier et al. 1992) was carried out with ARLEQUIN (Excoffier et al. 2005) to determine the partitioning of variation within and between populations. Analysis was performed including all populations as well as excluding population ST.

Evolutionary relationships among haplotypes were assessed by reconstructing minimum spanning tree generated with the software TCS (Clement et al., 2000). Duplications of the 34 bp minisatellite repeat unit were treated as a single mutational event and as a single nucleotide indels and gaps were treated as fifth state. Although minisatellites are usually excluded from phylogeographic studies due to the potential homoplasy (e.g. Olsen and Schaal 1999, Dumolin-Lapègue et al. 1997), this particular minisatellite, as well as the 32 bp minisatellite described at the same locus were successfully used in Norway spruce for delineating Baltico-Nordic and central European lineages (Sperisen et al. 2001), justifying their phylogeographic use.

Unweighted pair-group method with arithmetic mean (UPGMA) (Sneath and Sokal 1973) based on Cavalli-Sforza chord distance (Cavalli-Sforza and Edwards, 1967) was used to construct a tree by successive clustering using an average-linkage method of clustering. Cavalli-Sforza chord distance assumes that there is no mutation and that all gene frequency changes are by genetic drift alone. Calculations were carried out using the software package PHYLIP 3.61 (Felsenstein, 2003) and UPGMA dendrograme was reconstructed using TREEVIEW (Page 1996).

Mantel test (Mantel 1967) was used to test the correlation between linear geographic and genetic distances between populations using the software IBD (Bohonak 2002). Two genetic distances (Cavalli-Sforza chord distance and pairwise population F_{ST}) were used. Undefined values which occurred when identical, monomorphic populations were involved in the pairwise calculations were converted to 0.0001 for correlations. All calculations were performed in the overall sample as well as excluding population ST.

Phylogeographic analysis was further performed by visual overlay of the haplotype tree upon geography and by interpreting the patterns of congruence or lack of congruence between the geographical distribution of haplotypes and their genealogical

relations assessed from haplotype tree (Schaal et al. 1998). A pattern of congruence is obvious if closely related haplotypes are geographically restricted and occur in proximity to each other. Such congruence would indicate a longstanding pattern of highly restricted gene flow (Schaal et al. 1998). The most ancient haplotype should be geographically widespread, whereas the most recent haplotypes should be at the tips of the haplotype tree and be localized geographically (Templeton et al. 1995, Schaal et al. 1998).

Effective number of female migrants per generation, as an indirect estimate of average levels of the seed flow relative to genetic drift under infinite island model, was calculated for haploid data as $Nm = ½ (1/F_{ST} – 1)$, where N is female effective population size and m is female migration rate (Slatkin 1993). This estimate was calculated in overall samples, in overall samples excluding population ST and for each pair of populations.

Results

Intraspecific variability in the partial mitochondrial *nad*1 intron b/c in Serbian spruce was detected. Variability was exclusively due to the variable copy number of the 34 bp minisatellite, as sequence data revealed no base substitutions. In total, five haplotypes were found and three of them were private (Table 2). Haplotype A, harbouring only one minisatellite copy was predominant, as eight out of ten populations were fixed for this type. The other four haplotypes were found only in two populations, VS and ST. Those populations shared a single common haplotype (B) and other three haplotypes were private and found only in population VS (J) or ST (C and D). In VS, haplotype A was also the most frequent one, as other haplotypes were found in only three trees. Surprisingly, this haplotype A was not found at all in ST. In this population, the predominant haplotype was B and relative abundance of haplotypes decreased with increasing number of minisatellite copies (D was found in a single tree).

>Table 2 approximately here

However, exclusive private haplotype J (10 minisatellite copies) was not found in ST, but in a single tree from the population VS. This tree was additionally characterized by atypical outcome of the fragment separation by capillary electrophoresis, even after repeated separate extractions, PCRs, enzyme digestions and fragment separations. Apart from two high peaks representing 5' and 3' fragments of the full sequence, additional nine peaks of the same height, corresponding to the 5' fragments were detected. The lengths of those fragments differed in the number of the minisatellite copies, implying the presence of fragments harbouring 1-9 and 11 minisatellite copies. Although several possible explanations were proposed for similar finding in *P. abies* (Sperisen et al. 2001), such pattern is probably a common PCR artifact occurring during the amplification of regions with high number of tandem repeats, as low signal peak corresponding to the fragment having eleven minisatellite repeats was detected.

The highest gene diversity was detected in population ST (H_S = 0.468), it was lower in VS (H_S = 0.279) and zero in all other populations fixed for a single haplotype A (Table 2). H_S, H_T and G_{ST} in overall sample were 0.075, 0.225 and 0.668, respectively (Table 3). Those values decreased to 0.031, 0.033 and 0.070, respectively, in case when population ST was omitted from the sample. Haplotypic richness in overall sample was 3.007.

>Table 3 approximately here

AMOVA revealed that a high and almost similar amount of genetic variation is found among (56.8 %) and within populations (43.2 %). If population ST is excluded, only 2.9 % of the genetic variation will be assigned to the variation among population.

Minimum spanning tree revealed stepping-stone pattern of haplotype occurrence, e.g. emergence of longer haplotypes by sequential increase in minisatellite copy number. Such pattern implies that haplotype A might be the ancestral one (Figure 2).

>Figure 2 approximately here

Expectedly, UPGMA analysis (Figure 3) revealed clustering of all populations fixed for haplotype A which shared the same branch (D^2 = 0.000) and the highest genetic distance was found between this cluster and a single population ST (D^2 = 0.485). The genetic distance between VS and ST (D^2 = 0.446) was more pronounced than between VS and all other populations fixed for haplotype A (D^2 = 0.039). Same

clustering of populations was reveled by reconstructing Neighbor Joining tree (data not shown).

>Figure 3 approximately here

No significant correlation between geographic and genetic distances was detected (r = 0.0285, $P \leq 0.351$ for Cavalli-Sforza chord distance; r = 0.0261, $P \leq 0.371$ for pairwise F_{ST}). If population ST was excluded from the sample, correlation coefficients decreased and also were not significant.

Phylogeographic signal in the haplotype distribution was not detected, as N_{ST} values were identical to G_{ST} values in the overall sample (0.668) and when omitting population ST (0.070). N_{ST} and G_{ST} values were not statistically different. Visual overlay of the haplotype tree upon geographical locations of populations revealed congruence of genealogical and geographical pattern of genetic variation. Haplotype A (presumably the ancestral one) was found to be geographically widespread, while derivative (e.g. tip) haplotypes were found to be localized geographically, as they were present exclusively in a single population ST and only in three trees from the neighboring population VS. Such congruence of genealogical and geographical structure implied longstanding, highly restricted gene flow.

Effective number of female migrants per generation (e.g. seed flow) was found to be very limited, although differed noticeably depending on the sample set. Seed flow was extremely low ($Nm < 1$), while it was rather high ($Nm > 7$) when excluding population ST. Pairwise estimates of seed flow between populations differed as well, ranging from $Nm < 1$ between ST and all other monomorphic populations and expectedly to infinite seed flow among pairs of populations due to complete fixation.

Discussion

Population genetics aspect

The retention of an ancestral molecular polymorphisms, as observed in *Pinus* (Syring et al. 2007), is likely in *Picea* as well (Bouillé and Bousquet 2005). The variability at the

mitochondrial *nad*1 intron b/c due to the duplications of the 34 bp minisatellite is probably an ancestral feature shared among several *Picea* lineages – the variable number of the 34 bp minisatellite was firstly described in range-wide sample of *P. abies* (Sperisen et al. 2001), it was later found in *P. crassifolia* (Meng et al. 2007) and presumably is present in *P. jezoensis* (Aizawa et al. 2007). Such shared polymorphisms between species may be expected only in cases of explosive radiation and/or lateral gene flow (cf. Karvonen et al. 1994, Bouillé and Bousquet 2005).

High genetic diversity at a single mitochondrial locus, as found in *P. omorika*, was probably associated with a long evolutionary history, which allows genetic variation to accumulate within lineages (cf. Chiang and Schaal 1999). Most DNA sequences are assumed to diverge little over the ice ages suggesting that only few new mutations will distinguish post-glacial haplotypes and implying that geographically distinct lineages within the species could be of Pliocene origin (Hewitt 2000). Taking into account the fact that mtDNA is characterized by the lowest rate of evolution out of all tree plant genomes (Palmer et al. 2000), the presence of several haplotypes in *P. omorika* prior to the LGM seems likely and only two rare haplotypes (D and J), each found in a single tree, might be of the post-glacial origin.

However, we did not expect a pronounced genetic diversity at this maternal locus in our material, especially because the *nad*1 intron b/c was previously found to be monomorphic in some *Picea* species with a limited natural range, such as in *P. chihuahuana* (Jaramillo-Correa et al. 2006) and *P. rubens* (Jaramillo-Correa and Bousquet 2003). Since a single haplotype A is found to be the predominant and almost all other haplotypes are found in a single population, five haplotypes detected in *P. omorika* amounted to a low average estimates of H_S (0.075), but to a very high among population differentiation (G_{ST} = 0.668). Such markedly mtDNA population subdivision was comparable to G_{ST} values estimated in widely distributed *Picea* species using one or even more different mtDNA markers: *P. jezoensis* - 0.901 (Aizawa et al. 2007), *P. abies* - 0.676 (Sperisen et al. 2001), *P. mariana* - 0.671 (Jaramillo-Correa et al. 2004) and *P. crassifolia* - 0.512 (Meng et al. 2007).

A single, highly polymorphic *P. omorika* population is confined to an area of less than one ha and is comprised of only 400 trees. Exclusion of this population from the analyses resulted in a distinct lowering in all population genetics parameters, e.g. G_{ST} and H_S decreased to 0.070 and 0.031, respectively. Latter value was even lower than in

P. rubens, where H_S detected in five populations was 0.120 (Jaramillo-Correa and Bousquet 2003). Such a chance of among population differentiation was also found in *P. abies* in the Alpine region when highly polymorphic populations were omitted (Gugerli et al. 2001). In their study, F_{ST} value decreased from 0.41 to 0.06.

As shown in our study, a single, small forest caused extremely pronounced differences in estimates of the genetic diversity in *P. omorika*. It seems that in assessing genetic diversity in a species, special care should be devoted to small populations and patches of trees, as they might represent the remains of populations currently experiencing extinction and they might harbour important information on genetic diversity in species *per se*. The risk of misleading assumptions on genetic diversity due to inadequate sampling is even higher in species widely distributed in the past which are currently confined to small areas and consequently, tend to become depauperate (Hamrick and Godt 1996).

Whether Serbian spruce is genetically depauperate or not is an intricate question. Ballian et al. (2006) concluded that Serbian spruce would have reduced allozyme diversity, while Kuittinen et al. (1991) described a relatively high genetic diversity at allozyme loci. However, if one compares identical markers, differences between both studies become marginal revealing surprisingly high diversity for a narrow endemic, as also revealed by nuclear EST-SSRs (Aleksić et al. submitted). Contrary to these results, Nasri et al. (2007) suggested that based on cpDNA SSRs, Serbian spruce is one of the few examples of genetically depauperate tree species. Our results rather corroborate the view that Serbian spruce is not genetically depauperate species (Kuittinen et al. 1991, Aleksić et al. submitted).

Expectedly, the estimates of seed flow varied from very low values for the overall sample to infinitely large seed flow between each pair of eight populations fixed for the haplotype A. As a selection can be ruled out for the given mtDNA locus, such extreme variation is probably due to historical events (see below) which caused biologically misleading estimate of *Nm* based on a single locus (cf. Templeton 1998). In our study, it was surprising to find monomorhic populations in close vicinity to the genetically highly diverse population ST - the maximal geographical distance between population pairs was 17 km and all distances averaged to only 6 km. The predominant haplotype in ST is found only in two trees in VS, suggesting that the effective gene flow by seeds must be assessed as very limited. At least as far as maternal gene flow is concerned, isolation-

by-distance seems to be very effective in this conifer and, in dependence of the paternal gene flow, this has to be taken into consideration by delineation of conservation units of the species.

Phylogeographic inferences

As the emergence of the minisatellite motif within a certain haplotype must precede the occurrence of haplotypes harbouring increased number of minisatellite copies (Aleksić et al. in preparation) generated by slipped-strand mispairing (as suggested by Levinson and Gutman 1987), it seems likely that haplotype A is the ancestral within our sample, despite the existence of haplotypes with no 34 bp minisatellite copies reported in *P. crassifolia* (Meng et al. 2007). Although the inferences on evolutionary relationships among haplotypes based on micro- and minisatellite regions must be taken cautiously due to the potential homoplasy (cf. Olsen and Schaal 1999, Dumolin-Lapègue et al. 1997), this particular minisatellite, as well as the 32 bp minisatellite described at the same mtDNA locus were successfully used in Norway spruce for delineating Baltico-Nordic and central European lineages (Sperisen et al. 2001).

Surprisingly, our study revealed that the predominant, presumably ancestral haplotype A is absent in population ST characterized by the highest genetic diversity of all populations studied. Such finding may simply be due to the sampling. However, based on our sample size, it is unlikely that haplotype A would also be a common type in ST. The observed pattern of genetic structuring in Serbian spruce corresponded well to one of the examples described in Cruzan and Templeton (2000) where a disjunct population is assumed to originate either from long distance dispersal or via migration from a cryptic refugium. Hence, we discuss those biological reasons as causative.

(1) If Serbian spruce was confined to a single refugium in the Balkan during the last glaciation, it is reasonable to assume that such population was comprised of several mtDNA haplotypes, as found in ST, since the emergence of four different haplotypes through mutations during the last 12 ky is not likely (cf. Hewitt 2000). Subsequently, population fixed for a single haplotype A (disjunct population in Cruzan and Templeton 2000) might emerge from this population via long distance dispersal and it might be fragmented later into several smaller populations resulting in a patchy meta-population

structure found today. However, in this case it is rather difficult to explain the loss of the ancestral haplotype A in population ST, as this haplotype is assumed to be the most abundant one and drift effects are known to be less effective on abundant haplotypes.

(2) Serbian spruce might survive LGM in at least two isolated refugial populations representing two different gene pools of pre-glacial origin, as indicated by AMOVA. One gene pool might have been monomorphic and confined to the Mt. Tara and/or neighboring region. The presence of *P. omorika* in this region prior to LGM is supported by findings of fossil *P. omorika* pollen in 12 ky old peat sediments at 1.100 m a.s.l. (Gigov 1956). Therefore, it seems that monomorphic *P. omorika* populations survived LGM *in situ* and they were not exposed to large geographical displacements since the time of their establishment. Although in that case high genetic diversity is expected (Petit et al. 2005, Meng et al. 2007), due to the fact that Serbian spruce rather occupies smaller areas due to the presence of more competitive species (fossil *P. excelsa* pollen was found at several locations in Serbia, Gigov 1956), the genetic diversity might be reduced by (i) ancient colonization involving leading edge expansion well ahead of the main front (cf. Hewitt 2000) - similar conclusions were drawn for almost monomorphic *P. crassifolia* populations found at Qinghai-Tibetan plateau (Meng et al. 2007), or simply (ii) due to the severe past bottleneck effect. Later on, this region might remain the marginal area of the species natural range and therefore it might not be enriched with new variants, as found in several marginal populations in *P. mariana* (Gamache et al. 2003). With the onset of the last interglacial, expansion of monomorphic population probably occurred in all directions, but it was presumably limited to the Mt. Tara and neighboring region due to the absence of favorable habitats and competition by other species. Some of the newly established populations might went extinct and single trees and tree patches frequently found today within this area might represent the remains of those populations. Analysis of a single specimen and three trees (i.e. population MI) in close vicinity to populations ST and CS, respectively, revealed the identical ancestral haplotype A (data not shown).

The other gene pool, today represented by the highly polymorphic population ST, might have been confined to the south or south-east region in relation to the Serbian spruce present range. Fossil *P. omorika* pollen and needles are found at Aegean coast during the early Pleistocene (Ravazzi 2002) and in south-east Serbia immediately after the LGM (Gigov 1956). Therefore, it is reasonable to assume that during the ice ages

those populations, contrary to populations from the Mt. Tara region, survived glaciations *ex situ*, e.g. in regions south of the 45° latitude and they were exposed to larger geographical displacements, although Petit et al. (2005) postulated that within this region populations expanded little or not at all during the Quaternary. As a corollary, migrating populations were probably exposed to more frequent bottleneck-drift effects, implying that ancestral haplotype A could be lost long before the LGM. However, the possibility that high genetic diversity in ST might result from the recent admixture of several lineages fixed for different haplotypes can not be ruled out, especially taking into account ecological features of this species. With the onset of the warming climate, the post-glacial retreat of populations from the second gene pool was possible only in north-west direction and a zone of the secondary contact of two gene pools is expected at the southern edge of the species current range, resulting in the present genetic structuring in *P. omorika*. However, since the highly polymorphic population ST is found to be nested between monomorphic ones, the phylogeographic signal was not detected nor statistically supported.

The Balkan, although characterized by an exceptionally high genetic diversity, is still referred to as 'poorly known', due to the scarce data on genetic diversity within this region (Hewitt 2000, Petit et al. 2005). As shown in our study, even small number of *P. omorika* natural populations scattered within an area of approx. 10.000 km^2, can be assigned to two distinctive lineages of pre-glacial origin with different evolutionary histories and past migrations. The existence of two gene pools in *P. omorika* is also supported by recent studies using allozymes (Ballian et al. 2006) and cpDNA microsatellites (Nasry et al 2007). Within the western, Bosnian region, the geographical trend among populations was not observed (Ballian et al. 2006), probably due to the fact that this region is the contact zone where populations from different gene pools are mixed. On the other hand, clear north-south geographical trend was observed among eastern, Serbian populations using cpDNA SSRs (Nasry et al. 2007), as northern populations were found to be fixed or nearly fixed for a common haplotype, while southern populations were more polymorphic. Although population ST was not used in the above mentioned study, high diversity within southern populations is due, at least partially, to the presence of this population. The population in question – Studenac, was legally protected by the Republic Institute for Nature Conservation of Serbia in 1993.

Acknowledgements

This work was financially supported by the Republic of Austria and Bioversity International. It comprises a portion of the PhD dissertation of JM Aleksić. The authors thank M. Mengl, L. Weißenbacher, M. Lakić and D. Milekić for their help in the field, B. Heinze, B. Fussi, M. Mengl for fruitful discussions on the laboratory issues and K-M Schweinzer for help in making the map. All samples were collected with permission and support of the National Park 'Tara' (Serbia), Ministry of Agriculture, Forestry and Water Management and Forest Enterprise 'Šume Republike Srpske' (Republic of Srpska, Bosnia and Herzegovina).

References

Aizawa M, Yoshimaru H, Saito H, Katsuki T, Kawahara T, Kitamura K, Shi F, Kaji M (2007). Phylogeography of a northeast Asian spruce, *Picea jezoensis*, inferred from genetic variation observed in organelle DNA markers. *Mol Ecol* **16**: 3393-3405.

Aleksić JM , Schueler S, Mengl M, Geburek Th (submitted). EST-SSRs developed for other *Picea* species reveal high genetic variation in *Picea omorika.*

Avise JC (2000). *Phylogeography: the history and formation of species*. Harvard University Press, Cambridge, Massachusetts, London, England.

Ballian D, Longauer R, Mikić T, Paule L, Kajba D, Gömöry D (2006). Genetic structure of a rare European conifer, Serbian spruce (*Picea omorika* (Pančić) Purk.). *Pl Syst Evol* **260**: 53-63.

Bennett KD (1990). Milankovitch cycles and their effects on species in ecological and evolutionary time. *Paleobiology* **16**: 11-21.

Bohonak AJ (2002). IBD (Isolation by distance): a program for analysis of isolation by distance. *J Hered* **93**: 153-154.

Bouillé M, Bousquet J (2005). Trans-species shared polymorphisms at orthologous nuclear gene loci among distant species in the conifer *Picea* (Pinaceae): implications for the long-term maintaince of genetic diversity in trees. *Am J Bot* **92**: 63-73.

Búžek Č, Kvaček Z, Holý F (1985). Late Pliocene palaeo-environment and correlation of the Vildštejn floristic complex within central Europe. *Rozpravy Československé Akademie Věd* **95**: 1-72.

Burschel P (1965). Die Omorikafichte. *Forstarchiv* **36**: 113-131.

Castelloe J, Templeton AR (1994). Root probabilities for intraspecific gene trees under neutral coalescent theory. *Mol Phylogenet Evol* **3**: 102-113.

Cavalli-Sforza LL, Edwards AWF (1967). Phylogenetic Analysis: models and estimation procedures. *Evolution* **32**: 550-570.

Chiang TY, Schaal BA (1999). Phylogeography of ten Northern American *Hylocomium splendens* based on nrDNA ITS sequences. *Mol Ecol* **8**: 1037-1042.

Clement M, Posada D, Crandall K (2000). TCS: a computer program to estimate gene genealogies. *Mol Ecol* **9**: 1657-1660.

Conifer Specialist Group (1998). *Picea omorika*. In: IUCN 2007. *2007 IUCN Red List of Threatened Species* (http://www.iucnredlist.org/search/details.php/30313/summ).

Cruzan MB, Templeton AR (2000). Paleoecology and coalescence: phylogeographic analysis of hypotheses from the fossil record. *Trends Ecol Evol* **15**: 491-495.

Čolić DB (1965). Poreklo i sukcesija šumskih zajednica sa Pančićevom omorikom (*Picea omorika* Panč.) na planini Tari. *Zaštita prirode* **29-30**: 65-90.

Čolić DB (1966). Požar kao ekološki faktor u sukcesiji zajednica Pančićeve omorike i redukovanju njenog areala. *Zaštita prirode* **33**: Beograd.

David A, Keathley D (1996). Inheritance of mitochondrial DNA in interspecific crosses of *Picea glauca* and *Picea omorika*. *Can J For Res* **26**: 428-432.

Dumolin-Lapègue S, Demesure B, Fineschi S, Corre VLe, Petit RJ (1997). Phylogeographic structure of white oaks throughout the European continent. *Genetics* **146**: 1475-1487.

Eisenhut G (1961). Untersuchungen über die Morphologie und Ökologie der Pollenkörner heimischer und fremdländischer Waldbäume. *Forstwiss Forsch* **15**:1-68.

El Mousadik A, Petit RJ (1996). High level of genetic differentiation for allelic richness among populations of the agran tree [*Agrania spinosa* (L.) Skeels] endemic to Marocco. *Theor Appl Genet* **92**: 832-839.

Excoffier L, Smouse PE, Quattro JM (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**: 479-491.

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

Farjon A (2001). World checklist and bibliography of conifers. Ed. 2. Royal Botanic Gardens, Kew.

Felsenstein J (2003). PHYLIP**:** Phylogeny Inference Package version 3.61. University of Washington, Seattle. http://evolution.gs.washington.edu/phylip.html.

Fukarek P (1967). Pančićevo otkriće omorike i njeno dalje proučavanje. In: Josifović M (ed) *Pančićev zbornik u spomen 150-godišnjice njegovog rođenja*. Srpska Akademija Nauka i Umetnosti, Odeljenje prirodno-matematičkih nauka: Beograd. pp 27-67.

Gajić M, Vilotić D, Karadžić D, Mihajlović Lj, Isajev V (1994). Omorika – *Picea omorika* (Pančić) Purkyně na području Nacionalnog parka Tara (monografska studija). Nacionalni park Tara – Bajina Bašta and Šumarski fakultet: Beograd.

Gamache I, Jaramillo-Correa JP, Payette S, Bousquet J (2003). Diverging patterns of mitochondrial and nuclear DNA diversity in subarctic black spruce: imprint of a founder effect associated with postglacial colonization. *Mol Ecol* **12**: 891-901.

Ge X-J, Zhou X-L, Li Z-C, Hsu T-W, Schaal BA, Chiang T-Y (2005). Low genetic diversity and significant population structuring in the relict *Amentotaxus argotaenia* complex (Taxaceae) based on ISSR fingerprinting. *J Plant Res* **118**: 415-422.

Gigov A (1956). Dosadašnji nalazi o postglacijalnoj istoriji šuma Srbije. *Institut za ekologiju i biogeografiju, Zbornik radova* **7/3**: 1-25.

Gugerli F, Sperisen C, Büchler U, Magni F, Geburek Th, Jeandroz S, Senn J (2001). Haplotype variation in a mitochondrial tandem repeat of Norway spruce (*Picea abies*) populations suggests a serious founder effect during postglacial re-colonization of the western Alps. *Mol Ecol* **10**: 1255-1263.

Hamrick JL, Godt MJW (1996). Conservation genetics of endemic plant species. In: Avise JC, Hamrick JL (eds) *Conservation genetics: case histories from nature,* Chapman Hall: New York. pp 281-304.

Hamrick JL, Godt MJW, Murawski DA, Loveless MD (1991). Correlations between species and allozyme diversity: implications for conservation biology. In: Falk DA, Holsinger KE (eds) *Genetics and conservation of rare plant*, Oxford University Press: New York. pp 75-86.

Hewitt G (2000). The gentic legacy of the Quaternary ice ages. *Nature* **405**: 907-913.

Hewitt G (2001). Speciation, hybrid zones and phylogeography – or seeing genes in space and time. *Mol Ecol* **10**: 537-549.

Huang S, Chiang YC, Schaal BA, Chou CH, Chiang TY (2001). Organelle DNA phylogeography of *Cycas taitungensis*, a relict species in Taiwan. *Mol Ecol* **10**: 2669-2681.

Jaramillo-Correa JP, Bousquet J (2003). New evidence from mitochondrial DNA of a progenitor-derivative species relationship between black spruce and red spruce (Pinaceae). *Am J Bot* **90**: 1801-1806.

Jaramillo-Correa JP, Beaulieu J, Bousquet J (2004). Variation in mitochondrial DNA reveals multiple distant glacial refugia in black spruce (*Picea mariana*), a transcontinental North American spruce. *Mol Ecol* **13**: 2735-2747.

Jaramillo-Correa JP, Beaulieu J, Ledig FT, Bousquet J (2006). Decoupled mitochondrial and chloroplast DNA population structure reveals Holocene collapse and population isolation in a threatened Mexican-endemic conifer. *Mol Ecol* **15**:2787-2800.

Karvonen P, Szmidt AE, Savolainen O (1994). Length variation in the internal transcribed spacers of ribosomal DNA in *Picea abies* and related species. *Theor Appl Genet* **89**: 969-974.

Král D (2002). Assessing the growth of *Picea omorika* [Panč.] Purkyně in the Masaryk Forest Training Forest Enterprise at Křtiny. *J For Sci* **48**: 388-398.

Kuittinen H, Muona O, Kärkkäinen K, Borzan Ž (1991). Serbian spruce, a narrow endemic, contains much genetic variation. *Can J For Res* **21**:363-367.

Kuittinen H, Savolainen O (1992). *Picea omorika* is a self-fertile but outcrossing conifer. *Heredity* **68**: 183-187.

Ledig FT, Hodgskiss PD, Krutovskii KV, Neale DB, Eguiluz-Piedra T (2004). Relationships among the spruces (*Picea*, pinaceae) of southwestern North America. *Syst Bot* **29**: 275-295.

Levinson G, Gutman G (1987). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol* **4**: 203-221.

Mantel N (1967). The detection of desease clustering and generalize regression approach. *Cancer Res* **27**: 209-220.

Meng LA, Yang R, Abbott RJ, Miehe G, Hu T, Liu J (2006). Mitochondrial and chloroplast phylogeography of *Picea crassifolia* Kom. (Pinaceae) in the Qinghai-Tibetan Plateau and adjacent highlands. *Mol Ecol* **16**: 4128-4137.

Myers N, Mittermeier RA, Mittermeier CG, Fonseca GAB da, Kent J (2000). Biodiversity hotspots for conservation priorities. *Nature* **403**: 853-858.

Nasri N, Bojovic S, Vendramin GG, Fady B (2007). Population genetic structure of the relict Serbian spruce, *Picea omorika*, inferred from plastid DNA. *Pl Syst Evol* (early online).

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

Newton AC, Allnutt TR, Gillies ACM, Lowe AJ, Ennoss RA (1999). Molecular phylogeography, intraspecific variation and the conservation of tree species. *Trens Ecol Evol* **14**: 140-145.

Olsen KM, Schaal BA (1999). Evidence on the origin of cassava: Phylogeography of *Manihot esculenta*. *Proc Natl Acad Sci USA* **96**: 5586-5591.

Ostojić D (2005). Ekološki činioci prirodnog održavanja i obnove cenopopulacija Pančićeve omorike u NP Tara. PhD thesis, Faculty of forestry, Belgrade.

Page RDM (1996). TREEVIEW: An application to display phylogenetic trees on personal computers. *Comp Applic Biosci* **12**: 357-358.

Palmer JD, Adams KL, Cho Y, Parkinson CL, Qiu Y-L, Song K (2000). Dynamic evolution of plant mitochondrial genomes: Mobile genes and introns and highly variable mutation rates. *Proc Natl Acad Sci USA* **97**: 6960-6966.

Petit RJ, Hampe A, Cheddadi R (2005). Climate changes and tree phylogeography in the Mediterranean. *Taxon* **54**: 877-885.

Pons O, Petit RJ (1995). Estimation, variance and optimal sampling of gene diversity I. Haploid locus. *Theor Appl Genet* **90**: 462-470.

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

Ravazzi C (2002). Late Quaternary history of spruce in southern Europe. *Rev Paleobot Palinol* **120**: 131-177.

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

Schaal BA, Hayworth DA, Olsen KM, Rauscher JT, Smith WA (1998). Phylogenetic studies in plants: problems and prospects. *Mol Ecol* **7**: 465-474.

Slatkin M (1993). Isolation by distance in equilibrium and nonequilibrium populations. *Evolution* **43**: 1349-1368.

Sneath PHA, Sokal RR (1973). *Numerical taxonomy.* Freeman: San Francisco.

Sperisen C, Büchler U, Gugerli F, Mátyás G, Geburek Th, Vendramin GG (2001). Tandem repeats in plant mitochondrial genomes: application to the analysis of population differentiation in the conifer Norway spruce. *Mol Ecol* **10**: 257-263.

Syring J, Willyard A, Cronn R, Liston A (2005). Evolutionary relationships among *Pinus* (Pinaceae) subsections inferred from multiple low-copy nuclear loci. *Am J Bot* **92**: 2086-2100.

Taberlet P, Fumagalli L, Wust-Saucy AG, Cossons JF (1998). Comparative phylogeography and postglacial colonization routes in Europe. *Mol Ecol* **7**: 453-464.

Templeton AR, Routman E, Phillips CA (1995). Separating population structure from population history: A cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. *Genetics* **140**: 767-782.

Templeton AR (1998). Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. *Mol Ecol* **7**: 381-397.

Templeton AR (2001). Using phylogeographic analyses of gene trees to test species status and processes. *Mol Ecol* **10**: 779-791.

Tucić B, Stojković B (2001). Shade avoidance syndrome in *Picea omorika* seedlings: a growth-room experiment. *J Evol Biol* **14**: 444-455.

Weir BS (1996). *Genetic data analysis II*, 2nd edn. Sinauer: Sunderland, MA.

Table 1. Natural populations of Serbian spruce

Population – abbreviation	Longitude	Latitude	Mean altitude (m)	Total area (ha)
Topli Do – TD[1]	43º 59’ 31’’	19º 18’ 54’’	900	700[1]
Bilješke stene – BS[1]	43º 59’ 18’’	19º 19’ 12’’	1000	
Ljuti Breg – LB	43º 55’ 38’’	19º 20’ 26’’	1200	5[3]
Bilo – BI	43º 55’ 38’’	19º 20’ 49’’	1150	14[3]
Crvene Stene – CS	43º 55’ 38’’	19º 22’ 30’’	1000	46[3]
Studenac – ST	43º 53’ 24’’	19º 20’ 40’’	1300	1
Vranjak – VR	43º 51’ 54’’	19º 24’ 09’’	750	4
Zmajevački Potok - ZP	43º 51’ 35’’	19º 25’ 39’’	900	4
Veliki Stolac - VS[2]	*43º 54’ 45’’*	*19º 17’ 24’’*	*1420*	*50*
Gostilja - GO[2]	*43º 51’ 36’’*	*19º 20’ 06’’*	*1250*	*10*

[1] - populations TD and BS are found within the 700 ha area at Mt. ‘Zvijezda’; [2] - populations from the Republic of Srpska, Bosnia and Herzegovina (other populations are from Serbia); data in italic were taken from Ballian et al. (2006); [3] – exact areas of populations available at the National Park ‘Tara’; total areas of other populations were evaluated in the field.

Table 2. Haplotype characterization and the number of haplotypes (*nh*), haplotypic richness (*R*) and gene diversity (H_S) per population

Haplotype		A	B	C	D	J		
Length (bp)		787	821	855	889	1093		
34 bp copies[1]		1	2	3	4	10		
Population	*nh*	Haplotype frequences					*R*	H_S
8 pop[2]	1	1.000					1.000	0.000
VS	3	0.850	0.100			0.050	3.000	0.279
ST	3		0.700	0.250	0.050		3.000	0.468
Mean	1.4	0.885	0.080	0.025	0.005	0.005	3.007	0.075

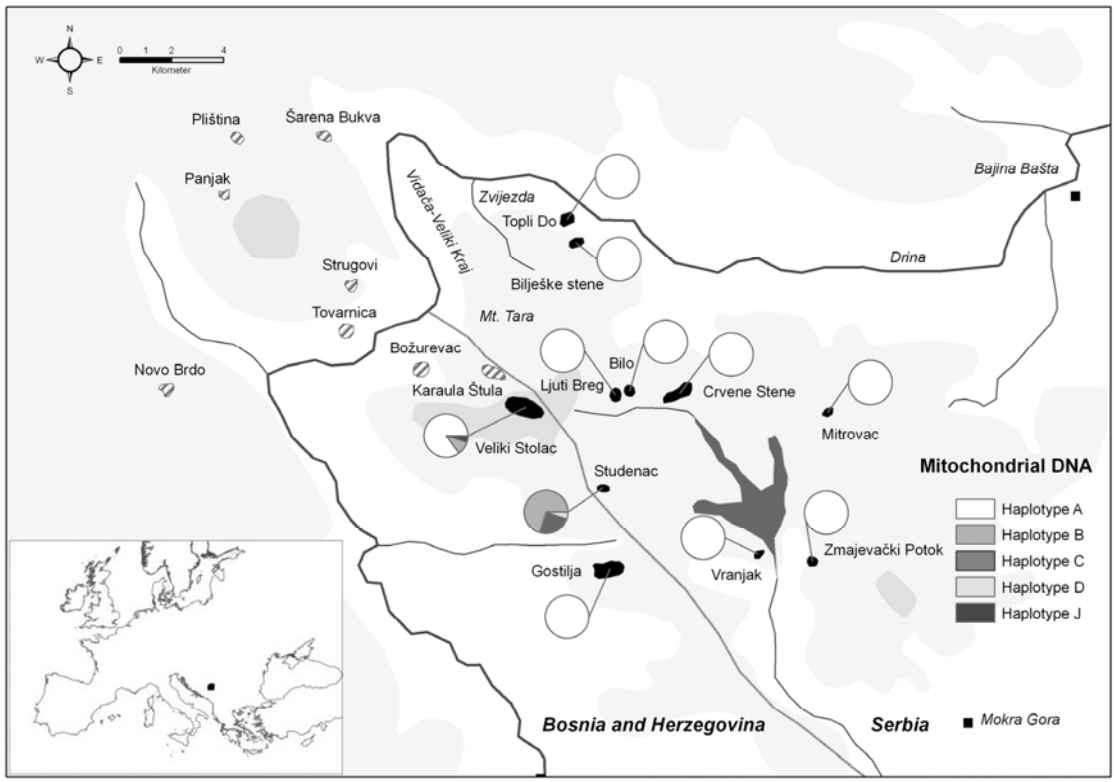
bp – base pairs; [1] - number of 34 bp minisatellite copies within the haplotype; [2] - eight populations fixed for a single haplotype A: TD, BS, LB, BI, CS, VR, ZP and GO.

Table 3. Genetic diversity parameters in overall sample and in a sample without population ST

	H_S	H_T	G_{ST}	N_{ST}
9 pop + ST	0.075	0.225	0.668	0.668
sd	0.052	0.164	nc	nc
9 pop - ST	0.031	0.033	0.070	0.070
sd	0.031	0.032	nc	nc

H_S – gene diversity; H_T – total gene diversity; G_{ST} – among population differentiation based on haplotype composition; N_{ST} – among population differentiation including the genetic distance among haplotypes; 9 pop + ST and 9 pop – ST – analyses performed in overall sample and in sample with omitted population ST, respectively; sd – standard error.

Figure 1. Locations and areas of majority of *Picea omorika* natural populations in Serbia and Bosnia and Herzegovina and the distribution of the mitochondrial DNA haplotypes detected in this study



Ten natural populations used in this study are marked by the black color of their areas. The atypical population 'Mitrovac' (the only one found on a swampy terrain) was also used, but it was considered as a patch of trees, because only three trees were remnant and the natural regeneration was absent. Three populations found south from this core area are not presented.

The relative abundance of five mtDNA haplotypes detected in this study is presented for each population.

Figure 2. UPGMA dendrograme based on Cavalli-Sforza genetic distances among populations

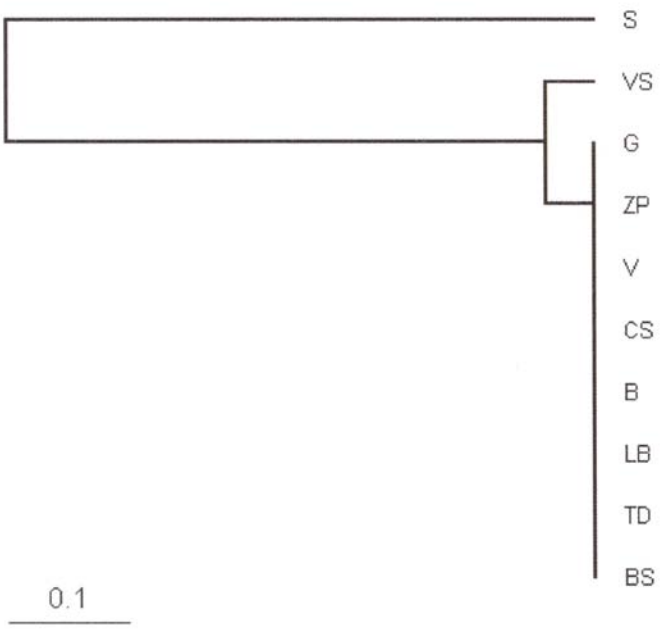
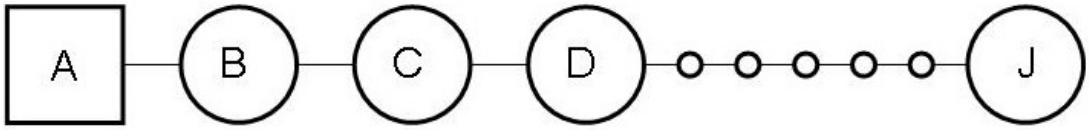


Figure 3. Evolutionary relationships among haplotypes



Haplotypes with increased number of a 34 bp minisatellite copies evolved from each other in a stepping-stone manner. Duplications of the minisatellite repeat unit are presented as single mutational events. However, recurrent mutations can not be ruled out and the risk of inaccurate homology assessment is still present.

3.3 Length mutations in two mitochondrial group II introns in *Picea omorika*: new insights into spruce phylogeny

Aleksić MJ, Geburek T, Fussi B and Heinze B

Abstract

Two mitochondrial group II introns - nad1i477 and nad5i230, were used to test a hypothesis of a close relationship between an European narrow endemic and a Tertiary relict - *P. omorika*, and two North American spruces - *P. mariana* and *P. rubens*, revealed by crossability studies and supported by two recently published molecular phylogenies. Our results support the view that the recognition of microstructural changes leading to the occurrence of length mutations can improve the assessment of primary homology (i.e. alignment) and consequently, the phylogenetic analysis itself, and demonstrates that incorporation of length mutations in genealogical and phylogenetic reconstructions can substantially improve the resolution, especially in cases when length mutations play a major role in sequence evolution, as found at nad1i477. The domain IV in this intron is found to harbour two hypervariable large indels (up to 1300 bp) characterized by essentially different organization of variability and secondary structure. Indel A, present only in group A spruce species (*P. omorika*, *P. abies* and the majority of Asian species) is found to form a long hairpin suggesting that the loss of such a structure might result in the loss of the whole indel A, as found in group B spruce species. Similarly, group B spruce species (*P. schrenkiana* and *P. smithiana* and all American species, including *P. mariana* and *P. rubens*) are found to harbour indel B, organized in a stem-loop structure, implying that the loss of this structure would lead to the loss the whole indel B, as found in group A spruce species. Alternatively, indel B may have been gained by recombination. The concordance of such deep division among spruce species found at nad1i477 corresponds well to the five haplotypes described at the second, less variable mitochondrial intron nad5i230, and to geographical distribution of all species used in this study. The only species which did not follow the detected pattern was *P. breweriana,* characterized as an isolated, relict species, similarly as in case with *P.*

omorika. Both species are found to be distantly related to *P. mariana* and *P. rubens*. A hypothesis of the eastern Asian origin of spruces is supported and based on our results, Wright's 'the most primitive species with generalized morphological features' - *P. koyamai,* still might be the basal spruce.

Key words: *Picea omorika*, group II introns, length mutations, indel coding, genealogical and phylogenetic analysis

Introduction

The genus *Picea* A. Dietr. is the third largest genus within the family *Pinaceae* (Farjon 2001). Its monophyletic origin is generally accepted (Wright 1955, Prager et al. 1976, Frankis 1988, Sigurgeirsson and Szmidt 1993, Ran et al. 2006) as well as the view that morphological (and genetic) variation within the genus is limited and comparable to a single series in *Pinus* (Wright 1955, Sigurgeirsson and Szmidt 1993). However, this genus is still referred to as systematically and taxonomically recalcitrant. Morphological, anatomical and chemical comparisons, as well as cross-compatibility and molecular studies (see references in Discussion) did not yet resolve the delineation of species, genus subdivision, nor its origin.

Serbian spruce [*Picea omorika* (Panč.) Purk.] is maybe the best example to ilustrate the depth of this inconclusiveness. This species was discovered in 1875 and soon after, Willkomm (1887) designated a whole section after this species. Since than, *P. omorika* has been placed into the section/subsection *Omorika* in all clasification schemes (see Schmidt-Vogt 1977), based on either Willkomm's two-section (with sections *Picea* and *Omorika* as subsections within the same section) or Mayr's three-section system (sections *Morinda*, *Casicta* and *Omorika*, Mayr 1890).

Fossil findings of *P. omoricoides* Weber, a close fossil relative of *P. omorika,* are confined to Europe where this species was probably widely distributed since the Tertiary (Ravazzi 2002). With the onset of the last glacial maximum, *P. omorika* was restricted to the several refugial populations within the northern Balkans (Aleksić and Geburek, submitted). After the retreat of the glaciers, range expansion of *P. omorika* was limited and currently more than 30 natural populations are scattered within the area of approximately 10.000 km^2. Although this species is not critically endangered, it is IUCN red-listed (Conifer Specialist Group 1998) due to the limited natural range.

Northern American species *P. breweriana* shares several morphological features with *P. omorika*, it is also classified into the section/subsection *Omorika* and it has been usually characterized as an isolated lineage (e.g. Wright 1955, Sigurgeirsson 1992, Ledig 2004). However, molecular phylogenies based on chloroplast (cpDNA) (Sigurgeirson and Szmidt 1993) and combined cpDNA and mitochondrial (mtDNA) markers (Ran et al. 2006) translated the center of origin of spruces into the North America and pinpointed *P. breweriana* as the basal species, although Li (1953) and Wright (1955) highlighted Eastern Asia as the center of origin for spruces, because 24 out of 35 *Picea* species recognized by recent classifications and checklists (Schmidt 1998, Farjon 2001) are confined to this region and despite the fact that the center of diversity is usually the center of origin for many species.

On the other hand, *P. omorika* differs morphologically from the North American spruces *P. rubens* and *P. mariana,* classified into the section/subsection *Picea*, although the latter species is characterized by several features rare within the genus, such as cone serotiny and tolerance of swampy soils, as reported also in *P. omorika* (Rushforth 1987 in Sigurgeirsson and Szmidt 1993). However, the close relationship between an European narrow endemic and a relict from the Arcto-Tertiary flora, and two species widely distributed throughout North America, argumented on the basis of crossability studies (Wright 1955, Fowler 1982 and references therein, Ledig 2004) and supported by molecular data (Sigurgeirson and Szmidt 1993, Nkongolo 1999, Smith and Klein 1996, Ledig et al. 2004, Ran et al. 2006), seems to be accepted quite unconditionally and present, the trans-continental distribution of those species is explained by vicariance. Furthermore, *P. omorika* was assumed to provide a connection between the two North American spruce complexes, i.e. the *P. mariana-rubens* complex and the *P. glauca-engelmannii-sitchensis-mexicana* complex (Ledig et al. 2004).

Therefore, it seems necessary to shed more light on evolutionary relationships between *P. omorika* and other, especially North American spruces. Molecular phylogenies at lower taxonomic levels are usually based on more variable non-coding regions from all three plant genomes (Laroche et al. 1997, Feliner and Rossello 2007, Shaw et al. 2007). Apart from the internal transcribed spacers of the nuclear ribosomal DNA, which are the most popular non-plastid regions for species level phylogenetic studies (Feliner and Rossello 2007), chloroplast and mitochondrial introns have been widely used (e.g. Kelchner 2002, Shaw et al. 2007, Freudenstein and Chase 2001). Recent advances in understanding group II intron structure and function revealed that they are structurally constrained with a mosaic of highly conserved elements altering with sequence stretches that might be more or less freely evolving (Kelchner 2000, Löhne and Borsch 2005). Conserved regions of all group II introns enable folding intro the same, typical secondary structure with six main stem-loop domains radiating from a central wheel (Michel et al. 1989). Such structure is essential for their function - the proper splicing of gene transcripts prior to translation of the mRNA into the protein, and mutations that disrupt secondary and tertiary interactions are likely to be eliminated by strong counter-selection (Michel and Ferat 1995). Variable regions in introns (and other non-coding segments), on the other hand, are characterized by relatively high numbers of length mutations in addition to base substitutions (Freudenstein and Chase 2001, Clegg et al. 1994, Golenberg et al. 1993, Gielly and Taberlet 1994). Those regions are usually confined to the loops of the secondary structure (Kelchner 2000, Löhne and Borsch 2005), and the loop of domain IV is frequently characterized by pronounced length variability (Malek et al. 1997, Won and Renner 2003, Löhne and Borsch 2005). Mutational hot-spots

have been described within this loop (e.g. Löhne and Borsch 2005), making any homology assessment very difficult or even impossible (Kelchner 2000).

Laroche et al. (1997) found a positive correlation between the number of substitutions and the number of insertions/deletions (indels) per site in six mitochondrial introns and made implications for their potential phylogenetic use. To date, growing evidence suggests that length mutations in both cpDNA and mtDNA introns are reliable phylogenetic characters with often low levels of homoplasy, although the latter have been rarely used in plants (e.g. Wissinger et al. 1991, Bakker et al. 2000, Freudenstein and Chase 2001). Incorporation of indel characters in phylogenetic analysis is becoming more important because not using such length mutations is equivalent to discarding data (Simmons and Ochoterena 2000, Kelchner 2000). Simmons and Ochoterena (2000) provided the most elaborate treatment of coding indels for phylogenetic utilization. Based on their two alternative ways of indel coding, several alignment algorithms and software implementations are currently available (see Müller 2006 and references therein).

However, the phylogenetic utility of mtDNA introns and their length mutations is not straight-forward and the risk of inaccurate homology assessment is still present (see Kelchner 2000). Gu and Li (1995) pointed out that mutational mechanisms leading to microstructural changes have to be considered for the correct primary homology assessment. Kelchner (2000, 2002) reviewed mutational mechanisms directing molecular evolution in non-coding regions in chloroplast (and mitochondrial) group II introns and highlighted that awareness and identification of those mechanisms are essential for improving both (i) the alignment, as the hypothesis of primary homology (cf. Simmons and Ochoterena, 2000) and (ii) the phylogenetic analysis itself, in which the hypothesis of primary homology is tested by congruence (cf. Simmons and Ochoterena, 2000).

Slipped-strand mispairing is suggested as the mechanism which generates length mutations (Levinson and Gutman 1987) resulting in short indels (e.g. Laroche et al. 1997, Kelchner 2000) and tandem minisatellite regions (e.g. Sperisen et al. 2001, Godbout 2005). The latter regions, as well as minute inversions associated with hairpins, are the most homoplasious characters in molecular phylogenetics (Kelchner 2000). Additional obstacles in accurate homology assessment are indels of unusual size that contain sequence content not readily identifiable in origin (Kelchner 2000). Such indels are resulting from intramolecular recombination within a discrete non-coding region or on a genomic scale (e.g. Laroche and Bousquet 1999, Graham et al. 2000, Kelchner 2000, Löhne and Borsch 2005). Apparently they do not annihilate possible secondary structure formation, and the same holds for the deletions of loops, bulges and even parts of the stems (Kelchner 2000).

The initial aim of this study was to test the hypothesis of a close relationship between an European narrow endemic and a Tertiary relict - *P. omorika*, and two North American spruce species - *P. mariana* and *P. rubens*. We utilized the available data on two mitochondrial group II introns (nad1i477 and nad5i230) to demonstrate that (i) the detection of mutational mechanisms leading to microstructural changes in group II introns substantially improves the homology assessment, (ii) secondary structure reconstructions in nad1i477 can reveal mutational mechanisms leading to the observed organization of variability in mtDNA, and that (iii) incorporation of phylogenetic signal in length mutations can substantially improve genealogical and phylogenetic inferences, especially in cases of introns in which length mutation are important source of mutations. A *de novo* phylogeny of spruces, however, was not the aim of this study.

2. Materials and methods

The nad genes encode subunits of complex I of the respiratory chain, the NADH-ubiquinone-oxidoreductase. In tracheophytes, four group II introns are present in both the nad1 gene (Chapdelaine and Bonen 1991, Conklin et al. 1991, Kubo et al. 2000, Notsu et al. 2002, Unseld et al. 1997, Wissinger et al. 1991) and the nad5 gene (Groth-Malonek et al. 2005). Two group II introns - nad1i477 and nad5i230, confined to the nad1 and nad5 genes, respectively, were used in this study. The nomenclature of organelle introns adopted here follows Dombrovska and Qiu (2004).

Nad1i477

Intron nad1i477 has been reported to be present in a large number of seed plants by Gugerli et al. (2001), Won and Renner (2003) and Dombrovska and Qiu (2004). This intron was most likely acquired in the common ancestor of vascular plants and has been vertically inherited ever since, with secondary losses in a small number of plant lineages, such as non-Pinaceae conifers (Gugerli et al. 2001) and several unrelated angiosperms (Gugerli et al. 2001, Dombrovska and Qiu 2004).

The size range of the nad1i477 is from 896 bp in *Arabidopsis* to 2844 bp in *Ophioglossum*. In conifers, this intron is approximately 3000 bp long (Demesure et al. 1995,

Gugerli et al. 2001, Won and Renner 2003) and exhibits pronounced inter- and intraspecific variability.

Two minisatellites (32 and 34 bp) have been described in *P. abies,* resulting in high intraspecific variability (Sperisen et al. 2001). Only the latter (34 bp) minisatellite has been reported in *P. crassifolia* (Meng et al. 2006) and *P. omorika* (Aleksić and Geburek, submitted). Variability was found in *P. mariana, P. glauca* and *P. pungens* as well (Jaramillo-Correa et al. 2003), and although it was mainly due to the length mutations, neither of the two minisatellites mentioned above was reported. In *P. rubens* (Jaramillo-Correa and Bousquet 2003) and *P. chihuahuana* (Jaramillo-Correa et al. 2006), which have limited natural ranges, this locus was found to be invariable.

Owing to high inter- and intraspecific variability at this mitochondrial nad1i477 in spruces, poorly resolved phylogenies with low bootstrap support for many clades resulted when phylogenetic relationships were examined (but see Bousquet et al. 2007); in other studies, this locus has been omitted altogether (Ran et al. 2006).

We retrieved all available sequences of the *Picea* nad1i477 from Gene Bank (www.ncbi.nlm.nih.gov/Genbank) in January 2008, and we also used sequences of *P. omorika* submitted to Gene Bank by ourselves recently (Aleksić and Geburek submitted) (Table 1 in supplementary material). Multiple accessions were available for 6 species, resulting in a total of 41 sequences corresponding to 17 *Picea* species. Among them, sequences for the whole intron were available for 16 species (including *P. omorika*), while *only* partial sequences (corresponding to the variable region which can be amplified with primers described by Sperisen et al. 2001) were available for two species, *P. omorika* (additional specimens) and *P. crassifolia*. Nad1i477 sequences were not available for the other 15 Asian *Picea* species recognized by Farjon (2001), while the reported length of the partial intron sequence in *P. obovata* (Sperisen et al. 2001) as well as the restriction fragment profile in *P. alcoquiana* (Aizawa et al. 2008) imply the organization and variability in those species (see later).

Nad5i230

Nad5i230 is characterized as a conventional *cis*-arranged intron in the nad5 gene (Groth-Malonek et al. 2005). Among mosses, intron size ranges from 831 bp in *Encalypta streptocarpa* to 875 bp in *Sphagnum fallax*. This intron is of similar size in *Arabidopsis thaliana* (829 bp), while in gymnosperms its size exceeds 1.000 bp (Wang et al. 2000, Ran et al. 2006).

All nad5i230 intron sequences in mosses, hornworts and tracheophytes were found to deviate from the group II consensus structure (Groth-Malonek et al. 2005). The authors reported a high degree of sequence conservation in the 5' intron region and around intron domains V and VI though, indicating stable vertical transmission.

In conifers, this intron is characterized by lower intra- and interspecific variability in comparison to nad1i477 (Ran et al. 2006). However, variability due to indels and base substitutions has been found in *P. mariana* (Jaramillo-Correa et al. 2003), *P. glauca* and *P. pungens* (Jaramillo-Correa et al. 2003, Ran et al. 2006). Jaramillo-Correa et al. (2003) reported a variable region characterized by the insertions of a five bp repeat motif. Based on the most likely order of mutational events leading to the occurrence of five distinctive length variants confined to this region, Ran et al. (2006) described five haplotypes (A-E) in 34 spruce species and used this intron (and two cpDNA loci) for their reconstruction of spruce phylogeny.

We retrieved all available *Picea* sequences of the nad5i230 from the Gene Bank in January 2008 (Table 2 in supplementary material). In total, 46 sequences were available for 34 species. Single accessions were available for 26 species, while multiple accessions (20 in total) were available for 8 species.

Genealogical and phylogenetic analysis

Due to the different nature of intra- and interspecific variability at nad1i477 and nad5i230 and incomplete samples for the former intron, genealogical and phylogenetic analyses were firstly performed for each locus separately, using all available sequences per each locus. Since the sequences for both introns were available for 16 species (including *P. omorika* and *P. mariana-rubens*), only in those species combined (composite) hybrid sequences were reconstructed from both introns, and used for genealogical and phylogenetic analyses. In cases where multiple accessions were available in one species at one locus and only one sequence at the second locus, all sequences at the variable locus were combined with a single sequence at the second locus for the reconstruction of the hybrid sequences and all retained in the analyses. Since both loci are within the mtDNA, complete linkage is expected, although mitochondrial DNA is known to exist in multiple forms, and recombination cannot be excluded based on our current knowledge.

For nad1i477, possible, ancestral haplotypes within each species were detected by aligning multiple sequences within each species - otherwise, all sequences per species were used. The six domains of the secondary structure were identified in all species based on the

delineation reported in *P. mariana* (accession AY057956) and *P. rubens* (AY057951) (supplementary material in Won and Renner 2003). The alignment was performed per each domain separately using CLUSTAL W2 (Larkin et al. 2007). All alignments were refined manually and recognized conserved regions were used as anchor points. If alternative alignments were possible, indels were positioned in a way to detect possible duplications of multi- and mononucleotide repeat units confined to the neighboring 5' region (frequent type b and a gaps, respectively, in Kelchner 2000).

For nad5i230, although multiple accessions were available for 8 species, only sequences for the whole intron reported by Ran et al. (2006) were used, due to the possible incomplete intron sequences reported by other authors. Therefore, two accessions per species were available for four species (*P. wilsonii*, *P. schrenkiniana*, *P. likigianensis* and *P. chihuahuana*) and all were retained.

Genealogical relations among haplotypes were assessed by reconstructing a minimum spanning tree (MST) using TCS (Clement et al. 2000). Gaps were coded manually in the following way: simple duplications of recognized repeat units of variable size were treated as a single mutational event (e.g. single nucleotide duplications), as well as short deletions (up to 30 bp), although the latter might result from multi-mutational events. Complex indels causing ambiguous alignments were excluded from the analysis. Missing nucleotides were treated as a fifth state.

Indels were firstly treated as missing data and than coded using Simmons and Ochoterena (2000) simple indel coding (SIC) and 'complex indel coding' (CIC). The latter coding method is implemented in SeqState (Müller 2006) as modified complex indel coding (MCIC). The phylogenetic analysis was performed using PAUP*4.0b10 (Swofford 2002) with heuristic search. The search options were 500 random addition sequence replicates, tree-bisection-reconnection (TBR) branch-swapping, MULTREES and a maximum of 1000 trees saved per round. All characters were equally weighted. To evaluate relative robustness of the clades found in the most parsimonious trees, a bootstrap analysis (Felsenstein 1985) was performed with 1000 replicates using the same heuristic search settings, except that a maximum of 100 trees were saved per round.

Nad1i477 domain IV and the origin of spruces

For the reconstructions of the secondary structure of the hipervariable domain IV in nad1i477, *P. omorika* and *P. mariana* sequences were used. In the latter species, accession AY057957 was used because this sequence is identical to all four *P. rubens* sequences due

to the progenitor-derivative relation between those species (Jaramillo-Correa et al. 2003) and therefore this sequence represents both species. Only here, a *P. omorika* haplotype harbouring ten 34 bp minisatellite copies was employed, in order to check the effects of these repeats on folding and organization of domain IV. Secondary structure was recovered using MFOLD (Zuker et al. 1999).

To reveal the possible origin of spruces, sequences of representatives of both *Pinus* subgenera - *Pinus sylvestris* (accession AJ223312) and *Pinus cembra* (AF160261) were used, as well as the sequence of *Cathaya argyrophylla* (DQ983604). Sequences of the domain IV in the above mentioned species were plotted against themselves and against each other using DOTMATCHER (http://emboss.sourceforge.net/) to reveal possible similar regions within and between sequences. Threshold was set to 30, while window size ranged from 20 to 60.

3. Results

Nad1i477

Seven nad1i477 sequences were available in *P. omorika* - one whole intron sequence harbouring five 34 bp minisatellite repeats (2143 bp) and six partial sequences harbouring one to five, and ten, 34 bp minisatellites. Therefore, the assessed size range of nad1i477 in *P. omorika* was from 2007 bp to 2313 bp. Since all sequences were identical except for the variable number of minisatellite copies (base substitutions were not detected), only the sequence harbouring one 34 bp minisatellite copy was retained in further analyses, as the likely ancestral haplotype (see Discussion). According to the length of the partial sequence, this haplotype was labeled as 787, to keep the analogy with the labeling used in *P. abies* by Sperisen et al. (2001).

Only partial sequences were available in *P. crassifolia,* two per each of the five size variants, harbouring one to five 34 bp minisatellite copies*.* Similarly as in *P. omorika*, sequences harbouring one 34 bp minisatellite were assumed as the ancestral haplotype. Those sequences were identical to the corresponding region within the 1932 bp whole intron sequences in *P. asperata, P. jezoensis* and *P. wilsonii* - sequences in those species were identical and all harboured one 34 bp minisatellite. Therefore, all were labeled as the haplotype 712 according to the length of the partial sequence. Several single nucleotide indels and point mutations within the last 50 bp at the 3’ ends in all *P. crassifolia* sequences

were considered as likely artifacts due to lower sequencing precision at 3' ends - this region was found to be conserved in all other spruces. Haplotype 712 was also reported in *P. obovata* (Sperisen et al. 2001) and it is probably present in *P. alcoquiana* (see later).

Four sequences were available in *P. abies.* Apart from the consensus sequence (see Sperisen et al. 2001), only the whole intron sequence for haplotype 815 (AY289611) was further used, since the other two sequences harboured unidentified nucleotides (low quality sequence). The length of this haplotype was 2035 bp.

In *P. mariana,* four whole intron sequences yielded 3014 bp aligned sequence. Two sequences were identical (accessions AY057953 and AY057956) and 2987 bp long, while two other sequences were 2982 bp (AY057957) and 2990 bp long (AY057954). The most divergent sequence was AY057954, which harboured seven mutations absent in other sequences.

All four *P. rubens* whole intron sequences were 2982 bp long and were identical among themselves and to the *P. mariana* sequence AY057957. Two *P. glauca* whole intron sequences were identical among themselves and to *P. engelmanii* sequences and all were 2291 bp long. A *P. sitchensis* sequence, 3001 bp long, differed from the last two species in three duplications and a single base substitution. Whole intron sequences in *P. pungens*, *P. chihuahuana* and *P. mexicana* were all 2974 bp long and differed in two T/G transversions (T in the former species, G in the two latter species) at positions 482 and 1044. The whole intron sequences in *P. breweriana, P. smithiana and P. schrenkiana* were 2863, 2759 and 2998 bp long, respectively.

Alignment of sequences corresponding to the domains I, II and III in all species revealed that they were identical, except for mutations detected at three positions. The first variable site was found within domain II at position 426 and it was characterized by two indels and a single base substitution found only in two species - *P. smithiana* and *P. schrenkiana*. Two mutations were found within domain III - one G/T transversion found only in *P. pungens* at position 482 and a five bp duplication found only in *P. breweriana* at position 509. Domains V and VI were identical in all sequences.

Pronounced differences were detected only within domain IV. A strikingly different organization of variability in two different species groups was revealed here. Group A or Eurasian spruce species (e.g. European species *P. omorika* and *P. abies* and, hereafter Asian A spruce species - *P. asperata, P. jezoensis, P. wilsonii*, *P. crassifolia*) harbour a variable region comprised of three to six newly discovered tandem repeats which were approx. 30 bp long and similar, and a 34 bp minisatellite region described by Sperisen et al. (2001). An additional 32 bp minisatellite was found in all group A spruce species upstream of the 34 bp minisatellite region. *P. obovata*, characterized as haplotype 712 (Sperisen et al.

2001) and *P. alcoquiana,* characterized by an identical pattern of restriction fragment lengths as in *P. jezoensis*, are most likely group A spruce species as well. Surprisingly, this whole region (hereafter indel A) was absent in group B spruce species. Instead, this group of spruce species comprised of *P. schrenkiana* and *P. smithiana* (hereafter Asian B spruce species) and all American species harboured a 1281 bp long aligned stretch (hereafter indel B), absent in group A spruce species. A scheme of domain IV according to this interpretation is presented in Figure 1.

Both groups of spruce species harboured a 45 bp long stretch (hereafter fragment qts) positioned downstream of indel A and upstream of indel B (a single base substitution was found only in *P. schrenkiana*). This stretch was used as the anchor region to align whole intron sequences in spruce species of both groups. The length of the aligned sequence was extraordinary long - 3465 bp, due to the large indels A and B (the aligned sequence without those indels was only half length - 1840 bp). In total, 36 variable sites were detected (11 parsimony informative sites and 25 singletons). Additional shorter indels (one to 31 bp) were found at 30 positions, resulting in a total of 66 variable sites. Almost half (32) of all variable sites were confined to indel B - 6 parsimony informative sites, 12 singletons and 14 indels. Apart from the newly discovered tandem repeats, the variability within indel A (as shown in Figure 1) was due to a single parsimony informative site and one single nucleotide indel. However, the alignment of those tandem repeats was not straight-forward, because although they are found to be similar to each other and to the qt fragment and to the 34 and 32 bp minisatellites, some of them are found to harbour shorter tandem repeats (three to four bp) and minute inversions, making the homology assessment difficult (see later). Only one of three possible alignments is presented in Figure 1.

Nad5i230

In four spruce species, two accessions per species were available, and differences within each species consisted of one five bp duplication and one base substitution (*P. wilsonii*), one base substitution (*P. schrenkiana*), one five bp duplication (*P. likigianensis*) and one single nucleotide indel and two base substitutions (*P. chihuahuana*).

The alignment of all sequences resulted in a 1192 bp long sequence with 10 variable sites (10 parsimony informative sites and 6 singletons). Indels (one to 15 nucleotides) were detected at three positions, resulting in a total of 13 variable sites. The first variable site was found only in *P. chihuahuana 2* at position 386 and it was a single nucleotide indel. The second variable site was found at position 509 and it was characterized by two consecutive length mutations found in six species - one 10 bp insertion was found in *P. omorika*, *P.*

maximoviczii, *P. likigiaensis 2*, *P. wilsonii 2* and both *P. schrenkiana* sequences. In the latter four species, an additional five bp deletion was found, but in *P. likigianensis 2* this deletion was shifted for one base upstream in comparison to the *P. wilsonii 2* and both *P. schrenkiana* sequences. The third variable site was found at position 580 and it corresponds to the variable region reported by Jaramillo-Correa et al. (2003) and used by Ran et al. (2006). Adjacent nucleotides upstream of this region enabled the recognition of a 4 bp repeat motif (5' ACTT 3') found to be repeated two times in tandem in all spruces. However, mutational events leading to the occurrence of five haplotypes (A to E, labeling as in Ran et al. 2006) were found to differ from those described by Ran et al. (2006) and all haplotypes can be better explained by assuming evolution from each other in a stepping-stone manner (see Figure 2). Haplotype D, which did not harbour any additional indels except for the described four bp repeat motif repeated two times, is found to be the basal for other spruces (as suggested by Ran et al. 2006), since it was found in almost all other representatives of Pinaceae and some non-Pinaceae conifers. However, haplotype E is found to evolve from the haplotype D by two slipped-strand mutations, while further length mutations resulted in haplotypes A, C and B.

Genealogical and phylogenetic analysis

Interestingly, genealogical analysis at nad1i477 revealed that the division of spruce species into groups A and B can be detected even when indels A and B were omitted from the analysis (data not shown). Due to the fact that manual coding of complex indels A and B in the overall sample was almost impossible, further genealogical analysis was performed in each group separately. In that way, coding of indels A and B enabled a better resolution within each group (see later).

Genealogical analysis of the whole intron sequences in 34 *Picea* species at nad5i230 revealed the astonishing fact that the emergence of all spruces can be explained by a rather simple stepping-stone pattern (Figure 3). Generally, inclusion of all informative sites revealed the grouping of spruce species corresponding well to the grouping based only on the variable site characterized by 5 bp indels (see Figure 2). However, within each of five groups of spruce species (haplotypes A to E), better resolution was achieved by using the information in whole sequence. Among species harbouring haplotype D, the complex *P. koyamae*, *P. koraiensis*, *P. meyeri*, *P. orientalis*, *P. obovata* and *P. retroflexa* is found to be more ancient, due to the location at the centre of the haplotype tree. *P. omorika* is found to be a tip haplotype of this complex, while *P. breweriana* is found to be a tip haplotype of two complexes: (i) the *P. jezoensis*, *P. abies* and *P. alcoquiana*, and (ii) the *P. asperata*, *P.*

crassifolia and *P. torano* complex. This reticulate network was lost when *P. breweriana* was omitted from the analysis. Only the latter complex is found to be connected through an intermediate haplotype to species harbouring haplotype E. Within this group, the comlex *P. purpurea*, *P. spinulosa*, *P. brachytyla*, *P. likigiaensis*, *P. farreri* and *P. smithiana* is found to take the central position and to be connected through an intermediate haplotype to all American spruce species. Among them, those harbouring haplotype A (*P. engelmanii* and *P. chihuahuana 1*) are found to (i) stand in a central position, and (ii) to be connected to species harbouring haplotypes C (*P. mariana* as a central haplotype and *P. rubens* as a tip haplotype) and a rare haplotype B (*P. glauca 2*). Only in *P. glauca 1* the grouping based on the partial variable region (haplotype A) deviated from the grouping based on the whole intron sequences (grouped with *P. mariana* and *P. rubens* harbouring haplotype B). However, a rare haplotype B, found only in *P. glauca 2*, was also placed together with the latter two species as a tip haplotype.

Genealogical analysis in 16 species based on hybrid sequences from both introns (indels A and B omitted) enabled a slightly better resolution in the overall sample in comparison to the results presented in Figure 3 (data not shown). However, within each group of spruce species, inclusion of additional information in length mutations confined to indels A and B enabled further inferences on evolutionary relations among spruce species. Within group A spruce species, ambiguous alignment of short tandem repeats within indel A caused alternative assumptions on evolutionary relations and only a haplotype tree based on the alignment presented in Figure 1 is shown in Figure 4a. Within group B spruce species, indel B enabled their subdivision into the three complexes (i) *P. sitchensis-engelmannii-glauca* (with *P. mariana* AY057954 as a tip haplotype), (ii) *P. pungens-chihuahuana-mexicana* and (iii) *P. mariana-rubens* complex (Figure 4b). But, it is rather surprising to see that at one end of the haplotype tree, all three complexes are connected to *P. breweriana,* while at the other end of the haplotype tree they are connected to *P. smithiana-schrenkiana.* The number of mutational steps separating *P. breweriana* and *P. smithiana-schrenkiana* from all other American spruces is similar (approx. 20).

Surprisingly, the concordance of the deep division of spruces described at nad1i477, of the five haplotypes described at nad5i230 and the geographical distribution of spruce species is almost ideal (see Figure 3). All species harbouring haplotype D are group A species confined to Europe and north-eastern Asia. Species harbouring haplotype E are classified into both groups, A and B, and are confined to the north-eastern Asia and Himalayan-Hengduan region (HH), implying that the profound mutational changes at nad1i477 occurred within this group. Species harbouring haplotypes A, C and B are all American, group B spruce species. The only species which does not follow the described

pattern is *P. breweriana*. This American group B spruce species is characterized by haplotype D at nad5i230 and is grouped with the European and Asian group A species.

Phylogenetic reconstructions based on single loci revealed that nad1i477 is more informative than nad5i230 due to the higher number of parsimony informative sites (and indels), although relations among spruces were not resolved (data not shown). On the other hand, the resolution in phylogenetic reconstructions based on hybrid sequences comprised of both introns is found to strongly depend on the treatment of the length mutations. When they were treated as missing data, poorly resolved phylogenies with low bootstrap support were obtained (Figure 5a). Coding length mutations either by SIC or MCIC substantially improved resolution, due to the fact that, apart from large indels A and B, additional phylogenetic information in 30 and 13 shorter indels confined to nad1i477 and nad5i230, respectively, was utilized. The division of spruce species into two groups was revealed and two sub-clades, corresponding to those groups, are formed. The bootstrap support for this node was 80 % and 100 % in SIC and MCIC, respectively. A 50 % majority-rule consensus of 533 trees reconstructed by utilizing length mutations by SIC is presented in Figure 5b. Tree length was 86, CI = 0.8605 and RI = 0.8182. Expectedly, *P. omorika* and *P. mariana-rubens* were grouped into different clades.

Nad1i477 domain IV and the origin of spruces

Secondary structure of the domain IV in *P. omorika* and *P. mariana-rubens* confirmed the essential structural difference between those species and implies possible structural peculiarities resulting in the occurrence of indels A and B.

In *P. omorika,* folding of the domain IV revealed 40 different structures of similar free energy (approx. ΔG = - 550). In one such folding, the tandem repeat region spanning over 600 nucleotides was found to form a long hairpin radiating from a loop (Figure 6a). Interestingly, the 5’ and 3’ ends of this structure were found to correspond to the 5’ and 3’ ends of indel A, indicating that the loss of this hairpin might result in the loss of the whole indel A, as found in group B spruce species. Furthermore, fragments q1 and p1, (present in all group A spruce species), were found to form a basal part of this long hairpin, implying their possible role in the formation of this structure. Fragments q4 and r1 were found to fold to themselves, each forming a single hairpin radiating from a smaller loop, indicating that the loss of those smaller hairpins might result in a sequence with only few tandem repeats within indel A, as found in Asian A spruce species.

In *P. mariana-rubens*, folding of the domain IV did not reveal a similar pattern for indel B. In all six recovered structures with similar free energy (approx. ΔG = - 300), a long stem-loop structure comprising the whole indel B and an adjacent 300 bp stretch downstream of this indel was formed (data not shown). The loss of such a stem-loop structure resulting in the loss of indel B (as found in group A spruces) is probably likely, but in that case the additional 300 bp stretch, which was found to be conserved in all spruces, would be lost, too. However, in all group B spruce species, a stretch of 30 bp at the 3' end of indel B was found to be almost identical to the conserved ts fragment. This stretch was labelled as t1s1 and the alignment with the ts fragment is presented in Figure 8b. Furthermore, t1s1 was found to easily pair with ts and introducing constraint into the folding, e.g. forcing t1s1 to pair with ts, revealed eight different structures characterized by substantially lower free energy (ΔG = - 250). In all those structures, indel B formed a stem-loop structure with two long hairpins radiating from a loop, while the position of additional smaller loops, bulges and hairpins was slightly different - only the structure with lowest free energy (ΔG = - 112.40) is presented in Figure 6b.

Dot plot analysis confirmed the distinctive differences between *P. omorika* and *P. mariana-rubens* domain IV sequences. In *P. omorika*, plotting of the domain IV sequence to itself confirmed the presence of several shorter fragments of similar sequences spanning over 300 bp, corresponding to indel A (Figure 7a). Pairwise alignment of all tandem repeats positioned within the indel A as shown in Figure 3, is presented in Figure 8. In P. *mariana-rubens*, comparison of the domain IV sequence to itself did not reveal a similar pattern (Figure 7c), while plotting this sequence against *P. omorika* confirmed that 5' ends (first 250 bp) and 3' ends (last 1000 bp) were identical in those species and that the difference between them can be attributed to the presence of indels A and B confined to the middle part of both sequences (Figure 7b).

The plotting of *Pinus cembra* and *Pinus sylvestris* sequences to themselves and to each other revealed that (i) regions resembling tandem repeats regions are present in each species, although longer and more abundant repeats are found in the former species, and (ii) that indels of variable size (100 to 300 bp) causing the length difference between those species are found at three positions (Figure 7j).

The comparison of the domain IV sequences in *P. omorika* and *P. mariana-rubens* to *Pinus cembra* and *Pinus sylvestris* (Figure 7 d, e, g and h) revealed that (i) 5' ends (first 250 bp) and 3' ends (last 600 bp) are identical in all species, although a 100 bp insertion is present only in *Pinus cembra* within the 3' end, and that (ii) *P. omorika* sequence is more similar to both *Pinus cembra* and *P. sylvestris*, than *P. mariana-rubens*. Based on this analysis, it is not possible to distinguish whether *P. omorika* is more similar to *Pinus cembra* or to *Pinus sylvestris*.

The plotting of the *C. argyrophyla* sequence against the sequences in both spruce species and pines revealed that although the 5' and 3' ends are similar in all species, *C. argyrophyla* has an almost identical sequence to *P. mariana-rubens*, except for a 400 bp insertion present in *C. argyrophyla* at position of approximately 300 bp, and a 600 bp insertion present in *P. mariana-rubens* at position of approximately 1500 bp (Figure 7 f, i, k and l).

3. Discussion

Mitochondrial group II introns nad1i477 and nad5i230

This study demonstrates that the utilization of mitochondrial group II introns in phylogenetic reconstructions is not straight-forward. Even in our incomplete spruce sample (19 species, including *P. obovata* and *P. alcoquiana*), the length difference at nad1i477 is found to be almost 2000 bp. Such pronounced size variation is associated almost completely with the domain IV, as only three short indels are found within other domains (II and III). Since Jacquier (1996) estimated that the group II introns would need no less then 600 nucleotides to maintain all structural features involved in proper splicing, it seems likely that the pronounced size variation in this intron might be a result of two phenomena commonly associated with mtDNA genome. The first one is multipartite nature of mtDNA introns (Knoop et al. 1991, Pereira de Souza et al. 1991, Wissinger et al. 1991, Knoop et al. 1997, Knoop 2004) and consequently, the possible occurrence of the domain IV sequence of one intron within another intron (S. Kelchner, pers. comm.). Second is intramolecular recombination (Houchins et al. 1986, André et al. 1992, Malek et al. 1997) associated with the deletion of some elements of secondary structure.

Our alignment reveals an astonishing fact that the domain IV in spruces contains two hipervariable large indels, similarly as two mutational hot spots described in basal angiosperms (Löhne and Borsch 2005). Those indels are characterized by essentially different organization of variability, and secondary structure reconstructions revealed that they form a long hairpin (indel A) or a stem-loop structure (indel B). Indel A is characterized by the presence of several tandem repeats of approx. 30 bp, while within indel B, tandem repeats are found at 5' and 3' ends. Short, tandemly repeated sequences often predict to fold into thermodynamically stable secondary structures, as shown within the animal control region (Lunt et al. 1998) and Pinaceae nrDNA ITS1 (Gernandt and Liston 1999, Gernandt et

al. 2001, Campbell et al. 2005). Therefore, the formation of stable structures, such as hairpins, was expected, and an increased number of repeats (such as ten 34 bp minisatellite copies in *P. omorika*) is found to contribute only to the increased length of the hairpin and therefore, does not annihilate secondary structure.

The formation of those structures implies that their loss might result in loss of indel A, as found in group B spruce species, or loss of indel B, as found in group A spruce species. Thus, it seems likely that the intramolecular recombination triggered by the presence of repeated elements and followed by the deletion of hairpins or stem-loop structures was operating within this intron. It has been shown that repeated sequences usually cause recombination in mtDNA (Houchins et al. 1986, André et al. 1992) and repeated elements in the large and variable domain IV in the secondary structure of the *rps*3 intron have been shown to increase the recombinational activity in the mtDNA genome (Malek et al. 1997).

Apart from large indels A and B, an important role of length mutations in evolution of nad1i477 is also supported by the finding of almost identical number of base substitutions (36) and shorter indels up to 30 bp (30), which are probably due to slipped-strand misspairing, as suggested by Levinson and Gutman (1987).

The second mtDNA intron used in this study - nad5i230 has been already characterized by less pronounced variability (Ran et al. 2006). Our results reveal that this intron harbours three and ten times less base substitutions and indels, respectively, in comparison to nad1i477. However, the recognition of the four bp repeat motif confined to the 5' region upstream of the hypervariable region resulted in alternative inferences on evolutionary relations among haplotypes (e.g. species). This four bp motif is found to be present in *A. thaliana* and *C. panzhihuaensis*, while in almost all Pinaceae conifers and majority of spruces (haplotype D), it is repeated two times. Therefore, if this region is more prone to length mutations due to the specific structural features (e.g. 'trigger' sequences in Kelchner 2000), additional length mutations due to slipped-strand mispairing are likely, leading to the occurrence of a five bp repeat unit within haplotype E. Species harbouring haplotype E are found to represent an important internal group, rather than a tip group (cf. Ran et al. 2006) in which profound structural changes at nad1i477 occurred leading to the occurrence of all group B spruce species (see later).

Genealogical and phylogenetic inferences

The genus *Picea* is often characterized as systematically and taxonomically recalcitrant and its infrageneric classification, species delineation as well as the origin can still be referred as

unresolved (cf. Sigurgeirsson and Szmidt 1993, Karvonen et al. 1994, Ran et al. 2006), mainly due to the (i) limited morphological (and genetic) variation within the genus (Wright 1955, Sigurgeirsson and Szmidt 1993, Karvonen et al. 1994), and (ii) the limitations and drawbacks of all methods used.

Earlier taxonomies based on variable number of differential morphological characters are biased by morphological convergence and parallelism (Wright 1955, Wang et al. 2000, Ran et al. 2006) as well as high interspecific crossability (Wright 1955, Fowler 1983 and references therein, Hoffmann and Kleinschmit 1979 in Vidaković 1991, Ledig et al. 2004). Crossability studies, also used as a measure of phylogenetic relations among spruces (Wright 1955, Fowler 1980, 1983, Ledig 2004), rely on the premise: the closer the species are, the more readily they cross. However, estimates of crossability were found to be highly variable on individual tree bases (Wright 1955, Fowler 1983) and, as already pointed out by Ledig et al. (2004), relationships based on crossability may not be an accurate measure of total genetic similarity or difference. It has been shown that *P. mariana* and *P. glauca* crosses constantly failed when using trees originating from the part of the range where those species are sympatric (see Fowler 1983 and references therein), while crosses were successful when using *P. glauca* trees originating well outside from the *P. mariana* range (Fowler 1983). A similar pattern was observed in *P. mariana* and *P. rubens* (Manley 1975 in Fowler 1983, Gordon 1976), suggesting that crossing barriers have evolved in regions of sympatry in response to natural selection against hybrids. Our study supports the finding of intraspecific hybridization between *P. mariana* and *P. glauca*, because a single *P. mariana* haplotype (AY057954) characterized by the highest number of mutations out of four reported haplotypes in this species, is found to group with *P. glauca-engelmanii-sitchensis* complex and it is found to be a *P. glauca* tip haplotype (other *P. mariana* and *P. rubens* haplotypes are found to cluster together). On the other hand, *P. glauca* is the only species unambiguously grouped with *P. engelmanii* and *P. sitchensis* based on the nad1i477 and partial nad5i230 sequence (their close relationship is also reported by Weng and Jackson 2000 and Ledig et al. 2004), while based on the nad5i230 whole intron sequences this species was grouped with the *P*. *mariana-rubens* complex. As pointed out by Weng and Jaskson (2000), previously, *P. glauca* has been traditionally grouped with the latter two species. At least in our case, such inconsistent position of *P. glauca* seems to be due to the utilization of the specific and diverse plant material used as a source for two mtDNA loci by different authorities (e.g. Germano-Presby et al., unpublished vs. Ran et al. 2006).

P. omorika was found to easily hybridize with a majority of American spruces (Wright 1955, Mikkola 1969, 1972, Fowler 1983 and references therein, Gordon 1976, Ledig et al. 2004) and based on this finding, it was assumed to provide a connection between the two North American spruce complexes, i.e. the *P. rubens-mariana* complex and the *P. glauca-*

engelmannii-sitchensis-mexicana complex (Ledig et al. 2004). Based on our genealogical analysis, *P. omorika* and all American spruces (except *P. breweriana*) were positioned at the opposite ends of the haplotype tree based on nad5i230 and they were grouped into different groups based on nad1i477. However, successful crosses between *P. omorika* and some Asian spruces (Wright 1955, Hoffmann and Kleinschmit 1979 in Vidaković 1991), all characterized as group A-haplotype D spruce species at nad1i477 and nad5i230, respectively, likely represent accurately relations among those species, while crosses between *P. omorika* and *P. abies* (Wright 1955, Fowler 1969 in Nkongolo 1999, Hoffmann and Kleinschmit 1979 in Vidaković 1991) probably reflect a similar pattern as described in *P. mariana* and *P. glauca* and/or *P. rubens*. Furthermore, different genome sizes and organization have been reported in *P. omorika* and *P. abies* and their hybrids have never been observed in nature (Siljak-Yakovlev et al. 2002). Therefore, it seems that evolutionary relations based on crossability studies must be interpreted with cautions, especially because small genetic changes can result in reproductive isolation and crossability is not the sole determinant of reproductive isolation (cf. Ledig et al. 2004).

Similarly, the limitations of molecular phylogenies caused by specific features of all three plant genomes as well as potential flaws of all methods used are well known (cf. Sigurgeirsson and Szmidt 1993, Wang et al. 1999, Wang et al. 2000, Wei and Wang 2003, Ledig et al. 2004, Campbell et al. 2005, Shaw et al. 2005, Ran et al. 2006, Nkongolo 1999). The mitochondrial genome has been rarely used in phylogenetic reconstructions at species level (e.g. Bakker et al. 2000, Freudenstein and Chase 2001) due to the slow rate of sequence evolution and frequent recombinations and length mutations (Palmer et al. 2000, Knoop 2004, Kubo and Mikami 2007). This study reveals that nad5i230 provides low phylogenetic resolution due to the low number of variable sites and length mutations, although it has been used for phylogenetic reconstructions in spruces (Ran et al. 2006). However, the alignment and consequently, phylogenetic inferences are substantially improved by recognition of microstructural changes resulting in length mutations, as already suggested by Gu and Li (1995) and Kelchner (2000). This holds particularly for the second group II mtDNA intron used in this study - nad1i477, in which length mutations are found to play a major role in sequence evolution.

The domain IV in group II introns is usually characterized by pronounced length variability (Malek et al. 1997, Won and Renner 2003, Löhne and Borsch 2005) and the homology assessment in such regions is reported to be difficult or even impossible (Kelchner 2000). However, the detection of a 45 bp stretch within nad1i477 domain IV which is found to be present in all spruce species enabled the recognition of two large indels and consequently, the division of all species into two groups. Whenever those large indels and other, shorter indels were excluded from the phylogenetic analysis (treated as missing data),

poor resolution and low bootstrap support was obtained for all clades, as already reported in this locus by Bousquet et al. (2007). More importantly, species unequivocally characterized as group A (confined to Eurasia) or group B species (confined to Himalayan-Hengduan region and North America) did not form two separate clades, but they were rater mixed (although this clade was not resolved in our study). Similar mixing of group A and group B spruce species was observed in two published spruce molecular phylogenies, i.e. Sigurgeirsson and Szmidt (1993) and Ran et al. (2006). In the former case, the limitations of the method used pointed out by the authors themselves probably resulted in altered homology assumptions, while in latter case ancient hybridization ('chloroplast capture') could also be involved, as many of the related species are parapatric.

Coding indels, either by the simple or modified complex indel coding methods, substantially improved the phylogenetic resolution. The deep division of spruce species into two groups is recovered, since two clades were obtained and this node was supported by 80% or 100% bootstrap value, depending on the method used for indel coding. Therefore, we support the view that not coding length mutations is equivalent to discarding data (cf. Simmons and Ochoterena 2000), because treating indels as missing data actually led to misleading assumptions on evolutionary relations among spruces. Ogden and Rosenberg (2007) showed that in cases when differences in alignments are present, coding gaps outperformed their treatment as missing data, while Simmons et al. (2007) revealed that SIS and MCIC performed equally well, as found in this study also.

Therefore, apart from the differential rates of molecular sequence evolution in introns, they seem to have differential rates of length mutations (Shaw et al. 2007 vs. Golenberg et al. 1993, Clegg et al. 1994, Gielly and Taberlet 1994). This finding suggested that length mutations might not be an important source of mutations in all introns and consequently, the utilization of phylogenetic signal in length mutations might be limited to some loci. Furthermore, the trend towards larger substitution-to-indel ratios with increasing taxonomical distance was already reported for other mitochondrial introns (Laroche et al. 1997, Laroche and Bousquet 1999) and for the non-coding chloroplast DNA region (Golenberg et al 1993), suggesting that the phylogenetic utility of length mutations might be justified at some loci at genus level. Furthermore, multiple indel events at the same site might hamper identification between more distantly related taxa (Laroche et al 1997). However, in our case, even incorporating indels in phylogenetic reconstructions did not terminally resolve evolutionary relations among spruces. This finding was expected, due to the limited variation in the genus *Picea* (Wright 1955, Sigurgeirsson and Szmidt 1993) and the lowest rate of evolution in mtDNA in comparison to other genomes (Palmer et al. 2000, Knoop 2004, Kubo and Mikami 2007). But, on the other hand, the fully resolved phylogeny of spruces was not the aim of this study.

Nad1i477 indels A and B and the origin of spruces

Originally, two minisatellites (34 and 32 bp) confined to the nad1i477 have been described by Sperisen et al. (2001) in *P. abies* and their common origin was assumed. Our study reveals several other tandem repeats of approx. 30 bp with similar but not identical sequences upstream of the reported minisatellite regions. All repeats, including both minisatellites, are found to comprise indel A. Due to the fact that (i) all repeats, including the 32 and 34 bp minisatellites are found to be relatively easily aligned among themselves and to the qt fragment, and (ii) the presence of shorter indels, minute inversions, base substitutions and shorter motifs repeated in tandem within some of those repeats (see Figure 8), it seems likely that all fragments confined to indel A are of common origin and that independent mutations within each of those fragments led to their further differentiation. It seems possible that the newly discovered tandem repeats represent the repeats of an ancestral minisatellite which was 'active' in a common ancestor of spruces and whose sequence is probably similar to the qt fragment which is found in all spruce species. High rate of mutations in this region may have led to the occurrence of a 34 bp minisatellite, followed by the emergence of a 32 bp minisatellite. The variability in former minisatellite has been reported in the *P. abies* Baltico-Nordic lineage (Sperisen et al. 2001), *P. crassifolia* (Meng et al. 2006) and *P. omorika* (Aleksić and Geburek, submitted), while variability in the 32 bp minisatellite has been reported only in *P. abies* central European lineage (Sperisen et al. 2001). Such deep division in the *P. abies* into two lineages has been assumed previously (cf. Karvonen et al. 1994, Campbell et al. 2005). Furthermore, this finding suggests that (i) the emergence of the minisatellite motifs must have preceded the occurrence of haplotypes with increased number of copies, and that (ii) the presence of a certain minisatellite motif might provide insights into evolutionary relations among species, since spruce species harbouring variability in the 34 bp minisatellite are probably closely related due to the retention of such ancestral polymorphism. Shared polymorphisms between species may be expected only in cases of explosive radiation and/or interspecific hybridization (cf. Karvonen et al. 1994, Bouillé and Bousquet 2005, Syring et al. 2007).

Indel B is found to differ substantially from the indel A in sequence content and secondary structure organization. This indel is found to be present only in group B spruce species (and *C. argyrophyla*, a sister group of *Picea*, Wang et al. 2000) and a BLAST search did not reveal other similar sequences in Gene Bank (data not shown). Therefore, although it is possible that the common ancestor of spruces harboured both indels and that the indel A was lost in group B spruce species, while indel B was lost in group A spruce species,

alternatively, only indel A may have been present in a common ancestor of spruces and therefore, the loss of indel A and actually the gain of the indel B by recombination due to the presence of tandem repeats at its 5' and 3' ends may have resulted in the occurrence of all group B spruce species. As already mentioned, such profound structural changes in nad1i477 are assumed to have occurred within the group of spruce species characterized by haploytpe E in nad5i230. Among those spruce species, confined to the north-eastern Asia and Himalayan-Hengduan region, the complex *P. purpurea-spinulosa-brachytyla-likigiaensis-smithiana* is found to take the central position in the haplotype tree, suggesting that a group B species basal for all American spruce species (except for *P. breweriana*) is likely to be found within this complex. *P. breweriana* (as well as *P. omorika*) is found to represent an isolated lineage, as already reported by several authors (Wright 1955, Sigurgeirsson 1992, Sigurgeirsson and Szmidt 1993, Ledig 2004).

Our results suggest that a group A spruce species are likely the basal group in spruces. The dot plot analysis reveals higher sequences similarity between a group A species (*P. omorika*) and both pines - *P. cembra* and *P. sylvestris* than between a group B species (*P. mariana-rubens*) and pines. Within the group A spruce species characterized by the basal haplotype D in nad5i230, the complex *P. koyamai-koraiensis-meyeri-orintalis-obovata-retroflexa* is found to take the central position in the haplotype tree, suggesting that the basal spruce is likely to be found within this complex. The likely choice would be *P. koyamai*, 'the most primitive species with generalized morphological features' (Wright 1955). In that case, the Asian origin of spruce is likely, as suggested more than 50 years ago by Li (1953) and Wright (1955), due to (i) the predominance of spruce in Asia, and (ii) the fact that the southward migration from this region along known mountain routes could explain the distribution of nearly all spruces known today. Wright (1955) suggested two independent migrations of *Picea* to America which gave rise to eastern and western *Picea* complexes, respectively, while Nienstaedt and Teich (1972) hypothesized that the ancestral Asian *Picea* produced red, black and white spruce and that the other American *Picea* evolved from white spruce lineage.

Therefore, this study based on two mtDNA loci contradicts not just the assumed origin of spruces reported in two recently published spruce molecular phylogenies (Sigurgeirsson and Szmidt 1993, Ran et al. 2006), but also a close relationship between *P. omorika* and *P. mariana-rubens*. Based on different morphological features, fossil findings confined to two different continents, current (and past) wide geographical allopatry and two mtDNA group II introns, it is highly unlikely that *P. omorika* and *P. mariana-rubens* are closely related species. The reproductive barriers between those species were not developed probably because those species were never exposed to natural hybridization and selection against hybrids, while reports of the common feature in *P. mariana* and *P. omorika* in relation to the

tolerance of swampy soils has to be taken with caution, because the former species is reported to grow on poorly drained environments (cf. Weng and Jackson 2000), while the only one *P. omorika* population found in such terrain is currently experiencing extinction, since only three trees were recently reported in this population and natural regeneration in this stand is absent (Aleksić and Geburek, submitted). Furthermore, the presence of serotiny in *P. omorika* (Rushfort 1987 in Sigurgeirsson and Szmidt 1993) is probably misinterpreted due to the lack of knowledge in this species, because fire does not facilitate opening of cones and seed dispersal and only empty cones can remain attached at trees in natural populations for three to four years (Čolić 1966 in Aleksić 2008).

Acknowledgement

This work was financially supported by the Republic of Austria and Bioversity International. It comprises a portion of the PhD dissertation of JM Aleksić.

References

Aizawa M, Yoshimaru H, Saito H, Katsuki T, Kawahara T, Kitamura K, Shi F, Kaji M (2007). Phylogeography of a northeast Asian spruce, *Picea jezoensis*, inferred from genetic variation observed in organelle DNA markers. *Mol Ecol* **16**: 3393-3405.

Aleksić MJ (2008). Genetic strucutre of natural populations of Serbian spruce [*Picea omorika* (Panč.) Purk.]. PhD thesis, University of Natural Resources and Applied Life Sciences, Vienna, Austria.

Aleksić MJ, Geburek T (submitted). Was there more than one refugial population of an endemic Serbian spruce in the Balkans during the last glaciation?

André C, Levy A, Walbot V (1992). Small repeated sequences and the structure of plant mitochondrial genomes. *Trends Genet* **8**: 128-132.

Bakker FT, Culham A, Pankhurst CE, Gibby M (2000). Mitochondrial and chloroplast DNA-based phylogeny of *Pelargonium* (Geraniaceae). *Am J Bot* **87**: 727-734.

Bouillé M, Bousquet J (2005). Trans-species shared polymorphisms at orthologous nuclear gene loci among distant species in the conifer *Picea* (Pinaceae): implications for the long-term maintaince of genetic diversity in trees. *Am J Bot* **92**: 63-73.

Bousquet J, Isabel N, Pelgas B, Cottrell J, Rungis D, Ritland (2007). 3 spruce. In (ed) Kole C., *Genome mapping and molecular breeding in plants, v. 7 - Forest trees*. Springer-Verlag Berlin Heidelberg. p 93-114.

Campbell CS, Wright WA, Cox M, Vining TF, Major CS, Arsenaults MP (2005). Nuclear ribosomal DNA internal transcribed spacer 1 (ITS1) in Picea (Pinaceae): sequence divergence and structure. *Mol Phylogenet Evol* **35**: 165-185.

Chapdelaine Y, Bonen L (1991). The wheat mitochondrial gene for subunit I of the NADH dehydrogenase complex: a trans-splicing model for this gene-in-pieces. *Cell* **65**: 465-472.

Clegg MT, Gaut BS, Learn Jr GH, Morton BR (1994). Rates and patterns of chloroplast DNA evolution. *Proc Natl Acad Sci USA* 91: 6795-6801.

Clement M, Posada D, Crandall K (2000). TCS: a computer program to estimate gene genealogies. *Mol Ecol* **9**: 1657-1660.

Conklin PL, Wilson RK, Hanson MR (1991). Multiple *trans*-splicing events are recquired to produce a mature nad1 transcript in a plant mitochondrion. *Genes Dev* **5**: 1407-1415.

Conifer Specialist Group (1998). *Picea omorika*. In: IUCN 2007. *2007 IUCN Red List of Threatened Species* (http://www.iucnredlist.org/search/details.php/30313/summ).

Demesure B, Sodzi N, Petit RJ (1995). A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Mol Ecol* **4**: 129-131.

Dombrovska O, Qiu YL (2004). Distribution of introns in the mitochondrial gene nad1 in land plants: phylogenetic and evolutionary implications. *Mol Phylogen Evol* **32**: 246-263.

Farjon A (2001). *World checklist and bibliography of conifers*. 2nd edition. Royal Bot Gard, Kew, London, UK.

Feliner GN, Rosselló JA (2007). Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. *Mol Phylogen Evol* **44**:911-919.

Felsenstein J (1985). Conidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**: 783-791.

Fowler DP (1980). Hybridization of black spruce and Serbian spruce. *Can For Serv Marit For Cent Inf Rep* M-X-112.

Fowler DP (1982). The hybrid black x Sitka spruce, implications to phylogeny of the genus *Picea*. *Can J For Res* **13**: 108-115.

Frankis MP (1988) Generic inter-relations in Pinaceae. *Notes R bot Gdn Edinb* **45**: 527-548.

Freudenstein JV, Chase MW (2001). Analysis of mitochondrial nad1b-c intron sequences in *Orchidaceae*: utility and coding of length-change characters. *Syst Bot* **26**: 643-657.

Gernandt DS, Liston A (1999). Internal transcribed spacer region evolution in Larix and Pseudotsuga (Pinaceae). *Am J Bot* **86**: 711-723.

Gernandt DS, Liston A, Piñero D (2001). Variation in the nrDNA ITS of *Pinus* subsection *Cembroides*: implications from molecular systematic studies of pine species complexes. *Mol Phylogenet Evol* **21**: 449-467.

Gielly L, Taberlet P (1994). The use of chloroplast DNA to resolve plant phylogenies: noncoding versus rbcL sequences. *Mol Biol Evol* **11**: 769-777.

Godbout J, Jaramillo-Correa, Beaulieu J, Bousquet J (2005). A mitochondrial DNA minisatellite reveals the postglacial history of jack pine (*Pinus banksiana*), a broad-range North American conifer. *Mol Ecol* **14**: 3497-3512.

Golenberg EM, Clegg MT, Durbin ML, Doebley J, Ma DP (1993). Evolution of a noncoding region of the chloroplast genome. *Mol Phylogenet Evol* **2**: 52-64.

Gordon AG (1976). The taxonomy and genetics of Picea rubens and its relations to Picea mariana. *Can J Bot* **54**: 781-813.

Graham SW, Reeves PA, Burns ACE, Olmstead RG (2000). Microstructural changes in noncoding choroplast DNA: interpretation, evolution, and utility of indels and inversions in basal angiosperm phylogenetic inference. *Int J Plant Sci* **161**: S83-S96.

Groth-Malonek M, Pruchner D, Grewe F, Knoop V (2005). Ancestors of trans-splicing mitochondrial introns support serial sister group relationships of hornworts and mosses with vascular plants. *Mol Bio Evol* **22**: 117-125.

Gu X, Li WH (1995). The size distribution of insertions and deletions in human and rodent pseudogenes suggests the logarithmic gap penalty for sequence alignment. *J Mol Evol* 40: 464-473.

Gugerli F, Sperisen C, Büchler U, Brunner I, Brodbeck S, Palmer JD, Qiu Y-L (2001a). The evolutionary split pf Pinaceae from other conifers: evidence from an intron loss and a multigene phylogeny. *Mol Phylogenet Evol* **21**: 167-175.

Houchins JP, Ginsburg H, Rohrbaugh M, Dale RMK, Schardl CL, Hodge TP, Londsdale DM (1986). DNA sequence analysis of a 5.27-kb direct repeat occuring adjacent to the regions of S-episome homology in maize mitochondria. *EMBO J* **5**: 17-29.

http://emboss.sourceforge.net/

Gugerli F, Sperisen C, Büchler U, Magni F, Geburek Th, Jeandroz S, Senn J (2001b). Haplotype variation in a mitochondrial tandem repeat of Norway spruce (*Picea abies*) populations suggests a serious founder effect during postglacial re-colonization of the western Alps. *Mol Ecol* **10**: 1255-1263.

Jacquier A (1996). Group II introns: elaborate ribozymes. *Biochimie* **78**: 474-428.

Jaramillo-Correa JP, Bousquet J (2003a). New evidence from mitochondrial DNA of a progenitor-derivative species relationship between black spruce and red spruce (Pinaceae). *Am J Bot* **90**: 1801-1806.

Jaramillo-Correa JP, Bousquet J, Beaulieu J, Isabel N, Perron M, Bouillé M (2003b). Cross-species amplification of mitochondrial DNA sequence-tagged-site markers in conifers: the nature of polymorphism and variation within and among species in *Picea*. *Theor Appl Genet* **106**: 1353-1367.

Jaramillo-Correa JP, Beaulieu J, Ledig FT, Bousquet J (2006). Decoupled mitochondrial and chloroplast DNA population structure reveals Holocene collapse and population isolation in a threatened Mexican-endemic conifer. *Mol Ecol* **15**:2787-2800.

Karvonen P, Szmidt AE, Savolainen O (1994). Lenght variation in the internal transcribed spacers of ribosomal DNA in Picea abies and related species. *Theor Appl Genet* **89**: 969-974.

Kelchner SA (2000). The evolution of noncoding chloroplast DNA and its application in plant systematics. *Ann Miss Bot Gard* 87: 482–498.

Kelchner SA (2002). Group II introns as phylogenetic tools: structure, function, and evolutionary constraints. *Am J Bot* **89**: 1651-1669.

Knoop V (2004). The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. Curr Genet 46: 123-139.

Knoop V, Schuster W, Wissinger B, Brennicke A (1991). Trans splicing integrates an exon of 22 nucleotides into nad5 mRNA in higher plant mitochondria. *EMBO J* **10**: 3483-3493.

Knoop V, Altwasser M, Brennicke A (1997). A tripartite group II intron in mitochondria of an angiosperm plant. *Mol Gen Genet* **255**: 269-276.

Kubo T, Mikami T (2007). Organization and variation of angiosperm mitochondrial genome. *Physiol Plantarum* **129**: 6-13.

Kubo T, Nishizawa S, Sugawara A, Itchoda N, Estiati A, Mikami T (2000). The complete nucleotide sequence of the mitochondrial genome of sugar beet (*Beta vulgaris* L.) reveals a novel gene for $tRNA^{Cys}$ (GCA). *Nucleic Acids Res* **28**: 2571-2576.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947-2948.

Laroche J, Bousquet J (1999). Evolution of the mitochondrial *rps3* intron in perennial and annual angiosperms and homology to *nad5* intron 1. *Mol Biol Evol* **16**: 441-452.

Laroche J, Li P, Maggia L, Bousquet J (1997). Molecular evolution of angiosperm mitochondrial introns and exons. *Proc Natl Acad Sci USA* **94**: 5722-5727.

Ledig FT, Hodgskiss PD, Krutovskii KV, Neale DB, Eguiluz-Piedra T (2004). Relationships among the spruces (*Picea*, pinaceae) of southwestern North America. *Syst Bot* **29**: 275-295.

Levinson G, Gutman G (1987). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol* **4**: 203-221.

Li HL (1953). Present distribution and habitats of the conifers and taxads. *Evolution* **7**: 245-261.

Löhne C, Borsch T (2005). Molecular evolution and phylogenetic utility of the *petD* group II intron: a case study in basal angiosperms. *Mol Biol Evol* **22**: 317-332.

Lunt DH, Whipple LE, Hyman BC (1998) Mitochondrial DNA variable number tandem repeats (VNTRs): utility and problems in molecular ecology. *Mol Ecol* **7**: 1441-1455.

Malek O, Brennicke A, Knoop V (1997). Evolution of trans-splicing plant mitochondrial introns in pre-Permian times. *Proc Natl Acad Sci USA* **94**: 553-558.

Mayr H (1890). *Monographie der Abietineen des Japanischen Reiches*. M. Rieger'sche Universitäts-Buchhandlung, München.

Meng LA, Yang R, Abbott RJ, Miehe G, Hu T, Liu J (2006). Mitochondrial and chloroplast phylogeography of *Picea crassifolia* Kom. (Pinaceae) in the Qinghai-Tibetan Plateau and adjacent highlands. *Mol Ecol* **16**: 4128-4137.

Michel F, Ferat J-L (1995). Structure and activities of group II introns. *Annual Rev Biochemistry* **64**: 435-461.

Michel F, Umesono K, Ozeki H (1989). Comparative and functional anatomy of group II catalytic introns - a review. *Gene* **82**: 5–30.

Mikkola L (1969). Observations on interspec ific sterility in Picea. *Ann Bot Fenn* **6**: 285-339.

Mikkola L (1972). Crossability between *Picea omorika* (Pancic) Purkyne and *P. glauca* (Moench) Voss. *Ann Bot Fenn* **9**: 33-36.

Müller K (2006). Incorporating information from length-mutational events into phylogenetic analysis. *Mol Phylogenet Evol* **38**: 667–676.

Nienstaedt H, Teich A (1972). Genetics of white spruce. *Forest Service Research Paper, USDA* WO-15.

Nkongolo KK (1999). RAPD and cytological analyses of *Picea* spp. from different provenances: genomic realtions among taxa. *Hereditas* **130**: 137-144.

Notsu Y, Masood S, Nishikawa T, Kubo N, Akiduki G, Nakazono M, Hirai A, Kadowaki K (2002). The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: frequent DNA sequence acquisition and los during the evolution of flowering plants. Mol Genet Genomics 268: 434-445.

Ogden TH, Rosenberg MS (2007). How should gaps be treated in parsimony? A comparison of approaches using simulation. *Mol Phylogenet Evol* 42: 817- 826.

Palmer JD, Adams KL, Cho Y, Parkinson CL, Qiu Y-L, Song K (2000). Dynamic evolution of plant mitochondrial genomes: Mobile genes and introns and highly variable mutation rates. *Proc Natl Acad Sci USA* **97**: 6960-6966.

Pereira de Souza A, Jubier M-F, Delcher E, Lancelin D, Lejeune B (1991). A trans-splicing model for the expression of the tripartite nad5 gene in wheat and maize mitochondria. *The Plant Cell* **3**: 1363-1378.

Prager EM, Fowler DP, Wilson AC (1976). Rates of evolution in conifers (Pinaceae). *Evolution* **30**: 637-649.

Ran J-H, Wei X-X, Wang X-Q (2006). Molecular phylogeny and biogeography of *Picea* (Pinaceae): implications for phylogeographical studies using cytoplasmic haplotypes. *Mol Phylogenet Evol* **41**: 405-419.

Ravazzi C (2002). Late Quaternary history of spruce in southern Europe. *Rev Paleobot Palinol* **120**: 131-177.

Schmidt PA (1998). *Picea* A. Dietr. P. 1-14. In (eds) Schütt P, Schuck HJ, Lang U, Roloff A, Enzyklopädie der Holzgwächse. 14. Erg. Lfg., 12/95. Ecomed-Verlag, Landsberg.

Schmidt-Vogt H (1977). *Die Fichte, Ein Handbuh in zwei Bänden*. Verlag Paul Parey, Hamburg und Berlin.

Sigurgeirsson A, Szmidt A (1993). Phylogenetic and biogeographic implications of choroplast DNA variation in *Picea*. *Nord J Bot* **13**: 233-246.

Sigurgeirsson A (1992). Insights into the evolution of *Picea* inferred from chloroplast DNA. PhD thesis, Swedish University of Agricultural Sciences, Umeå, Sweden.

Shaw J, Lickey EB, Schilling EE, Small RL (2007). Comparison of whole choroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *Am J Bot* **94**: 275-288.

Siljak-Yakovlev S, Cerbah M, Coulaud J, Stoian V, Brown SC, Zoldos V, Jelenic S, Papes D (2002). Nuclear DNA content, base composition, heterochromatin and rDNA in *Picea omorika* and *Picea abies*. *Theor Appl Genet* **104**: 505-512.

Simmons MP, Müller K, Norton A (2007). The relative performance of indel-coding methods in simulations. *Mol Phylogenet Evol* **44**: 724-740.

Simmons MP, Ochoterena H (2000). Gaps as characters in sequence based phylogenetic analyses. *Syst Biol* **49**: 369–381.

Smith DE, Klein AS (1996). Phylogenetic inferences on the relationship of North American and European *Picea* species based on nuclear ribosomal 18S sequences and the internal transcribed spacer 1 region. *Mol Phylogenet Evol* 3: 17-26.

Sperisen C, Büchler U, Gugerli F, Mátyás G, Geburek Th, Vendramin GG (2001). Tandem repeats in plant mitochondrial genomes: application to the analysis of population differentiation in the conifer Norway spruce. *Mol Ecol* **10**: 257-263.

Syring J, Willyard A, Cronn R, Liston A (2005). Evolutionary relationships among *Pinus* (Pinaceae) subsections inferred from multiple low-copy nuclear loci. *Am J Bot* **92**: 2086-2100.

Swofford DL (2002). PAUP*: Phylogenetic analysis using parsimony (* and related methods). Version 4.0b10. Sinauer Associates, Sunderland, MA.

Unseld M, Marienfeld JR, Brand P, Brennicke A (1997). The mitochondrial genome of Arabidopsis thaliana contains 57 genes in 366,924 nucleotide. *Nat Genet* **15**: 57-61.

Vidaković M (1991). *Conifers, morphology and variation*. 2nd edition. Grafički zavod Hrvatske, Zagreb.

Wang X-Q, Tsumura Y, Yoshimaru H, Nagasaka K, Szmidt AE (1999). Phylogenetic relations of Eurasian pines (*Pinus*, Pinaceae) based on chloroplast *rbcL*, *matK*, *rpl20-rps18* spacer, and *trnV* intron sequences. *Am J Bot* **86**: 1742-1753.

Wang X-Q, Tank DC, Sang T (2000). Phylogeny and divergence times in Pinaceae: evidence from three genomes. *Mol Biol Evol* **17**: 773-781.

Wei X-X, Wang X-Q (2003). Phylogenetic split of *Larix*: evidence from paternally inherited cpDNA *trnT-trnF* region. *Plant Syst Evol* **239**: 67-77.

Weng C, Jackson ST (2000). Species differentiation of North American spruce (*Picea*) based on morphological and anatomical characteristics of needles. *Can J Bot* **78**: 1367-1383.

Willkomm M (1887). *Forstliche Flora von Deutchland und Oesterrich*. CF. Winter'sche Berlagshandlung, Leipzig.

Wissinger B, Schuster W, Brennicke A (1991). *Trans*-splicing in Oenothera mitochondria: *nad1* mRNAs are edited in exon and *trans*-splicing group II intron sequences. *Cell* **65**: 473–482.

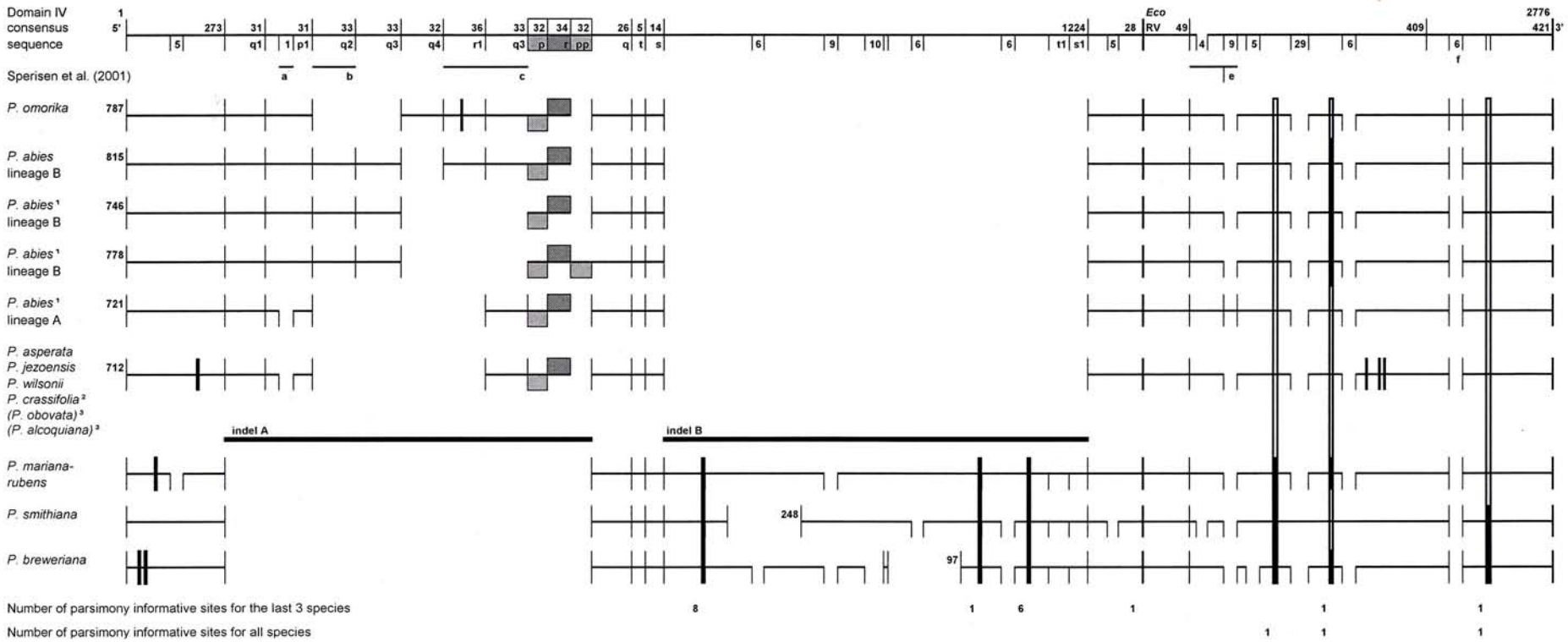
Won H, Renner S (2003). Horizontal gene transfer from flowering plants to Gnetum. *Proc Natl Acad Sci USA* **100**: 10824-10829.

Wright JW (1955). Species crossability in spruce in relation to distribution and taxonomy. *For Sci* **1**: 319-349.

www.ncbi.nlm.nih.gov/Genbank

Zuker M, Mathews DH, Turner DH (1999). In (eds) Barciszewski J, Clark BFC. Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide in RNA biochemistry and biotechnology. NATO ASI Series, Kluwer Academic Publishers.

Figure 1. Indels A and B within the nad1i477 domain IV consensus sequence in 11 spruce species including *P. omorika* and *P. mariana-rubens*



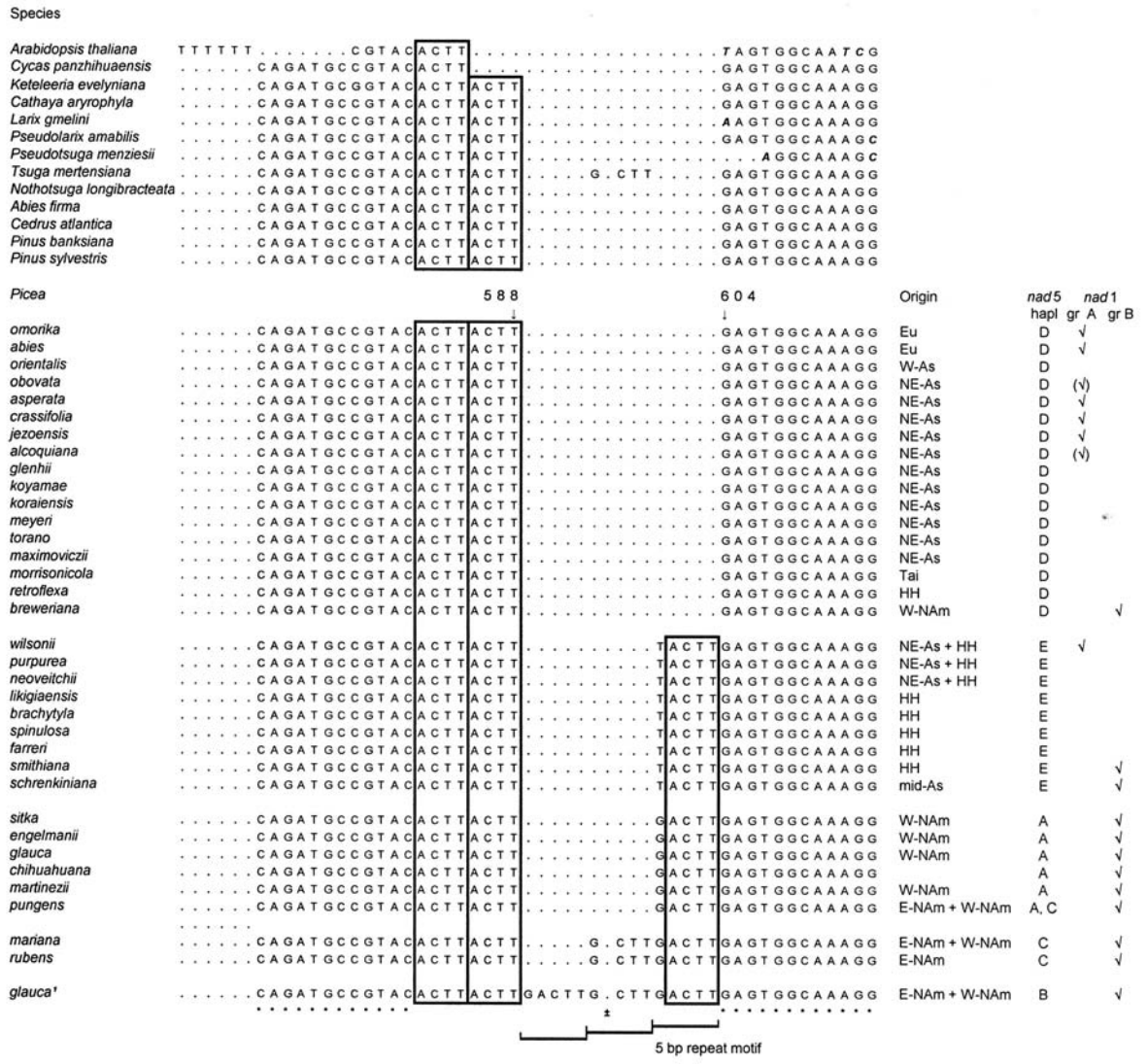
The aligned length (2776 bp) of the domain IV in 11 spruce species. Horizontal lines delineate detected fragments - their labels are given below the sequence and their lengths are given above the sequence (for instance, the length of the fragment q1 is 31 bp). Only newly discovered tandem repeats of approx. 30 bp are labelled, as well as fragments q, t, s and t1s1. The numbers below the consensus sequence boxed from the left and right sides represent the lengths of deletions whose position can be traced in certain spruce species by the discontinuity within its sequence (for instance, a five bp deletion is present only within the first fragment of 273 bp in *P. mariana-rubens*); r - 34 bp minisatellite reported by Spersen et al. (2001); pp and p - 32 bp minisatellites reported by Sperisen et al. (2001) and newly discovered 32 bp minisatellite, respectively; fragments +a, +b, +c and +e reported by Sperisen et al. (2001) are presented as well. The latter fragment is found to be duplicated only in *P. abies* haplotype 721, while a four bp deletion within this fragment is found only in *P. smithiana*; according to this nomenclature, a six bp fragment found to be duplicated only in *P. omorika* is labelled as +f; base substitutions in a single species are presented as horizontal thick black bars in that species, while parsimony informative sites are presented as horizontal black bars in all species; *Eco* RV - restriction enzyme cutting site; indel A - characterized by several newly discovered tandem repeats of approx. 30 bp and previously reported 34 and 32 bp minisatellites is present only in group A spruce species; indel B - characterized by only occasional duplications of repeat motifs up to 30 bp and by relatively frequent deletions is present only in group B spruce species;

[1] - *P. abies* haplotypes from the Baltico-Nordic (lineage A) and central European (lineage B) regions reconstructed based on the *P. abies* consensus sequence (Sperisen et al. 2001) and haplotype 712 (*P. asperata*, *P. jezoensis* and *P. wilsonii*). Haplotype 746 differs from the haplotype 815 in a 69 bp stretch (fragments r1 and q3), while 32 bp minisatellite was reported downstream of the 34 bp minisatellite region in haplotype 778 (Sperisen et al. 2001). Haplotype 721 differs from the haplotype 712 in one 9 bp duplication (+e in Sperisen et al. 2001);

[2] - in *P. crassifolia* only partial nad1i477 sequence are available. A sequence harbouring one 34 bp minisatellite (e.g. ancestral sequence) is identical to the corresponding region within the haplotype 712 in *P. asperata*, *P. jezoensis* and *P. wilsonii*, implying to the same haplotype in this species;

[3] - haplotype 712 was reported in *P. obovata* (Sperisen et al. 2001), while the pattern of the restriction fragment lengths in *P. alcoquiana* (Aizawa et al. 2008) is identical to that in *P. jezoensis* (Meng et al. 2006) implying to the same haplotype 712 in former species.

Figure 2. Variable region in nad5i230 characterized by duplications of a five bp repeat motif in 34 *Picea* species, a four bp repeat motif in *Arabidopsis thaliana*, Pinaceae, non-Pinaceae conifers and comparison to the grouping of spruce species based on nad1i477 and in relation to their geographical distribution



A four bp repeat motif 5’ ACTT 3’ is found to be present in all *Picea* species tested. In *A. thaliana* and *C. panzhihuaensis* this motif is present only once, while in *K. evelyniana*, *C. argyrophyla* and other Pinaceae conifers is repeated in tandem two times. Base substitutions (bold and italic) and additional duplications and deletions are found in some species.

The haplotype D with two 4 bp repeats is most likely the basal one in all spruce species. Those species are characterized as group A spruce species based on nad1i477 and they are confined to Europe (Eu) and Asia, except for *P. breweriana* characterized as group B species and confined to western North America (W-NAm). Although the spruce sample was incomplete at latter locus, other spruce species characterized by haplotype D are presumably group A spruce species.

The haplotype E emerged presumably by two slipped-strand mutations - first is additional duplication of a four bp motif and the second is duplication of a single nucleotide (T) at second four bp repeat unit. In that way, a five bp repeat motif (5’ TACTT 3’) emerged. Haplotype E is present in species confined to north-eastern Asia (NE-As) and Himalayan-Hengduan region (HH). Spruce species harbouring haplotype E are classified into both groups based on nad1i477 and presumably profound structural changes at this locus occurred within this group of species.

A single nucleotide transversion (T/G) within the third five bp repeat unit resulted in the occurrence of a repeat motif 5’ GACTT 3’ and an emergence of the haplotype A. Additional duplication of this five bp motif and presumably a deletion of a single nucleotide (labelled as ±) resulted in the emergence of a haplotype C, while additional duplication of a five bp repeat motif resulted in a rare haplotype B found only in *P. glauca 2* - this species is also characterized by abundant haplotype A (*P. glauca 1*). All species characterized by haplotypes A, C and B are American - western-North American (W-NAm) and/or eastern-Nort American (E-NAm) group B spruce species.

The only species which did not follow the described pattern is *P. breweriana* (e.g. haplotype D - group B species confined to western North-America) and therefore the reverse mutation, e.g. deletion of five bp repeat unit(s) is assumed in this species.

Figure 3. Haplotype tree based on the whole nad5i230 sequences in 34 *Picea* species, their characterization as haplotypes A - E based on the nad5i230 hipervariable region with five bp indels, geographical distribution and assignment to groups A and B based on nad1i477

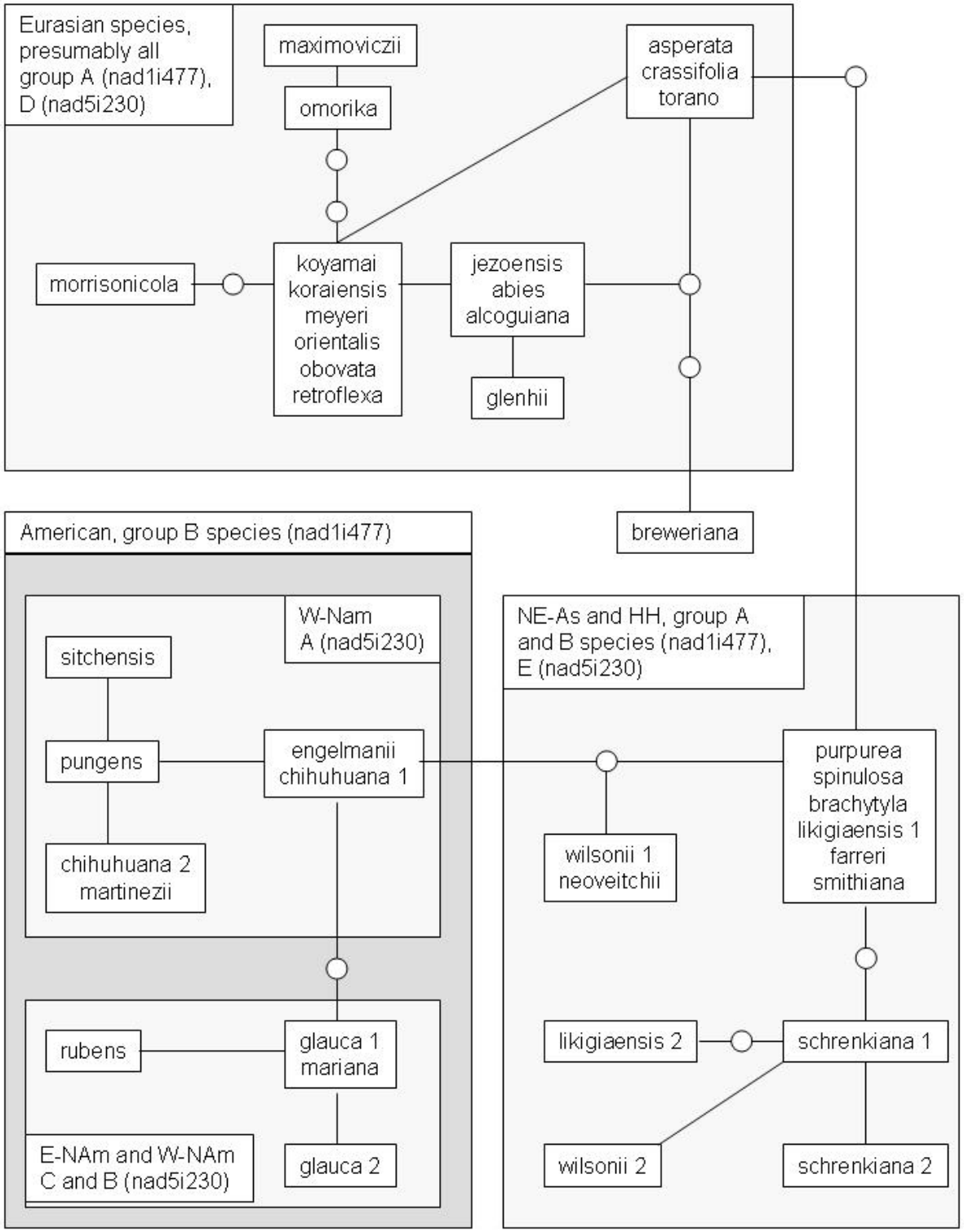
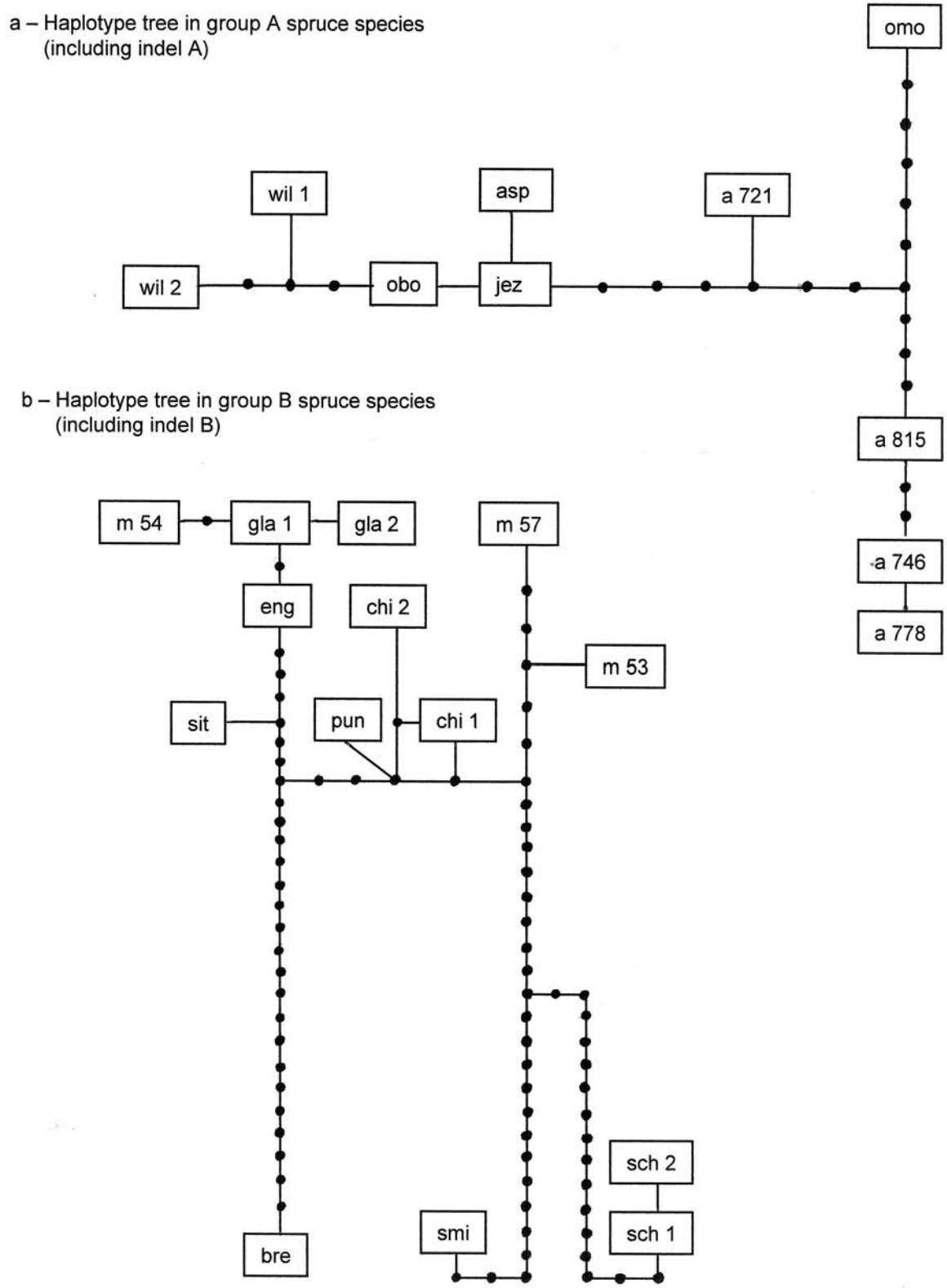


Figure 4. Haplotype trees in group A and B spruce species based on composite sequences of nad1i477 (including indels A and B) and nad5i230

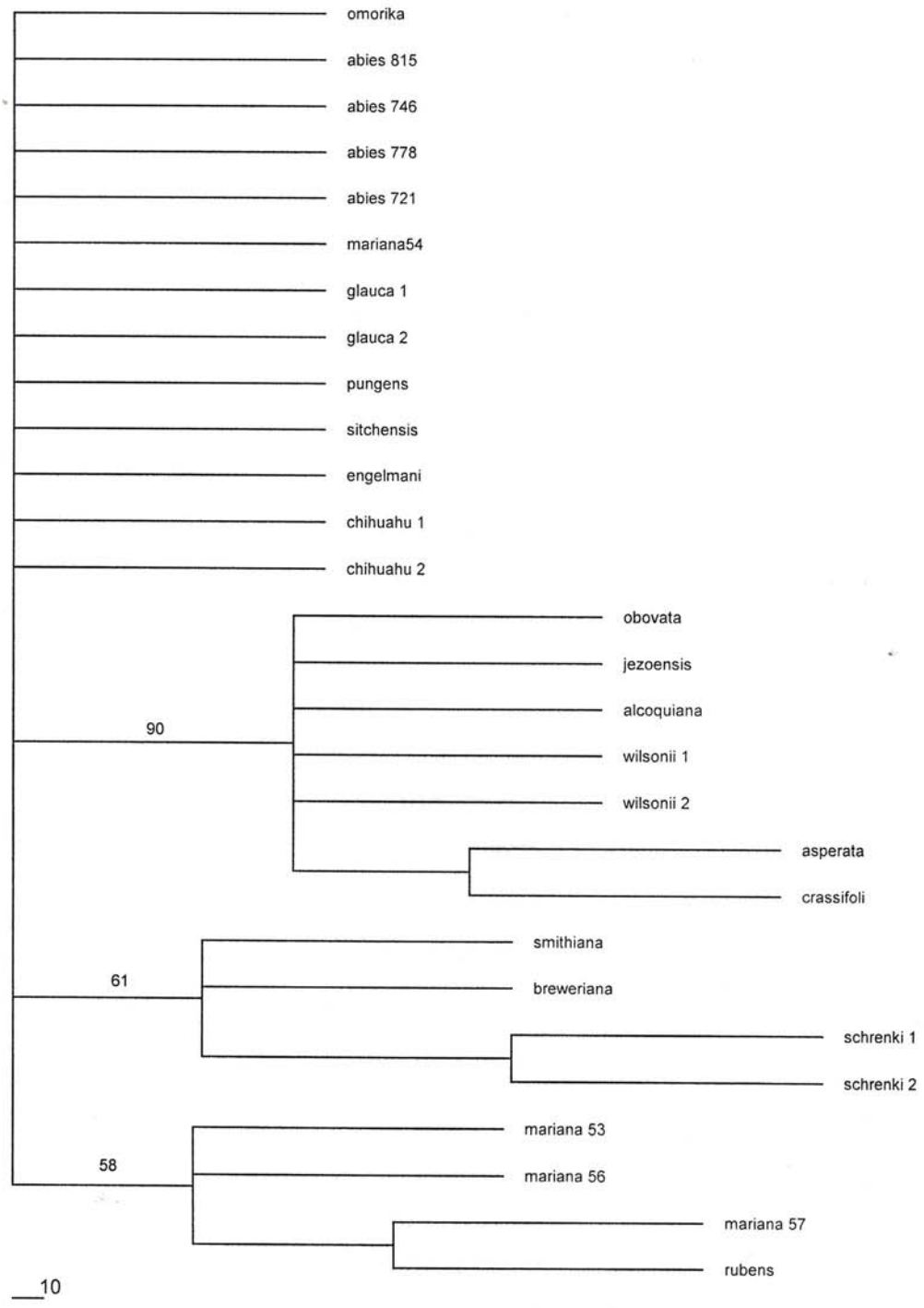


a 721, a 815, a 746 and a 778 - *P. abies* haplotypes 721, 815, 746 and 778, respectively; m 53, m 54 and m 57 - *P. mariana* accessions AY057953, AY057954 and AY057957, respectively.

Figure 5. Phylogenetic reconstructions based on composite nad1i477 and nad5i230 sequences and different methods of indel coding

a) Indels treated as missing data - 50% majority rule consensus tree

tree length 78, CI = 0.577, RI = 0.616, numbers above branches denote bootstrap values above 50% after 1000 replicates



b) Indels coded by Simple Indel Coding - 50% majority rule tree

tree length 86, CI = 0.8605, RI = 0.8182, values above branches denote bootstrap values above 50% after 1000 replicates

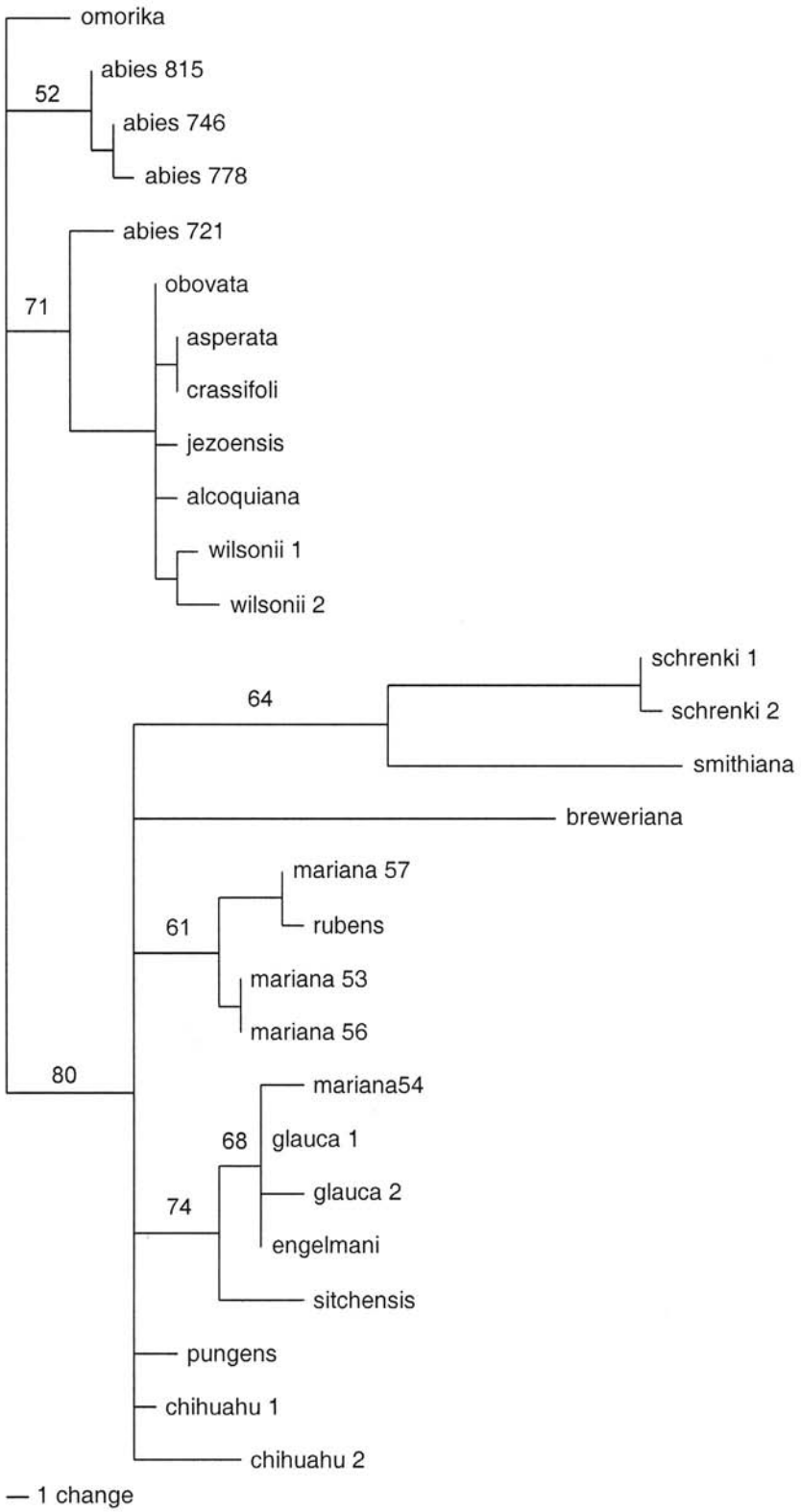


Figure 6. Secondary structure of the hypervariable indels A and B

a - indel A

b - indel B

a - indel A is present in group A spruce species and is characterized by several tandem repeats (approx. 30 bp) organized in a long hairpin (5’ and 3’ ends of this indel are labeled with the black bar). Fragments q1 and p1 (present in all group A spruce species) and qt (present in all species) are responsible for the formation of the basal part of a hairpin, while fragments present only in *P. omorika* and *P. abies* (e.g. q3, q4 and r1) are found to fold to themselves forming smaller hairpins radiating from a smaller loop, implying that the loss of those smaller hairpins might result in losses of fragments q3, q4 and r1, as found in other group A spruce species.

b - indel B is present in group B spruce species. Forced pairing of fragments ts and t1s1 at 5’ and 3’ ends of this indel, respectively, resulted in the formation of a stem-loop structure with two long hairpins radiating from a loop (5’ and 3’ ends of this indel are labeled with the black bar). Within this structure, characterized by the absence of tandem repeats and only occasional duplications of motifs up to 30 bp, smaller loops, bulges and hairpins are frequent.

Figure 7. Dotplot comparison of *P. omorika*, *P. mariana-rubens*, *P. sylvestris*, *P. cembra* and *C. argyrophylla* domain IV sequences

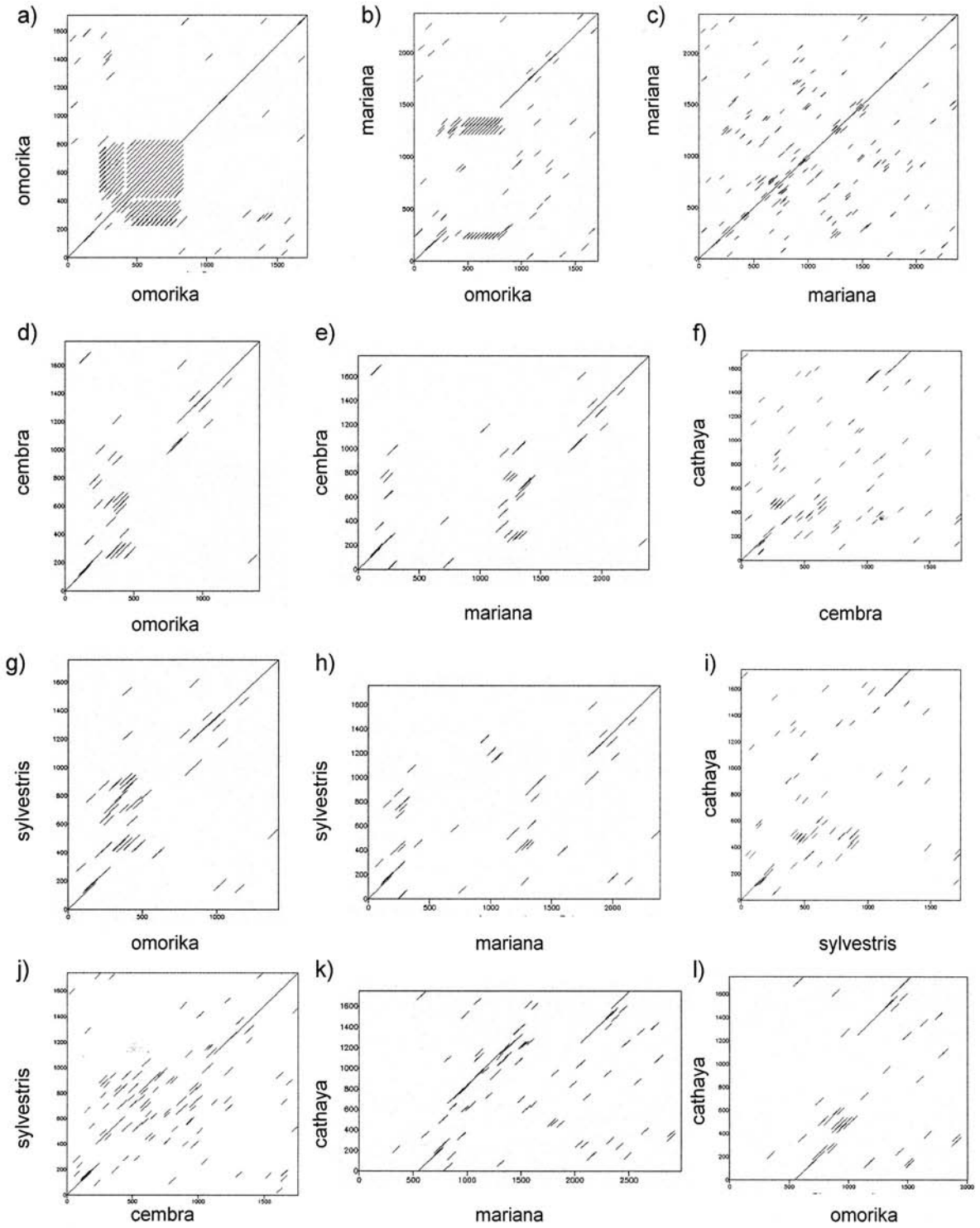
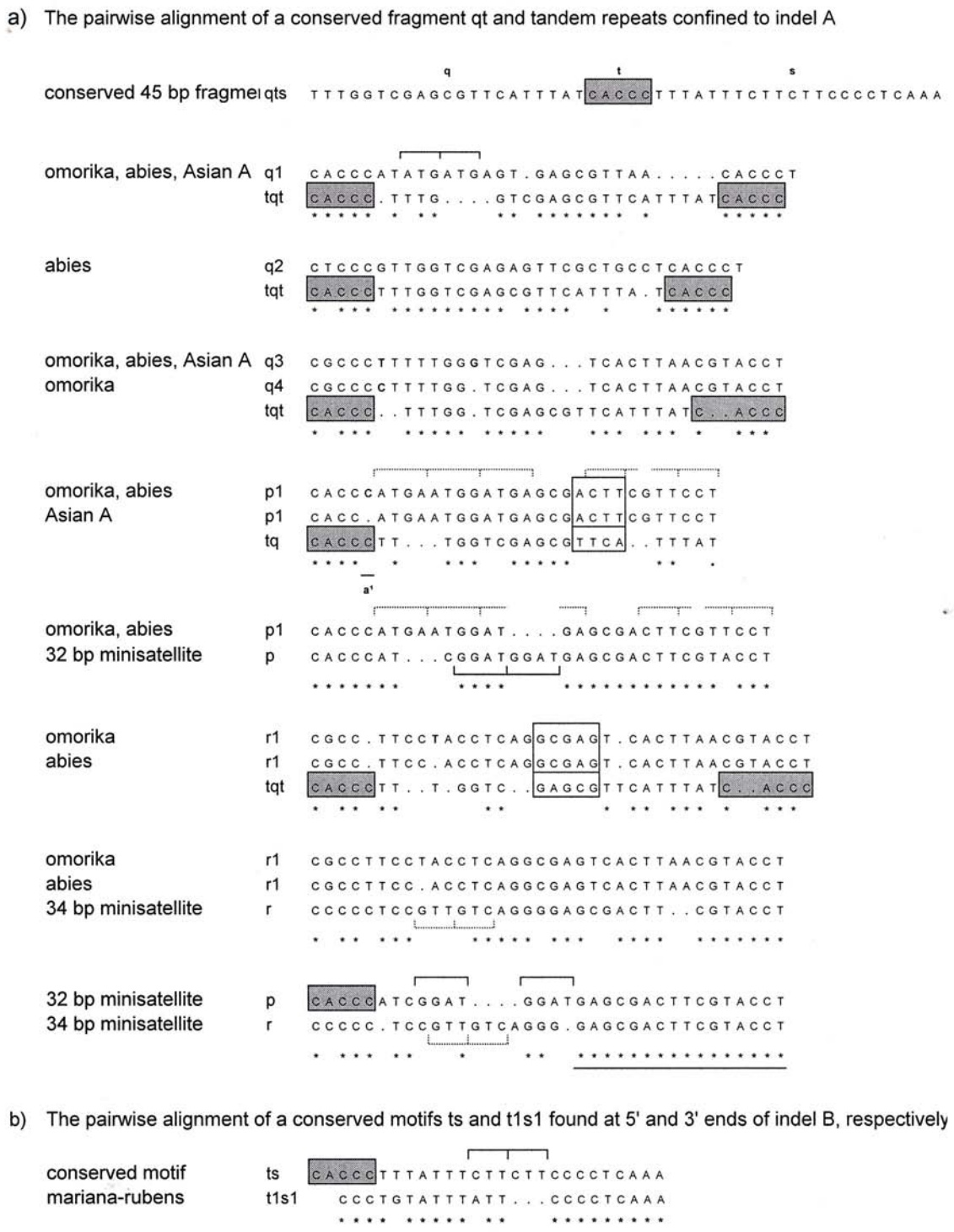


Figure 8. The pairwise alignment of fragments confined to indels A and B



a - the motif qt is conserved in all spruce species and probably represents a trigger sequence which was producing tandem repeats by slipped-strand mispairing (see Kelchner 2000). Therefore, this motif is probably the motif of an ancestral minisatellite which was 'active' in a common ancestor of spruces and the presence of several tandem repeats (e.g. qt-qt-qt etc.) is likely. Later on, tandem repeats emerging from this conserved motif most likely evolved independently, since several indels and shorter repeat motifs are found within some of those repeats. Duplications of such shorter repeats are marked by full lines, while duplications harbouring additional base substitution and/or indels are marked with dotted lines. Inverted fragments are boxed. Therefore, it seems that all tandem repeats confined to indel A are exceptionally prone to mutations, resulting in almost impossible homology assessment. However, it seems likely that during the evolution of those ancestral tandem repeats, the 34 bp minisatellite motif emerged and became an 'active' (e.g. trigger) minisatellite producing copies in several *Picea* species, such as *P. omorika* (Aleksić and Geburek submitted), *P. crassifolia* (Meng et al. 2006), *P. abies* Baltico-Nordic lineage (Sperisen et al. 2001) and presumably some other group A spruce species. Since the 16 bp at 3' ends of 34 and 32 bp minisatellites are indentical (underlined with a black line) and each minisatellite is found to harbour duplications of shorter motifs, is seems likely that a 32 bp minisatellite evolved from the 34 bp minisatellite and became an 'active' one in *P. abies* central European lineage. The common origin of those minisatellites is assumed by Sperisen et al. (2001) and up to authors best knowledge, the 32 bp minisatellite has not been reported in other *Picea* species, implying its relatively recent occurrence.

b - Fragment ts (conserved in all spruce species) is found at 5' end of indel B, while similar fragment t1s1 is found only in group B spruce species at 3' end of this indel. A three bp repeat motif (5' CTT 3') is found to be repeated two times within ts, while in t1s1 is repeated only once and additional transversion (C/A) is found. Another sequence similar to ts and t1s1 fragments is found upstream of the t1s1 fragment within the indel B (at position of approx. 1.300 bp in *P. mariana-rubens* sequence) and its presence is shown in Figure 7b.

Supplenentary material

Table 1. The source material for nad1i477 - the list of species, their Gene Bank accession numbers and the authors

Species	Gene Bank access. n.	Author
P. abies	AY289611	Germano-Presby,J., Roy,R.A. and Klein,A.S.
P. abies isolate A	AF142641	Grivet,D., Jeandroz,S. and Favre,J.M.
P. abies isolate D	AF142642	Grivet,D., Jeandroz,S. and Favre,J.M.
P. asperata	AY153790	Germano-Presby,J., Klein,A.S. and Thorner,A.R.
P. breweriana	AY153786	Germano-Presby,J., Klein,A.S. and Thorner,A.R.
P. chihuahuana	AY153784	Germano-Presby,J., Klein,A.S. and Thorner,A.R.
*P. crassifolia** voucher R.Y hlm1	AY786580	Yang,R. and Liu,J.
*P. crassifolia** voucher R.Y hlm2	AY786581	Yang,R. and Liu,J.
*P. crassifolia** voucher R.Y hlm3	AY786582	Yang,R. and Liu,J.
*P. crassifolia** voucher R.Y hlm4	AY786583	Yang,R. and Liu,J.
*P. crassifolia** voucher R.Y hlm5	AY786584	Yang,R. and Liu,J.
*P. crassifolia** isolate A	EF524073	Meng,L.-H., Yang,R., Richard,A., Georg,M., Hu,T.-H. and Liu,J.-Q.
*P. crassifolia** isolate B	EF524074	Meng,L.-H., Yang,R., Richard,A., Georg,M., Hu,T.-H. and Liu,J.-Q.
*P. crassifolia** isolate C	EF524075	Meng,L.-H., Yang,R., Richard,A., Georg,M., Hu,T.-H. and Liu,J.-Q.
*P. crassifolia** isolate D	EF524076	Meng,L.-H., Yang,R., Richard,A., Georg,M., Hu,T.-H. and Liu,J.-Q.

*P. crassifolia** isolate E	EF524077	Meng,L.-H., Yang,R., Richard,A., Georg,M., Hu,T.-H. and Liu,J.-Q.
P. engelmannii	AY153788	Germano-Presby,J., Klein,A.S. and Thorner,A.R.
P. glauca isolate 494	AY057955	Germano-Presby,J., Klein,A.S. and Thorner,A.R.
P. glauca isolate 64	AY057958	Germano-Presby,J., Klein,A.S. and Thorner,A.R.
P. jezoensis	AY153791	Germano-Presby,J., Klein,A.S. and Thorner,A.R
P. mariana isolate 3293	AY057953	Germano-Presby,J., Klein,A.S. and Thorner,A.R.
P. mariana isolate 4274	AY057954	Germano-Presby,J., Klein,A.S. and Thorner,A.R.
P. mariana isolate 4962	AY057956	Germano-Presby,J., Klein,A.S. and Thorner,A.R.
P. mariana isolate 5004	AY057957	Germano-Presby,J., Klein,A.S. and Thorner,A.R.
P. mexicana	AY153785	Germano-Presby,J., Klein,A.S. and Thorner,A.R.
P. omorika	AY153792	Germano-Presby,J., Klein,A.S. and Thorner,A.R.
*P. omorika** haplotype A	EU649707	Aleksic,J.M., Mengl,M. and Geburek,T.
*P. omorika** haplotype B	EU649708	Aleksic,J.M., Mengl,M. and Geburek,T.
*P. omorika** haplotype C	EU649708	Aleksic,J.M., Mengl,M. and Geburek,T.
*P. omorika** haplotype D	EU649710	Aleksic,J.M., Mengl,M. and Geburek,T.
*P. omorika** haplotype E	EU649711	Aleksic,J.M., Mengl,M. and Geburek,T.
*P. omorika** haplotype J	EU649712	Aleksic,J.M., Mengl,M. and Geburek,T.

P. pungens	AY153783	Germano-Presby,J., Klein,A.S. and Thorner,A.R.
P. rubens	AY057949	Germano-Presby,J., Klein,A.S. and Thorner,A.R.
isolate 2019		
P. rubens	AY057950	Germano-Presby,J., Klein,A.S. and Thorner,A.R.
isolate 2022		
P. rubens	AY057951	Germano-Presby,J., Klein,A.S. and Thorner,A.R.
isolate 2032		
P. rubens	AY057952	Germano-Presby,J., Klein,A.S. and Thorner,A.R.
isolate 2505		
P. schrenkiana	AY153782	Germano-Presby,J., Klein,A.S. and Thorner,A.R.
P. sitchensis	AY153787	Germano-Presby,J., Klein,A.S. and Thorner,A.R.
P. smithiana	AY153781	Germano-Presby,J., Klein,A.S. and Thorner,A.R.
P. wilsonii	AY153789	Germano-Presby,J., Klein,A.S. and Thorner,A.R.

* - species in which only partial nad1i477 sequences were available - those fragments can be amplified by primers described in Sperisen et al. (2001)

Table 2. The source material for nad5i230 - the list of species, their Gene Bank accession numbers and the authors

Species	Gene Bank access. n.	Author
Arabidopsis thaliana	X60045	Knoop,V., Schuster,W., Wissinger,B. and Brennicke,A.
Cycas panzhihuaensis	AF143425	Wang,X.Q., Tank,D.C. and Sang,T.
Keteleeria evelyniana	AF143418	Wang,X.Q., Tank,D.C. and Sang,T.
Cedrus atlantica	DQ983607	Qiao,C.Y., Ran,J.H., Li,Y. and Wang,X.Q.
Larix gmelini	AF143417	Wang,X.Q., Tank,D.C. and Sang,T.
Tsuga mertensiana	AF143421	Wang,X.Q., Tank,D.C. and Sang,T.
Pseudotsuga menziesii	AF143416	Wang,X.Q., Tank,D.C. and Sang,T.
Abies firma	AF143419	Wang,X.Q., Tank,D.C. and Sang,T.
Pinus banksiana	AF143412	Wang,X.Q., Tank,D.C. and Sang,T.
Pinus sylvestris	EU072471	Pyhajarvi,T., Salmela,M. and Savolainen,O.
Picea		
alcoquiana	DQ358170	Ran,J.-H., Wei,X.-X. and Wang,X.-Q.
asperata	DQ358171	Ran,J.-H., Wei,X.-X. and Wang,X.-Q.
brachytyla	DQ358172	Ran,J.-H., Wei,X.-X. and Wang,X.-Q.
breweriana	DQ358203	Ran,J.-H., Wei,X.-X. and Wang,X.-Q.
chihuahuana	DQ358173	Ran,J.-H., Wei,X.-X. and Wang,X.-Q.
chihuahuana	DQ415981	Ran,J.-H., Wei,X.-X. and Wang,X.-Q.
crassifolia	DQ358174	Ran,J.-H., Wei,X.-X. and Wang,X.-Q.
crassifolia isolate G	EF524078	Meng,L.-H., Yang,R., Richard,A., George,M., Hu,T.-H. and Liu,J.-Q.

crassifolia	EF524079	Ran,J.-H., Wei,X.-X. and Wang,X.-Q.
isolate E		
engelmannii	DQ358175	Ran,J.-H., Wei,X.-X. and Wang,X.-Q.
farreri	DQ358176	Ran,J.-H., Wei,X.-X. and Wang,X.-Q.
glauca	AY196183	Jaramillo-Correa,J., Bousquet,J., Beaulieu,J., Isabel,N., Perron,M. and Bouille,M.
glauca	AY196184	Jaramillo-Correa,J., Bousquet,J., Beaulieu,J., Isabel,N., Perron,M. and Bouille,M.
glauca	DQ358177	Ran,J.-H., Wei,X.-X. and Wang,X.-Q.
glehnii	DQ358178	Ran,J.-H., Wei,X.-X. and Wang,X.-Q.
jezoensis	DQ358179	Ran,J.-H., Wei,X.-X. and Wang,X.-Q.
koraiensis	DQ358180	Ran,J.-H., Wei,X.-X. and Wang,X.-Q.
koyamae	DQ358205	Ran,J.-H., Wei,X.-X. and Wang,X.-Q.
likiangensis	DQ358181	Ran,J.-H., Wei,X.-X. and Wang,X.-Q.
likiangensis	DQ358182	Ran,J.-H., Wei,X.-X. and Wang,X.-Q.
mariana	AY196179	Jaramillo-Correa,J., Bousquet,J., Beaulieu,J., Isabel,N., Perron,M. and Bouille,M.
mariana	AY196180	Jaramillo-Correa,J., Bousquet,J., Beaulieu,J., Isabel,N., Perron,M. and Bouille,M.
mariana	AY196182	Jaramillo-Correa,J., Bousquet,J., Beaulieu,J., Isabel,N., Perron,M. and Bouille,M.
mariana	DQ358183	Ran,J.-H., Wei,X.-X. and Wang,X.-Q.
martinezii	DQ415982	Jaramillo-Correa,J., Beaulieu,J., Ledig,F.T. and Bousquet,J.
maximowiczii	DQ358184	Ran,J.-H., Wei,X.-X. and Wang,X.-Q.
meyeri	DQ358186	Ran,J.-H., Wei,X.-X. and Wang,X.-Q.
morrisonicola	DQ358185	Ran,J.-H., Wei,X.-X. and Wang,X.-Q.
neoveitchii	DQ358187	Ran,J.-H., Wei,X.-X. and Wang,X.-Q.
obovata	DQ358188	Ran,J.-H., Wei,X.-X. and Wang,X.-Q.
omorika	DQ358189	Ran,J.-H., Wei,X.-X. and Wang,X.-Q.
orientalis	DQ358190	Ran,J.-H., Wei,X.-X. and Wang,X.-Q.

pungens	DQ358191	Ran,J.-H., Wei,X.-X. and Wang,X.-Q.
purpurea	DQ358192	Ran,J.-H., Wei,X.-X. and Wang,X.-Q.
retroflexa	DQ358194	Ran,J.-H., Wei,X.-X. and Wang,X.-Q.
rubens	AY196181	Jaramillo-Correa,J., Bousquet,J., Beaulieu,J., Isabel,N., Perron,M. and Bouille,M.
rubens	DQ358195	Ran,J.-H., Wei,X.-X. and Wang,X.-Q.
schrenkiana	DQ358196	Ran,J.-H., Wei,X.-X. and Wang,X.-Q.
schrenkiana	DQ358197	Ran,J.-H., Wei,X.-X. and Wang,X.-Q.
sitchensis	DQ358198	Ran,J.-H., Wei,X.-X. and Wang,X.-Q.
smithiana	AF143414	Wang,X.Q., Tank,D.C. and Sang,T.
smithiana	DQ358199	Ran,J.-H., Wei,X.-X. and Wang,X.-Q.
spinulosa	DQ358204	Ran,J.-H., Wei,X.-X. and Wang,X.-Q.
torano	DQ358200	Ran,J.-H., Wei,X.-X. and Wang,X.-Q.
wilsonii	DQ358201	Ran,J.-H., Wei,X.-X. and Wang,X.-Q.
wilsonii	DQ358202	Ran,J.-H., Wei,X.-X. and Wang,X.-Q.

4 Conclusion

Two European spruce species – *P. abies* and *P. omorika,* differ not just in relation to the range of their natural distributions, but also in relation to the available knowledge and the intensity with which those species are studied. While *P. abies* is widely distributed throughout Europe and is one of the most studied spruce species, *P. omorika* is endemic to an extremely small area of approx. 10.000 km^2 in western Bosnia and Herzegovina and eastern Serbia and still it can be referred to as a poorly known species, despite the fact that more than 600 articles about this species have been published during the last 100 years or so. It seems that one can not give an unambiguous answer even to some fundamental questions, such as:

- Is *P. omorika* morphologically and genetically uniform or it is as diverse as other widely distributed spruce species?
- What is the reproductive strategy in this species, and how effectively does it distribute its genes through pollen and seed?
- Are selfs and inbreds abundant in nature due to the high self-fertility or they are eliminated as effectively as in other conifers?
- Is this species closely related to Asian or American spruce species?
- Are current conservation measures adequate for the preservation of the remaining genetic resources in this species, or natural regeneration and range expansion in Serbian spruce can be improved?

The genetic diversity in *P. omorika* based on both nuclear and mitochondrial loci, is found to be comparable to that observed in other widely distributed spruce species. This finding is rather surprising, especially at mtDNA locus nad1i477, due to the fact that the mitochondrial genome is characterized by the lowest rate of sequence evolution of all three plant genomes and this locus was found to be monomorphic in other spruce species with limited natural range. Therefore, based on two plant genomes, *P. omorika* cannot be characterized as a genetically depauperate species, despite being a narrow endemic. Furthermore, previous reports on morphological uniformity in this species are probably due to limited sample sizes, because variability in several morphological traits has been reported as well.

On the other hand, the utilization of the mtDNA locus nad1i477 highlighted the importance of the sampling scheme in assessing genetic diversity in species *per se*. As shown in this dissertation, two separate gene pools are detected in *P. omorika* natural populations and only a single, small population is found to represent a second gene pool. When this population was omitted from the analysis, the genetic diversity parameters in

remaining natural populations indicated genetic uniformity in this species. Therefore, in cases of formerly widely distributed species currently confined to small areas, exhaustive sampling is recommended when assessing genetic diversity, because small populations or even patches of trees might harbour important information on genetic diversity in species *per se*.

Reconstructions of the secondary structure in mitochondrial group II intron nad1i477 enabled the recognition of the microstructural changes leading to the occurrence of two large indels (up to 1.300 bp) at this locus, implying that a similar mechanism might operate at other mtDNA non-coding regions. The recognition of mutational mechanisms (e.g. duplications and indels up to 30 bp) leading to the length variability at both mtDNA loci – nad1i477 and nad5i230, improved the assessment of primary homology and consequently, the phylogenetic inferences, while incorporating phylogenetic signal in length mutations by coding indels either by simple or complex indel coding improved the resolution, especially at nad1i477 in which the length mutations are found to play a major role in sequence evolution.

Based on two mtDNA loci, *P. omorika* is found to be closely related to *P. abies* and other Asian spruces characterized as a group A spruce species, rather than to *P. smithiana*, *P schrenkiana* and all American spruces, characterized as group B spruce species. Based on several lines of evidence, such as morphological features, current and past distribution and the mitochondrial genome, it is highly unlikely that *P. omorika* is closely related to *P. mariana* and *P. rubens,* despite reports on successful crosses between those species and their close relations reported in two recently published molecular phylogenies of spruces. Therefore, it seems that crossability studies have to be interpreted with cautions, as well as molecular phylogenies based on loci in which the homology assessment is ambiguous.

This dissertation reveals, at least partially, the complex structuring of genetic diversity within the Balkan – the biodiversity hotspot, frequently referred to as a poorly known region. In trees with long life spans, the current genetic structuring might still reflect the pattern and the speed of the past migrations, and the maternally inherited mitochondrial genome is one of the most powerful tools in studying phylogeography in plants. As shown in *P. omorika*, even within an area of approx. 10.000 km^2, natural populations in this species can be assigned to two different gene pools characterized by different evolutionary history and past migrations, suggesting that similar patterns might be expected in other endemic and/or relict species confined to this region. The relict tree populations confined to the Mediterranean Basin (and consequently, the Balkan) are considered as an '*evolutionary heritage of disproportionate significance for the conservation of European plant diversity'* (Petit et al. 2005). The remaining natural populations of *P. omorika* certainly belong to this category.

In Serbia, the majority of *P. omorika* natural populations are confined to the National Park 'Tara' and they are characterized as untouched forests left for the free development.

They correspond to the category I of the Protected Area Management Categories of IUCN and to the category 1.1 according to the MCPFE classes of protected and protective forests. The main objective of conservation in such areas is biodiversity conservation with no active intervention. Other natural populations and tree clusters are in private property and are protected by the law. However, recently, the introduction of conservation through minimum intervention (IUCN category II, MCPFE category 1.2), at least at some sites, is proposed due to the finding that natural regeneration can be improved by the occasional removal of individual trees of *P. omorika*'s main competitors, such as Norway spruce, fir and beech. The results of this dissertation corroborate this view, especially because until now the available data on genetic structure and gene dispersal in this species were either unknown or contradictory. The finding of extraordinary limited seed flow in *P. omorika* implies that for the successful expansion of the species, minimum intervention management measures are necessary.

As already mentined, the presence of two gene pools in *P. omorika* is revealed in this dissertation. The first gene pool, confined to the Mt. Tara and the neighboring region was probably not exposed to large geographical displacements, while the another gene pool is assumed to be exposed to frequent migrations due to the climate oscillations. In the past, the individuals from this gene pool were probably abundant in south and south-east regions in relation to the *P. omorika* current range and they were probably present at as distant areas as the Aegean coast. Population 'Studenac', the representative of the second gene pool, has been legally protected as late as 1993. This population is comprised of only 419 trees and due to the fact that such small forests can be easily erased by fire or other catastrophes, its loss would results in a substantial decrease in genetic diversity in species *per se*. Therefore, the recommendation for the preservation of the remaining genetic resources in *P. omorika*, which are already narrowed by the past historical events, is to introduce conservation through minimum intervention at some sites and to use all trees from the population 'Studenac' for the establishment of an *ex situ* collection.

5 Appendix

Nuclear EST-SSR

NCBI accession number: CN480896

Locus: WSOO19.F22

Forward primer: 5’ - AAGCGTTTCTCATTTTCTTGG - 3’

Reverse primer: 5’ - GGGCCCAGAACTAACAATGA - 3’

Repeat type: dinucleotide repeat

Repeat motif: AT

Flanking region: 349 bp

GGGCCCAGAACTAACAATGAAATTCATTAAAAAAACTTTTGAGTAAGAACACATTGTCAG
GTGTAAAGCATTTCCCAAAACAACTTGAGATATGCCAAAGCTCAAGTTACACCCGGTGTT
GACATATAATTAGCTTTAACTATATGAAAGGAGAATTAACAATTTCACACAATTCTGGTAA
TTCTGCATATTGAATATATATATATATATATATATATATAACTTGTACTATGACACCTATCT
TACTCATGCCTTTTTCTTGTGCTTCAGAGGCAGAAGGAAAATGAATAACAAGATGAAGAG
AAAAGCGAAAACATATTACATCCAGAAATATTAATAAAAACAAGAATAGTGCCAAGAAAA
TGAGAAACGCTT

Nuclear EST-SSR

NCBI accession number: CN480899

Locus: WS0022.B15

Forward primer: 5’ - TTTGTAGGTGCTGCAGAGATG -3’

Reverse primer: 5’ - TGGCTTTTTATTCCAGCAAGA - 3’

Repeat type: dinucleotide repeat

Repeat motif: AG

Flanking region: 160 bp

TGGCTTTTTATTCCAGCAAGAAATGAAAAATAACATGATATATTATGGAGCTTTTTTCATG
AAGCTGACCTAATATGTGCCCTCAACTTCACCTGAAGAGAGAGAGAGAGAGAGAGAGA
GTGCATAAAAAGAATATGGTCAGATTTTATGCAGATTTATATTACATCTCTGCAGCACCTA
CAAA

Nuclear EST-SSR

NCBI accession number: CN480900

Locus: WS0023.B03

Forward primer: 5’ - AGCAGCTGGGGTCAAAGTT - 3’

Reverse primer: 5’ - AAAGAAAGCATGCATATGACTCAG - 3’

Repeat type: dinucleotide repeat

Repeat motif: AT

Flanking region: 156 bp

AAAGAAAGCATGCATATGACTCAGTTAAAATCACTTTTAGCTACCTCAGCTGCATCAGTC
AAGCTTCCTTTATAAAAGGGCAACCTAAGTTCATATATATATATATATATATAAAGTGCAA
CCAGAATCTCATATTTATCTTATCCAAGCATTACGCAACTTTGACCCCAGCTGCT

Nuclear EST-SSR

NCBI accession number: CN480898

Locus: WS0053.K16

Forward primer: 5’ - ACATATCATGGTTGCGATGC - 3’

Reverse primer: 5’ - CCACAGCCCCTAAAATGTGA - 3’

Repeat type: dinucleotide repeat

Repeat motif: AT

Flanking region: 200 bp

ACATATCATGGTTGCGATGCGAAGTTTTCAGACGTGTAACAATACAATATTTTCAAATTAT
AAGATATATATATATATATATATATATATACCGTATAAAATCGTGGAGAACATAGTGTGTT
CTCAAATCGCATCTCTATATTCTAAGCAATCTTAGATCAACTCCCACTATAGTCCTTACAT
CAAGCTCACACTAAGCTTTTACTTCACATTTTAGGGGCTGTGG

Nuclear EST-SSR

NCBI accession number: CN480903

Locus: WS0073.H08

Forward primer: 5’ - TGCTCTCTTATTCGGGCTTC - 3’

Reverse primer: 5’ - AAGAACAAGGCTTCCCAATG - 3’

Repeat type: dinucleotide repeat

Repeat motif: AT

Flanking region: 193 bp

TGCTCTCTTATTCGGGCTTCATAGCACAAAAGTAGCATTATCCTAATACATGGAACAAGT
TCAATGGAAGCTCGGTATATTATACAAATCAATATATATCTATATATAGATATATATATATA
TATATATATATATATATGAATAAATGATGATTTATTCTTGAGAACCAGACCTGACTTCAGAAT
CCAGATCAACACAATAAACATTGGGAAGCCTTGTTCTT

Mitochondrial haplotype A
NCBI accession number: EU649707
Authors: Aleksić JM, Mengl M, Geburek T

Forward primer: 5’ - CTCTCCCTCACCCATATGATG - 3’

Reverse primer: 5’ - AGATCCCCATATATTCCCGG - 3’

Repeat type: 34 bp minisatellite
Repeat motif: 5’ - CCCCCTCCGTTGTCAGGGGAGCGACTTCGTACCT - 3’
Number of repeats: 1
Total length: 787 bp

CTCTCCCTCACCCATATGATGAGTGAGCGTTAACACCCTCACCCATGAATGGATGAGCG
ACTTCGTTCCTCGCCCCTTTTGGTCGAGTCACTTAACGTACCTCGCCTTCCTACCTCAG
GCGAGTCACTTAACGTACCTCGCCCTTTTTGGGTCGAGTCACTTAACGTACCTCACCCA
TCGGATGGATGAGCGACTTCGTACCTCCCCCTCCGTTGTCAGGGGAGCGACTTCGTAC
CTTTTGGTCGAGCGTTCATTTATCACCCTTTATTTCTTCCCCTCAAAGGGGCTTTGTAAG
GACTTGTCTGGAGATATCTGTCTCACTAAGCGGTAACCTCGCTACCTCACCCCCTAACG
ACTGCTCTATCTTAGTGTTCGTTCGCGCGACAAGGATACAATAAAGAGCATTTATGCGCT
CGACCGACGTGAAATTCATAGTCTTAGGAGAGGACACAACACCTTAGAACTTGAACATT
TTCACATCCTTTCAGGACCTTTTAATGAAGGGAAGCCCCGTAGCGAAATCCCTTTCCTTT
GCGTAGCTCAGGTAAGGTTATCCTTACCTTCGCGTTCAGGCTTTGAGGGGAAGAAATAT
AGGGACCTTCCCTTCGCTATGCGTAAGACCTAGACCACCGTAGCGAAGGGAAGGCAAG
ATCTATAGTTACAGAATTACGGCACTATGGATCGATCTACGTTGCCGTAACAAATTCACA
AATGAATTCTAGATCTAGATAGTGCCCAATTTAGAACCTTGCGGAGGGACCGTAGCTCC
GGGAATATATGGGGATCT

Mitochondrial haplotype B

NCBI accession number: EU649708

Authors: Aleksić JM, Mengl M, Geburek T

Forward primer: 5’ - CTCTCCCTCACCCATATGATG - 3’

Reverse primer: 5’ - AGATCCCCATATATTCCCGG - 3’

Repeat type: 34 bp minisatellite

Repeat motif: 5’ - CCCCCTCCGTTGTCAGGGGAGCGACTTCGTACCT - 3’

Number of repeats: 2

Total length: 821 bp

CTCTCCCTCACCCATATGATGAGTGAGCGTTAACACCCTCACCCATGAATGGATGAGCG
ACTTCGTTCCTCGCCCCTTTTGGTCGAGTCACTTAACGTACCTCGCCTTCCTACCTCAG
GCGAGTCACTTAACGTACCTCGCCCTTTTTGGGTCGAGTCACTTAACGTACCTCACCCA
TCGGATGGATGAGCGACTTCGTACCTCCCCCTCCGTTGTCAGGGGAGCGACTTCGTAC
CTCCCCCTCCGTTGTCAGGGGAGCGACTTCGTACCTTTTGGTCGAGCGTTCATTTATCA
CCCTTTATTTCTTCCCCTCAAAGGGGCTTTGTAAGGACTTGTCTGGAGATATCTGTCTCA
CTAAGCGGTAACCTCGCTACCTCACCCCCTAACGACTGCTCTATCTTAGTGTTCGTTCG
CGCGACAAGGATACAATAAAGAGCATTTATGCGCTCGACCGACGTGAAATTCATAGTCT
TAGGAGAGGACACAACACCTTAGAACTTGAACATTTTCACATCCTTTCAGGACCTTTTAA
TGAAGGGAAGCCCCGTAGCGAAATCCCTTTCCTTTGCGTAGCTCAGGTAAGGTTATCCT
TACCTTCGCGTTCAGGCTTTGAGGGGAAGAAATATAGGGACCTTCCCTTCGCTATGCGT
AAGACCTAGACCACCGTAGCGAAGGGAAGGCAAGATCTATAGTTACAGAATTACGGCAC
TATGGATCGATCTACGTTGCCGTAACAAATTCACAAATGAATTCTAGATCTAGATAGTGC
CCAATTTAGAACCTTGCGGAGGGACCGTAGCTCCGGGAATATATGGGGATCT

Mitochondrial haplotype C

NCBI accession number: EU649709

Authors: Aleksić JM, Mengl M, Geburek T

Forward primer: 5’ - CTCTCCCTCACCCATATGATG - 3’

Reverse primer: 5’ - AGATCCCCATATATTCCCGG - 3’

Repeat type: 34 bp minisatellite

Repeat motif: 5’ - CCCCCTCCGTTGTCAGGGGAGCGACTTCGTACCT -3’

Number of repeats: 3

Total length: 855 bp

CTCTCCCTCACCCATATGATGAGTGAGCGTTAACACCCTCACCCATGAATGGATGAGCG
ACTTCGTTCCTCGCCCCTTTTGGTCGAGTCACTTAACGTACCTCGCCTTCCTACCTCAG
GCGAGTCACTTAACGTACCTCGCCCTTTTTGGGTCGAGTCACTTAACGTACCTCACCCA
TCGGATGGATGAGCGACTTCGTACCTCCCCCTCCGTTGTCAGGGGAGCGACTTCGTAC
CTCCCCCTCCGTTGTCAGGGGAGCGACTTCGTACCTCCCCCTCCGTTGTCAGGGGAGC
GACTTCGTACCTTTTGGTCGAGCGTTCATTTATCACCCTTTATTTCTTCCCCTCAAAGGG
GCTTTGTAAGGACTTGTCTGGAGATATCTGTCTCACTAAGCGGTAACCTCGCTACCTCA
CCCCCTAACGACTGCTCTATCTTAGTGTTCGTTCGCGCGACAAGGATACAATAAAGAGC
ATTTATGCGCTCGACCGACGTGAAATTCATAGTCTTAGGAGAGGACACAACACCTTAGA
ACTTGAACATTTTCACATCCTTTCAGGACCTTTTAATGAAGGGAAGCCCCGTAGCGAAAT
CCCTTTCCTTTGCGTAGCTCAGGTAAGGTTATCCTTACCTTCGCGTTCAGGCTTTGAGG
GGAAGAAATATAGGGACCTTCCCTTCGCTATGCGTAAGACCTAGACCACCGTAGCGAAG
GGAAGGCAAGATCTATAGTTACAGAATTACGGCACTATGGATCGATCTACGTTGCCGTA
ACAAATTCACAAATGAATTCTAGATCTAGATAGTGCCCAATTTAGAACCTTGCGGAGGGA
CCGTAGCTCCGGGAATATATGGGGATCT

Mitochondrial haplotype D
NCBI accession number: EU649710
Authors: Aleksić JM, Mengl M, Geburek T

Forward primer: 5’ - CTCTCCCTCACCCATATGATG - 3’

Reverse primer: 5’ - AGATCCCCATATATTCCCGG - 3’

Repeat type: 34 bp minisatellite
Repeat motif: CCCCCTCCGTTGTCAGGGGAGCGACTTCGTACCT
Number of repeats: 4
Total length: 889 bp

CTCTCCCTCACCCATATGATGAGTGAGCGTTAACACCCTCACCCATGAATGGATGAGCG
ACTTCGTTCCTCGCCCCTTTTGGTCGAGTCACTTAACGTACCTCGCCTTCCTACCTCAG
GCGAGTCACTTAACGTACCTCGCCCTTTTTGGGTCGAGTCACTTAACGTACCTCACCCA
TCGGATGGATGAGCGACTTCGTACCTCCCCCTCCGTTGTCAGGGGAGCGACTTCGTAC
CTCCCCCTCCGTTGTCAGGGGAGCGACTTCGTACCTCCCCCTCCGTTGTCAGGGGAGC
GACTTCGTACCTCCCCCTCCGTTGTCAGGGGAGCGACTTCGTACCTTTTGGTCGAGCGT
TCATTTATCACCCTTTATTTCTTCCCCTCAAAGGGGCTTTGTAAGGACTTGTCTGGAGAT
ATCTGTCTCACTAAGCGGTAACCTCGCTACCTCACCCCCTAACGACTGCTCTATCTTAGT
GTTCGTTCGCGCGACAAGGATACAATAAAGAGCATTTATGCGCTCGACCGACGTGAAAT
TCATAGTCTTAGGAGAGGACACAACACCTTAGAACTTGAACATTTTCACATCCTTTCAGG
ACCTTTTAATGAAGGGAAGCCCCGTAGCGAAATCCCTTTCCTTTGCGTAGCTCAGGTAA
GGTTATCCTTACCTTCGCGTTCAGGCTTTGAGGGGAAGAAATATAGGGACCTTCCCTTC
GCTATGCGTAAGACCTAGACCACCGTAGCGAAGGGAAGGCAAGATCTATAGTTACAGAA
TTACGGCACTATGGATCGATCTACGTTGCCGTAACAAATTCACAAATGAATTCTAGATCT
AGATAGTGCCCAATTTAGAACCTTGCGGAGGGACCGTAGCTCCGGGAATATATGGGGA
TCT

Mitochondrial haplotype E
NCBI accession number: EU649711
Authors: Aleksić JM, Mengl M, Geburek T

Forward primer: 5’ - CTCTCCCTCACCCATATGATG - 3’

Reverse primer: 5’ - AGATCCCCATATATTCCCGG - 3’

Repeat type: 34 bp minisatellite
Repeat motif: 5’ - CCCCCTCCGTTGTCAGGGGAGCGACTTCGTACCT - 3’
Number of repeats: 5
Total length: 923 bp

CTCTCCCTCACCCATATGATGAGTGAGCGTTAACACCCTCACCCATGAATGGATGAGCG
ACTTCGTTCCTCGCCCCTTTTGGTCGAGTCACTTAACGTACCTCGCCTTCCTACCTCAG
GCGAGTCACTTAACGTACCTCGCCCTTTTTGGGTCGAGTCACTTAACGTACCTCACCCA
TCGGATGGATGAGCGACTTCGTACCTCCCCCTCCGTTGTCAGGGGAGCGACTTCGTAC
CTCCCCCTCCGTTGTCAGGGGAGCGACTTCGTACCTCCCCCTCCGTTGTCAGGGGAGC
GACTTCGTACCTCCCCCTCCGTTGTCAGGGGAGCGACTTCGTACCTCCCCCTCCGTTGT
CAGGGGAGCGACTTCGTACCTTTTGGTCGAGCGTTCATTTATCACCCTTTATTTCTTCCC
CTCAAAGGGGCTTTGTAAGGACTTGTCTGGAGATATCTGTCTCACTAAGCGGTAACCTC
GCTACCTCACCCCCTAACGACTGCTCTATCTTAGTGTTCGTTCGCGCGACAAGGATACA
ATAAAGAGCATTTATGCGCTCGACCGACGTGAAATTCATAGTCTTAGGAGAGGACACAA
CACCTTAGAACTTGAACATTTTCACATCCTTTCAGGACCTTTTAATGAAGGGAAGCCCCG
TAGCGAAATCCCTTTCCTTTGCGTAGCTCAGGTAAGGTTATCCTTACCTTCGCGTTCAGG
CTTTGAGGGGAAGAAATATAGGGACCTTCCCTTCGCTATGCGTAAGACCTAGACCACCG
TAGCGAAGGGAAGGCAAGATCTATAGTTACAGAATTACGGCACTATGGATCGATCTACG
TTGCCGTAACAAATTCACAAATGAATTCTAGATCTAGATAGTGCCCAATTTAGAACCTTG
CGGAGGGACCGTAGCTCCGGGAATATATGGGGATCT

Mitochondrial haplotype J

NCBI accession number: EU649712

Authors: Aleksić JM, Mengl M, Geburek T

Forward primer: 5’ - CTCTCCCTCACCCATATGATG - 3’

Reverse primer: 5’ - AGATCCCCATATATTCCCGG - 3’

Repeat type: 34 bp minisatellite

Repeat motif: 5’ - CCCCCTCCGTTGTCAGGGGAGCGACTTCGTACCT - 3’

Number of repeats: 10

Total length: 1093 bp

CTCTCCCTCACCCATATGATGAGTGAGCGTTAACACCCTCACCCATGAATGGATGAGCG
ACTTCGTTCCTCGCCCCTTTTGGTCGAGTCACTTAACGTACCTCGCCTTCCTACCTCAG
GCGAGTCACTTAACGTACCTCGCCCTTTTTGGGTCGAGTCACTTAACGTACCTCACCCA
TCGGATGGATGAGCGACTTCGTACCTCCCCCTCCGTTGTCAGGGGAGCGACTTCGTAC
CTCCCCCTCCGTTGTCAGGGGAGCGACTTCGTACCTCCCCCTCCGTTGTCAGGGGAGC
GACTTCGTACCTCCCCCTCCGTTGTCAGGGGAGCGACTTCGTACCTCCCCCTCCGTTGT
CAGGGGAGCGACTTCGTACCTCCCCCTCCGTTGTCAGGGGAGCGACTTCGTACCTCCC
CCTCCGTTGTCAGGGGAGCGACTTCGTACCTCCCCCTCCGTTGTCAGGGGAGCGACTT
CGTACCTCCCCCTCCGTTGTCAGGGGAGCGACTTCGTACCTCCCCCTCCGTTGTCAGG
GGAGCGACTTCGTACCTTTTGGTCGAGCGTTCATTTATCACCCTTTATTTCTTCCCCTCA
AAGGGGCTTTGTAAGGACTTGTCTGGAGATATCTGTCTCACTAAGCGGTAACCTCGCTA
CCTCACCCCCTAACGACTGCTCTATCTTAGTGTTCGTTCGCGCGACAAGGATACAATAA
AGAGCATTTATGCGCTCGACCGACGTGAAATTCATAGTCTTAGGAGAGGACACAACACC
TTAGAACTTGAACATTTTCACATCCTTTCAGGACCTTTTAATGAAGGGAAGCCCCGTAGC
GAAATCCCTTTCCTTTGCGTAGCTCAGGTAAGGTTATCCTTACCTTCGCGTTCAGGCTTT
GAGGGGAAGAAATATAGGGACCTTCCCTTCGCTATGCGTAAGACCTAGACCACCGTAG
CGAAGGGAAGGCAAGATCTATAGTTACAGAATTACGGCACTATGGATCGATCTACGTTG
CCGTAACAAATTCACAAATGAATTCTAGATCTAGATAGTGCCCAATTTAGAACCTTGCGG
AGGGACCGTAGCTCCGGGAATATATGGGGATCT

6 Table of abbreviations

A - adenine

BLAST - Basic Local Alignment Search Tool

bp – base pairs

C – cytosine

°C - degree Celsius

cf. - confer

CI - consistency index

CIC – complex indel coding

cm – centimetre

cpDNA – chloroplast deoxyribonucleic acid

DNA – deoxyribonucleic acid

dNTP - deoxyribonucleotide triphosphate

EST – expressed sequence tags

e.g. – exempli gratia

et al. – et alii

G – guanosine

ΔG - free energy

ha – hectare

HWE - Hardy-Weinberg expectations

IAM – infinite allele model

i.e. – id est

ITS – internal transcribed spacer

IUCN – the International Union for Conservation of Nature

indel – insertion/deletion

kb - kilo bases

kg – kilogram

ky - kilo years

LD - linkage disequilibrium

m – metre

mDNA - messenger deoxyribonucleic acid

MCIC - Modified Complex Indel Coding

MCPFE – Ministerial Conference on the Protection of Forests in Europe

min - minute

mm – millimetre

mM - mili Mole

mtDNA – mitochondrial deoxyribonucleic acid

MYA - million years ago

µm – micrometer

µM - micro Mole

NADH dehydrogenase - nicotineamide adenine dinucleotide dehydrogenase

ng - nanogram

nrDNA – nuclear ribosomal deoxyribonucleic acid

PCR - Polymerase Chain Reaction

pers. comm. - personal communication

RFLP - Restriction Fragment Length Polymorphism

RI - retention index

s – second

SANU – Serbian Academy of Sciences and Arts

SIC – simple indel coding

SMM – stepwise mutation model

SSR – simple sequence repeat

syn. – synonym

T – thymine

TBR branch-swapping - tree-bisection-reconnection branch-swapping

U/µl - units per microliter

7 Curriculum Vitae

Personal data

Name	Jelena M Aleksić
Date of Birth	07 11 1970
Place of Birth	Belgrade
Country	Serbia
Nationality	Serbian
University	University of Natural Resources and Applied Life Sciences Vienna

Education

2008 - Ph.D. at University of Natural Resources and Applied Life Sciences, Vienna, Austria as a holder of the Bioversity International fellowship on forest genetic resources and OEAD grant student

2005 - M.Sc. at Faculty of Agriculture, Department of Genetics and Plant Breeding, University of Novi Sad (not finalized through the public defense due to the onset of the Bioversity International fellowship) - acknowledged as PhD due to the implementation of the Bologna system of education in Serbia

Topic: *Ex situ* conservation of genetic resources of elms - *Ulmus minor* Mill. and *Ulmus glabra* Pall.

1999 - Dipl. Biol. at Faculty of Biology, Department of Plant Physiology, University of Belgrade - acknowledged as M.Sc. due to the implementation of the Bologna system of education in Serbia

Topic: *In vitro* vegetative propagation of *Digitalis lanata* Ehrh.

Working languages

English (very good)
Russian (very good)
Serbian and other Slavic languages

Professional experience

Work in the Laboratory for tissue culture for 7 years

Teaching assistant on subject Plant Physiology at the Department of Plant Physiology, Faculty of Biology, University of Belgrade during academic year 1999/2000

Work at the Institute of Forestry, Belgrade (2000-2002) and at the Institute of Lowland Forestry and Environment (2002-2004) as a holder of a scholarship of the Ministry of Science and Environment of Republic of Serbia

Employed as a research assistant at the Institute of Lowland Forestry and Environment in July 2004

Teaching assistant on subject Ornamental Plant Breeding for three years (since 2002/2003) at the Department of Horticulture, Faculty of Agriculture, University of Novi Sad

Work in the Laboratory for the DNA analysis at BFW, Vienna since July 2005

Published papers (in Serbia):

1. Radojević Lj., Jevremović S. and Aleksić J., 1998: Stimulation of shoot propagation of *Digitalis lanata* Ehrh. by use of thidiazuron in *in vitro* tissue culture. Pharmaceutical archive 6: 968-969.

2. Kovačević B., Orlović S. and Aleksić J., 2002: Possibilities of application of biotechnology in forestry. Symposium on biotechnology in Vojvodina. Proceedings from scientific meeting: 127-131.

3. Aleksić Jelena and Orlović Saša, 2003: Micropropagation of elms (*Ulmus* spp.). XV Symposium of Yugoslav Society of Plant Physiologists, Vrdnik. Programme and abstracts: 81-81.

4. Aleksić J., Orlović S., Pilipović A. and Pekeč S., 2003: Conservation of elms in Yugoslavia. International elm conference, Valsain, Spain. Programme and Abstracts: 61-61.

5. Aleksić J. and Orlović S., 2005: Conservation of genetic resources of elms (*Ulmus* spp.). Savremena poljoprivreda 54 (3-4): 1-5.

6. Aleksić J. and Orlović S., 2004: *Ex situ* conservation of genetic resources of *Ulmus minor* Mill. and *U. laevis* Pall. Genetika, 36 (3): 221-227.

Participation in scientific meetings:

1. Second Congress of Yugoslav Pharmaceuticsts with International Participation, October 1998, Belgrade, Yugoslavia.
2. Symposium on biotechnology in Vojvodina, September 12 – 13, 2002, Novi Sad, Yugoslavia.
3. XV Symposium of Yugoslav Society of Plant Physiologists, May 31 – June 3, 2003, Vrdnik, Yugoslavia.
4. International conference on sustainable agriculture and European integration processes, September 19 – 24, 2004, Novi Sad, Serbia and Montenegro.
5. Third Congress of Serbian Geneticists, Subotica, November 30-December 4, 2004, Subotica, Serbia and Montenegro.
6. IPGRI Training Workshop on Forest Biodiversity, 13–24 June 2005, Pushkino, Russian Federation.

Vienna July 2008

Dissertations and dissertation print of the University of Natural Resources and Applied Life Sciences Vienna

Module 1: Dissertations

Imprint of this document:

Elßlergasse 17, A-1130 Vienna
Phone. +43 (0)1 877 04 26, Fax: +43 (0)1 876 40 04
Dr.-Simoneit-Straße 36, D-45473 Mülheim a. d. Ruhr
E-Mail: verlag@guthmann-peterson.de
http://www.guthmann-peterson.de