UNIVERSITY OF AGRICULTURAL SCIENCES AND VETERINARY MEDICINE CLUJ-NAPOCA
DOCTORAL SCHOOL
FACULTY OF HORTICULTURE

Eng. Valentina FLORAN

SUMMARY OF PhD THESIS

Spatial genetic analysis on Scots pine (*Pinus sylvestris* L.) natural populations

Scientific Supervisor:
Prof. dr. Radu SESTRAŞ

Cluj-Napoca
2011

SUMMARY

The PhD thesis entitled "**Spatial genetic analysis on Scots pine (*Pinus sylvstris* L.) natural populations"** tries to reconstruct the post glacial colonization routes of an economically important and widely distributed conifer, Scots pine.

Natural selection, mutation, recombination, demographic history and chance, all have a role in evolution. In natural populations, the outcome of these forces is seen as adaptations, differences between geographic varieties, and as genetic diversity in populations—both at the phenotypic and molecular levels. In this thesis I wanted to examine the roles of the evolutionary forces shaping molecular genetic diversity in trees, with emphasis on a boreal conifer, Scots pine (*Pinus sylvestris*).

Pinus sylvestris (Scots pine) is the most widely distributed Eurasian conifer. It is found in a range of environments, soils, and climates from arid, mountainous areas of Spain and Turkey to subarctic forests of northern Scandinavia and Siberia. A large part of its current distribution was covered by the continental ice or was otherwise uninhabitable during the last glacial maximum (LGM) which occurred about 20,000 years ago (WILLIS et al. 1998; SVENDSEN et al. 1999). During the LGM, Scots pine was present in the ice-free regions of Southern and Central Europe, but it is also possible that it survived in Siberia and in more eastern parts of Europe (WILLIS and van ANDEL 2004). After the LGM, Scots pine expanded to new areas as the habitats became more suitable for it. Based on the European pollen records of the genus *Pinus*, its distribution started to expand about 16,000 years ago in Southern European regions, and it reached northernmost Scandinavia 7,800 years ago (WILLIS et al. 1998). However, little is known about the origin of the populations that colonized the northern parts of Europe.

Phylogeographic history and past population size changes have a dominant role in molecular diversity of *P. sylvestris*. The effect of the Last Glacial Maximum (37 000–16 000) was observed in the distribution of mitochondrial DNA variation. Not much evidence of positive natural selection was found in pines or trees in general. This is in contrast to strong natural selection that is observed at the phenotypic level. Positive selection is difficult to prove, especially when the genome is still affected by demographic history. Mutation–drift equilibrium may rarely be reached in tree populations.

The paper highlights the present knowledge of Scots pine (*Pinus sylvestris* L.) diversity, historical and geographical distribution, based on mitochondrial and chloroplast DNA data. The observed differences in the estimates of genetic differentiation between different types of genomes suggest that both pollen and seed contribute significantly to gene flow within species. Organelles' diversity represents an important criterion which could be later applied in planning for future forest management and breeding through a better understanding of adaptation strategies of different Scots pine haplotypes. Research on organelles' diversity could lead to important practical applications in areas such as traceability and eco-certification of forest products, and the identification of plant populations for conservation. This analysis would provide valuable references when facing current day problems with climate change, species adaptation, and loss of forest with negative effects on biodiversity.

The cold periods of the Pleistocene had a dramatic impact on most of the species in temperate regions (Webb 1992) responding through migrations to regions where climatic condition allowed them to survive (Taberlet et al. 1998). Spatial information is an important element to be considered in order to understand genetic resources (Heywood 1991), habitat connectivity and distribution of the organisms.

A major interest on understanding the spatial genetic structure on Scots pine is motivated by the importance of such knowledge in tree breeding and conservation. Therefore, a lot of effort has been devoted to study the spatial genetic structure in Scots pine (e.g., Petit et al 2004; Pyhäjärvi et al 2008; Naydenov et al. 2007), however, most of those

works cover large extensions of Scots pine distribution at the cost of having a very dispersed sampling and relatively small sample sizes per site.

The European pine pollen fossil records are exhaustive (HUNTLEY and BIRKS 1983), but for the following reasons, they only provide an approximation of the species distribution. First, pollen disperses long distances, and second, pollen production tends to be low in severe environments (SARVAS 1962). Macro- and mega fossil data, which are considered to be more reliable than pollen data, suggest that *P. sylvestris* survived in Hungary and the Czech Republic more than 30,000 years ago (reviewed by WILLIS and van ANDEL 2004). Based on genetic data, it is highly probable that there were some isolated populations in the Iberian and Italian Peninsulas during the LGM (SINCLAIR et al. 1999; SORANZO et al. 2000; CHEDDADI et al. 2006).

Genetic diversity studies can be applied in forest breeding and management, which are already challenged by the climate change, in order to preserve and improve Scots pine populations. Genetic studies would provide valuable references for future requirements, which already today are expressed as forest fragmentation and species extinction, which in extension affect biodiversity.

Survival in glacial refugia and postglacial colonization are thought to be important determinants of the genetic structure we see in the present populations. When combined with pollen data, Tollefsrud et al. (2008) identified five main sources for expansions of a forest tree in Europe during the Holocene: the Russian plains, the eastern Alps, the Bohemian massif, the West Carpathians, and the southern part of the East Carpathians. Genetic diversity was relatively high in most source areas both in northern and central Europe. Therefore, colonization across vast areas in northern Europe did not lead to a large loss in genetic diversity, but to an increase in population differentiation, a pattern concordant with pollen data, which suggest that colonization, was rapidly taking place supplemented with long-distance dispersal. In central Europe, diversity was kept over much shorter distances, possibly as a result of population bottlenecks. Along with the migration routes in central

Europe, population admixture from different refugia probably took place, increasing genetic diversity (Tollefsrud et al. 2008).

The other option is that response to local conditions, for example climate–day length combination, has emerged during and after the colonization. To resolve the relative importance of these two processes resulting in adaptation, it is important to know the colonization history. The history of colonization of forest trees also helps to deduce the climate and habitat conditions in different parts of Eurasia during the Ice Age. The extensive study of PETIT et al. (2003) on the postglacial history of 25 European trees and shrubs clarified the history of European forests. It has been concluded that even if the refugia for most species were located in the same regions, the routes and dynamics of re colonization differ from species to species (FERRIS et al. 1998; PETIT et al. 2002).

In *Pinaceae*, the different modes of inheritance of mitochondrial (maternal), chloroplast (paternal) (NEALE & SEDEROFF 1989 and citations therein) and nuclear (biparental) DNA can be exploited to differentiate between pollen and seed mediated components of gene flow or in inferring population history at several time-scales.

A lower effective population size, a clinal mode of evolution, and different means of gene flow of the organelles genomes have contrasting effects on the level of differentiation compared to the nuclear genome (PETIT et al. 1993; ENNOS 1994), especially in wind pollinated species such as *P. sylvestris*. The pollen flow is also extensive. For example, ROBLEDO- ARNUNCIO & Gil (2005) estimated that 4.3% of fertilizing pollen in an isolated Spanish population came from more than 30 km distance. Presumably due to effective pollen flow, the species is not genetically highly structured (GULLBERG *et al.* 1985, KARHU *et al.* 1996, Muona & Harju 1989, SZMIDT & MUONA 1985). Even populations that are separated by thousands of kilometers show little genetic differentiation at neutral markers (DVORNYK *et al.* 2002). As a conclusion, it can often be assumed that *P. sylvestris* has a large, panmictic population. *P. sylvestris* has several useful characteristics from a population genetics viewpoint. Low genetic differentiation is useful because population structure can cause false positives in association analysis (ARANZANA *et al.*

2005, NEALE & SAVOLAINEN 2004). However, the large genome and gene families make finding the loci causing adaptive phenotypic variation a demanding task.

In this study, we use data on mitochondrial markers to distinguish between the alternative locations of glacial refugia and postglacial colonization routes of Scots pine by examining the distribution of mitochondrial haplotypes with special emphasis on resolving the descent of Northern European populations.

Because trees are both economically and ecologically important group of plants, it is important to understand how the natural selection has affected their evolution. This helps to understand and predict how and at what pace the trees would respond to environmental changes in the future. I wanted to examine the demographic history of *P. sylvestris* and especially study how postglacial colonization of central-southern and northern Europe has affected the molecular variation in *P. sylvestris*. I also wanted to study how differences in the histories of natural populations can be observed in their genomic variation. A goal is also to get better estimates of scaled mutation and recombination rates and how they vary in the *P. sylvestris* genome, because they are essential in understanding the molecular evolution at the genomic level.

Material and methods

Only a brief outline of the methods is presented in this summary. Details about sampling, molecular and statistical methods are given in the original thesis. Generally, samples from natural populations were preferred in order to minimize possible anthropogenic effects on molecular diversity of *P. sylvestris.* 69 populations and 714 individual trees were sampled. Sampling was concentrated on central-south Europe (Romania and Hungary) and northern European populations (Sweden) that were not well covered in previous studies on *P. sylvestris* post-glacial colonization history (SINCLAIR et al. 1999, SORANZO et al. 2000).

DNA was extracted from needles and buds using CTAB protocol (DOYLE and DOYLE, 1987). The population structure of Scots pine was investigated using maternally

inherited mitochondrial DNA and paternally inherited chloroplast DNA polymorphisms. DNA variation in *P. sylvestris* mitochondria is low (SORANZO 1999). Range-wide variation was investigated using maternally inherited mitochondrial two markers (mtDNA) and paternally inherited chloroplast 14 markers (cpDNA) DNA markers. Only one known variable position, *nad1* had been found before (SORANZO et al. 2000). The two mitochondrial regions were amplified and sequenced to find polymorphisms. No single point mutation was observed, only variable insertion-deletion positions were found in nad1.

Genetic diversity may decrease during sequential bottlenecks or in isolated populations due to smaller Ne. In addition to basic F-statistics based analysis of geographic distribution, the amount of nucleotide variation in different populations was compared. Genetic and geographical structure by clustering of nucleotide diversity was studied with Bayesian method, implemented in BAPS software (CORANDER *et al.* 2003), which partitions samples into groups based on allele frequency data. Distribution of genetic variation within and among populations was described with several FST estimates. The pattern of isolation by distance in mtDNA was tested by comparing genetic differentiation and geographical distance between populations (Mantel test), because according the IBD model, there should be positive correlation between the two (WRIGHT 1943). Phylogeographic structure of mtDNA data was inspected by comparing two FST estimates, GST and NST (PONS & PETIT 1996). The two estimates differ because, GST is based only on allele frequencies, but NST also takes into account the distance between haplotypes. If NST is higher than GST, it indicates that alleles inside a population are more closely related than alleles compared among populations.

Parameters of molecular diversity were calculated using the following programs: Arlequin, Contrib, Permut. At the mitochondrial DNA level less diversity was observed in all population compared with chloroplast DNA. Mean genetic diversity ranged between 0 in populations from Suceava, Valcea, Vracea, Velemer and 0.522 in Fenyőfő, Ungaria. Genetic differentiation within population (Gst) was 0.184 and N_{ST} =0.184 for Romanian and Hungarian populations, which indicates that there is no phylogeographic pattern in

population differentiation but seemed to be a phylogeographic pattern even though the G_{ST}-N_{ST} comparison did not have power to indicate it and the small differentiation between populations is due to the orography of the country (Carpathians mountains).

In Swedish populations the genetic diversity was very low or no diversity Hs=0.000 in 32 populations and pretty high 1.000 in Eckersholm (74E) population. A differentiation between populations across the country suggesting a phylogeoraphic pattern of isolation by distance was observed in the values of G_{ST}= 0.194 and N_{ST}= 0.197.

Clustering of populations using spatial genetic analysis revealed two main migration routs in recolonization of Romanian Scots pine forests. One is from the south-west Hungarian population colonizing through valleys and Danube the out Carpathian territory and one is from north-west Hungarian population which colonized the interior Carpathian populations .

According to comparative studies (TABERLET et al. 1998) based on phylogeography of ten species, we also found two main postglacial colonization routes which meet in Central Sweden One is from the northeast, and the other one from the southwest.

Overall, there was a considerable difference in the geographic distribution and amount of molecular variation between mitochondrial and chloroplast DNA. This was not surprising, because these genomes have different modes of inheritance. Chloroplast DNA is paternally inherited and is therefore dispersed long distances by pollen.

Patterns of post-glacial colonization are clear from mtDNA, because the alleles do not spread as fast as in markers that have pollen mediated gene flow. The drawback of using mitochondrial and chloroplast markers is that the whole mitochondrial genome is effectively only one locus, since there is no recombination. Altogether three mitochondrial haplotypes were found and 230 chloroplast haplotypes in Romania and 240 in Swedish populations see graphs in the thesis.

In addition to generation time, the time to the most recent common ancestor of a sample depends also on Ne. The larger the Ne, the further back in history the sample is carrying information from. Mitochondrial and chloroplast DNA (in monoecious species)

have half the Ne of nuclear markers and therefore are affected by more recent events than nuclear DNA. In *P. pinaster* and *Quercus suber* even the chloroplast genomes have been claimed to reflect the geographic events that took place 15 million years ago (MAGRI *et al.* 2007). In many species, this is close to timescales where interpretation of data requires taking into account also speciation events. Considering the pace of changes in the environment, the generation times and genetic diversity in trees, the situation where they would reach equilibrium is very unlikely. Changes in environment happen faster relative to coalescence events in species with long generation time, like trees compared to annual species, for example *A. thaliana*.

The geographic distribution of mitotypes appeared to be essentially non-random, with clusters of mitotypes geographically quite well delineated (Figure 28, 29, 30, 31 in the thesis). The cosmopolitan haplotype AA was largely distributed and found in most of the populations sampled. Geographic structuring was evident in Swedish populations and also in Romanian even though the parameters showed no differentiation.

It has also been suggested that large populations and populations with high levels of diversity should be prioritized for *in situ* conservation because of their presumed increased capacity for adaptation and for favoring the ecosystem recovery after drastic environmental changes.

Overall, the results of this study suggest that, despite periodic interstadial fragmentation episodes, Scots pine biology provides for the long-term maintenance of high within population and low among-population diversity at neutral genetic markers.

Any attempt to restore and enhance gene flow among populations or to design seed zones for reforestation should take into account the population structure detected in this and previous genetic studies. In particular, the precautionary principle suggests that the novel portrait depicted by mtDNA markers should be taken into account even if we have no evidence of variation in quantitative characters between the mitotype groups. For instance, seed transfer from the northern to the central or southern stands and vice versa should be avoided.

Further studies including additional sampling and the assessment of quantitative traits appear necessary in order to better determine the adaptive value of particular phenotypic traits and to compare the genetic structure of adaptive characters to that derived from neutral genetic markers.

BIBLIOGRAFIE

1. BENNETT KD., PC. TZEDAKIS, K.J. WILLIS, 1991, Quaternary Refugia of North European Trees. J Biogeogr 18:103-115.
2. BIRKS HJB., 1989, Holocene Isochrone Maps and Patterns of Tree-Spreading in the British-Isles. J Biogeogr 16:503-540.
3. CHEDDADI R.., J.L DE BEAULIEU., J. JOUZEL, V. ANDRIEU-PONEL, J.M. LAURENT, M. REILLE, D. RAYNAUD, A. BAR-HEN, 2005, Similarity of vegetation dynamics during interglacial periods. P Natl Acad Sci USA 102:13939-13943.
4. CORANDER J., P. WALDMANN, M.J. SILLANPÄÄ, 2003, Bayesian analysis of genetic differentiation between populations. Genetics 163:367–374.
5. CRITCHFIELD WB., JR. LITTLE EL, 1966, Geographic Distributions of the Pines of the World. USDA For Serv Misc Public 991, Washington DC.
6. DOYLE JJD. şi JL. DOYLE, 1987, A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin of the American Mathematical Society 19:1-15.
7. DVORNYK V., A. SIRVIO, M. MIKKONEN, O. SAVOLAINEN, 2002, Low nucleotide diversity at the pal1 locus in the widely distributed *Pinus sylvestris*. Mol Biol Evol 19:179-188.
8. ECKERT AJ., 2006, Influence of substrate type and microsite availability on the persistence of foxtail pine (*Pinus balfouriana*, Pinaceae) in the Klamath Mountains, California. Am J Bot 93:1615-1624.
9. ENNOS RA., WT. SINCLAIR, MT. PERKS, 1997, Genetic insights into the evolution of Scots pine, *Pinus sylvestris* L., in Scotland. Botanical Journal of Scotland, 49:257–265.
10. FLORAN Valentina and María Rosario GARCÍA GIL 2011, Genetic structure analysis of the Romanian Scots pine forest using organelles markers, New Phytologist, Comunicat.

11. FLORAN Valentina, Stefana GANEA, R. SESTRAŞ, María Rosario GARCIA GIL, 2010, Genetic Variability in Populations of Scots Pine from Romania and Sweden, Buletin of Horticulture Science, Cluj Napoca, Romania
12. FLORAN Valentina, R.E. SESTRAŞ, María Rosario GARCÍA GIL, 2011, Organelle genetic diversity and phylogeography of Scots pine (*Pinus sylvestris* L.), Not Bot Hort Agrobot Cluj, 2011, 39(1):317-322.
13. GODWIN MR., 1956, Research on Improved Merchandising of Agricultural Products. J Farm Econ 38:1346-1353.
14. GUGERLI F., M. ANZIDEI, A. MADAGHIELE, U. BUCHLER, C. SPERISEN, J. SENN, and GG. VENDRAMIN, 2001, Chloroplast microsatellites and mitochondrial nad1 intron 2 sequences indicate phylogeographic relationship of Swiss stone pine (*Pinus cembra*), Siberian stone pine (*P. sibirica*), and Siberian dwarf pine (*P. pumila*). Mol. Ecol. 10:1489-1497.
15. HUNTLEY B., 1988, Europe. In: HUNTLEY. B. şi WEBB 111. T. (Eds.Vegetation History, 341-383. Kluwer. Dordrecht
16. HUNTLEY B., 1990, European post-glacial forests: compositional changes in response to climatic change. .J. Veg. Sci. 1:507- 518.
17. HUNTLEY B., HJB. BIRKS, 1983, An atlas of past and present pollen maps of Europe: 0-13000 years ago. *Cambridge University Press, Cambridge, UK.*
18. IBRAHIM KM., RA. NICHOLS, GM. HEWITT, 1996, Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. Heredity 77: 282-291.
19. JARAMILLO-CORREA JP., J. BOUSQUET, J. BEAULIEU, N. ISABEL, MB. PERRON, 2003, 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.
20. LAGERCRANTZ U., H. ELLEGREN, L. ANDERSSON, 1993, The Abundance of Various Polymorphic Microsatellite Motifs Differs between Plants and Vertebrates. Nucleic Acids Res 21:1111-1115.

21. LODHI MA., Z. GUANG-NING, FN. WEEDEN, BI. REISCH, 1994, A simple and efficient method for DNA extraction from grapevine cultivars, Vitis species and Ampelopsis, Plant. Molecular Biology Reporter 12(1):6-13.
22. LOWE N., 2011, Discover life.
23. MIROV NT., 1967, The genus *Pinus*. The Ronald Press Company, New York.
24. MUELLER DR., WF. VINCENT, MO. JEFFRIES, 2003, Break-up of the largest Arctic ice shelf and associated loss of an epishelf lake. Geophys Res Lett 30:
25. NAYDENOV K., S. SENNEVILLE, J. BEAULIEU, F. TREMBLAY, J. BOUSQUET, 2007, Glacial vicariance in Eurasia: mitochondrial DNA evidence from Scots pine for a complex heritage involving genetically distinct refugia at mid-northern latitudes and in Asia Minor. BMC Evol Biol 7:233.
26. NEALE DB., KA. MARSHALL, RR. SEDEROFF, 1989, Chloroplast and Mitochondrial-DNA Are Paternally Inherited in Sequoia-Sempervirens D Don Endl. *P* Natl Acad Sci USA 86:9347-9349.
27. NEALE DB., RR. SEDEROFF, 1989, Paternal Inheritance of Chloroplast DNA and Maternal Inheritance of Mitochondrial-DNA in Loblolly-Pine. Theor Appl Genet 77:212-216.
28. PALMÉ A.E., Q. Su, A. RAUTENBERG, F. MANNI, M. LASCOUX, 2003, Postglacial recolonization and cpDNA variation of silver birch, *Betula pendula*. Mol Ecol. 12:201-12.
29. POP RODICA MA., D. PAMFIL, IOANA MARINA GABOREANU, 2003, The Efficiency of Different DNA Isolation and Purification in Ten Cultivars of Vitis vinifera. Bul. USAMV 59:259-261.
30. POWELL W., M. MORGANTE, R. MCDEVITT, GG. VENDRAMIN, JA. RAFALSKI, 1995, Polymorphic Simple Sequence Repeat Regions in Chloroplast Genomes - Applications to the Population-Genetics of Pines. P Natl Acad Sci USA 92:7759-7763.
31. PRAVDIN L.F., 1964, Sosna obyknovennaya. Izdatel'stvo Nauka, Moskva [Translated Israel progr. scient. transl., Jerusalem, as Scots Pine].

32. PROVAN J., N. SORANZO, NJ. WILSON, JW. MCNICOL, GI. FORREST, J. COTTRELL, W. POWELL, 1998, Gene-pool variation in Caledonian and European Scots pine (Pinus sylvestris L.) revealed by chloroplast simple-sequence repeats. P Roy Soc Lond B Bio 265:1697-1705.
33. PYHÄJÄRVI TANJA, MARIA ROSARIO GARCIA-GIL, T. KNÜRR, M. MIKKONEN, W. WACHOWIAK, and O. SAVOLAINEN, 2007, Demographic history has influenced nucleotide diversity in european *Pinus sylvestris* populations, Genetics Society of America.
34. SINCLAIR WT., JD. MORMAN, RA. ENNOS, 1999, The postglacial history of Scots pine (*Pinus sylvestris* L.) in Western Europe: evidence from mitochondrial DNA variation. Mol Ecol 8:83-88.
35. SORANZO N., J. PROVAN, W. POWELL, 1999, An example of microsatellite length variation in the mitochondrial genome of conifers. Genome 42:158-161.
36. SORANZO N., R. ALIA, J. PROVAN, W. POWELL, 2000, Patterns of variation at a mitochondrial sequence-tagged-site locus provides new insights into the postglacial history of European *Pinus sylvestris* populations. Mol Ecol 9:1205-1211.
37. SVENDSEN JI., VI. ASTAKHOV, DY. BOLSHIYANOV, I. DEMIDOV, JA. DOWDESWELL, V. GATAULLIN, C. HJORT, HWL. HUBBERTEN, EJ. MANGERUD, M. MELLES, P. MOLLER, M. SAARNISTO, MJ. SIEGERT, 1999, Maximum extent of the Eurasian ice sheets in the Barents and Kara Sea region during the Weichselian. Boreas 28:234-242.
38. TABERLET P., L. FUMAGALLI, AG. WUST-SAUCY, JF. COSSON, 1998, Comparative phylogeography and postglacial colonization routes in Europe. Mol Ecol 7:453-464.
39. TANTĂU I., M. REILLE, JL. DE BEAULIEU, S. FARCAS, 2006, Late Glacial and Holocene vegetation history in the southern part of Transylvania (Romania): pollen analysis of two sequences from Avrig. J Quaternary Sci 21:49-61.
40. TURNOCK D., 2002, Ecoregion-based conservation in the Carpathians and the land-use implications. Land Use Policy 19:47-63.

41. VENDRAMIN GG., L. LELLI, P. ROSSI, M. MORGANTE, 1996, A set of primers for the amplification of 20 chloroplast microsatellites in Pinaceae. Mol Ecol 5:595-598.
42. WILLIS KJ., KD. BENNETT, HJB. BIRKS, 1998, The late Quaternary dynamics of pines in Europe. p. 107-121. In: Richardson DM (Eds.). Ecology and Biogeography of Pinus. Cambridge University Press, Cambridge.
43. WILLIS KJ., TH. VAN ANDEL, 2004, Trees or no trees? The environments of central and eastern Europe during the Last Glaciation. Quaternary Sci Rev 23:2369-2387.
44. WILLYARD A., J. SYRING, DS. GERNANDT, A. LISTON, R. CRONN, 2007, Fossil calibration of molecular divergence infers a moderate mutation rate and recent radiations for pinus. Mol Biol Evol 24:620-620.
45. SZMIDT AE., and XR. WANG, 1993, Molecular systematics and patterns of geographic differentia-tion in geographic varieties of Pinus sylvestris (L.) and P. densiflora (Sieb. et Zucc.). Theor Appl Genet: 159-165.