ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE

FAKULTA LESNICKÁ A DŘEVAŘSKÁ

KATEDRA DENDROLOGIE A ŠLECHTĚNÍ LESNÍCH DŘEVIN

Zakládání druhé generace semenných sadů borovice lesní

Molekulárně - genetická část

Establishment of the second-generation seed orchards of Scots pine

Molecular – genetics part

Disertační práce

Ph.D. Thesis

Autor: **Ing. Jiří Korecký**

Školitel: doc. Ing. Milan Lstibůrek, MSc., Ph.D.

2012

Statement

I declare that I wrote this Ph.D. thesis on my own, quoted references being used.

Prohlášení

Prohlašuji, že jsem předloženou disertační práci vypracoval samostatně s použitím citované literatury.

10. 11. 2012 ………………………..

Jiří Korecký

Acknowledgements

I wish to thank to my supervisor Dr. Milan Lstibůrek for his support that he provided me during my studies.

My great thanks belong to Dr. Jaroslav Klápště for his priceless help with genetic data analysis. I would also like to thank to Dr. Yousry El-Kassaby for his great help and support combined with his unique enthusiasm, which made my life easier. I am very grateful to the colleagues Lenka Flašková and Dr. Tomáš Fér from Department of Botany, Charles University, who kindly allowed me to take advantage of their well equipped laboratory.

Last but not least, many thanks to my family and my friends for their patience and understanding which they provided during my education process.

Abstract

PhD thesis entitled **Establishment of the second-generation seed orchards of Scots pine. Molecular – genetics part** was developed during the years 2009 – 2012. This thesis is part of comprehensive and long-term operational and scientific project which deals with the establishment of second-generation seed orchards within the Czech Republic and also in the larger context of Europe. These activities have been financially supported by several research grants to doc. Ing. Milan Lstibůrek, MSc., Ph.D.

The proposed PhD thesis is partly based on conclusions of a PhD thesis written by Ing. Jan Kaňák, Ph.D and defended in 2011. Ing. Kaňák investigated phenology of flowering in Scots pine (*Pinus sylvestris* L.) seed orchards and compared various methods of clonal identity evaluation. The most important output of his work is considered to be the evaluation of morphological parameters of progeny trees. These data were utilized for phenotypic pre-selection of subsequently genotyped progenies.

In this thesis, several objectives were specified in order to logically follow the activities of the project. The results of these objectives will be necessary for the continuation and completion of the entire project in the future. High quality DNA extraction is an indispensable prerequisite for microsatellite genotyping. DNA was extracted using extraction kits, however it was still necessary to set up conventional extraction protocols for specific plant material, i.e. Scots pine dormant vegetative buds. DNA fragments (microsatellites) were amplified using polymerase chain reaction (PCR). Testing of potentially appropriate Scots pine microsatellite primers in addition to optimization of PCR protocols for each potential primer (e.g., determination of reaction mixture components and ideal annealing temperature of the primers) could be considered an additional purpose of this thesis. Naturally, genotyping of all individuals involved in the project is considered to be the core output.

Genotypic data were entered into the pedigree reconstruction software CERVUS. Assigned mothers of progeny trees were compared to the recorded mothers. If identification was in agreement, i.e. assigned mother corresponded

with the recorded mother, thus implied the genotypic profile of the father too (performed parent pair analysis were transformed into confirmed mother – assigned father scenario). Revealed relationships among individuals and their parents were summarized into several bar charts. Additionally, effective population size of an analyzed subset of individuals was estimated. Detailed results can be found in Appendix 3.

A secondary objective of the thesis (as an alternative for pedigree reconstruction by CERVUS) was a sib-ship assessment of progeny trees using software COLONY. Estimated relationships were used as input data for various models for genetic parameters' estimation. The outputs from the models (pedigree model, combined model using pedigree and markers information and four various marker-based pairwise relationship models) such as heritability and variance components were compared.

Since all objectives were accomplished, it can be concluded that the molecular-genetic phase of the project "Establishment of the second-generation seed orchards of Scots pine" was successfully completed. All aspects are set up for the last two stages of the activity, the application of appropriate algorithm of candidate trees' selection and the design of future seed orchard.

Keywords: Second-generation seed orchard, Scots pine (*Pinus sylvestris* L.), microsatellites (SSRs), pedigree reconstruction

Anotace

Disertační práce **Zakládání druhé generace semenných sadů borovice lesní. Molekulárně - genetická část** byla vypracována v průběhu let 2009 – 2012. Práce je součástí uceleného dlouhodobého záměru zakládání semenných sadů lesních dřevin vyšších generací v českých a obecně celoevropských podmínkách. Tato aktivita byla a je podpořena několika vědeckými projekty různých kategorií, jejichž odpovědným řešitelem je doc. Ing. Milan Lstibůrek, MSc., Ph.D.

Předložená disertace volně navazuje na práci Ing. Jana Kaňáka, Ph.D., která byla úspěšně obhájena v roce 2011. Ing. Kaňák se zabýval studiem fenologie kvetení v semenných sadech a srovnával různé metody ověřování identity klonů. Nejdůležitějším praktickým výstupem pak bylo kompletní vyhodnocení morfologických charakteristik testovacích výsadeb. Tyto výstupy byly využity jako podkladová vstupní data pro fenotypovou předselekci jedinců polosesterských potomstev, jejichž genotyp byl následně analyzován pomocí mikrosatelitových markerů.

Disertační práce sleduje několik vytčených cílů, které jsou nejenom logicky navazujícími aktivitami celé koncepce, ale také nezbytným předpokladem pro pokračování a budoucí úspěšné dovršení celého projektu.

Aby bylo možno určit genotyp rodičovských stromů a vybraných jedinců polosesterských potomstev, bylo třeba izolovat z každého jedince DNA v dostatečném množství a kvalitě. Deoxyribonukleová kyselina byla získána pomocí komerčně dostupného izolačního kitu, obecný extrakční protokol však bylo třeba optimalizovat pro předmětný rostlinný materiál, tj. dormantní vegetativní pupeny borovice lesní.

Dalším dílčím cílem byl výběr souboru vhodných mikrosatelitových primerů a následná optimalizace polymerázové řetězové reakce (určení vhodného poměru složek reakční směsi, teploty annealingu atd.) pro každý vytipovaný primer. Navazujícím a naprosto stěžejním krokem celého projektu bylo za takto optimalizovaných podmínek provést určení genotypu všech vybraných jedinců.

Informace o genotypu stromů byla využita při rekonstrukci rodokmene pomocí programu CERVUS, kdy byla nejprve ověřena evidenční příslušnost jedince z polosesterských potomstev k uvedené matce. V případě pozitivního ověření evidence vyplynul z analýzy také genotypový profil otcovského stromu. Odhalené rodičovské vazby byly kumulativně zpracovány do podoby přehledných grafů a detailně jsou uvedeny v přílohách jako Appendix 3.

Dalším ze stanovených cílů práce a určitou alternativou k rekonstrukci rodokmene pomocí programu CERVUS bylo, při využití programu COLONY, odhalení vzájemných příbuzenských vazeb jedinců polosesterských potomstev. Tyto údaje sloužily jako vstupní při určení kvantitativních genetických parametrů s využitím různých modelů a jejich následným vzájemným porovnáním. Konkrétně šlo o model s využitím rodokmenu, model kombinující informaci o mateřském klonu a mikrosatelitové genotypy a čtyři různé varianty modelů založených čistě na informaci získané z molekulárních markerů.

Splněním všech vytyčených cílů disertační práce byla zakončena molekulárně-genetická část projektu. Vlastní založení semenného sadu druhé generace je možno navázat následnými kroky, tj. aplikovat vhodný selekční algoritmus výběru nejvhodnějších jedinců pro semenný sad druhé generace a navrhnout optimální prostorové schéma budoucího semenného sadu.

Klíčová slova: semenný sad druhé generace, borovice lesní (*Pinus sylvestris* L.), mikrosatelity (SSRs), rekonstrukce rodokmene

Table of contents

List of tables ... ix

List of figures ... x

I. Literature review ... 1

1. History and presence of the genus *Pinus* ... 1
2. Scots pine (*Pinus sylvestris* L.) ... 4
   - 2.1. Scots pine within the Czech typological system ... 5
   - 2.2. The reproductive system of Scots pine ... 8
   - 2.3. The karyotype and cytogenetics of Scots pine ... 10
3. Seed orchards ... 12
   - 3.1. Pollen contamination in seed orchards ... 13
   - 3.2. Breeding without Breeding (BwB) approach ... 15
   - 3.3. Seed orchard designs ... 19
4. Genetic markers ... 21
   - 4.1. Potential microsatellite genotyping difficulty ... 26
5. Pedigree reconstruction ... 28
   - 5.1. The exclusion approach ... 28
   - 5.2. The categorical and the fractional technique ... 29
   - 5.3. Parental reconstruction ... 31
   - 5.4. The review of parentage analysis software ... 31

II. Objectives of the dissertation ... 36

III. Materials and methods ... 37

Location of research work and contribution of the author ... 37

Materials: Parental and half-sib progeny trees ... 39

6. DNA extraction and optimization of the extraction protocol ... 42
   - 6.1. The collection of organic tissue and extraction of DNA ... 42
   - 6.2. Grinding of samples ... 44
   - 6.3. Isolation and determination of the quality of DNA ... 45
   - Tested scenarios ... 46

6.4. Results of extraction protocol optimization 48
6.5. Discussion of DNA extraction protocol and its optimization 58
7. Selection of primers for the DNA analysis 59
7.1. Results and discussion of Scots pine primer selection 62
8. PCR protocols optimization and genotyping 63
8.1. PCR protocol 63
8.2. Product visualization and genotyping 65
8.3. Results and discussion of optimization of genotyping process 69
9. Pedigree reconstruction 71
9.1. Description of individuals entering into pedigree reconstruction 72
9.2. Results and discussion of pedigree reconstruction 73
10. Genetic parameters comparison among different models 81
10.1. Results and discussion of genetic parameters comparison 84

Conclusions 91
References 93
Appendices 105

List of tables

Table 1	Natural occurrence of pines (predominantly *Pinus sylvestris* L.) in the Czech Republic	6
Table 2	Characteristics of seed orchards and their half-sib progeny plots	40
Table 3	Tested scenarios and their characteristics	46
Table 4	The time variable aspect of DNA isolation	57
Table 5	Markers used for the microsatellite analysis by Genomac	60
Table 6	Markers used for the microsatellite analysis at UBC	61
Table 7	Optimized PCR conditions	64
Table 8	Demonstration of genotyping results of the gel B2	68
Table 9	Demonstration of multiloci genotypes	70
Table 10	Allele frequency analysis	73
Table 11	Description of family structure at each site separately and across both sites for pedigree reconstruction (unassigned individuals excluded)	78
Table 12	Description of familial structure at each site separately and across both sites	84
Table 13	Variance components (additive, environment, and total phenotypic and their standard errors) and heritability estimates comparison	87
Table 14	Pearson's product-moment (below diagonal) and Spearman's rank-order (above diagonal) correlations between individuals' breeding values	88

List of figures

Figure 1	Comparison of selection schemes (El-Kassaby and Lstibůrek 2009)	17
Figure 2	Influence of concentration of RNase over the purity of DNA	48
Figure 3	Yield of DNA in series A	49
Figure 4	Yield of DNA in series B	50
Figure 5	Yield of DNA in series C	51
Figure 6	Yield of DNA in series D	52
Figure 7	Scenario with the lysis for 50 min	53
Figure 8	Concentration of DNA in the variant A3	54
Figure 9	Concentration of DNA in the variant A4	55
Figure 10	Image of the polyacrylamide gel B2 obtained using LI-COR	66
Figure 11	Analysis of the gel B2 (the same as Fig. 10) using software Saga™	66
Figure 12	Parental gametic contribution as revealed by pedigree reconstruction	77
Figure 13	Familial structure	77
Figure 14	Male gametic contribution on each site and on both sites	79
Figure 15	Cumulative paternal (male) gametic contribution for all offspring assessed to their parents	80
Figure 16	Graphic output of breeding values' correlation among different cases of estimation	89

I. Literature review

1. History and presence of the genus *Pinus*

Gymnosperms represent an important component of the plants being dominant in many ecosystems (Friesen et al. 2001). *Pinus* is the most common genus of the family *Pinaceae*, which in turn is the largest family within the *Coniferales* (CABI 2002). The family *Pinaceae* includes eleven genera such as *Abies, Cathya, Cedrus, Keteleeria, Larix, Picea, Pinus, Pseudolarix, Pseudotsuga, Nothotsuga* and *Tsuga* (Farjon 1998). Without a doubt, *Pinus* is the most ecologically and economically significant conifer genus in the world (Richardson 1998). The phylogenetic tree of this family of conifers provided by Hart (1987) was based on the assessment of many morphological traits. The phylogenetic tree designed by Price et al. (1987) is based on immunological characteristics, thus the radioimmunoassay comparison of seed proteins. Recently the phylogenetic investigation based on the analysis of cpDNA gene sequences, nDNA and ribosomal DNA was conducted (Liston et al. 2003).

Traditional classifications of the genus *Pinus* are based on morphological characters of the foliar and reproductive parts of trees. The first modern classification of this genus was proposed in the 1920s (White et al. 2007). Liston et al. (1999), Wang et al. (2000), Gernandt (2005) and Syring et al. (2005) carried out further investigation on intergeneric relationships using sequences of the chloroplast, mitochondrial and low-copy nuclear genes.

The genus *Pinus,* including more than 100 pine species, is very important and often a dominant component of tree vegetation over large areas of the Northern Hemisphere (Mirov 1967). This reflects the evolution of the species in the ancient northern supercontinent of Laurasia. Bird dispersal has not occurred between hemispheres, presumably because of bird predation of the seeds (CABI 2002). Musil (2003) mentioned that only *Pinus merkusii* occurs naturally on the Southern Hemisphere in the mountains of the Sumatra Island. Nevertheless Scots pine was introduced to many countries out of its natural

occurrence including Korea, China, Mexico and New Zealand (Boratyńsky 1991).

The genus *Pinus* is evolutionarily quite ancient. It is thought to have evolved in the early-middle Mesozoic era. Miller and Charles (1976) claimed *Pinaceae* appeared during the Triassic and Jurassic and radiated during the Cretaceous. The first known species of *Pinus* was found in deposits from the Mesozoic era (White et al. 2007). The region where the genus *Pinus* evolved has split, through plate tectonic movements, into eastern North America and Western Europe (CABI 2002). Nevertheless, populations in Central Europe are supposed to be relics from the Pleistocene. Palaeoecological evidence indicates that pine has been present in Europe throughout the Quaternary. The Quaternary history of the evolution of pines is generally characterized by rapid changes in the distribution and genetic structure of populations caused by expansion and contraction of glaciers (Molotkov and Patlaj 1991; White et al. 2007).

Based on the results of restriction site and sequence comparisons of the chloroplast genome and comparative morphological and biochemical data, two major lineages of pines have been recognized: subgenus *Strobus* (haploxylon or soft pines), and subgenus *Pinus* (diploxylon or hard pines), which is larger. Both of the main subdivisions contain a number of subdivisions that are variously treated as sections or subsections (CABI 2002). Following the taxonomic classification of species in the genus *Pinus* by Price et al. (1998), the taxa are divided hierarchically into subgenera, sections, subsections, groups and species. The further overview of the development of pine systematics is provided in Richardson (1998). *P. sylvestris* L. belongs to the subgenus *Pinus*, section *Pinus,* subsection *Pinus* (Price et al. 1998). The description following GRIN (2010) taxonomy recognizes four varieties: var. *sylvestris*, var. *hamata*, var. *mongolica* and var. *sylvestriformis*. The distribution of *Pinus sylvestris* L. ssp. *sylvestris* climatypes may be, according to Molotkov and Patlaj (1991), spaced out into these groups: Northern European climatypes, lowland climatypes in Western Europe, Mountain climatypes in the western

Mediterranean, climatypes of Central Europe, climatypes of the Balkans and climatypes of the former Soviet Union.

2. Scots pine (*Pinus sylvestris* L.)

Pinus sylvestris L. is called in Czech language "borovice lesní", and in English "Scots" (UK) or "Scotch" (USA) pine. *P. sylvestris* L. is the only species known to have spread into Northern Europe during the interglacial and is thus the only species of pine in Europe to demonstrate invasive behavior on these broad scales of time and space (Richardson 1998). It has the largest geographical distribution of any pine (Boratyńsky 1991) and has been estimated as the most widespread European pine (Nowakowska 2005). It extends from boreal habitats southwards into the deciduous forests of both the Atlantic and Central European forest regions (Richardson 1998). It occupies an area longitudinally from Scotland to the Pacific Coast of Siberia, spreading over a distance about 14 000 km (Boratyńsky 1991). Latitudinally, Scots pine spreads from Norway to Spain and from Arctic Siberia to Mongolia. It also occurs in the Mediterranean region (Mirov 1967). Kaňák (2001) claimed that there are two evolutionary variants of Scots pine in the Czech Republic: the pioneer variant, which can grow naturally as a monoculture in very poor soil, and the climax variant, which is found at higher elevations with Norway spruce, beech or silver fir.

2.1. Scots pine within the Czech typological system

The classification system currently in use in the Czech Republic ("Typological System of Forest Management Institute") is based on stable environmental factors, humus forms and plant communities of climax vegetation. This typological system has been applied in the Czech Republic since 1970 (Viewegh et al. 2008).

The Scots pine forest and natural clusters of Scots pine belong to the pine zone ("bory"), which is different from other vegetation zones in that it is defined not by climate but by soil conditions. The pine zones are apparently climate independent. Plíva (1987) calculated the proportion of the pine zone at 3,73% (ÚHÚL 2010b). The same figure was cited in the Report on Czech forestry 2009. Nevertheless, the final report of the National Forestry Inventory set the pine zone at 4,2% (ÚHÚL 2007). The pine zones, divided according to the Typological System of FMI, are presented in Table 1.

Table 1: Natural occurrence of pines (predominantly *Pinus sylvestris* L.) in the Czech Republic

Typological System of FMI			
Series *řada*	**Category** ***kategorie***	**Groups of forest type** ***SLT***	**Proportion of total area of SLT** (Plíva 1987)
		Pine zone ***bory***	3,73 (4,2% ÚHÚL 2007)
Nutrient-rich (**C**) *živná*	**S** - oligo-mesotrophica *středně bohatá* **F** - lapidosa mesotrophica *svahová kapradinová* **C** - subxerothermica *vysýchavá*	**0C** Pinetum serpentinicum *hadcový bor*	0,04%
Acid (**K**) *kyselá*	**K** - acidophila *kyselá* **N** - lapidosa acidophila *kamenitá*	**0K** Pinetum acidophilum *kyselý dubo-bukový bor*	1,3%
	N - lapidosa acidophila *kamenitá* **I** - ilimerosa acidophila *kyselá uléhavá* **S** - oligo-mesotrophica *středně bohatá*	**0N** Piceeto-Pinetum (lapidosum acidophilum) *smrkový bor*	0,4%
	M - oligotrophica *chudá*	**0M** (Querceto-) Pinetum oligotrophicum *chudý (dubový) bor*	0,77%
Extreme (**X**) *extrémní*	**Z** - humilis *extrémní*	**0Z** Pinetum relictum *reliktní bor*	0,23%
	X - xerothermica *xerotermní*	**0X** Pinetum dealpinum (xerothermicum) *dealpínský bor*	0,01%
	Y - saxatilis *skeletová*	**0Y** Pinetum saxatile *roklinový bor*	--
Stagnic (**P**) *oglejená*	**P** - variohumida acidophila *oglejená kyselá*	**0P** Pinetum quercino-abietinum variohumidum acidophilum *kyselý jedlodubový bor*	0,01%
	Q - Variohumida oligotrophica *oglejená chudá*	**0Q** Pinetum quercino-abietinum variohumidum oligotrophicum *chudý jedlodubový bor*	0,15%
	O - variohumida mesotrophica *oglejená středně bohatá*	**0O** Pinetum quercino-abietinum variohumidum mesotrophicum *svěží jedlodubový bor*	0,06%
Wet (**G**) *podmáčená*	**G** - paludosa mesotrophica *podmáčená - středně bohatá*	**0G** Piceeto-Pinetum paludosum (mesotrophicum) *podmáčený smrkový bor*	0,22%
	T - paludosa oligotrophica *chudá podmáčená*	**0T** Betuleto-Pinetum (paludosum oligotrophicum) *chudý březový bor*	0,21%
	R - turfosa *rašelinná*	**0R** Pinetum turfosum *rašelinný bor*	0,14%

The principal associates of *Pinus sylvestris* L. are *Quercus petraea* agg. and *Fagus* sylvatica (Viewegh et al. 2008). It is also possible to find significant occurrences of Scots pine in some types of the first vegetation zone (mainly in 1M, but as a mixture also in 1Q, 1K, 1I or 1S). In the poor categories (M,Q and R), Scots pine regularly occurs in the fifth vegetation zone (Mikeska et al. 2008).

Scots pine grows in a wide range of soil types, from dry, nutrient-poor soils to wet, nutrient-rich sites. In Central Europe, *Pinus sylvestris* L. characteristically occurs on sandy soils across the lowlands. This species is successful in mountainous habitats as well. From a geographical point of view, *P. sylvestris* L. is an interzonal tree (Mirov 1967). In the Czech Republic, Scots pine dominates or creates the major part of tree species only in special edaphic conditions such as deep sand, serpentine, limestone, peats and acidic rock outcrops (Viewegh 2003).

According to the Report on Czech forestry 2009, the species composition of Scots pine forest makes up 16.9% of all forest stands (MZE 2010; ÚHÚL 2010b). The current composition of pine forests is very close to the recommended composition. However, an improvement of the quality of present stands would be desirable. This could be achieved by artificial reforestation of genetically suitable planting material (Kaňák and Nárovcová 2004).

Pinus sylvestris L. is adapted to a wide range of climates as indicated by its very wide-ranging distribution. This species tolerates regular and irregular precipitation regimes. The best growth occurs where annual precipitation is about 500 – 700 mm (CABI 2002). With such a wide range of biogeographical and ecological distribution, it is not surprising that *P. sylvestris* L. is highly plastic and contains extensive genetic diversity, which is particularly high at the intra-population level (Wang et al. 1991). The high level of diversity within populations is the result of a generic system that effectively provides for the creation, storage and release of genetic variation (Richardson 1998).

2.2. The reproductive system of Scots pine

Pines are monoecious, meaning the male and female reproductive structures are created separately on the same tree (Musil 2003). Female strobili are formed predominantly high in the crown near the ends of branches. Male strobili are established on the bases of annual shoots in the lower part of the crown. Temporal separation between male and female gametes is weak in general (Mirov 1967; Koski 1991).

The reproductive system of pines is, with very few exceptions, sexual. Reproductive structures are megasporangiate strobili and microsporangiate strobili, often incorrectly referred as female and male flowers (Richardson 1998). Male catkins are mostly yellow and can rarely be pink (Molotkov and Patlaj 1991). The reproduction cycle requires approximately 27 months (Kozlowski and Pallardy 1997). Primordia are initiated within the buds in the late summer almost a year before flowering. Male strobili proceed to grow and are clearly visible. Female primordia remain tiny and are hardly visible (Koski 1991). The anatomical development of "flowers" continues following spring when the female strobili emerge from the buds. Within a period of days or weeks, anthesis occurs in the male strobili and the female strobili reach a receptive stage for pollination. Musil (2003) mentioned the occurrence of flowering in May or early June. Fertilization occurs one year after pollination in the second spring, after fertilization seeds develop, mature and are shed from the cone in the autumn (Richardson 1998). According to Musil (2003), cones become mature in the autumn, but the main release of the seeds from the cones occurs in the early spring of the third year.

The very large majority of pines are outbred organisms having considerable tree-to-tree genetic variation (CABI 2002). Many naturally occurring interspecific hybrids are known in pines. In terms of interspecific hybridization of pines, Mirov (1967) has divided barriers as external barriers including geographical isolation and internal barriers including difference in time of pollen ripening, failure of pollen tube to reach the ovule and failure of seedling to reach maturity. Multiple archegonia and multiple pollination events of *Pinus sylvestris* L. provide an opportunity for competition and selection

among embryos within the ovule, only one of which usually survives to germination. Competition among embryos may be the most important principle by which pines maintain highly outcrossing mating system (Koski 1991; Burczyk 1998). The potential increase of homozygotes due to selfing can be eliminated without the reduction in seed production. On the other hand, the lack of self-incompatibility mechanisms causes little competition among embryos when pollination is poor, which provides for some seed production even from selfing (Richardson 1998).

During the development, pine pollen creates two wings, air bladders that form between the intine and exine of the pollen grain (Koski 1991). It is a significant adaptation for anemophily, dispersal by wind. The potential for transport by wind, as well as pollen production, is tremendous (Richardson 1998). Pollen distribution may be measured by catching pollen on sticky traps at various distances from an isolated source (Parantainen and Pulkkinen 2003). Koski (1970) concluded that half of the pollen received by a given tree in a *P. sylvestris* L. stand probably originates from trees within a 50 m radius, if the area circumscribes at least 25 – 30 trees. Nevertheless the effective population size in *P. sylvestris* L. spans distances of tens to hundreds of kilometers (Koski 1970).

According to genetic studies of isozymes among the genus *Pinus*, in most cases normal segregation of alleles is the rule and studies of genotypic-frequency distributions within populations show no significant departures from the Hardy-Weinberg equilibrium (Sewell et al. 1999). Allozyme loci with allele combinations that tend to segregate together can be easily identified in the pine megagametophyte system, and the recombination fraction used to map linkage groups (Yazdani et al. 1995). Chloroplast DNA (unlike that in angiosperm) is inherited through the pollen parent, the mitochondrion deoxyribonucleic acid, whereas in almost all other plants and animals it is inherited maternally. The inheritance of nuclear DNA is naturally biparental (Dong and Wagner 1994). The recombination system of pines can be characterized as open. The rate of recombination depends on the organization of genes in chromosomes and the mechanics of meiosis. Recombination rates in

pines are high because of the large number of chromosomes and their large size, which favors crossover. In summary, the reproductive cycle of pines has several advantages for genetic analysis. Haplogenetics is possible because the megagamethophyte develops from a single meiotic product and the eggs have the same genotype as the nutritive tissue surrounding them (Varis et al. 2005).

2.3. The karyotype and cytogenetics of Scots pine

Pines are diploid organisms (2n = 24) with a haploid chromosome number of 12. The genomes of pines are characterized by their great size and complexity which makes genomic analysis difficult. The genome size of pine species was estimated at more than 20 000 Mbp (Schmidt et al. 2000), ranking between 18 – 40 Gbp (Morse et al. 2009). Ahuja and Neale (2005) published genome size of Scots pine 24,6 Gbp. The overview of nuclear DNA amounts in gymnosperm was provided for instance by Grotkopp et al. (2004) who estimated mean genome size of Scots pine 25,06 pg using laser flow cytometry.

Pines are known for very slow chromosomal evolution. All of them share the same number of chromosomal pairs which are morphologically similar (Guevara et al. 2005). The results of investigation provided by Hizume et al. (2002) confirmed very low chromosomal differentiation among the *Pinus* species. Scots pine chromosomes are either metacentric or submetacentric (Borzan 1991). Muratova (1997) claimed eleven pairs of chromosomes are metacentric while one pair was estimated to be submetacentric. Seven pairs of metacentric chromosomes have secondary constriction, and two pairs of chromosomes have more than one constriction.

The haploid karyotype of pines is uniform (it is difficult to distinguish one chromosome from another) and the karyotype is highly conserved in the genus. The amount of DNA in the pine karyotype is remarkably high when compared to, for instance, maize, eucalypt or human DNA. It is unlikely to be completely sequenced in the near future (Krutovsky et al. 2007). Most of the pine's DNA is in highly repeated, multi-copy sequences. Schmidt et al. (2000) and Guevara et al. (2005) claimed that 75% of the pine's genome corresponds

to highly repeated sequences. Much of the DNA in pines seems to be non-coding. Probably only 0,1% is expressed in mRNA (Richardson 1998).

Expected heterozygosity under the Hardy-Weinberg equilibrium ranges from 0 to about 0,33 for the investigated pine species. The average value of *Pinus sylvestris* L. based on investigations of 13 different populations is estimated to be 0,30. This means that the most individuals are expected to be heterozygous at about 30% of their loci (Richardson 1998).

3. Seed orchards

White et al. (2007) defined a seed orchard as a collection of selected clones or families established in one physical location. Populations in seed orchards are mainly aimed at providing mass quantities of genetically superior seed (Kaňák et al. 2009). The first seed orchard of conifers (European larch) in the Czech Republic was established in 1956 (Musil et al. 2007). The first seed orchard of Scots pine originated from 1973 (Rambousek 2003).

The main goal of seed orchards is to produce seeds of a higher genetic quality compared to the seeds originating from forest stands. Therefore, seed orchards are generally established from phenotypically superior trees, expecting that their offspring will retain a part of their superiority. However, the phenotypic similarity between both generations depends, in addition to the heritability of the respective traits, on the similarity of their genotypic constitution (Gömöry et al. 2003). The common treatise about seed orchards was described, for instance, by Eriksson et al. (2006) or White et al. (2007).

Two types of seed orchards exist: clonal seed orchards established by vegetative propagation, and seedling seed orchards created with open-pollinated or full-sib offspring from selected trees (Paule 1992). The seed orchards in the Czech Republic are all clonal seed orchards based on grafted plants selected from plus trees. Thus far, the Scots pine seed orchards in the Czech Republic are all the first generation (Kaňák et al. 2009).

First generation seed orchards are established from initial plus tree phenotypic selections. Therefore, information on respective genetic quality is missing in this early phase of tree improvement (Kaňák et al. 2009). Methodological procedures for the establishment of first generation seed orchards in the Czech Republic were published by Kaňák et al. (2008). Forward selection among their progenies of seed orchard trees based on selection indices can lead to the establishment of seed orchards in advanced generations (Kaňák et al. 2009). All seed orchards established in the Czech Republic are recorded in the national forestry register (Rambousek 2003). Based on this register, ERMA (available on-line at erma.uhul.cz), the area of 37 seed orchards

of Scots pine in the Czech Republic has been estimated at 125,5 ha in total (ÚHÚL 2010a). Out of these, 24 seed orchards with a total area of 85,7 ha are being managed by Lesy České republiky, state enterprise (Svoboda 2010).

The prediction of genetic response to selection is based on the theoretical assumption that the fusion of gametes in a seed orchard is random (Falconer and Mackay 1996). Nevertheless, many studies in coniferous seed orchards and consequent confirmations using genetic markers revealed that the theoretical assumption of panmixis is unrealistic, because flowering and phenology may vary considerably (Muona and Harju 1989; Lindgren 1994; Lindgren et al. 2004; Slavov et al. 2004; Slavov et al. 2005a).

3.1. Pollen contamination in seed orchards

Clonal seed orchards are the most important sources of forest reproductive material for several conifer species, including Scots pine. The formation of genetic structure of seed orchard offspring is determined by several factors and processes, such as the composition and size of the maternal population, fertility variation, and mating system (inbreeding and outbreeding, pollen dispersal within the seed orchard, pollen contamination, etc.).

Pollen contamination is detrimental to the genetic quality of seed orchard crops (Slavov et al. 2005a). The simplest method to directly measure pollen immigration is the paternity exclusion. Genotypes of a sample of progeny are compared to those of potential male parents from the investigated array and the progeny whose multiloci genotypes are incompatible with all of these parents are assumed to result from pollen immigration (Slavov et al. 2005b). Slavov et al. (2005a) proved that the estimation of alien pollen migration depends on an appropriate adjustment for mistyping.

Because allozymes show low allelic diversity (i.e., low effective numbers of alleles per locus), it is impossible to directly detect all seeds that were fertilized by non-orchard pollen. However, highly variable SSR markers make it possible to accurately measure pollen contamination. Slavov et al. (2005a) estimated pollen contamination in a clonal seed orchard of Douglas-fir

(*Pseudotsuga menziesii* (Mirb.) Franco) to be on average 35,3%. Robledo-Arnuncio et al. (2004) used microsatellites analysis in a *Pinus sylvestris* L. stand to estimate the average pollination distance of 17 – 22 m. Lindgren (1994) covered some tops of *Pinus sylvestris* L. growing in seed orchards to prevent against non-orchard pollination. The results were estimated using assay with allozymes. It was concluded that the covers had no evident effect on alien pollination probably due to the pollen transport by strong and turbulent wind, reaching the reproductive structures under the cover.

Robledo-Arnuncio and Gil (2005) investigated spatial patterns of pollen dispersal in *Pinus sylvestris* L. using chloroplast and nuclear microsatellites. They performed a total-exclusion paternity analysis on collected seeds from a population of 36 trees. A total of 24 seeds from each tree were investigated and the minimum effective pollen immigration from outside the stand was estimated at 4,3%. The rate of self-fertilization among individuals varied, ranging from 0 to 92%. Burczyk et al. (2004) investigated gene flow in forest trees. It was estimated that pollen contamination of seed orchards ranged from 1% to 91% depending on the age of trees, size and the isolation of seed orchards. In the past, it was common to establish a pollen dilution zone of approximately 150 meters, where the same tree species inside this area is eradicated. Nevertheless, it currently appears that these zones of conifer seed orchards are ineffective and an isolation of 500 to 1000 m would be necessary for at least some protection (White et al. 2007). Even these distances may not be sufficient, as shown by Di-Giovanni et al. (1996), who claimed that large amounts of pine pollen can be dispersed tens of kilometers away. The distance limit for conifer pollen dispersal could be up to 2000 km but it is unlikely that the pollen would remain viable. White et al. (2007) mentioned other options of how to deal with pollen contamination, such as pollen enrichment zones, increasing pollen production within the orchard or altering the floral phenology of orchard trees.

According to the Czech legislation act 149/2003 Coll. (Zákon o obchodu s reprodukčním materiálem lesních dřevin), seed orchards must be isolated or cultivated in such a way as to prevent or significantly reduce pollination from

trees growing outside of the seed orchard. Nevertheless there are not strict rules as to what an adequate distance means.

3.2. Breeding without Breeding (BwB) approach

The principal part of this thesis is based on an innovative approach to tree breeding called Breeding without Breeding. This method, invented by El-Kassaby and Lstibůrek (2009), allows to capture 75 – 85% of the genetic response attained through conventional selection schemes without the need for controlled pollination that is considered to be the most resource-demanding activity of a breeding program. BwB combines the initial selection of superior trees based on their phenotype and genotype, informative DNA markers for fingerprinting and pedigree reconstruction of offspring. Among existing half-sib families, elite genotypes could be consequently identified and selected for further genetic improvement (El-Kassaby and Lstibůrek 2009). In this particular case, this leads to the establishment of second-generation seed orchards.

Conventional tree improvement programs follow recurrent selection schemes consisting of base populations where initial phenotypic selections are made; breeding populations where crosses among the selected individuals are created and tested; and finally, the establishment of seed orchards from elite genotypes. They are structured around systematic repeated cycles of breeding, testing and selection requiring substantial planning and resources. An alternative to full-sib breeding and testing using the DNA markers has already been published (Lambeth et al. 2001; Grattapaglia et al. 2004), but this was proposed to apply to very specific scenarios, while BwB was developed for general application in many tree breeding programs worldwide. The BwB strategy requires much less effort and capitalizes on the assembly of natural crosses among selected parents (El-Kassaby and Lstibůrek 2009). El-Kassaby et al. (2006) claimed the logistically manageable number of parents to be about 30 individuals and that reduced pollen contamination levels are expected to increase the efficiency of the proposed approach.

El-Kassaby and Lstibůrek (2009) demonstrated the BwB method in a retrospective study of Douglas-fir parent trees and their progeny identified by site, replication and the male and female parents. Their study included theoretical expectations demonstrating expected minimum genetic gain compared to conventional approaches. According to their results on the half-sib (HS) model, where only female parents were considered known, the approximate captured genetic gain was 85% compared to the full-sib (FS) model, where both parents are known. The full-sib model can therefore substitute conventional selection schemes. The phenotypic-selection (PH) model, where all parents were considered unknown, produced surprisingly high gain, capturing 75% of the genetic gain attributable to the FS model.

The genetic gain and the retained diversity are connected by indirect proportion, therefore the observed decline in average genetic gain with increased diversity is not a reason for concern (El-Kassaby and Lstibůrek 2009). The balance between seed production and breeding value in seed orchards was studied, for instance, by Lindgren et al. (2004).

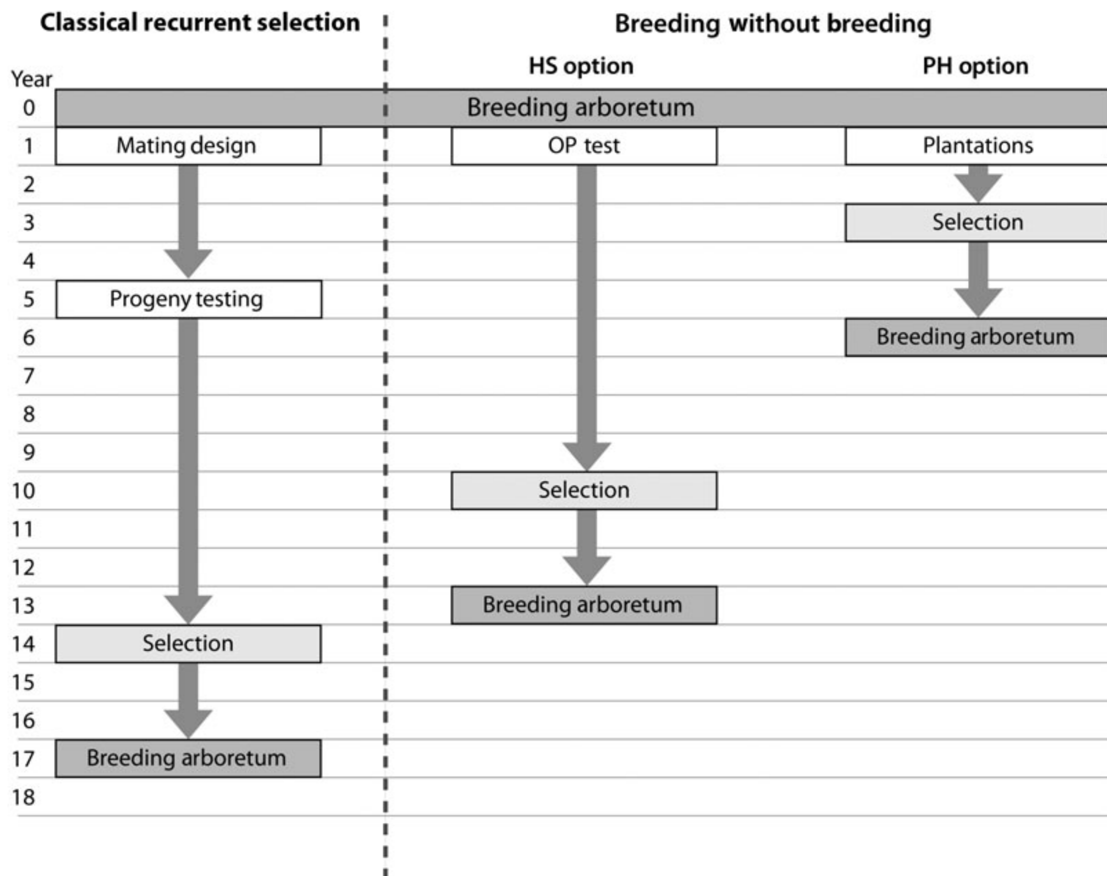


Fig. 1: Comparison of selection schemes (El-Kassaby and Lstibůrek 2009)

The scheme in Figure 1, originally published by El-Kassaby and Lstibůrek (2009), represents the timeline for a classical recurrent selection scheme and the two BWB strategies. This comparison assumes that the initial breeding arboretum is sexually mature and the plantations of mixed families have already been established. This time-scale chart could vary generally, but the time proportions would be maintained.

El-Kassaby and Lstibůrek (2009) assumed their retrospective study to be error-free. In general, the critical factors in pedigree reconstruction are the number of loci used and their polymorphism information content – PIC (Botstein et al. 1980; Yazdani et al. 1995), and the rate of genotyping error (Vandeputte et al. 2006; Wang 2007).

While most pedigree reconstruction methods account for genotyping error, the ability to estimate its magnitude and adjust its impact are essential (El-Kassaby and Lstibůrek 2009). A limited number of loci is very effective for the

pedigree reconstruction (Gerber et al. 2000). However, the higher the number of informative loci analyzed, the fewer the errors, including false assignment and false exclusion (Vandeputte et al. 2006). In addition, the immigration of non-orchard pollen into the seed orchards is expected to increase fingerprinting efforts (Pompanon et al. 2005). El-Kassaby and Lstibůrek (2009), however, expected that most of the individuals resulting from the pollen contamination would be eliminated during pre-selection based on their inferior qualities. This issue was further developed as a deterministic model for various values of gene flow, selection differential between the parental population and contamination source, heritabilities and intensity of pre-selection prior to genotyping. The results from various scenarios show some general trends such as decrease in either heritability or difference between the seed orchard and the base population was associated with higher proportion of offspring originated from contaminant pollen. The proportion of offspring resulting from contamination also increased with lower intensity of pre-selection (Lstibůrek et al. 2012).

The cost of BWB implementation is mainly dependent on the number of individuals requiring genotyping. El-Kassaby and Lstibůrek (2009) observed small differences in the genetic response amid the intensity of pre-selection, but a balance between the number of genotyped individuals and the accuracy of pedigree reconstruction should be considered. The numerous primers for many species have been already published, which reduces expenses.

3.3. Seed orchard designs

The main aim of the spatial design in seed orchards is to maximize genetic gain while maintaining acceptable levels of genetic diversity in collected seeds. The crucial parameter to determine is the appropriate number of clones (White et al. 2007). For the initial phase of establishing a seed orchard, many layouts were developed. At this stage, breeding programs start with initial, unrelated individuals' selections from wild stands. Thus relatively simple orchard layouts have served the purpose well. Later, genetic structure of breeding programs changed considerably with increased relatedness and inbreeding. It was necessary to start developing new, sophisticated approaches. A review of these is available by Giertych (1975) ex Lstibůrek and El-Kassaby (2010).

White et al. (2007) mentioned three broad classes of seed orchard design: modified randomized block designs, permutated neighborhood designs and systematic layouts. In modified randomized block designs, one ramet per clone is planted in each block. Within the block, clones are dislocated randomly but the minimum distance between related trees is taken into account. The permutated neighborhood design provides an isolation of every tree in the orchard from other ramets of the same clone. The original concept of the permutated neighborhood design was modifies by Bell and Fletcher (1978), who published their concept of computer-organized orchard layouts (COOL). Systematic orchard designs mean utilization of repeating blocks in which a particular clone still has the same position. It is the easiest way of establishing the seed orchard and, in particular cases, may be favored.

The permuted neighborhood design is considered to be the most efficient in randomizing the clones into the seed orchard, nevertheless Lstibůrek and El-Kassaby (2010) mentioned that the COOL design is only limited to situations where clones are unrelated. Furthermore the COOL is based on a local assignment. Therefore they proposed an innovative Minimum-Inbreeding design (global assignment method), taking relatedness and additional factors into account (Lstibůrek and El-Kassaby 2010). This could have been done by advances in reproductive biology, including better understanding of mating systems, pollination biology, fertility variation, and contamination.

Using MI seed orchard design, the spatial distribution of individual trees in the seed orchard is treated as a modified quadratic assignment problem. A powerful heuristical tabu search algorithm developed by Misevicius (2005) may be used. The MI design is expected to provide an optimal solution with respect to the information provided during the establishment of the seed orchard. The kind of distribution and the matrix of distances can be suitably transformed. It means that the relative efficiency of MI approach can be calculated for each specific set of assumptions in an existing situation. Expected levels of selfing in the COOL scheme were reduced to approximately one half of the completely randomized scheme. The MI design further reduced the selfing rate to app. one half of the COOL layout. For further description, see Lstibůrek and El-Kassaby (2010).

There are other ways the genetic gain can be increased when the seed orchard already exists, such as the stratified collection in seed orchards. For example, Funda et al. (2009) published an optimization protocol to maximize the genetic gain of a crop at any desired level of genetic diversity through the selection of a subset of the crop.

4. Genetic markers

Genetic marker is any visible character for which alleles at individual loci segregate according to Mendelian laws. Until the early 1970s, biochemical genetic markers such as terpenes and allozymes (a term that originates from a phrase “allelic variants of enzymes”) were utilized for forest tree species (White et al. 2007). Biochemical markers are not very representative of genes throughout the genome due to the fact that only a small number of different marker loci could be revealed. This limitation was overcome in the early 1980s with the development of molecular or DNA based genetic markers. Since mid1990s molecular markers such as microsatellites have been widely used in genetic studies of forest tree species (Dzialuk et al. 2005).

The discovery of molecular markers, such as methods of RAPD, AFLP and microsatellites significantly contributed to the field of population genetics in the study of genetic diversity, reproductive systems and genomic mapping (Jarne and Lagoda 1996). These markers differ in important features such as genomic abundance, level of polymorphism, locus specificity, technical requirements and cost (Guevara et al. 2005). Therefore, the crucial question is to choose the appropriate marker for each study. Because of the widespread occurrence, the large variability in repeated numbers and their distribution among many loci, microsatellites are highly informative markers appropriate for genome analysis (Schmidt et al. 2000). Microsatellites detecting variation at individual loci have been thought of as the “new allozymes”, therefore much of their use was in studies where allozymes have been used (Robinson and Harris 1999).

In the case of allozymes, a protein solution is electrophoresed through a gel so that an enzyme-specific reaction reveals one locus whose alleles may migrate differently due to differences in electric charge. Allozymes show a low level of polymorphism, usually about 1 – 5 alleles per locus. Furthermore, surveys of natural variation based on allozymes were often challenged by non-neutral evolution of some markers (Schlötterer 2004). A further limitation of using allozymes is that they may differ in metabolic function, thus it is clear

that the ideal genetic marker is one that represents non-coding DNA rather than a gene product (i.e. allozyme) that is exposed to selective processes (Parker et al. 1998). Nowakowska (2007) compared general characteristics of isoenzymes, RAPD, microsatellites, PCR-RFLP and STS markers from 7 different points of view.

Microsatellites, also known as simple sequences repeats (SSRs), short tandem repeats (STRs) or simple sequence length polymorphisms (SSLPs) (Robinson and Harris 1999), are valued in population genetics for their high variability. They have proven to be very useful for the purpose of unveiling genetic diversity in forest tree species (Scotti et al. 1999; Liewlaksaneeyanawin et al. 2004). Microsatellite loci consist of tandem repeats of a short nucleotide motif. An allele is defined as a number of these repeats (Viard et al. 1996; Nowakowska 2005). Microsatellites were found in both coding and noncoding DNA regions (Zane et al. 2002). The variability of microsatellite loci is more easily detectable than variation in a length of DNA fragment obtained through an amplification of the whole genome. These co-dominant markers can be amplified by a polymerase chain reaction (PCR) from a small amount of DNA (Echt and MayMarquardt 1997). The detection of within-population diversity and structure is made easy by the high polymorphism of this kind of marker that allows for identifying single trees, even when analyzing few loci. For the same reasons, this technique can be applied to paternity analysis (Lefort et al. 1999). Microsatellite markers are, in comparison with allozymes, generally recognized as neutral (Robinson and Harris 1999), so that selection and environmental pressure do not influence their expression directly (Varshney et al. 2005). Another advantage of microsatellites is that the analysis can be multiplexed, which means the PCR products of different markers can be run on the same gel, which saves time, work and money (Scotti et al. 1999).

For most biological systems, microsatellite markers are considered the most powerful tool for parentage analysis. These markers are co-dominant and show high degrees of polymorphism, which means high information content arising from their multi-allelic nature (Echt et al. 1999). Masi et al. (2003) and Hayden et al. (2008) summarized other positive features of microsatellites

such as high discriminatory power, robust and reproducible assay, relative abundance with uniform genome coverage, the necessity of a small amount of DNA template and the possibility of detection via automated systems. The techniques can be partially automated using dye labels and standard DNA sequencing equipment (Devey et al. 2002).

With the appropriate method microsatellites make it possible to determinate the genotype, whereas other methods, such as isoenzyme analysis, may not always lead to a clear determination of the genotype due to their lower allelic variability. Robinson and Harrison (1999) mentioned that the advantage of DNA-based methods is the ability to use dried plant material for analysis in comparison to allozymes which required fresh plant tissue for analyses. On the other hand, it is fair to mention that microsatellites have some disadvantages and restrictions. The method requires designing primer pairs specific to the microsatellite locus (forward and reverse primer), which is the major drawback of microsatellites (Zane et al. 2002). The development cost of these markers is usually high. Microsatellite markers can be developed in several ways, via genomic libraries, BAC/YAC libraries or cDNA libraries (Scott 2001). A review about microsatellite isolation was published by Zane et al. (2002). Microsatellite primers developed for one species are not often useful for related species. Therefore, markers for each species must be developed separately. Low transferability of primers among species belonging to the genus *Pinus* are due to the long-time development of the genus *Pinus* and divergence between pines species (Karhu et al. 2000). Chagné et al. (2004) confirmed that the transferability of molecular markers between species depends on the respective phylogenetic distance.

Development of primers is complicated by the large genome size and number of repetitive stretches of DNA. During PCR, multilocus amplification occurs and the product is unable to be scored. Identification of markers suitable for population studies is a crucial factor. Soranzo et al. (1998) stated that only 20% of found primer pairs showed specific amplification of the polymorphic locus. To overcome this problem, low-copy sequences called

expressed sequence tags (EST), which increase the likelihood of isolation of a single marker locus (Scott 2001), were used.

Several approaches were applied to eliminate highly repetitive regions of DNA in gene libraries. Elsik and Williams (2001) also confirmed that low-copy microsatellites provide a higher degree of polymorphism and have higher information content in comparison with genomic DNA microsatellites. A similar conclusion was also reached by Scotti et al. (1999). Microsatellites, which are present in the ESTs, show a higher degree of amplification in relation to genomic DNA microsatellites. Short and highly conservative EST-SSR sequences, which do not differ in species that are closely related phylogenetically, show a lower degree of variability and lower frequency of null alleles. The primers designed based on EST are therefore more suitable for use in different species of the genus *Pinus*.

Analysis of chloroplast microsatellite markers makes possible the elucidation of the relative contribution to the genetic structure of the population because of their uniparental inheritance (Provan et al. 2001). A chloroplast genome is transferred paternally to the offspring. Burczyk et al. (2005) investigated sets of chloroplast and nuclear microsatellites in the Scots pine population. The most common type of repeats in the genus *Pinus* is a trinucleotide pattern, followed by repetition of two nucleotides. Tetranucleotide repetitions occur very rarely but are highly polymorphic (Jarne and Lagoda 1996). Schmidt at al. (2000) detected the organization of sequence repeats by Southern hybridization. Common types of two nucleotide repeats are AT and AG pairs, while incidences of AC and CG repeats are much lower. The most frequent trinucleotide repeat motifs are AAG, AGC and AGG (Chagne et al. 2004).

Additionally, there are other possible constraints such as allelic dropout, null alleles or occurrence of stutter bands. All these aspects are examined in the following chapter 4.1.

New marker type has been on the scene over the last few years and has obtained high popularity. It is named the "Single Nucleotide Polymorphisms" (SNPs) and involves the change of one nucleotide for another. These changes

are typically found every 300 – 1000 bp in most genomes (Aitken et al. 2004). Some authors consider the insertion or deletion of one or more nucleotides (indels) also as SNPs, although they occur by a different mechanism (Vignal et al. 2002). SNPs are usually biallelic which means that there are two alleles in a population. It was stated that five SNP markers provide similar information about the genome (the comparable level of discrimination power) to one microsatellite marker (Beuzen et al. 2000).

SNPs compared to SSRs are more abundant and therefore provide denser coverage across the genome (lower recombination fraction among the SNPs and loci contributing to adaptive traits). Moreover they are abundant and distributed widely across the genome, which prevents from biases associated with few loci analysis (Aitken et al. 2004). SNPs are more stably inherited which makes them more suited for long-term selection. They are more feasible for high throughput; especially lately when rapid and cost-effective approach "Genotyping by Sequencing" (Elshire et al. 2011) suited for high diversity and large genome species could be used. Taking all this into account, single nucleotide polymorphism method might be more appropriate tool in future studies.

The DNA technologies are powerful tools in the form of revealing relationship (i.e. genotype) among tested individuals (CABI 2002). This fact helps to answer one of the most important question of forestry breeders – how to balance between sufficient level of genetic diversity while capturing the satisfactory amount of genetic gains.

4.1. Potential microsatellite genotyping difficulty

Microsatellite genotyping errors occur when the genotype determined after molecular analysis does not correspond to the real genotype of the individual under consideration. In practice, genotyping errors are defined as the differences observed between two or more molecular genotypes obtained independently from the same sample. Eradicating genotyping errors is not possible, mainly because molecular assays and manual sample handling are not 100% reliable. Genotyping errors can be generated at every step of the process (sampling, DNA extraction, molecular analysis, scoring, data analysis) due to a variety of factors such as chance, human error or technical artefacts (Bonin et al. 2004).

Microsatellite genotyping errors have the potential to undermine the conclusions of most downstream analyses (Hoffman and Amos 2005). They may arise in a number of ways. When the template DNA is of low quantity or quality, as is typical of studies employing noninvasive tissue-sampling, PCR amplification can become unreliable (Gagneux et al. 1997; Hoffman and Amos 2005). A common problem is the stochastic failure of one allele to amplify, leading to heterozygotes appearing to carry only one allele, referred to as 'allelic dropout'. Another source of artefact is 'misprinting', in which amplification products are generated that can be misinterpreted as true alleles (Taberlet et al. 1996; Hoffman and Amos 2005). Another problem could be a phenomenon called homoplasy which means that alleles are of the same size but differ in base-pair composition (Nowakowska 2005).

Even when large quantities of high-quality DNA extracted from tissue are available, genotyping errors still occur. These include null alleles resulting from non-amplification caused by binding site mutation (Dakin and Avise 2004). Mutation in the binding regions of one or both of the microsatellite primers may inhibit annealing of appropriate primers, which causes the reduction or loss of the particular PCR product (Callen et al. 1993). Other errors can be due to electrophoresis artefacts (Davison and Chiba 2003), mis-scoring of allele banding patterns, data entry and other clerical errors. Null alleles can lead to underestimates of heterozygozity (Scotti et al. 1999). The

most important source of error is probably the incorrect naming of alleles on autoradiographs or fluorescent profiles. In particular, the presence of ‘stutter bands’, generated by slippage of *Taq* polymerase during PCR, can make it difficult to score alleles reliably (Johansson et al. 2003), especially when there are large signal intensity differences between alleles or the lengths of two alleles in a heterozygote individual differ by only a few nucleotides. Robinson and Harrison (1999) stated that slippage leads to the production of differently sized products that differ by approximately 1 - 5 repeated units from the expected alleles. This results in ambiguity in the estimation of the allele size and the possibility of mistaking a heterozygote for a homozygote, if the two bands are so close on the gel that the ladders produced by the two alleles overlap (Scotti et al. 1999). Unfortunately, these problems remain even when genotyping is provided automatically. The possible solution, therefore, is to correct the genotypes manually (Ewen et al. 2000).

Hoffman and Amos (2005) claimed that a 1% error rate in allele naming would lead to almost a quarter of 12-locus genotypes containing at least one error. With this error rate, only about 62% of comparisons between the same individual typed twice would show the same genotype. Analogously, the estimation of parentage analysis can be heavily impacted by errors, especially when candidate fathers are excluded on the basis of only one mismatch (Marshall et al. 1998). Fortunately, there are some approaches to identifying genotyping errors and estimating their rates. The most obvious way includes taking a subset of individuals, re-genotyping the same individuals and then comparing (Ewen et al. 2000). The economical method seems to be a statistical testing of the data that already exist. One commonly used test for deviations comes from the Hardy–Weinberg equilibrium (Gomes et al. 1999), which reveals the homozygous excess resulting from either null alleles or allelic dropout. Further verification can be achieved by comparing known mother and offspring pairs and subsequently looking for mismatches (Marshall et al. 1998).

5. Pedigree reconstruction

Modern marker technology, such as the use of microsatellites, has major potential for pedigree reconstruction (Lambeth et al. 2001), especially in open-pollinated families of known parentage trees. This is complicated by the wind pollination typical for coniferous trees. Plomion et al. (2007) mentioned molecular markers have a general potential to estimate relatedness between individuals without any need of prior knowledge of the genetic relationship. Parentage analysis could be defined as an application of searching for the most robust relatedness among candidate parents and a target offspring (Blouin 2003). However, if it is not possible to estimate one single parent, than it is necessary to apply methods that may distinguish between non-excluded candidates. The other issue involves genotyping errors. These errors may influence a process of pedigree reconstruction. Nevertheless, the appropriate likelihood method can deal with this drawback as well.

5.1. The exclusion approach

The simplest approach to molecular pedigree analysis is the exclusion. If a co-dominant marker with Mendelian transmission (such as an SSR marker) does not share at least one mutual allele at any given locus between parent and offspring, this potential parent will be excluded as a real one. In the case when one parent pair is known, this approach serves as a powerful method for testing a putative relationship (Wilson and Ferguson 2002).

This technique reflects Mendel´s laws of inheritance and uses the principle of potential incompatibility between the putative parent and its progeny. It is the simplest technique of parentage analysis. The genotype of potential fathers is compared with the genotype of the offspring. If the genotype of one parent (mostly mother genotype) is known, it is possible to identify or detect incorrect scoring alleles, mutations or null alleles comparing maternal haplotype (Marshall et al. 1998). Parental candidates differing in their genotype from their potential offspring on one or more loci are excluded. For instance, if a

mother has the diploid genotype AA and her offspring AB at a particular locus, then the male with the genotype AC can be excluded, whereas the male with genotype BC cannot. This technique is very powerful when there are few candidate parents and highly polymorphic genetic markers are available (Jones and Ardren 2003).

The potential disadvantage of a strict exclusion is that genotyping errors, null alleles and mutations in flanking primer sequences can falsely exclude true parents (Gagneux et al. 1997). The more loci are investigated, the greater the likelihood that a dataset will contain errors or mutations. However, software for parentage analysis often specifies the number of mismatches necessary for a valid exclusion. One true parent will remain only in cases where a small set of parental candidates was scored and the loci are highly polymorphic. However, Chakraborty el al. (1988) noticed that even if highly polymorphic co-dominant markers have been utilized, the method of exclusion might be insufficient. More often, several parent candidates remain. In this case, without the specific characteristics of probability, the individual who is most likely the real parent cannot be determined. If the complete exclusion is not possible, there are two alternative approaches that are useful. Both calculate the likelihoods in the same way.

5.2. The categorical and the fractional technique

The categorical technique assigns the entire offspring to a particular male, whereas the fractional technique splits the offspring among all compatible males. The categorical technique seems to produce results that represent biological truth (offspring can have only one father), but the fractional assignment may produce better statistical properties for the evaluation of some hypotheses (Wilson and Ferguson 2002).

The categorical approach (likelihood-based) selects the most likely parent from a subset of non-excluded parents, which involves calculating a logarithm of the likelihood ratio (LOD score). The quantity calculated by multiplying the ratios of particular loci is called the overall likelihood ratio. LOD score 0

indicates that the putative father is as likely to be the true father as any other potential father selected from the population. The negative value of the LOD score occurs when the putative father shares with a putative progeny far fewer similar alleles than other individuals. The positive LOD score indicates that the tested individual is more likely to be the real father than any other randomly selected candidate. If the LOD score of the most likely parent is sufficiently high, it is possible to consider this individual as the true parent. Another statistical parameter derived from the likelihood ratio is a quantity delta, which is defined as the difference in the values of LOD scores between the most likely and the second most likely candidate parent. This parameter is another criterion for a parentage determination (Kalinowski et al. 2007).

The fractional likelihood assigned to each offspring of all non-excluded candidate parents is between 0 and 1. The proportion of an offspring allocated to a particular putative parent is proportional to its likelihood of parenting that particular offspring compared to all other non-excluded candidate parents. This approach assumes genotypes of all parents are known and only one parent is being investigated (Jones and Ardren 2003).

There is one statistical fact, which needs to be taken into account. The homozygous males share more alleles with their putative offspring than the heterozygous males, because the homozygous males have two copies of the same allele.

5.3. Parental reconstruction

Another approach to parentage analysis is to reconstruct parental genotypes from the group of offspring. These putative parent genotypes can be compared to the genotypes of real candidate parents. Therefore this technique is susceptible to score errors, null alleles and mutations (Jones and Ardren 2003).

In summary, results from parentage analysis using strict exclusion could provide the most robust estimates. In reality, due to imperfect input, parameters must be often used in the categorical or fractional approach.

5.4. The review of parentage analysis software

For studying the genetic relationships, a couple of computer programs were created. The best choice of a particular program depends on the structure of input data, the type of analysis requested and the properties of the population (Jones and Ardren 2003). Probably the most important difference among the software is the level to which they can operate with null alleles, mutations and scoring errors. The presence of null alleles may cause false exclusion of the true parent if the heterozygote is incorrectly scored as the homozygote. Null alleles could be recognized due to the significant allelic deviation from the Hardy-Weinberg equilibrium. In studies where one parent of the offspring is known, it is easier to detect null alleles. Incompatibility can occur between the known parent and its progeny in the case that the heterozygote with the null allele is incorrectly scored as a homozygote.

Ideally, large set of offspring with matched parents is available. In this case, the molecular techniques serve only to verify the relational links. If the number of known facts decreases, it becomes difficult to determine the pedigree. With knowledge of the maternal parents of each progeny and all sets of possible fathers, this method of parentage analysis is still very reliable. If all possible parents are not known, it may be impossible to identify the parents of all offspring and some parents may be selected incorrectly. Generally, the less input data for the analysis of pedigree is known, the more important are the

qualitative properties of molecular markers (Chakraborty et al. 1988). There are significant limitations that cannot be easily overcome. The information about total number of candidates in parental population plays an important role in determining the reliability of parentage analysis (Marshall et al. 1998). According to Jones and Ardren (2003), who reviewed different methods of parentage analysis, the software CERVUS and FAMOZ are quite sensitive to this parameter, while the program PATRI is able to tolerate the ambiguous size of a parental population.

The conservative approach of the software PROBMAX (Danzmann 1997) considers that each homozygote is potentially the heterozygote with the null allele. This procedure is supposed to be the most reliable if unambiguous results are generated. Other computer programs (CERVUS, FAMOZ, PAPA and PARENTE) consider null alleles as a type of mutation or scoring error. This approach is rather problematic, because it allows only one value of the error to be constant for all loci, and thus reduces the informative value of the loci in which null alleles do not occur.

The software CERVUS works with an algorithm that calculates the degree of deviation from the Hardy-Weinberg equilibrium and warns against the potential null alleles. Most studies assume that the alleles are inherited independently, that they are, in other words, unlinked. In the case of a small number of loci, the probability of their interactions is actually very small. If the parentage analyses were performed using the linked loci, it would reduce the likelihood of exclusion, because non-random spacing of alleles reduces the genetic variability among individuals. So far, no software currently works with linked loci, so it is best to try to avoid analysis of these loci (Jones and Ardren 2003).

Mutation and genotyping errors are another source of complication when the pedigree is studied. It is not easy to do reliable estimation of mutation rates, yet the factors determining them are uncertain (Whittaker et al. 2003).

Most software tools take these factors into account as one factor. The occurrence of these constraints leads to the fact that some individuals cannot be assigned to their biological parents. The mutation rate of microsatellites

varies in different species and might be the other drawback with those mentioned in chapter Potential microsatellite genotyping difficulty. relatively low. Highlighting their presence and revealing the discrepancies in the genotyping could be accomplished by comparing the genotype of offspring with the genotype of the known parent. A special algorithm overcoming this problem was developed (Marshall et al. 1998). For instance, the software CERVUS (Kalinowski et al. 2007) replaces the incorrectly scored locus with alternative genotype according to its expected frequency in the population. The software PARENTE (Cercueil et al. 2002) works on a similar principle, but only the specific allele is replaced instead of the entire locus genotype. The PAPA (Duchesne et al. 2002) program uses a different algorithm which allows a specific setting depending on the selected mutation model. Unfortunately it happens very often that there is not enough information to select an appropriate model. Other programs use simpler approaches. NEWPAD allows the user to set the number of disagreements that lead to the exclusion of parentage. The software KINSHIP, PATRI (Signorovitch and Nielsen 2002) and GERUD (Jones 2001) do not allow genotyping errors and mutations, but may serve as an indicator of the occurrence of such phenomena in the data set.

The majority of natural systems cannot determine the pedigree using the strict exclusion approach. It is therefore necessary to resort to the other methods of statistical analysis. The suitable software may be CERVUS, which allows the simulation of the expected statistical parameters. The parameter delta is the difference between the values of the two most likely fathers. CERVUS is able to determine the critical value, which is the milestone where the parent can be considered as a candidate for the true parent. This algorithm is very important, especially in the presence of mutations and genotyping errors.

The CERVUS analyses calculate the frequency of each allele at each locus in the population and take into account the Hardy – Weinberg equilibrium as well as the presence of null alleles. These analyses help to determine the suitability of loci for downstream analysis. These include: expected heterozygozity, polymorphic information content (Botstein et al. 1980), average exclusion

probabilities and, optionally, Hardy-Weinberg chi-square statistic and estimation of null allele frequency. Polymorphic information content (PIC) is a measure of informativeness related to expected heterozygosity and likewise is calculated from allele frequencies. It is a quantity commonly used in linkage mapping. PIC estimates the power of a marker for the detection of polymorphism (Guo and Elston 1999). Its value is dependent on the number of distinguished alleles per locus and on their frequency in population. The comparison between expected and observed heterozygozity allows estimation of the null allele frequency (Kalinowski et al. 2007).

High likelihood ratio indicates that the individual is a likely parent of the examined offspring. If the input genotypes are identified with 100% accuracy than the potential mismatch among the alleles of parents and offspring clearly excludes the parenthood. If the genotype is determined by scoring more loci, the probability of incorrectly identified or mutated allele is relatively high, although their frequency on a single locus is low. Errors may occur in the offspring genotype, maternal and paternal genotype or in various combinations of all three. Therefore, CERVUS used probabilistic equations for data processing, which take this into account. The advantage of this approach is that the individual with non-compliance on more loci is not from the set of potential parents excluded, but the likelihood that it is a true parent will be estimated as low. The frequency of mutations and null alleles can be adjusted by the user. The overall likelihood ratio of each individual is calculated by multiplying the ratios of particular loci. This approach assumes that the studied loci are inherited independently. If the likelihood ratio is calculated without any tolerance of errors, any error would lead to a ratio classified as zero and the LOD score would not be defined. The critical value of the LOD score and Delta cannot be assessed using the normal distribution; therefore it is required to use simulation of parentage analysis for determining the level of confidence and to determinate the limits of these parameters, in which the putative parent could still be estimated as the true parent. For the purposes of simulation, analysis must be entered into the software beside the genotypes as well as the number of parent candidates, the number of potential analyzed parents, the

proportion of analyzed loci and the error frequency. The putative parent whose Delta and LOD score exceeds the critical value, while fulfilling the criteria specified by the distribution of simulated analysis, is the proper parent with a strict (95%) or relaxed (80%) probability.

FAMOZ (Gerber et al. 2003) uses a similar algorithm as CERVUS. The software PAPA and GERUD are less suitable because they do not calculate the level of assignment. CERVUS and KINSHIP (Goodnight and Queller 1999) take into account the presence of null alleles. The main difference between them is that CERVUS works with the logarithm of likelihood ratio, while KINSHIP uses likelihood ratio. CERVUS works with the parameters of the entire population simultaneously; KINSHIP simulates the relationship parent – offspring individually. Nielsen et al. (2001) criticized the statistical parameter delta that it is determined only from the two most likely parents, ignoring the other candidate parents. PATRI uses the Bayesian method, when the probability of paternity is judged in relation to all potential fathers. PARENTE works in a similar fashion. NEWPAT (Wilmer et al. 1999) is based on an entirely different principle. The program creates a hypothetical set of possible fathers (based on the frequency of alleles) and tests what percentage of these fathers cannot be excluded.

II. Objectives of the dissertation

The main goal of this thesis is to estimate genetic diversity and carry out pedigree reconstruction in the half-sib progeny trial in Scots pine breeding program in the Czech Republic. This goes in line with accelerating tree improvement activities using the "Breeding without Breeding" strategy.

Specific objectives:

- Optimization of the DNA extraction protocol
- Selection of Scots pine microsatellite primers
- Optimization of the PCR conditions, microsatellite analysis - genotyping
- Pedigree reconstruction (including analysis and assessment of the microsatellite loci)
- Comparison of genetic parameters among pedigree free, sib-ship and combined markers-pedigree models

III. Materials and methods

Location of research work and contribution of the author

Department of Dendrology and Forest Tree Breeding, Czech University of Life Sciences Prague, Czech Republic

Abbreviation used in thesis: KDŠLD
Activities: sample collection, DNA extraction, optimization of extraction protocol, data analysis
Contribution of the author: Jiří Korecký extracted DNA, optimized the extraction protocol and analyzed the data (pedigree reconstruction). He collected samples in cooperation with Jan Kaňák and Milan Lstibůrek.

Genomac – genomic analysis company, Prague, Czech Republic

Abbreviation used in thesis: Genomac
Activities: genomic analysis (including primer selection and optimization of PCR protocols)
Contribution of the author: Jiří Korecký was responsible for the activities provided by Genomac.

Department of Botany, Charles University, Prague, Czech Republic

Abbreviation used in thesis: Department of Botany
Activities: grinding of plant tissue, measurement of DNA quality and concentration
Contribution of the author: Jiří Korecký grinded the plant tissue and measured characteristics of the DNA.

Department of Forest Sciences, University of British Columbia, Vancouver, Canada

Abbreviation used in thesis: UBC

Activities: selection of primers, optimization of PCR protocols, genotyping, data analysis

Contribution of the author: Jiří Korecký tested the primers, optimized PCR protocols and genotyped the data. He also participated on data analysis (genetic parameters comparison) with Jaroslav Klápště, Yousry El-Kassaby and Milan Lstibůrek.

Materials: Parental and half-sib progeny trees

Description of parental trees within the seed orchards and their corresponding half-sib progeny trees

Scots pine seed orchards in Western Bohemia (Scots pine of Western Bohemia provenience) are composed almost entirely from the same clones. Seed orchard Doubrava was established 5 years later than seed orchard Nepomuk – Silov and occupied almost three times bigger area. For the each seed orchard was established the corresponding progeny test. Individuals originated from seed orchard Doubrava with its progeny test on site Skelná huť and the trial for progeny from Nepomuk – Silov was established on the site determined as Nepomuk.

The total number of ramets per each clone is not constant, it ranges between 1 and 21 (Nepomuk-Silov) or 1 - 28 ramets (Doubrava) respectively. For other characteristics of seed orchards and their progeny such as geographic location and altitude, area in hectares, years of establishment and information about forest vegetation types, see Table 2.

Genetic analyses were carried out on half-sib progeny superior trees (circa 10% individuals of progeny test) from two geographically distinct trial plots (Table 2). Phenotypic pre-selection followed the methodical approach by Lstibůrek et al. (2011) to minimizing the fingerprinting effort, while meeting the prescribed effective population size.

Progeny trees on these plots were grown from seed originating out of two seed orchards, but there is no difference in the origin of seed orchard trees. The quantitative parameters such as high and diameter $d_{1,3}$ were measured. Branching, shape of trunk and mortality of particular trees were determined. For further information see Kaňák et al. (2009).

Table 2: Characteristics of seed orchards and their half-sib progeny plots

Seed orchard	**Doubrava**	**Nepomuk – Silov**
LČR Forest district	Plasy	Klatovy
Geographic coordinates	49°54'31.034"N, 13°26'33.605"E	49°28'52.587"N, 13°31'41.536"E
Forest type	3 I	4 S
Number of clones	87	45
Number of ramets	1165	410
Ramets per clone	1-28	1-21
Spacing	6 x 6 m	6 x 6 m
Total area (ha)	6.48	2.24
Year of establishment	1980	1975
Forest vegetation type	2	4
Altitude	380	480-500

Progeny test	**Skelná huť**	**Nepomuk**
LČR Forest district	Plasy	Klatovy
Geographic coordinates	49°55‘53.489“N, 13°6‘43.268“E	49°29‘40.735“N, 13°33‘5.702“E
Forest type group	5K5	3S1
Progeny size	85 clones, 320 ramets	38 clones
Number of plots	960 (+ 7 control)	163 (+ 6 control)
Number of replicates	3 (ramets)	4 (clones)
Spacing	0.7 x 1.4 m	0.7 x 1.4 m
Total area (ha)	1.23	0.81
Year of establishment	1994	1991
Forest vegetation type	5	3
Altitude	610	490

Progeny tests originated from the seed orchards were established in different conditions (Forest type group 5K5 versus 3S1, Table 2). Nevertheless, different forest type groups of both plots are very beneficial. For future utilization in operational forestry, it is better to use genetic entities with above-average characteristics on both locations (are therefore supposed to be more universal). Therefore Kaňák (2011) suggested to establish trial plots on

distinctive places but within the range of suitable conditions for particular forest tree species.

Selection of appropriate trees for genetic analysis was based on biometric evaluation of all half-sib progeny trees. Quantitative parameters such as maximal height, diameter at breast height (DBH), height to diameter ratio (HDR) and qualitative parameters including straightness and branching were measured. Statistical evaluation and estimation of breeding values were accomplished using linear mixed models in software ASReml (Gilmour et al. 2006). These evaluations (REML, BLUP analysis) were subject of the dissertation thesis by Jan Kaňák and already published (Kaňák et al. 2009).

6. DNA extraction and optimization of the extraction protocol

6.1. The collection of organic tissue and extraction of DNA

The deoxyribonucleic acid (DNA) for the genetic analysis was extracted from pine buds collected during their (post) dormancy from February to the beginning of April 2010. Buds were collected from the terminal branches of the highest parts of the crown. These terminal parts of trees are supposed to contain large buds appropriate for subsequent DNA extractions. Collected branches were immediately placed in lockable polyethylene bags and labeled clearly to secure their identity. These bags were kept in a cooling box during the day of collection (in the presence of solid carbon dioxide).

After samples being transported into the laboratory of KDŠLD, they were placed into an ultra-low freezer (- 80°C). Low temperature inactivates the function of nucleases and buds could be stored here relatively long (for months) without damaging the integrity of DNA. Covering scales were removed by sterile laboratory scalpel. The DNA was extracted from organic tissue following the common practice, e.g. Metzenberg, (2007).

The traditional way of DNA isolation is relatively time consuming and requires more equipped laboratory than the modern approach based on extraction kits. The conventional approach is called the CTAB method (Doyle and Doyle 1987). This method is based on the ability of CTAB (cetyltrimethylammonium bromide) to form a complex with nucleic acids, which is soluble at a high concentration of salt (0.7 M NaCl), but at a lower concentration (0.45 M NaCl) it creates a precipitate. CTAB also acts as a detergent, which releases DNA from the membranes and proteins. Another possible protocol was published by Dellaporta et al. (1983).

Nevertheless, in this case the The Invisorb® Spin Plant Kit for the extraction DNA was chosen, which allows rapid and efficient isolation of high-quality genomic DNA from buds of *Pinus sylvestris* L.. The purification procedure is rapid and requires no hazardous buffers, no organic extraction or alcohol precipitation. The method is based on a unique patented technology

without the use of toxic chaotropic salts. It enables fast and efficient isolation from fresh, frozen or dried plant material (STRATEC Biomedical 2011).

The core principle is the interaction between the negatively charged phosphate groups of the DNA and positively charged groups on the surface membrane in a spin filter placed in a tube, where the DNA is bound and, after the purification, released by the appropriate elution buffer. Binding or elution of the DNA depends on the concentration of salts and pH of the buffer.

The first step towards the isolation of the DNA is lysis of cells containing the desired nucleic acid. For the lysis of plant cells (containing cell walls) must be applied a mechanical force. The grinding of plant tissue frozen in nitrogen liquid can be done with the aid of an oscillating mill or a pestle.

6.2. Grinding of samples

Two different approaches of grinding the samples were tested. The simpler alternative used equipment of the KDŠLD consisted in the grinding of the sample by the aid of the sterilized plastic pestle into a tube. Each pestle must be sterilized before use by autoclaving (for 20 min at 120°C and at a pressure of 120 kPa). It is necessary to take extra care not to grind the sample with the same pestle again. The sample contaminated by the genetic material of the previous bud could result in incorrect genotyping.

The other approach being tested was grinding with an oscillating mill Retsch MM 400. Disintegration of frozen sample by oscillation was recommended by many authors (Vaughan and Russell 2004; Dzialuk et al. 2005). Two steel grinding balls were placed directly into the sample tube (sterilized following the same process as treatment of the pestles), the tubes were then positioned in a special box. This box was immersed in a liquid nitrogen (the temperature approximately - 196 °C) and cooled for several minutes. The chilled boxes were then attached to an oscillating mill. The grinding process ran for 2 minutes at a frequency of 25 cycles per second. In one series of crushing is possible to process up to 48 samples.

This method of homogenization shows off not only time saving, but also less intensive physical work of a laboratory technician. Provided samples were crushed with the pestle, the DNA must be isolated immediately. The reason is that these samples were homogenized into the lysis buffer, which is the first step of the isolation process. In case of processing buds using an oscillating mill, the lysis buffer is not added at this stage and therefore disintegrated samples may be stored in the ultra-cooling freezer again and the DNA extraction can be realized later.

6.3. Isolation and determination of the quality of DNA

The basic methodology of the DNA isolation was performed according to the manual Invisorb® Spin Plant Mini Kit, version 11.05 (STRATEC Biomedical 2011). Different time periods of the lysis phase were tested and the yield of the DNA in each time period was measured. The concentration of the DNA was measured in 30, 50, 70 and 90 minute period with the spectrophotometer Nanodrop 1000 at the Department of Botany.

The indicated parameters in this experiment were: (1.) the DNA concentration (ng/ul), and (2.) the purity of the DNA indicated by an absorption coefficient at wavelengths of 260 and 280 nm. The concentration of the nucleic acid sufficient for subsequent DNA analysis was determined to exceed 100 ng/ul.

The value of the absorbent coefficient of the pure DNA is claimed to be 1.8. The pure RNA is indicated by the absorbent ratio of 2. Proteins in the sample reduce this value below 1.7. Admixture of RNA, as well as the protein components in the sample, generally reduces the quality of genotyping outputs (Vondrejs and Storchová 1997). The producer of the extraction kit Invisorb declares the range of values of the absorbent ratio at which the DNA can be used for subsequent analysis between 1.6 and 2.0. It is recommended to inactivate the undesirable proteins by proteinase K. Nevertheless, the initial testing has proven very low content of proteins in these particular samples. Hence, the proteinase wasn't added to the sample, which simplified the extraction protocol. The RNA being presented in the samples was degraded by adding the RNase. The effect of 3 ul and 5 ul RNase in the sample was tested.

Another analyzed parameter was the influence of the initial centrifugation (12 000 rpm) over a total yield of the DNA. The initial centrifugation was recommended by researchers of Department of Botany as an additional step of extraction protocol to speed up the extraction process. If this step had no significant effect on the overall decrease of DNA yield, its implementation would be desirable. It would have saved time because it would prevent from clog of micro filters. Blockage of micro filters requires longer centrifugation during the extraction process.

The time demands in each tested scenario were evaluated. The variance of time required for grinding of the sample and extraction of the DNA was compared.

Tested scenarios

Table 3: Tested scenarios and their characteristics

Scenario	Disintegrated by	Grinding time (min)	Length of lytic phase (min)	Initial centrifugation
A1	oscillating mill	2	30	no
A2	oscillating mill	2	50	no
A3	oscillating mill	2	70	no
A4	oscillating mill	2	90	no
B1	oscillating mill	2	30	yes
B2	oscillating mill	2	50	yes
B3	oscillating mill	2	70	yes
C1	pestle	2	50	yes
C2	pestle	5	50	yes
C3	pestle	10	50	yes
D1	pestle	2	50	no
D2	pestle	5	50	no
D3	pestle	7	50	no
D4	pestle	10	50	no

The experiment included a total of 14 scenarios, each contained 12 samples. Scenarios varied in a way of grinding, grinding time, length of lytic phase and the initial centrifugation (Table 3). The group of samples coded **A1 - A4** was grinded in the oscillating mill, initial centrifugation of the lysate wasn't performed, lytic period: A1: 30 min, A2: 50 min, A3: 70 min, A4: 90 min. The variants **A3** and **A4** were also evaluated by yield of DNA during the first and second elution.

The variants **B1 - B3** were again grinded by the oscillating mill, in this series was conducted the initial centrifugation of the lysate. Lysis time: B1: 30 min, B2: 50 min, B3: 70 min.

Buds in variants **C1 - C3** were disintegrated with a plastic pestle, lysis time was in the entire series 50 min. The variants differ in the time of grinding of the buds (C1: 2 min, C2: 5 min, C3: 10 min.). All samples of series C was performed by centrifugation of the lysate.

Buds labelled **D1 – D4** were grinded by pestle, centrifugation of the lysate wasn´t performed.

Final elution of the DNA

Two DNA elution procedures of each sample were carried out. The first one took place in 150 ul of ultra-pure water, the second elution proceeded in the 100 ul of the elution buffer. The first elution lasted 15 min, the second one took 30 minutes.

Statistical analysis

Statistical analyses were performed using the software STATISTICA StatSoft ® 9. Paired difference (t-test) and the one-way ANOVA were followed by the Turkey test (StatSoft 2010).

6.4. Results of extraction protocol optimization

6.4.1. Influence of the concentration of RNase on DNA purity

The effect of concentration of the RNase on the value of the absorption coefficient was tested on samples of series A (see Fig. 2). In each of the four scenarios (A1, A2, A3 and A4), one half of the samples was treated by 3 ul of RNase and the other half by 5 ul of RNase. Significant difference in the absorption coefficient was revealed using the t-test (α= 0.05). The average absorption coefficient for 5 ul RNase 260/280 - 1.74 and for 3 ul RNase 260/280 - 1.8. These values are meaningful only as a comparative because proteins were not degraded by proteinase K. Nevertheless, the results confirm the general assumption that the lower amount of the active RNA in the sample, the lower the value of the absorption coefficient.

Since the difference in absorption coefficients shows to be significant, it was concluded that the optimized extraction protocol will conclude the step of addition 5 ul of RNase.

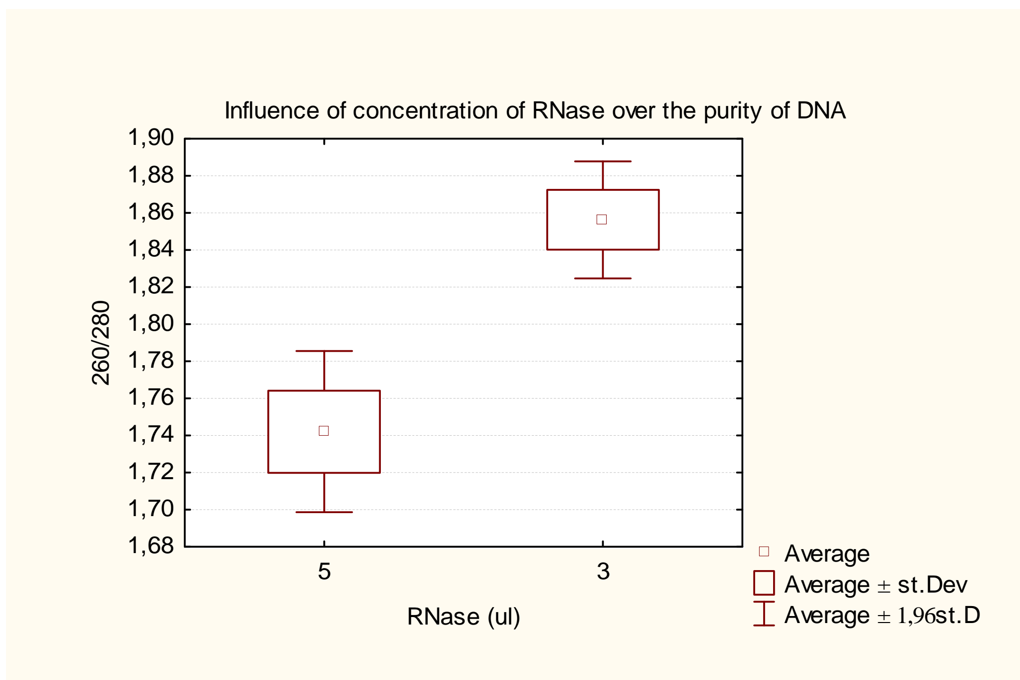


Fig. 2: Influence of concentration of RNase over the purity of DNA

6.4.2. Comparison of DNA gain using different scenarios of extraction

<u>Yield of DNA in series A</u>

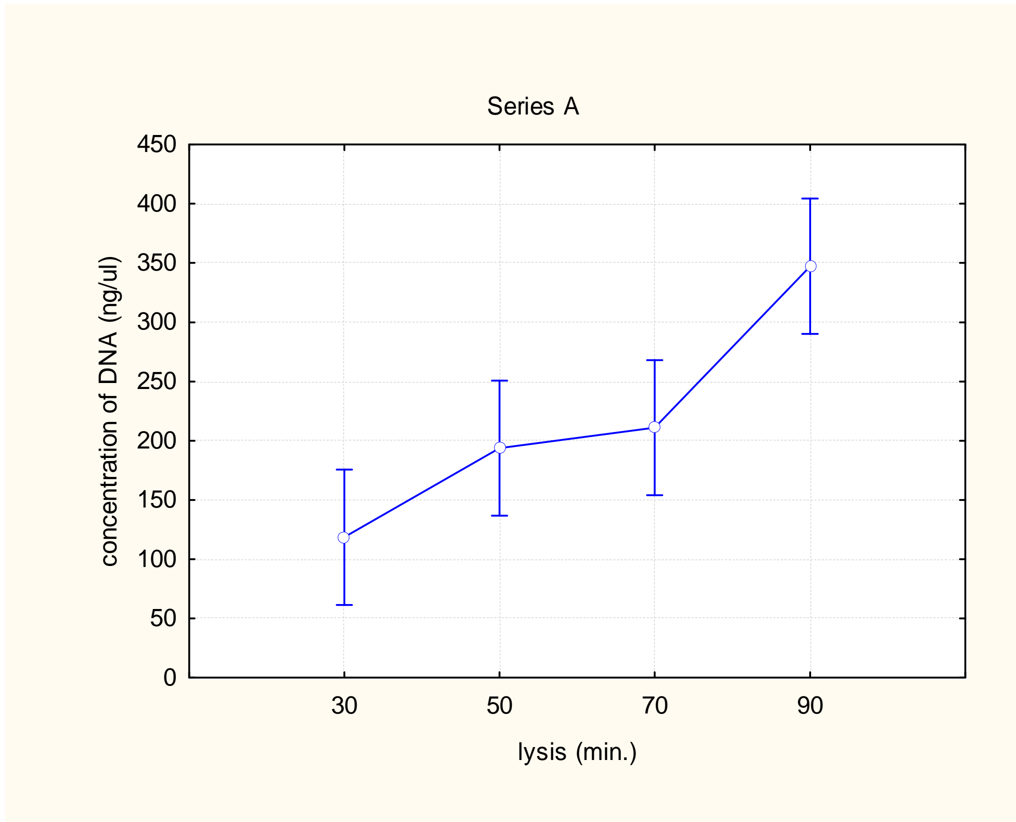


Fig. 3: Yield of DNA in series A

The lysis for 30 min led to the gain of DNA in an average concentration of 118.6 ng/ul. After adjusting, the deviations may not have always reached the required minimum value of 100 ng/ul. The lysis for 30 min was therefore evaluated as insufficient. The lysis for 50, 70 and 90 minutes was estimated as satisfactory, enabling to obtain more than determined amount of the DNA (100 ng/ul).

Yield of DNA in series B

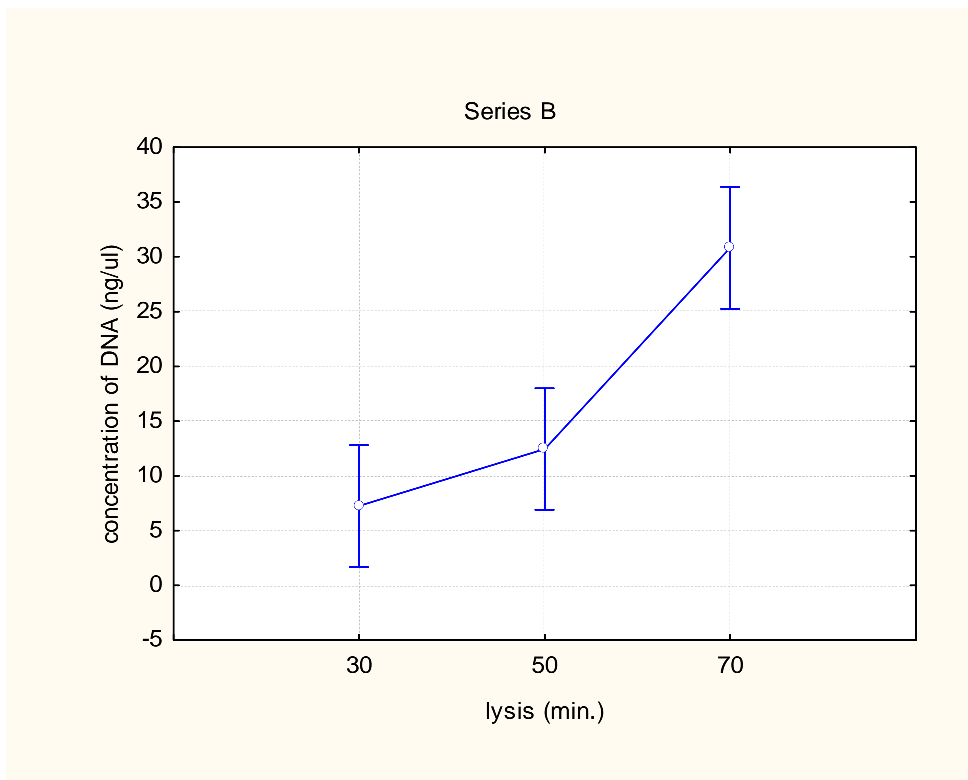


Fig. 4: Yield of DNA in series B

The lysis in the series B was carried out for 30, 50 and 70 minutes (see Fig. 4). Even when the lysis lasting for 70 minutes were conducted, the satisfactory concentration of the DNA wasn't obtained.

Yield of DNA in series C

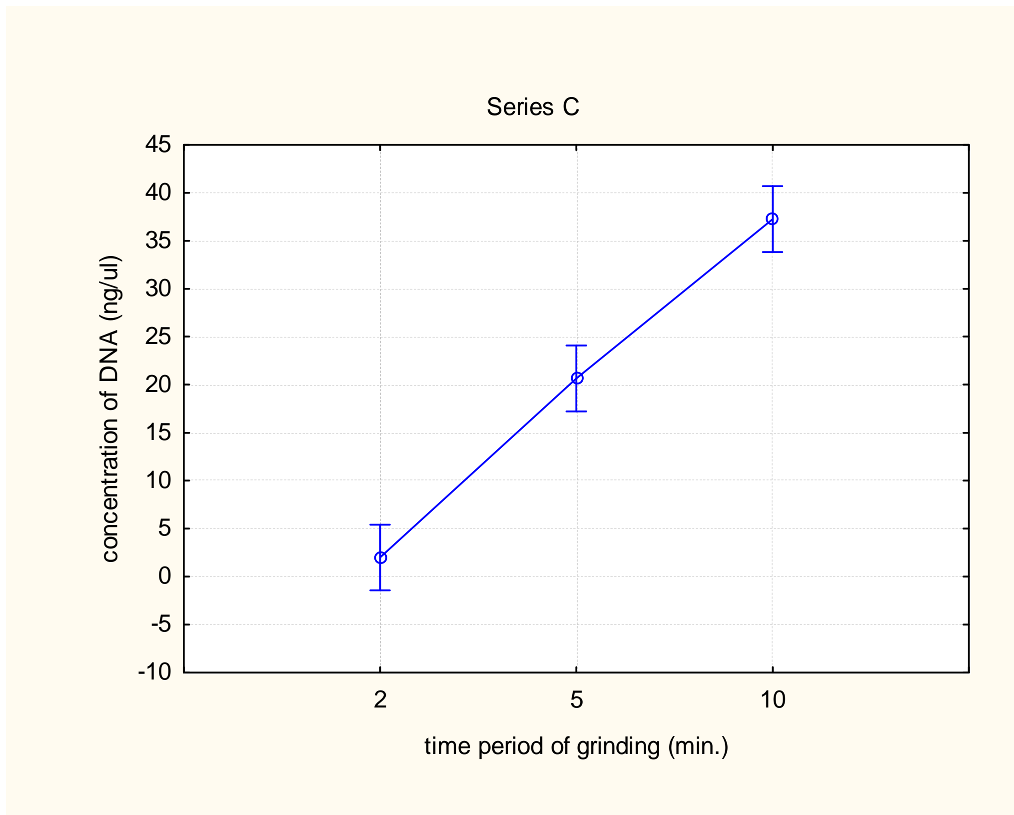


Fig. 5: Yield of DNA in series C

The gain of DNA is not sufficient at any time of grinding (2, 5 and 10 min) which is shown in Figure 5. The obtained yield was estimated lower than 100 ng/ul in all scenarios. The DNA yield at the time of grinding (2, 5 and 10 minutes) was found significantly different (Turkey post-hoc test).

Yield of DNA in series D

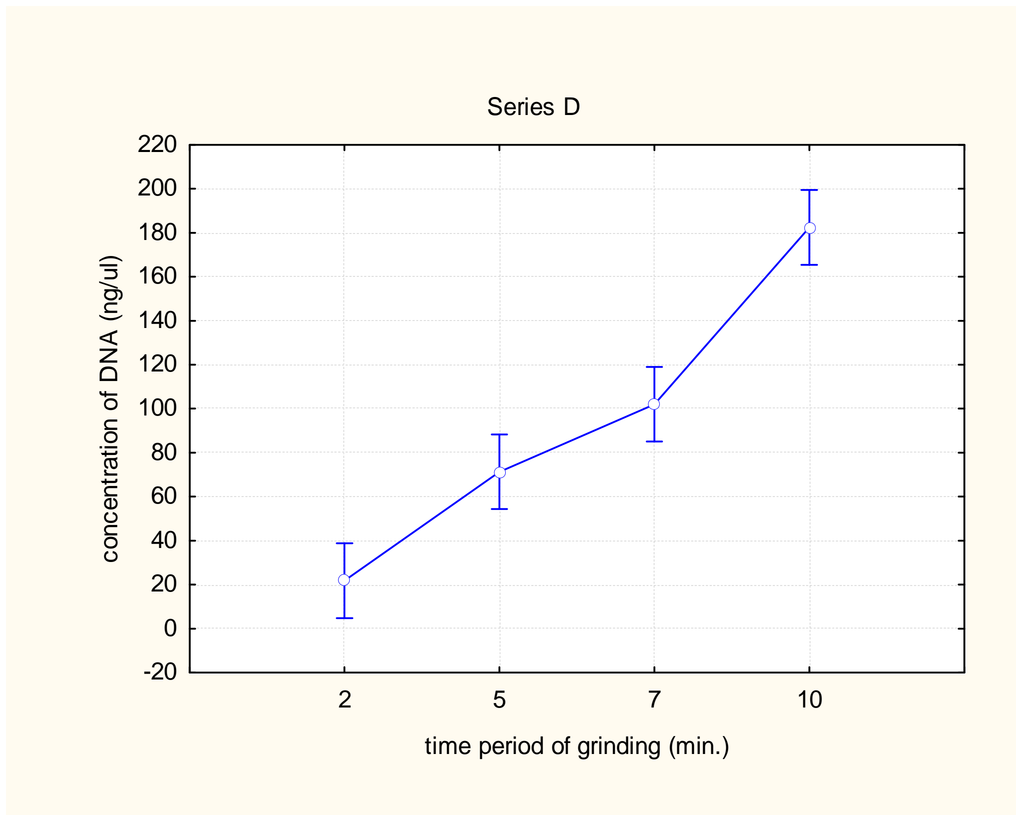


Fig. 6: Yield of DNA in series D

The DNA yield in scenarios of grinding by pestle without centrifugation of the lysate is shown in Figure 6. It is obvious that the sufficient yield (more than 100 ng/ul of the DNA) was achieved only by grinding for 10 minutes.

6.4.3. Mutual assessment of selected options with the lysis for 50 minutes

Scenarios with disintegration by oscillating mill and the lysis phase for 50 min (A2) compared with those grinded with the plastic pestle for 10 min and the length of the lysis phase 50 min as well (D4) are not significantly different. Both series were conducted without the initial centrifugation of the lysate. The concentration of DNA yield obtained in both cases was estimated more than the specified minimum value of 100 ng/ul. Variants with the initial centrifugation of the lysate grinding by oscillating mill (B2) and grinding with the pestle (C3) are not significantly different from each other; the gain of DNA was below the minimal limit. The chart of this mutual assessment is shown as Figure 7.

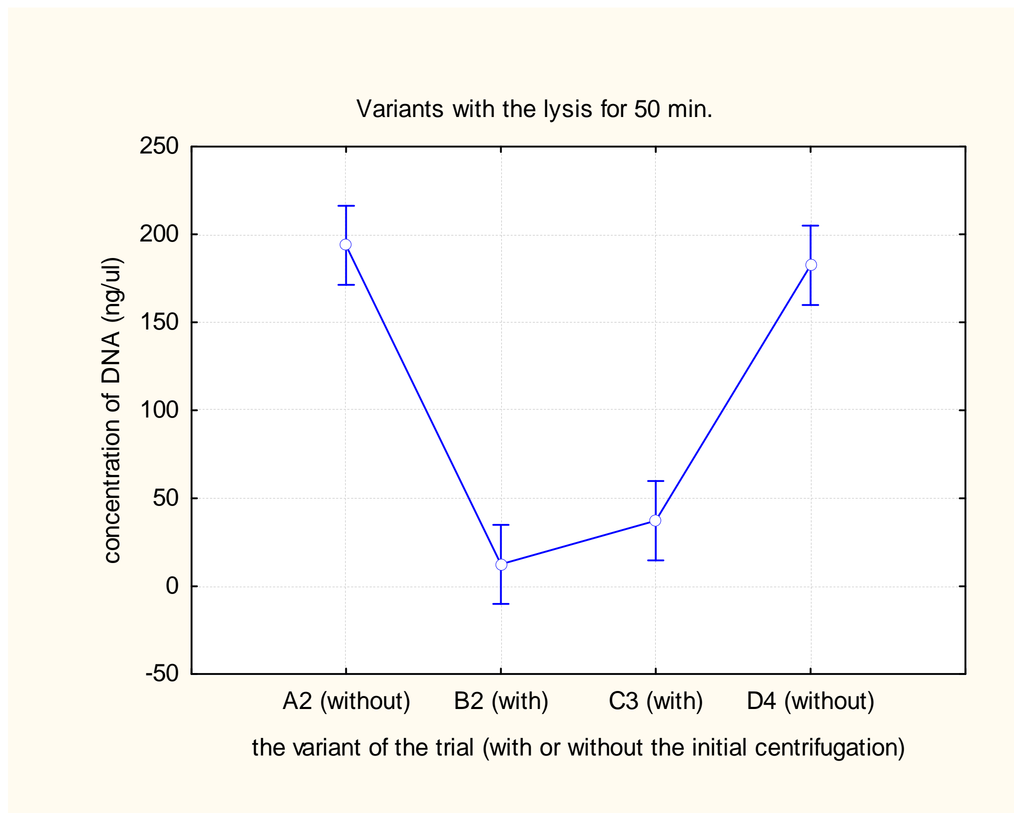


Fig. 7: Scenario with the lysis for 50 min

6.4.4. Comparison of the DNA yield in the first and the second elution

Concentration of the DNA in the first and the second elution was compared between the series A3 and A4. Concentrations of DNA obtained from the first and the second elution are significantly different which is also evident from Figure 8. The average concentration of DNA in the second elution of the A3 series was estimated 50 ng/ul. The concentration of the second elution did not reach the defined quantity.

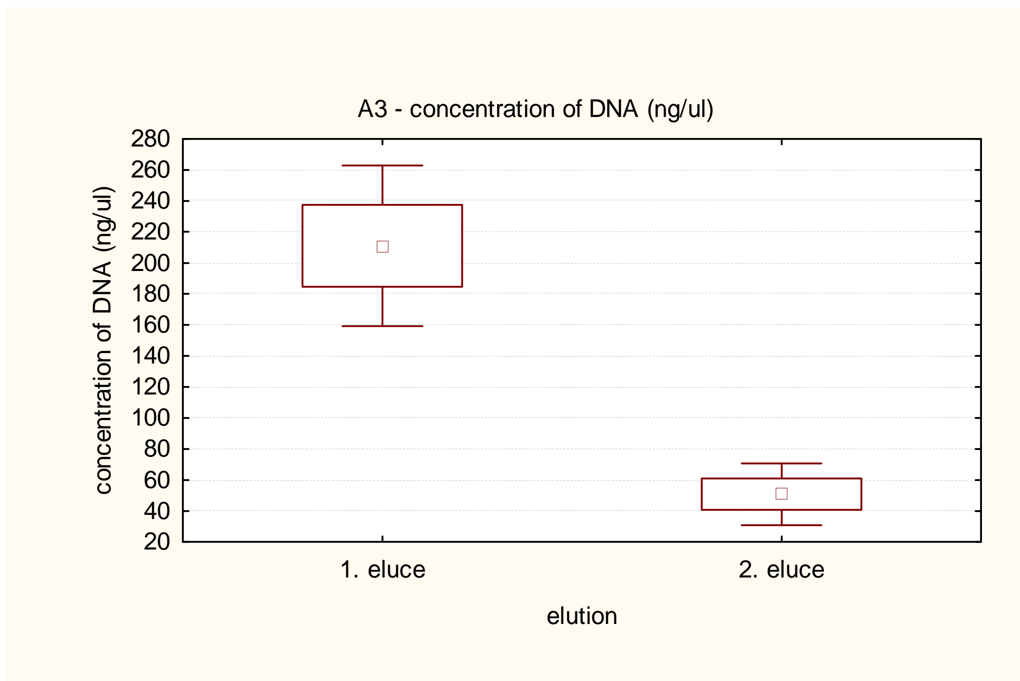


Fig. 8: Concentration of DNA in the variant A3

In case of scenario A4 (see Figure 9), the DNA yield during the first and the second elution are higher than the yields of the A3 series. Average concentration of DNA during the second elution was 170 ng/ul, which is more than the specified minimum value of 100 ng/ul.

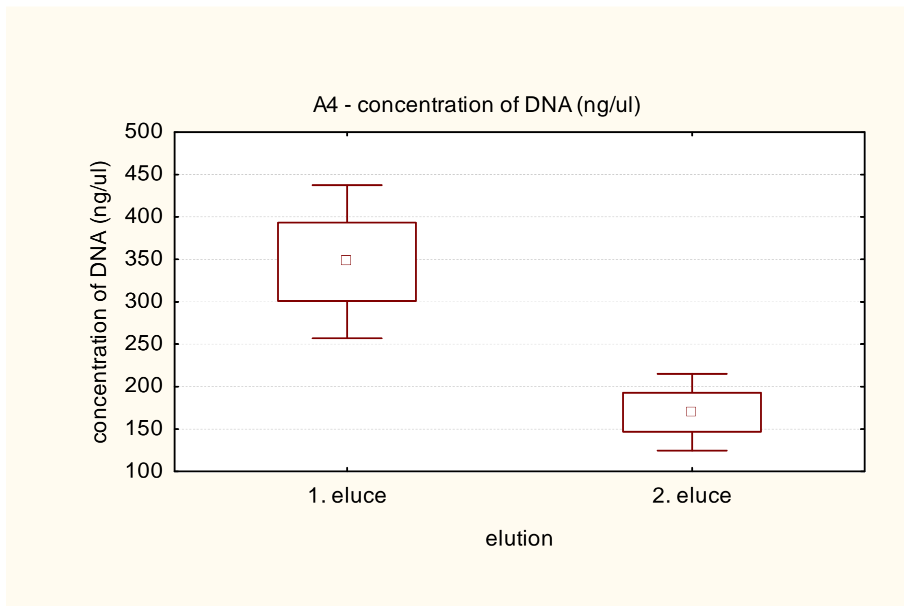


Fig. 9: Concentration of DNA in the variant A4

6.4.5. Timing consideration of the DNA isolation from plant material

Total time necessary to process a series of samples **=**

Time constant for all series (45 min) **+** Time variable phase

In Table 4, there are time-varying phases of all tested scenarios. Scenario A2 proved as the least time-consuming and concurrently with sufficient gain of DNA (crushing of samples in the oscillating mill and the lytic phase were concluded in 52 minutes).

In case buds have to be grind with the pestle, the D4 scenario was estimated as the most suitable one to reach demanded gain fo DNA (disintegration of the sample was done in 10 minutes).

The A2 scenario consisted of time-varying phase of extraction lasting 77 minutes. One series included extraction of the DNA from 12 samples. When the oscillating mill is not available, it is necessary to grind each single sample for 10 minutes with the pestle (i.e., 10 min for each of 12 samples). The time consumption in the time variable phase by the processing of one series was estimated 195 minutes.

Table 4: The time variable aspect of DNA isolation

Scenario	Time period of grinding (for series of 12 samples)	Time period of lysis phase	Accrual of time in scenarios without centrifugation	Time variable period of extraction (12 samples)
A1	2 (2)	30	25	57
A2	**2 (2)**	**50**	**25**	**77**
A3	2 (2)	70	25	97
A4	2 (2)	90	25	117
B1	2 (2)	30	0	32
B2	2 (2)	50	0	52
B3	2 (2)	70	0	72
C1	2 (24)	50	0	74
C2	5 (60)	50	0	110
C3	10 (120)	50	0	170
D1	2 (24)	50	25	99
D2	5 (60)	50	25	135
D3	7 (84)	50	25	159
D4	**10 (120)**	**50**	**25**	**195**

6.5. Discussion of DNA extraction protocol and its optimization

Scenarios with the initial centrifugation of the lysate were proven as ineffective, because the yield of the DNA did not reach in any modification the sufficient concentration. It could be concluded that the buds from Scots pine (it can be assumed that the validity of this fact can be extended to all morphologically similar buds of conifers) differ from the herbaceous plant material, in which the initial centrifugation does not affect the yield of the DNA. According to the experience of Lenka Flašková, researcher of the Department of Botany, (pers. comm.) the concentration of the DNA obtained from experiments with the initial centrifugation of the lysate is not significantly different from a scenario without the initial centrifugation and this additional step facilitates DNA extraction, because separated plant tissue (removed by initial centrifugation) does not block micro-filters and thus decreases time necessary for extraction process (congested micro-filters need to be centrifuged for longer). This approach did not seem to be applicable for Scots pine tissue and extraction without initial centrifugation despite longer extraction process is required. In case of crushing samples by oscillation mill and sufficient gain of DNA (i.e. scenario without initial centrifugation), the variant A2 is the most effective (Tab. 4; 50 min lytic period, total time of extraction 122 min (45 min time period constant for all scenarios and 77 min of variable time period specific for A2 scenario)).

Among "grinding by pestle" scenarios was found as the most effective D4 variant (Tab. 4; 10 min of grinding, total time of extraction 240 min). It can be concluded that the scenario "grinding by oscillation mill" (A2) should be clearly prefer as it is two times quicker and does not require hard manual grinding.

7. Selection of primers for the DNA analysis

Selection of suitable primers was conducted during the preliminary phase of the dissertation project. The isolation of microsatellite regions, their sequencing and the testing of primers can be time-consuming and expensive. Therefore one of the more efficient strategies for obtaining microsatellite markers is to review the literature (Scotti et al. 1999). Based on thematic peer-reviewed papers, 20 primers (loci) were found to be potentially suitable for *Pinus sylvestris* L. genotyping in the Czech Republic.

It was estimated, based on review of papers dealing with microsatellite analysis, that genotyping of 8-10 polymorphic loci should be adequate to provide statistically detectable differences in the pedigree reconstruction. Preliminary screening was performed on 20 loci, because practically all loci available in the literature and a central web database were specific to the particular species of genus *Pinus*. Attention was also focused on the loci of other species from the pine family (mainly *Pinus taeda* L. and *Pinus contorta* Dougl. ex Loud.), which were, in previous research, tested on *Pinus sylvestris* L.

Loci SPAC / SPAG published in Molecular Ecology (Soranzo et al. 1998) seem to be very promising, but unfortunately it was noted that some of them are difficult to score. Information about the number of alleles and heterozygozity was not available for some loci.

With the cooperation of Genomac, 10 markers presented in the Table 5 were chosen. Analyses of these loci were optimized as two multiplex PCRs.

Table 5: Markers used for the microsatellite analysis by Genomac

Locus	Repetition	Sequence of primers (5´-3´) forward/reverse	Primers described in details in:
SPAC 11.4	$(AT)_5(GT)_{19}$	TCACAAAACACGTGATTCACA GAAAATAGCCCTGTGTGAGACA	Soranzo et al. 1998
SPAC 11.6	$(CA)_{29}(TA)_7$	CTTCACAGGACTGATGTTCA TTACAGCGGTTGGTAAATG	Soranzo et al. 1998
SPAC 12.5	$(GT)_{20}(GA)_{10}$	CTTCTTCACTAGTTTCCTTTGG TTGGTTATAGGCATAGATTGC	Soranzo et al. 1998
LOP 1	$(TA)_{10}$	GGCTAATGGCCGGCCAGTGCT GCGATTACAGGGTTGCAGCCT	Liewlaksaneeyanawin et al. 2004
PtTX 2146	$(GCT)_4GCC(GCT)_7GCC(GCT)_8$	CCTGGGGATTTGGATTGGGTATTTG ATATTTTCCTTGCCCCTTCCAGACA	Auckland et al. 2002
PtTX 3025	$(CAA)_{10}$	CACGCTGTATAATAACAATCTA TTCTATATTCGCTTTTAGTTTC	Auckland et al. 2002
PtTX 3107	$(CAT)_{14}$	AAACAAGCCCACATCGTCAATC TCCCCTGGATCTGAGGA	Auckland et al. 2002
PtTX 4001	$(CA)_{15}$	CTATTTGAGTTAAGAAGGGAGTC CTGTGGGTAGCATCATC	Auckland et al. 2002
PtTX 4011	$(CA)_{20}$	GGTAACATTGGGAAAACACTCA TTAACCATCTATGCCAATCACTT	Auckland et al. 2002
SSrPT_ctg 64	$(CCG)_7$	GGAAGCTGTTACAAGTGCGG ATCGAGAAGAGAGGAAGGGC	Chagné et al. 2004

The later stage of the project, was performed at UBC (for details see chapter Location of research work and contribution of the author). At this department set of 12 primers from previously chosen 20 primers was optimized (Table 6) and these primers were used for DNA analyses. This step brought two main important aspects into this study – genotyping could be done taking all specifics and associations into consideration (e.g. assumed relatedness) which might contribute to the more accurate genotyping and the financial expense of genotyping might be lower.

Table 6: Markers used for the microsatellite analysis at UBC

Locus	Repetition	Sequence of primers (5´-3´) forward/reverse	Primers described in details in:
LOP 1	$(TA)_{10}$	GGCTAATGGCCGGCCAGTGCT GCGATTACAGGGTTGCAGCCT	Liewlaksaneeyanawin et al. 2004
LOP 3	$(TA)_{9}$	GTCTCCAGCCAGTTCACCTGC CAGTGGATCTGTCACCTCCTC	Liewlaksaneeyanawin et al. 2004
LOP 5	$(TA)_{33}$	AGCCGTAAAAGCTATCTTGTG GGCATACTTACATTTTAATAA	Liewlaksaneeyanawin et al. 2004
PtTX 2146	$(GCT)_4GCC(GCT)_7GCC(GCT)_8$	CCTGGGGATTTGGATTGGGTATTTG ATATTTTCCTTGCCCCTTCCAGACA	Auckland et al. 2002
PtTX 3025	$(CAA)_{10}$	CACGCTGTATAATAACAATCTA TTCTATATTCGCTTTTAGTTTC	Auckland et al. 2002
PtTX 3032	$(GAT)_{35}(GAC)_3GAT(GAC)_8…(GAC)_6AAT(GAT)_6$	CTGCCACACTACCAACC AACATTAAGATCTCATTTCAA	Auckland et al. 2002
PtTX 3107	$(CAT)_{14}$	AAACAAGCCCACATCGTCAATC TCCCCTGGATCTGAGGA	Auckland et al. 2002
PtTX 3116	$(TTG)_7…(TTG)_5$	CCTCCCAAAGCCTAAAGAATCATACAAG GCCTTATCTTACAGAA	Auckland et al. 2002
PtTX 4001	$(CA)_{15}$	CTATTTGAGTTAAGAAGGGAGTC CTGTGGGTAGCATCATC	Auckland et al. 2002
PtTX 4011	$(CA)_{20}$	GGTAACATTGGGAAAACACTCA TTAACCATCTATGCCAATCACTT	Auckland et al. 2002
SPAC 11.4	$(AT)_5(GT)_{19}$	TCACAAAACACGTGATTCACA GAAAATAGCCCTGTGTGAGACA	Soranzo et al. 1998
SsrPt_ctg 1376	$(AT)_{20}$	CGATATTATGGATTTTGCTTGTGA AAATGCATGCCAAACTTAAATAC	Chagné et al. 2004

7.1. Results and discussion of Scots pine primer selection

As it was stated earlier, development of the microsatellite primers is highly demanding and time-consuming process, requiring fully equipped molecular genetics laboratory. Fortunately, there are many of already developed primers for target species (Scots pine). Those primers are possible to order in the dried form, let them dissolve and add them to the PCR mixture.

Several peer-reviewed papers (Soranzo et al. 1998; Auckland et al. 2002; Chagne et al. 2004; Liewlaksaneeyanawin et al. 2004) were investigated to find the set of potentially suitable primers. Finally, 10 primers for genotyping by Genomac and 12 primers for genotyping at UBC were evaluated as suitable (based on their PIC, presence of null alleles and quality of allelic amplification) and their PCR conditions were optimized.

It is noteworthy to mention that there is set of primers which were independently chosen and optimized in both cases (LOP 1, PtTX 2146, PtTX 3025, PtTX 3107, PtTX 4001 and PtTX 4011), (for comparison see Table 5 and Table 6). The partial difference in selection of the most appropriate primers could be explained by different devices in both laboratories. PCR is very sensitive process which could produce qualitatively distinctive amplified fragments due to for example different trademark of thermal cyclers.

8. PCR protocols optimization and genotyping

8.1. PCR protocol

Genomic DNA as a component of reaction mixture was amplified by PCR using thermal cyclers (PE Applied Biosystems Gene Amp PCR System 9700 and Eppendorf Mastercycler gradient thermal cycler).

Each reaction mixture (10 µl) comprised of 1 µl PCR buffer (10 mm Tris, 1.5 mm $MgCl_2$, 50 mm KCl, pH 8.3), 1 mM dNTPs (0.25 mM each), 5 pmol of each forward and reverse primers, 0.3 pmol of M13 IRD-labeled primer (700 or 800 nm), 1 unit of Taq DNA polymerase (Roche), and 100 ng of genomic DNA. All primer sets contained one tailed primer with an additional 18 nucleotide tail complementary to infrared-labeled M13 primers, to allow incorporation of M13 labeled primers into PCR products for visualization on polyacrylamide gel. PCR reaction was carried out in PCR plates, each plate accommodated 60 genotypes.

PCR plates with reaction mixtures were initially denatured for 5 minutes at 94 °C to enable initial separation of DNA double strands. This step was followed by 25 - 35 cycles of 30 - 60 seconds of denaturation at 94°C, 30 - 60 seconds at a primer-specific annealing temperature, 30 - 60 seconds of extension at 72°C and a 10 minutes final extension at 72°C following completion of all amplification cycles (for details about PCR conditions see Table 7).

Table 7: Optimized PCR conditions

Locus	Number of cycles	Denaturation conditions (time, temperature)	Annealing conditions (time, temperature)	Extension conditions (time, temperature)
LOP 1	30	60 s, 94°C	60 s, 55°C	60 s, 72°C
LOP 3	30	60 s, 94°C	60 s, 48°C	60 s, 72°C
LOP 5	30	60 s, 94°C	60 s, 45°C	60 s, 72°C
PtTX 2146	15/15	60 s, 94°C	touchdown 30 s 57°C, decreasing 0.5°C/ 30 s 47°C	60 s, 72°C
PtTX 3025	15/20	30 s, 94°C	touchdown 30 s 60°C, decreasing 0.6°C/ 30 s 45°C	30 s, 72°C
PtTX 3032	15/15	60 s, 94°C	touchdown 30 s 59°C, decreasing 0.5°C/ 30 s 49°C	60 s, 72°C
PtTX 3107	15/15	60 s, 94°C	touchdown 30 s 60°C, decreasing 0.6°C/ 30 s 50°C	30 s, 72°C
PtTX 3116	15/20	60 s, 94°C	touchdown 30 s 55°C, decreasing 0.5°C/ 30 s 45°C	60 s, 72°C
PtTX 4001	15/15	60 s, 94°C	touchdown 30 s 60°C, decreasing 0.5°C/ 30 s 50°C	60 s, 72°C
PtTX 4011	15/15	60 s, 94°C	touchdown 30 s 60°C, decreasing 0.5°C/ 30 s 50°C	60 s, 72°C
SPAC 11.4	10/25	30 s, 94°C	touchdown 30 s 60°C, decreasing 1°C/ 30 s 50°C	30 s, 72°C
SsrPt_ctg 1376	10/15	30 s, 94°C	touchdown 30 s 60°C, decreasing 1°C/ 30 s 50°C	60 s, 72°C

8.2. Product visualization and genotyping

After PCR, all samples were mixed with 3 μl stop dye (LI-COR Inc., Lincoln, NE) and denatured at 95°C for 3 minutes. Denatured samples were separated by LI-COR 4300 automated sequencer based on molecular weight corresponding to the number of base pair repeats, on 25 cm long, 0.4 mm thick, 6% Long Ranger polyacrylamide gels (LI-COR Inc., Lincoln, NE).

The LI-COR system enables to detect DNA using infrared (IR) fluorescence. In a sequence reaction, the DNA polymerase incorporated forward or reverse primer labeled with dye into a set of chain-terminated fragments which separate according to size on an acrylamide gel. A diode laser excites the infrared dye on the DNA fragments as they migrate past the detector window. Two regions of the spectrum could be detected at the same time without spectral overlap (using two heptamethine cyanine dyes absorbing and fluorescing in absorption maxima about 700 and 800 nm) between detection channels, i.e. two independent image files are created from the same gel during electrophoresis. The raw image data are a series of bands displayed on the monitor in a format similar to an autoradiogram. Allele size was determined by 50 – 350 or 50 - 700 base pair sizing standards (depending on particular primer) loaded on the gel.

During analysis, lanes are found for each sample and bases are defined. The sequence data could be presented in the text (ASCII) or standard chromatogram curve format. Outputs obtained from the LI-COR were analyzed using automated microsatellite software Saga™ (LI-COR 2002) that is the software developed for the purposes of LI-COR's gel evaluation.

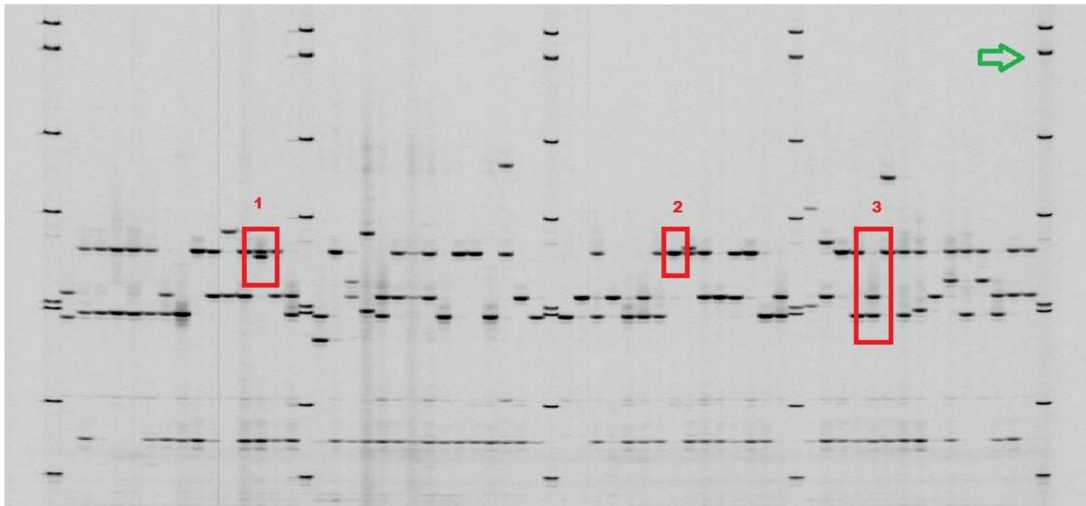


Fig. 10: Image of the polyacrylamide gel B2 (locus PtTX 2146) obtained using LI-COR

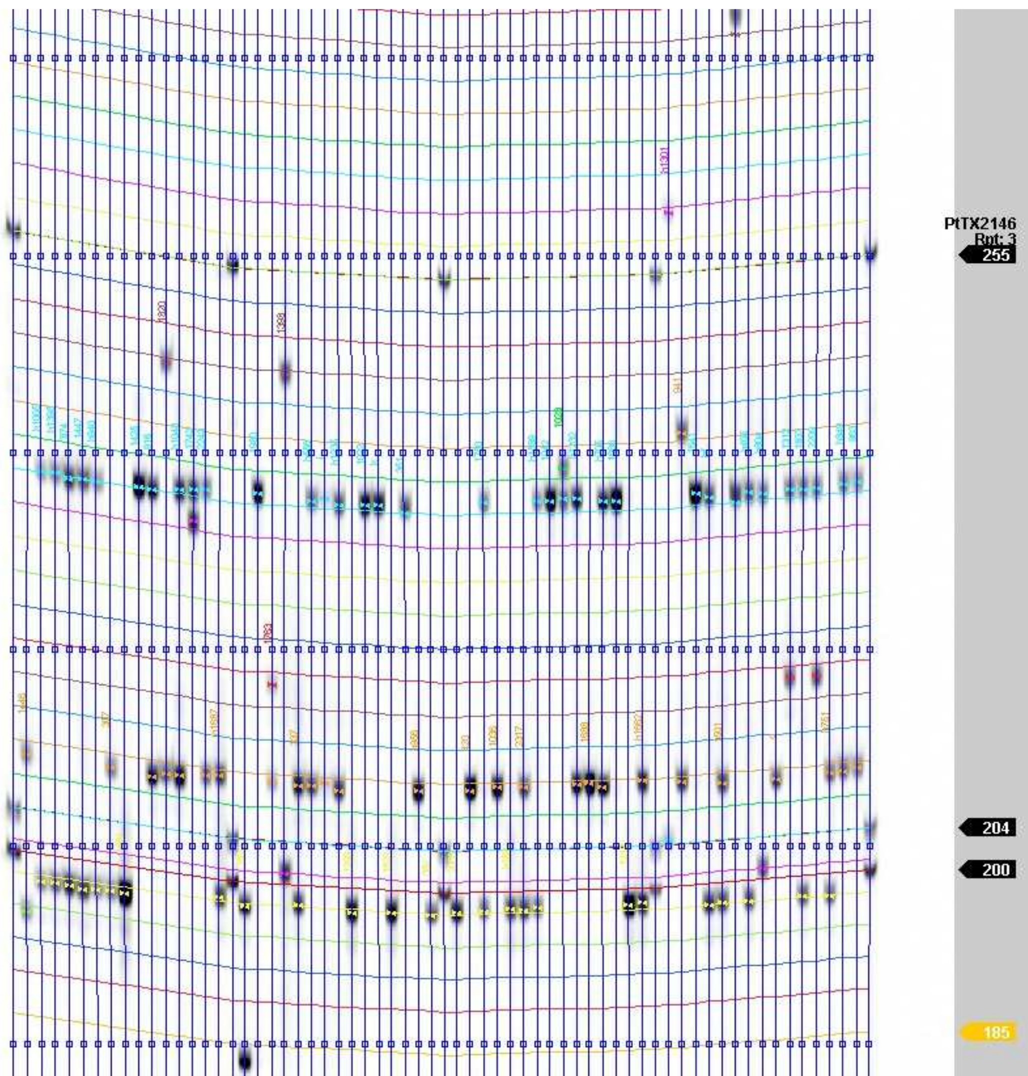


Fig. 11: Analysis of the gel B2 (the same as Fig. 10) using software Saga™

Comments on figures 10 and 11

Figure 10:

Each column represents genotype of single individual. Primer PtTX 2146 shows high level of polymorphism which enables to distinguish among many individuals based even only on this single locus.

Genotype of individual 1 (1742): alleles 231 and 234 bp – heterozygote differs in allele size about one repetition (3 bp)
Genotype of individual 2 (1042): alleles 234 bp – this individual appears to be homozygote at that locus
Genotype of individual 3 (1501): alleles 198 and 210 bp

The column emphasized by green arrow represents 50 – 350 bp size standard composed of 14 unambiguous labeled DNA fragments which help to determined size of amplified alleles.

Figure 11:

The same gel as figure 10 analyzed using software Saga™. Every individual was detected (blue vertical lines) and sizes of all potential alleles were set up (colored horizontal lines). Primary analysis was done automatically by the software, nevertheless gels required manual adjustment and precise re-analysis by experienced researcher. The genotypes determined from the gel B2 are shown in Table 8.

Table 8: Demonstration of genotyping results of the gel B2

Sample ID	Locus PtTX 2146		**Sample ID**	Locus PtTX 2146	
1446	195	210	**2053**	198	198
h1005	198	234	**420**	210	210
h1395	198	234	**1750**	198	234
874	198	234	**1036**	210	210
1447	198	234	**h153**	198	198
h940	198	234	**2317**	198	210
307	198	210	**h1589**	198	234
1681	198	198	**1042**	234	234
1425	234	234	**1028**	234	237
916	210	234	**h1232**	210	234
1820	210	246	**1688**	210	210
h1043	210	234	**h705**	210	234
1742	231	234	**1666**	234	234
2243	210	234	**1219**	198	198
h1687	198	210	**h1662**	198	210
1624	198	198	**h1301**	204	261
1980	234	234	**941**	210	240
1763	210	219	**1641**	234	234
1398	201	246	**hf**	198	234
337	198	210	**1501**	198	210
h556	210	234	**h1643**	234	276
i	210	234	**h405**	198	234
h1235	210	234	**h804**	201	234
1299	198	198	**o**	210	210
1022	234	234	**1312**	219	234
k	234	234	**1837**	198	234
1622	198	198	**2209**	219	234
351	234	285	**h751**	198	210
h956	210	210	**h849**	210	234
382	198	198	**963**	210	234

8.3. Results and discussion of optimization of genotyping process

The chapter PCR protocols optimization and genotyping is focused on the process of PCR optimization and subsequent procedure of genotyping at UBC.

Each reaction mixture comprised of defined volumes of reaction components (8.1. PCR Protocol) and each primer was optimized for thermal cyclers available at this department. Details of PCR conditions can be seen in the Table 7. Number of PCR cycles varied in range 25 – 35. Nine primes were treated using touchdown PCR procedure (Don et al. 1991) to set off amplification of target microsatellite alleles - bands. The decreasing of temperature was mostly 0.5 °C (0.6 and 1°C per cycle respectively). The first 15 cycles (except loci LOP 1, LOP 3 and LOP 5) were treated using touchdown procedure (Table 7).

Products of PCR reactions were separated by fluorescence sequencer LI-COR 4300. This part of the genotyping process is relatively time consuming and demanding as each PCR product (each primer) must be loaded separately on the polyacrylamide gel and each gel accommodates 60 individuals at the time. The option for acceleration and simplification of this process would be the multiplex PCR approach. Final results of genotyping are conveniently displayed in the form of multiloci table, where every single row represents genotype of one individual. Representative list of the multiloci genotypes is provided in Table 9.

Table 9: Demonstration of multiloci genotypes

ID	LOP 3		LOP 1		LOP 5		PtTX 3025		PtTX 4001		PtTX 2146		PtTX 3032		PtTX 3107		PtTX 3116		PtTX 4011		SPAC 11.4		SSRPT_CTG 1376	
1047	228	250	179	179	210	210	304	304	234	238	235	235	387	393	179	179	-1	-1	278	278	157	159	138	142
1049	228	250	178	190	212	-9	289	289	234	236	199	235	387	417	185	185	144	180	278	278	177	157	138	140
1054	228	228	176	182	210	214	304	325	236	238	235	235	351	414	182	182	174	180	278	284	173	157	136	136
1078	228	230	178	178	196	196	304	304	236	238	199	199	381	456	179	194	-1	-1	278	278	157	157	136	140
1080	228	228	178	-9	196	196	289	304	236	238	199	211	399	402	173	185	138	138	278	278	157	157	140	140
1099	228	228	178	182	210	226	292	304	234	236	205	211	408	426	173	179	165	180	278	278	157	161	138	140
1101	228	228	176	184	208	208	289	289	234	236	211	235	393	405	188	188	174	180	276	278	163	159	138	140
1108	230	250	182	184	210	210	289	289	236	240	199	199	399	420	182	-9	138	174	278	278	157	161	136	140
1109	-1	-1	178	178	206	210	289	304	238	240	199	250	417	483	173	182	138	138	278	298	175	157	136	140
1111	228	250	178	178	-1	-1	289	304	236	240	199	220	420	483	188	188	174	174	278	278	165	157	136	138
1112	228	228	-1	-1	202	210	289	289	236	240	199	235	384	480	179	182	192	195	276	278	171	157	140	140
1125	228	250	178	180	196	210	289	289	234	236	235	235	372	387	188	188	174	180	278	278	157	157	138	140
1129	228	250	182	179	208	212	-1	-1	234	236	235	235	387	396	185	185	174	180	278	278	157	159	138	140
1131	228	250	178	178	202	208	295	295	236	238	199	235	402	405	188	188	174	180	278	278	169	181	140	140
1132	228	250	178	186	202	-9	289	289	238	-9	199	235	387	435	185	185	144	174	278	278	169	161	140	140

Comment: Values of the genotypes correspond with the length of alleles (bp), parameter -1 means missing value, -9 null allele

9. Pedigree reconstruction

The complete study of parentage analysis means that the sampled offspring is assigned its true maternal and paternal parent. Existing techniques of parentage reconstruction attempt to reconstruct the pattern of parentage in a population (Jones and Ardren 2003). The pedigree of Scots pine individuals was determined with the aid of the software CERVUS (Kalinowski et al. 2007).

The data in input file represent each locus as two positions of subsequent columns (diploid genome) and the relevant line contains allelic size of all analyzed loci of the particular individual (e.g. Table 9). Entire set of input data is in common scenario represented by genotypes of the offspring and known parents (mothers) and genotypes of individuals from a group of potential fathers

The first step in the estimation of the pedigree requires the allele frequency analysis. CERVUS works with an approach based on the use of appropriate statistical characteristics in combination with simulations allowing the determination of the reliability of the outgoing parameters. The parentage analysis is based on an exclusion approach. To estimate the most likely parents, CERVUS uses two basic pieces of information, such as information about the allele frequency of offspring and the allele frequency of parental candidates, with a link of awareness about their homo- or heterozygozity.

9.1. Description of individuals entering into pedigree reconstruction

Parental trees

Parental DNA was obtained from seed orchards Plasy and Nepomuk. At least one ramet per each clone (according to the records) was analyzed. To evaluate the reliability of the records, many clones were genotyped at more ramets (up to 22 ramets per one clone), in total 204 parental trees were genotyped.

Subsequent analyses of parental genotypes showed some discrepancies between records and the real situation revealed by microsatellite analysis. Some trees were estimated to be mislabeled (belonging to the different clone) and even few trees with the genotype out of all recorded clones were found. These "extra clones" were added to the set of potential parents since the pedigree reconstruction is supposed to be more accurate.

Progeny trees

In total, 597 superior individuals (284 from Skelná huť and 313 from Nepomuk progeny test trials) were genotyped. Each progeny individual was meant to belong to the defined maternal half-sib (as we know the maternal parent according to the records).

For further details on parental trees and progeny plots, see section Parental and half-sib progeny trees on page 39.

9.2. Results and discussion of pedigree reconstruction

9.2.1. Allele frequency analysis

The results of allele frequency analysis are shown in Table 10. All loci are defined to be highly polymorphic, with the exception of the locus SsrPt_ctg 64, having low level of PIC and relatively high frequency of estimated null alleles. The column k in the Table 10 displays the number of alleles (allelic types) that were detected by genetic analysis. 49 alleles were found on the most polymorphic locus SPAC 12.5 in the population of 597 analyzed progeny trees and their potential parents inside seed orchard, i.e. 88 individuals with distinctive genotypes. The comparison of expected and observed heterozygozity implies rate of estimated null allele frequency. Loci SPAC 11.6 and SsrPt_ctg 64 were estimated to have highly significant deviation from Hardy-Weinberg equilibrium.

Table 10: Allele frequency analysis

Locus	k	HObs	HExp	PIC	F(Null)
SPAC 11.4	17	0.640	0.804	0.787	+0.1091
SPAC 11.6	46	0.418	0.954	0.952	+0.3919*
SPAC 12.5	49	0.847	0.943	0.940	+0.0378
LOP 1	14	0.473	0.663	0.616	+0.1676
PtTX 2146	21	0.757	0.773	0.744	+0.0087
PtTX 3025	12	0.508	0.534	0.478	+0.0205
PtTX 3107	14	0.331	0.371	0.784	+0.0655
PtTX 4001	14	0.770	0.740	0.711	-0.0237
PtTX 4011	12	0.508	0.695	0.649	+0.1566
SsrPt_ctg 64	11	0.272	0.809	0.349	+0.4958*

k – Number of alleles at the locus

Hobs – Observed heterozygozity

HExp – Expected heterozygozity

F(Null) – Estimated null allele frequency

***** indicating highly significant deviation from Hardy-Weinberg equilibrium

9.2.2. Pedigree reconstruction – assignment of parents

As revealed by clone identity verification, there is some inconsistency between seed orchard records and the real situation. Therefore it was decided instead of originally intended paternity analysis with defined mothers to perform parent-pair analysis with unknown sexes, i.e. every parental genotype was considered as potential maternal and paternal parent.

The analysis assigned both parents to 95.8% of progenies at the significant level of probability (35.7% with strict probability and 64.3% with relaxed probability). The value 4.2% of unassigned parents could be explained either by pollen contamination from the trees outside the seed orchard or the possible occurrence of another "extra clones" whose genotype wasn't revealed by clone identity verification as not all trees in the seed orchard were genotyped. The low value of unassigned parents fulfilled and even exceeded the theoretical expectation that the selection of phenotypically superior progenies (truncation selection) eliminates the influence of pollen contamination because they originated from superior parents within the seed orchard.

Lstibůrek et al. (2012) claimed that truncation selection increases the number of genotyped individuals from inside the seed orchard and reduced type II error (false exclusion). In situation where genotyping of pre-selected trees revealed gene contamination to be high the technique of supplemental mass pollination could be considered as an option (El-Kassaby and Lstibůrek 2009).

The significantly assigned parents of each individual were compared with the assumed mothers. There was the accordance of 77.1% with the records, i.e. one of the parents has the same genotype as the genotype of assumed mother. The rest of assigned parents did not match with the assumed mother. Therefore it could be stated that 22.9% individuals from the analyzed subset of progenies have different maternal contribution from the seed orchard than declared.

The subsequent analyses describing the family structure of progeny trees (Figure 12, 13 and 14) were done excluding individuals that did not match with

assumed mother. This approach enables to reveal paternal gametic contribution and its structure despite of pair analysis with unknown sexes since the maternal contribution is known and fixed. Estimation of parental gametic contribution in progeny of seed orchard's trees is the prerequisite for estimation of effective population size (diversity) and genetic worth. Effective population size can be explained as the size of ideal population comprised only unrelated and non-inbred individuals.

Briefly, the effective population size or effective status number N_e (Lindgren et al. 1997) of analyzed progeny was estimated using following equation:

$$N_e = \frac{1}{2\Theta} \qquad [1]$$

Where Θ is the average co-ancestry of all pairs of considered individuals including itself, i.e. their relatedness (Lindgren and Mullin 1997). Co-ancestry is the probability that any two alleles sampled randomly from two individuals are identical by descent (Malécot 1948). Concept of average co-ancestry Θ was introduced by Cockerham (1967):

$$\Theta = \sum_{i=1}^{N} \sum_{j=1}^{N} r_i r_j \, c_{ij} \qquad [2]$$

where N is the population size, r_i and r_j are parental representations of parent i and j and c_{ij} is co-ancestry between them (Lindgren and Mullin 1998). The kinship coefficients (co-ancestry) are as follows: c_{ij} = 0.5 for selfing, c_{ij} = 0.25 for full-sibling or parent-offspring pair, c_{ij} = 0.125 for half-sibling and c_{ij} = 0 for individuals without genetic relationship (unrelated and non-inbred individuals).

The effective population size for the set of analyzed individuals based on pedigree information was estimated N_e = 38.28.

It is important to mention that the pedigree structure greatly depends on the parental population size participated in mating. Lstibůrek et al. (2011) developed a probabilistic model to assist in calculating the appropriate sample

size needed to fulfill required parameters (effective population size). Based on their model they assume 33% higher genotyping effort to gain the same effective population size (N_e = 10) in scenario with 30 parents compared to the scenario with 60 parents respectively.

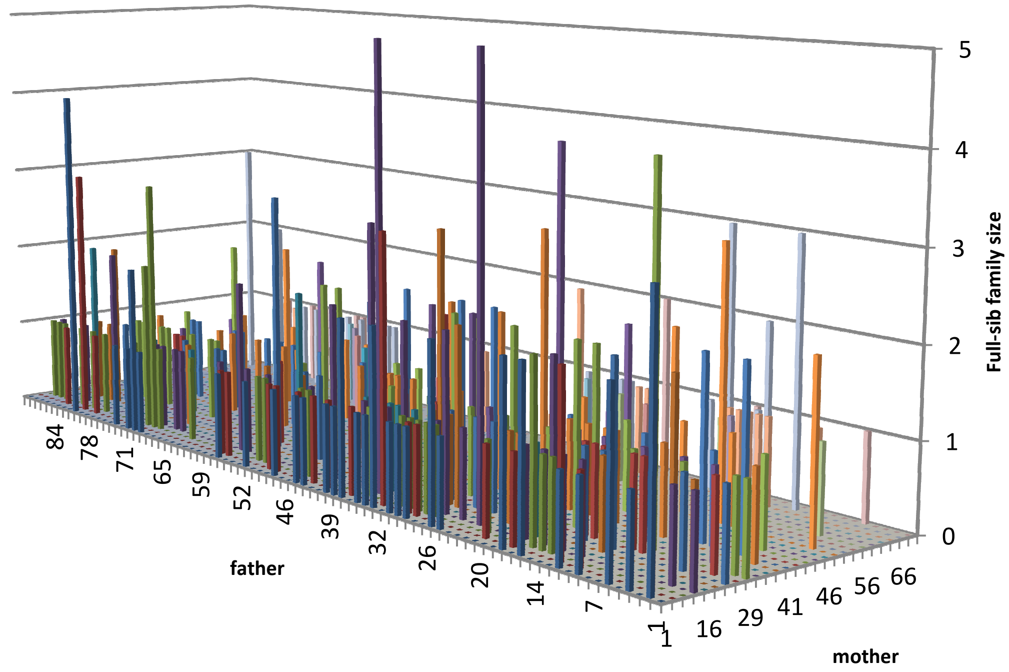


Fig. 12: Parental gametic contribution as revealed by pedigree reconstruction

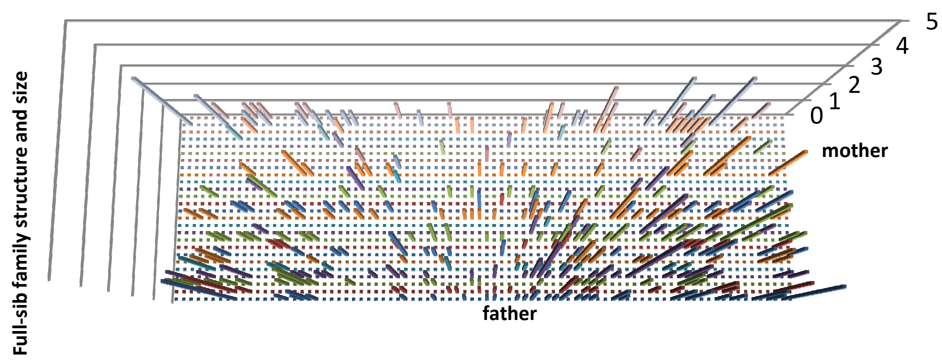


Fig. 13: Familial structure

Figure 12 pictured parental gametic contribution within subset of analyzed individuals. In total, 94.3% of all possible parental trees contributed as male into at least one progeny individual. For further details see Figure 14 and Table 11 in which other parameters were summarized. The parameters such as mean of half-sibs and their variances were estimated.

Table 11: Description of family structure at each site separately and across both sites for pedigree reconstruction - unassigned individuals excluded (total number of analyzed individuals bracketed)

Parameter	Skelná huť	Nepomuk	Both sites
# of individuals	217 (284)	208 (313)	425 (597)
# of defined maternal half-sibs	18	14	26
Mean of maternal half-sib size	12.056	14.857	16.346
Variance of maternal HS	100.879	121.670	167.835
# of potential male parents	88	88	88
# of contributing males (paternal HS)	75	67	83
Mean of paternal half-sib size	2.893	3.105	5.120
Variance of paternal HS	5.151	4.428	11.327

On the progeny trial Nepomuk (Fig. 14B) from which 208 individuals were analyzed, the most successful father contributed to 10 offspring, whereas the most successful paternal tree on the progeny trial plot Skelná huť (Fig. 14A, 217 progeny trees in total) passed its genetic information on 13 individuals. The most successful fathers at each site did not contribute much on the other site (10 versus 1 and 13 versus 3 respectively). Gametic contributions of all parents are pictured for all scenarios (site Nepomuk, site Skelná huť and both sites simultaneously on Figure 14).

All confirmed maternal and assigned parental trees, i.e. reconstructed pedigree, are enclosed as Appendix 3 of this thesis.

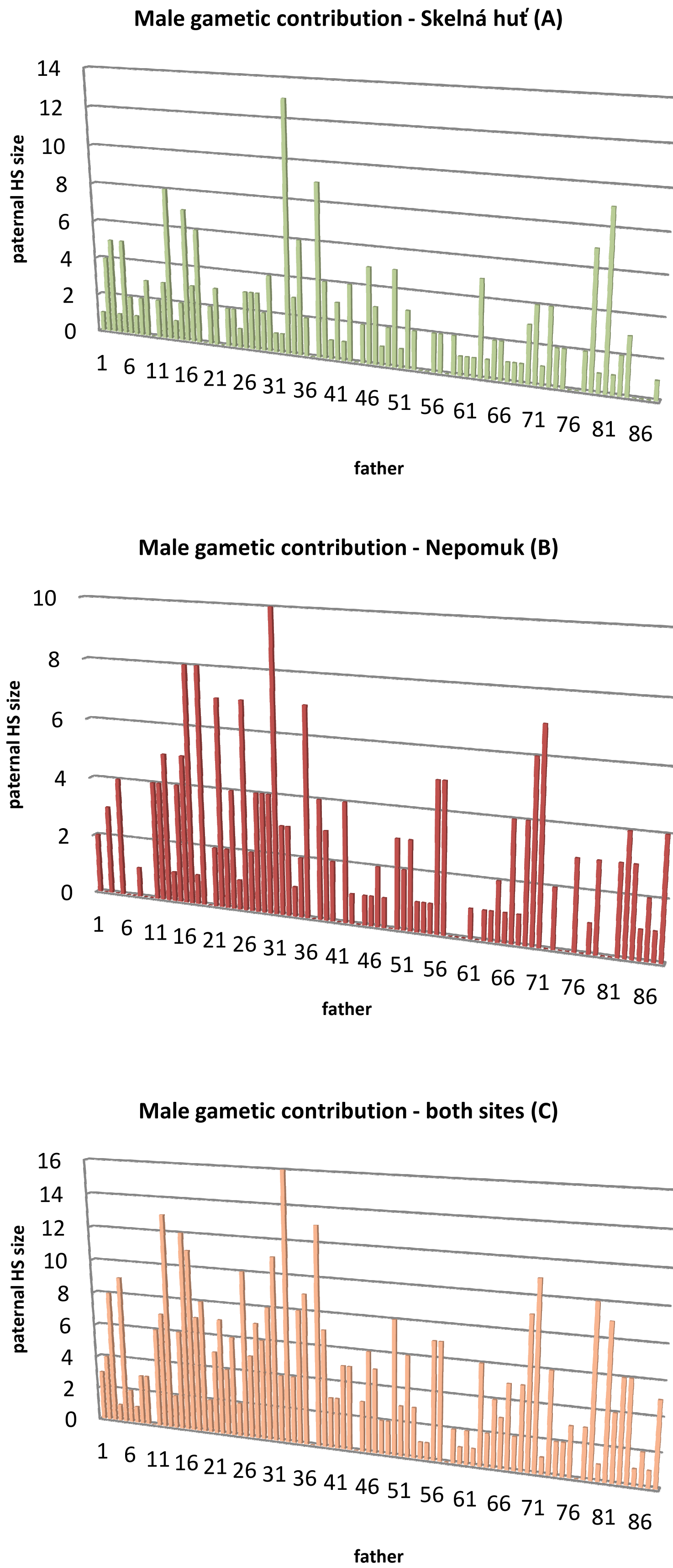


Fig. 14: Male gametic contribution on each site and on both sites

Figure 15 represents the cumulative male gametic contribution to their offspring (blue balance curve). The empirical observation of the cumulative parental contribution is known as „80/20 rule" commonly observed in many conifers seed orchards (Anonymous 1976). That empirical rule assumes that 80 % of seed orchard's crop is contributed by 20 % of parents. Lai et al. (2010) studied pollination dynamics of Douglas-fir seed orchard analyzed random bulk seed sample of 800 seeds. Their results are in concordance with the results presented here (Figure 15; the male contribution of 80 % of the progeny trees were produced by 51% seed orchard trees) as they revealed male gametic contribution for 80% crop of the seed orchard to be contributed by 46 % of fathers. Since it is not considered about the difference in design of Scots pine and Douglas-fir seed orchard and possibly distinctive pollination dynamics of those two species, it can be concluded that the results are comparable even that the analyzed samples were obtained by diverse ways (random bulk seed sampling versus intentional truncation selection approach). This discovery is very important and encouraging in view of the fact that phenotypic pre-selection does not cause decrease of paternal (male) gametic contribution in comparison to random selection.

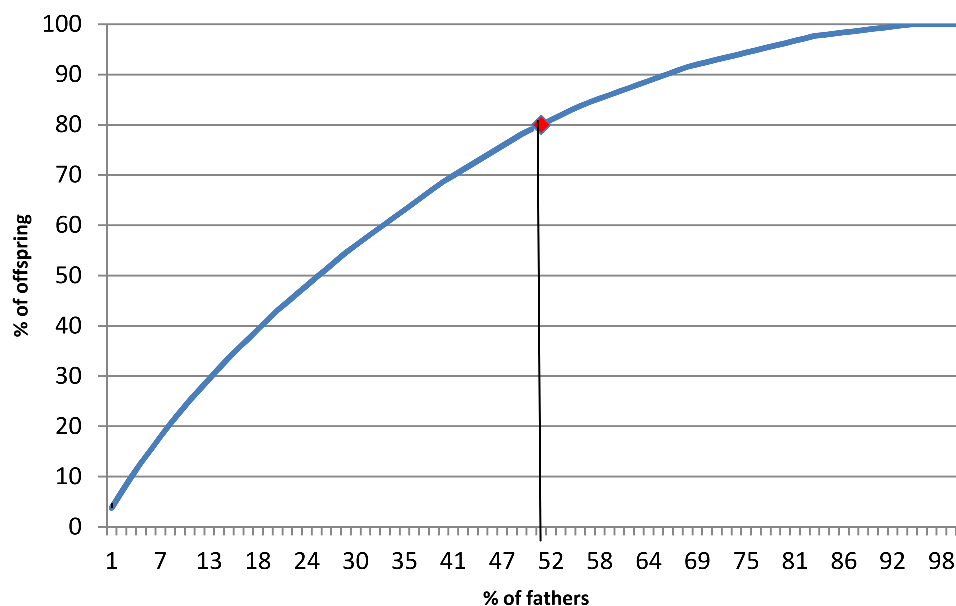


Fig. 15: Cumulative paternal (male) gametic contribution for all offspring assessed to their parents

10. Genetic parameters comparison among different models

The software COLONY (Jones and Wang 2010) implements a maximum likelihood method to estimate full-sib and half-sib relationships using individual multilocus genotypes. It was used to construct "pseudo pedigree" i.e. known maternal parents and unknown paternal parents were deduced according to observed relationships (Appendix 2) and full-sib families were identified for the subsequent sib-ship analysis. No selfing was assumed in the offspring generation as Scots pine is known for its high outcrossing rate ≈ 0.987±0.005 (Burczyk 1998).

Four pair-wise molecular relationship coefficients estimation methods were used to generate the all possible pair-wise marker based relationship matrices for their use as a substitute to the average relationship matrix (Queller and Goodnight 1989; Li et al. 1993; Lynch and Ritland 1999; Wang 2002). Pair-wise relationship matrices were estimated using SPaGeDi v. 1.3 software (Hardy and Vekemans 2002). SPaGeDi is the software designed to characterize special genetic structure of mapped individuals. It computes various statistics describing genetic relatedness or differentiation between individuals (details of the four pair-wise relationship methods used are described in Li et al. 1993; Lynch and Ritland 1999; Queller and Goodnight 1989; Wang 2002).

The pair-wise marker based relationship matrix produced for each method was used to replace the average relationship matrix in the quantitative genetics analyses. Pedigree records from the sib-shib analysis and molecular markers pair-wise relationship were used to estimate relationships combining both sources of information (Bömcke and Gengler 2009), i.e. the combined relationship estimates are computed as weighted pedigree-based and marker-based relatedness (Korecký et al. 2012).

Molecular markers based relationship estimates often produce not positive definite (PD) matrices due to internal inconsistencies, thus it was applied the "nearPD" function implemented in the R package "Matrix" to compute the nearest positive definite matrix and use it in the subsequent analyses (Knol

and Tenberge 1989; Cheng and Higham 1998; Higham 2002). Marker-based pair-wise relationship matrices were modified by removing negative elements and substituting them by zeros. The modified matrices were checked to be positive definite and the function “nearPD” in the R package “Matrix” was used to modify them as positive definite (Klápště et al. 2012).

Variance component estimates, heritability, and individuals’ breeding values and their level of precision (SE) were estimated with the restricted maximum likelihood (REML) animal model (Kruuk 2004) in ASReml v. 3.0 (Gilmour et al. 2002). The animal model can optimally accommodate variable family sizes (often found in open-family testing) and combine data from multiple locations and generations; furthermore, effects of non-random mating and selection are accounted for.

$$y = X\beta + Zu + e \qquad [3]$$

where **X** and **Z** are the incidence matrices relating to the fixed effect in vector β (population mean and site) and random effects in u (individual breeding values) to measurements in vector, y and e is vector of residuals following $E \sim N(0, \sigma_e^2)$. Vector of breeding values follows $\mathrm{Var}(u) = \mathbf{A}\sigma_a^2$, where σ_a^2 is additive genetic variance and **A** is 1) the average numerator relationship matrix in the scenario based on sib-ship reconstruction which is considered as reference model, 2) combined relationship matrix in the scenario where coefficients of relationship are estimated on both “pseudo-pedigree” records and marker information and 3) marker-based relationship matrix in scenarios based on pair-wise relationships estimated with only molecular markers. Narrow-sense heritability was estimated as follows:

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2} \qquad [4]$$

where σ_a^2 is the additive genetic variance and σ_e^2 is the residual variance. The accuracy of the predicted breeding values across all relationship estimators and the sib-ship analysis were calculated as follows:

$$r = \sqrt{1 - \frac{PEV}{(1+F_i)\sigma_\alpha^2}} \qquad [5]$$

Where PEV is the 'prediction error variance' of predicted breeding values (Gilmour et al. 1995) and F_i is the inbreeding coefficient for the ith individual.

10.1. Results and discussion of genetic parameters comparison

Since the offspring used to establish the two open-pollinated progeny tests originated from the seed orchards composed of the identical clones, the selected individuals from both sites in the sib-ship analysis to construct their family membership were used.

Appendix 2 shows the formation of full- and half- sib-ship composition confirming offspring authenticity; however, it does not assign offspring to a specific parental parent. The full description of familial composition at each and across sites is summarized at Table 12. Maternal (20 vs. 14) and paternal (62 vs. 21) distribution across the two sites, based on the selected offspring, showed increased representation in Skelná huť as compared to Nepomuk, indicating that the phenotypic selection was concentrated on fewer maternal parents. While Skelná huť has a greater parental representation, the sib-ship distribution in Nepomuk is better as indicated by greater mean full-sib family (3.5 vs. 2.6) and mean female (0.08 vs. 0.05) and male (0.05 vs. 0.02) contributions (Table 12).

Table 12: Description of familial structure at each site separately and across both sites

Parameter	Skelná huť	Nepomuk	Across both sites
# of individuals	284	313	597
Full-sib family size	1 - 30	1 - 20	1 - 30
Mean full-sib family size	2.6	3.5	3.0
Variance of FS size	15.376	14.071	15.122
# of contributing females	20	14	29
Mean female contribution	0.05	0.08	0.04
Variance in female contribution	0.0013	0.0010	0.0004
# of contributing males	62	21	66
Mean male contribution	0.02	0.05	0.02
Variance in male contribution	0.001	0.003	0.001

The differences in paternal and maternal representation between the two sites were also reflected in the estimated quantitative genetics parameters (Table 13A). The sib-ship analysis which was considered as the reference model produced consistently height heritability for each individual site and across sites models (0.28-0.29). The combined marker-pedigree estimator produced similar height heritability estimates for both sites; however, individual site's estimates varied between 0.25 and 0.37, reflecting sites differences in both phenotypic variability (site Nepomuk doubles that of site Skelná hut') (Table 13A). It is interesting to note that even the differences in variance components partitioning and their level of precision, as expressed by their standard errors, between the two sites were mirrored between the two analyses. Nepomuk clearly shows greater genetic parameters' precision which is expressed by lower standard errors, which was manifested in both analyses. The similar pattern is shown in the reliability of breeding value estimates where site Nepomuk provided better estimates only when marker based relationship matrix was used (0.48-0.59 vs. 0.3-0.56) and yielded slightly opposite estimates in the combined marker-pedigree model. With the exception of the Wang estimator, the remaining estimators generally produced the most reliable estimates of breeding values for across site analysis (0.47 – 0.66) (Table 12). This could be caused by either the better parental representation or the relative large environmental variance of site 1 (Table 13A).

The four pair-wise relationship methods, universally produced higher heritability estimates for Nepomuk (range: 0.18 – 0.26) as compared to Skelná hut' (range: 0.05 – 0.15), mirroring sites variation in both additive and environmental variances (Table 13B-1 and -B2). Across sites, height heritability estimates were lower than that from the sib-ship and markers-pedigree analyses (range: 0.08 -0.14 vs. 0.28) (Table 13).

Pearson's product moment and Spearman's rank correlation between individuals' breeding values were significant for all analyses (Table 14); however, across sites analyses produced higher corrections and were almost identical between the sib-ship and combined pedigree-markers analysis. It is noteworthy to mention, while they are significant, the product moment and

rank-order correlations between the sib-ship and all pair-wise relationships were lower than that observed between that of the sib-ship and combined pedigree-molecular markers (Table 14). The similar tendencies show graphic output of breeding values' correlation among different cases of estimation (Figure 16).

Table 13: Variance components (additive, environment, and total phenotypic and their standard errors) and heritability estimates comparison among sib-ship, combined marker and pedigree information (A), and all marker-based pair-wise relationships (B-1 and B-2).

A	**Sib-ship**			**Combined markers-pedigree**		
Variance	**Site 1**	**Site 2**	**Both sites**	**Site 1**	**Site 2**	**Both sites**
Additive	873.18 (461.01)	1963.2 (978.39)	1399.4 (528.62)	1247.1 (515.23)	1769.7 (832.5)	1460.5 (486.55)
Environmental	2224.7 (376.68)	4778.5 (745.51)	3592.8 (402.06)	2107.6 (322.75)	5188.3 (617.75)	3793.3 (327.65)
Phenotype	3097.9 (295.79)	6741.6 (645.66)	4992.2 (347.24)	3354.7 (372.47)	6957.9 (682.7)	5253.9 (393.86)
Heritability	0.28 (0.135)	0.29 (0.129)	0.28 (0.094)	0.37 (0.124)	0.25 (0.104)	0.28 (0.078)

B-1	**Li et al.**			**Wang**		
Variance	**Site 1**	**Site 2**	**Both sites**	**Site 1**	**Site 2**	**Both sites**
Additive	161.70 (221.92)	1151.25 (548.50)	454.79 (220.44)	351.16 (293.29)	1315.31 (628.26)	405.50 (232.82)
Environmental	2851.98 (316.98)	5376.40 (598.57)	4396.0 (312.00)	2671.69 (342.01)	5238.35 (651.80)	4452.9 (329.96)
Phenotype	3013.68 (259.18)	6527.65 (553.68)	4850.79 (291.75)	3022.8 (261.95)	6553.7 (554.07)	4858.4 (290.28)
Heritability	0.05 (0.073)	0.18 (0.079)	0.09 (0.044)	0.12 (0.095)	0.21 (0.091)	0.08 (0.047)

B-2	**Lynch & Ritland**			**Queller & Goodnight**		
Variance	**Site 1**	**Site 2**	**Both sites**	**Site 1**	**Site 2**	**Both sites**
Additive	324.43 (261.12)	1407.06 (548.77)	582.06 (234.99)	447.24 (245.34)	1694.70 (654.14)	679.56 (265.22)
Environmental	2691.19 (322.32)	5082.87 (571.85)	4258.3 (310.29)	2544.48 (291.49)	4934.23 (608.32)	4211.7 (313.67)
Phenotype	3015.62 (260.89)	6489.93 (553.19)	4840.3 (291.95)	2991.7 (261.18)	6628.9 (575.69)	4891.2 (299.18)
Heritability	0.12 (0.085)	0.23 (0.078)	0.12 (0.047)	0.15 (0.079)	0.26 (0.089)	0.14 (0.052)

Table 14: Pearson's product-moment (below diagonal) and Spearman's rank-order (above diagonal) correlations between individuals' breeding values among combined markers-pedigree (combined), sib-ship, and all marker-based pair-wise relationships for site 1 (A), site 2 (B), and both sites (C). (All correlations are significant at $P<0.0001$).

A	Combined	Sib-ship	Lynch & Ritland	Li et al.	Queller & Goodnight	Wang
Combined	1	0.7741368	0.6523275	0.6413032	0.717808	0.6422076
Sib-ship	0.8137823	1	0.4259441	0.4239663	0.323599	0.4463596
Lynch & Ritland	0.7134013	0.5118029	1	0.7904112	0.698399	0.7400609
Li et al.	0.6953154	0.5105861	0.8353239	1	0.700202	0.8477621
Queller & Goodnight	0.7639396	0.4189993	0.7484779	0.7477958	1	0.7031984
Wang	0.7006251	0.5271351	0.8024982	0.8818282	0.748552	1

B	Combined	Sib-ship	Lynch & Ritland	Li et al.	Queller & Goodnight	Wang
Combined	1	0.860264	0.704173	0.650113	0.712246	0.618048
Sib-ship	0.896965	1	0.571955	0.548204	0.549255	0.54714
Lynch & Ritland	0.778376	0.65683	1	0.806416	0.787371	0.737454
Li et al.	0.739174	0.634502	0.86205	1	0.762752	0.874863
Queller & Goodnight	0.793893	0.661802	0.829087	0.816178	1	0.721922
Wang	0.70877	0.631921	0.805464	0.909655	0.797985	1

C	Combined	Sib-ship	Lynch & Ritland	Li et al.	Queller & Goodnight	Wang
Combined	1	0.910125	0.728922	0.712154	0.745479	0.691103
Sib-ship	0.912541	1	0.633866	0.603775	0.603467	0.602734
Lynch & Ritland	0.741719	0.641534	1	0.837067	0.79605	0.809186
Li et al.	0.71535	0.609054	0.87708	1	0.802638	0.919658
Queller & Goodnight	0.774724	0.642922	0.826218	0.829267	1	0.777822
Wang	0.70487	0.614115	0.846454	0.942414	0.812507	1

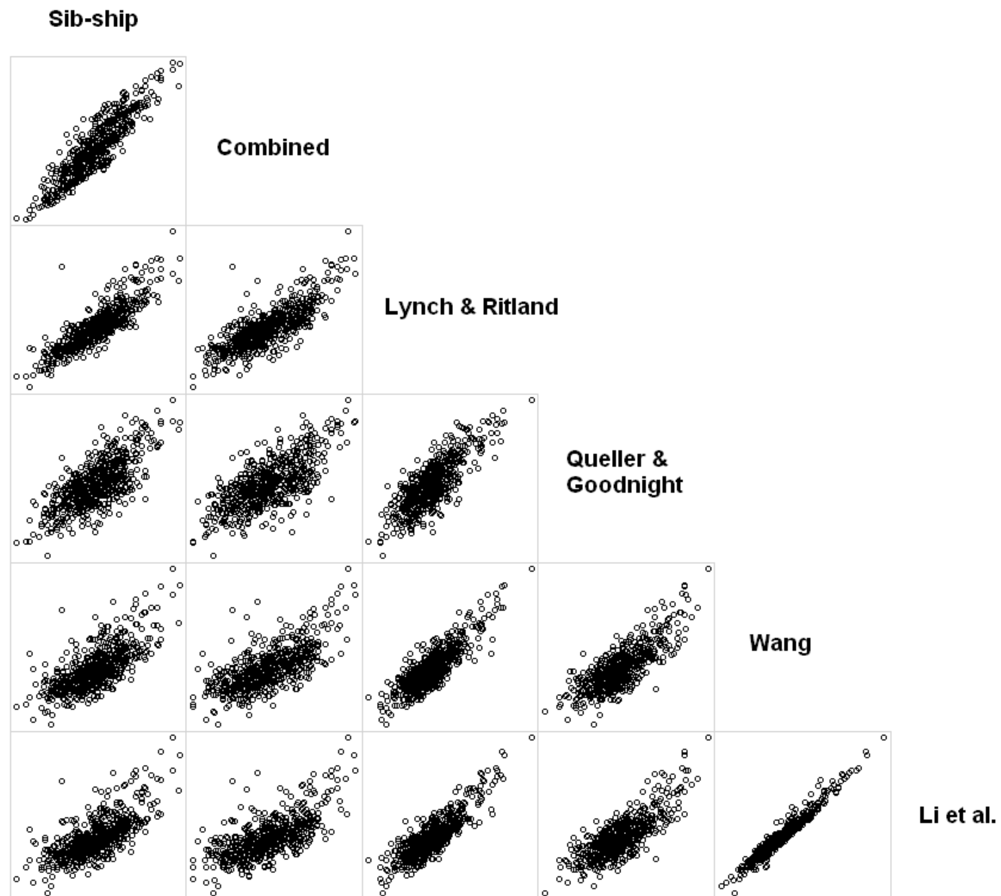


Fig. 16: Graphic output of breeding values' correlation among different cases of estimation

The development of molecular markers and their combination with advanced statistical approaches enables to merge the advantages of traditional OP testing with more advantaged full-sib family mating schemes (El-Kassaby and Lstibůrek 2009; El-Kassaby et al. 2011).

The animal models can optimally accommodate variable family size and combined data from multiple locations and generations; furthermore they reflect effects of non-random mating and selection. For this reason animal models provide better platform to pedigree reconstruction approaches in comparison to the simple half-sib evaluation which is based on simplified assumptions.

The combination of pedigree-based relatedness (A matrix) with marker-based relationship (G matrix) (Bömcke and Gengler 2009) was done to

minimize the impact of variable parental representation on marker-based relationships. Such a combined analysis leads to reliable ranking of individuals, when assessed by their respective additive genetic value (Blonk et al. 2010). High correlation coefficients between the sib-ship and combined analysis confirm their reasoning. On the other hand, the four pedigree-free approaches resulted in significant decrease in both variance estimation, as well as their respective standard errors. While the correlation coefficients were positive and high, we could not clearly explain the differences in variance estimates among the classical and the pedigree-free approaches.

It must be clarified that the resulting offspring structure as revealed by pedigree reconstruction is merely a function of the number of parents participated in mating; their respective reproductive phenology and output, as well as the contamination level (gene flow).

Conclusions

This PhD thesis is meant to be the molecular-genetics part of the project entitled the "Establishment of the second-generation seed orchards of Scots pine". This thesis utilized the phenotypic evaluation of Scots pine progeny trials which was completed in previous years (Kaňák 2011). The evaluated progeny trees comprised a base population for phenotypic pre-selection (truncation selection). Approximately 10% of trees were selected for subsequent genotyping utilizing microsatellite analysis.

Revealing of genomic information of the trees enables to transform half-sib progeny trials into full-sib progeny structure and subsequently apply "Breeding without Breeding" strategy (El-Kassaby and Lstibůrek 2009) within this breeding program. This approach is an innovative method which was used in the Czech Republic for the first time in practice. The results of this thesis confirmed many theoretical expectations and assumptions such as level of contamination (Lstibůrek et al. 2012) and the number of individuals which are necessary to genotype to obtain satisfactory population size for subsequent selection (Lstibůrek et al. 2011).

The laboratory procedures, such as DNA extraction protocol and PCR conditions for selected microsatellite primers were optimized for this particular species. The protocol for DNA extraction was optimized (as the best scenario was evaluated grinding by an oscillating mill for 2 minutes followed by a lytic phase for 50 min) and an addition of 5 ul of RNase per sample.

The pedigree reconstruction chapter deals with parentage assignment of analyzed progenies. The comparison of assumed maternal trees according to the records and assigned parental individuals enables the transformation of the results of parent-pair analysis with unknown sexes into the scenario with confirmed mother and assigned father (77.1% of all genotyped progenies).
The rest of individuals with both assigned parents (95.8% of all progeny trees in total) did not match with the assumed mother. The relatively high mismatch can be explained by inconsistency between records of the seed orchard's design and the real situation revealed by genotyping. Therefore, for

subsequent evaluations such as description of family structure, evaluation of full-sib family size and dynamics of male gametic contribution only individuals with confirmed maternal tree were used. The effective population size was also estimated.

In the chapter Genetic parameters comparison among different models software COLONY generated full-sib and half-sib structure among progenies. Three cases were used to estimate variance components and heritability (pedigree case, pedigree-free case and combined case). Four pair-wise molecular relationship coefficients methods were used to generate relationship matrices (pedigree-free case) for their use as a substitute to the average relationship matrix used for pedigree case. For combined case pedigree records from the sib-ship analysis and molecular markers pair-wise relationship were used to estimate relationships combining both sources of information (Bömcke and Gengler 2009). In each of the three cases, the same animal model was used to compute variances except when marker-based relationship (i.e., pedigree-free) was used, in this case the pedigree matrix was substituted with that computed based on the pairwise relationships matrix. The results showed high correlation in estimated genetic parameters between the pedigree and combined cases. In contrast, the pedigree-free approaches resulted in a significant decrease in both variance estimation and their standard errors. It was concluded that the combined model is the best approach as it represents the historical (pairwise) and contemporary (pedigree) relationships among the tested individuals.

It can be concluded that the results described in this thesis constitutes the molecular-genetics background for the latter stage of the long-term project of establishment the second generation Scots pine seed orchard in the Czech Republic. This project is unique in the Czech Republic's conditions due to his complexity and combination of molecular genetics aspects which help to generate the data with quantitative genetics methods enable to process them. This approach can be used in operational forestry for any target tree species and it can increase efficiency by decreasing time and financial constraints of breeding activities.

References

Ahuja MR, Neale DB. 2005. Evolution of genome size in conifers. *Silvae Genetica* **54**: 126-137.

Aitken N, Smith S, Schwarz C, Morin PA. 2004. Single nucleotide polymorphism (SNP) discovery in mammals: a targeted-gene approach. *Molecular Ecology* **13**.

Anonymous. 1976. Twentieth annual report on cooperative tree improvement and hardwood research program. North Carolina State University, Raleigh, USA.

Auckland LD, Bui T, Zhou Y, Shepherd M, Williams CG. 2002. *Conifer microsatellite handbook*. Corporate Press, Raleigh, N.C.

Bell GD, Fletcher AM. 1978. Computer organized orchard layouts (COOL) based on the permutated neighborhood design concept. *Silvae Genetica* **27**: 223-225.

Beuzen ND, Stear MJ, Chang KC. 2000. Molecular markers and their use in animal breeding. *Veterinary Journal* **160**.

Blonk RJW, Komen H, Kamstra A, van Arendonk JAM. 2010. Estimating Breeding Values With Molecular Relatedness and Reconstructed Pedigrees in Natural Mating Populations of Common Sole, Solea Solea. *Genetics* **184**: 213-219.

Blouin MS. 2003. DNA-based methods for pedigree reconstruction and kinship analysis in natural populations. *Trends in Ecology & Evolution* **18**: 503-511.

Bonin A, Bellemain E, Eidesen PB, Pompanon F, Brochmann C, Taberlet P. 2004. How to track and assess genotyping errors in population genetics studies. *Molecular Ecology* **13**: 3261-3273.

Boratyńsky A. 1991. Range of natural distribution. in *Genetics of Scots Pine* (ed. M Giertych), pp. 19-27. Akadémiai Kiadó, Budapest.

Borzan Ž. 1991. Karyotypes and cytogenetics. in *Genetics of Scots Pine* (ed. M Giertych), pp. 19-27. Akadémiai Kiadó, Budapest.

Botstein D, White RL, Skolnick M, Davis RW. 1980. Construction of a genetic-linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* **32**: 314-331.

Burczyk J. 1998. Mating system variation in a Scots pine clonal seed orchard. *Silvae Genetica* **47**: 155-158.

Burczyk J, Difazio SP, Adams WT. 2004. Gene flow in forest trees: How far do genes really travel? *Forest Genetics* **11**: 179-192.

Burczyk J, Dzialuk A, Wasilewska M, Chybicki I. 2005. Chloroplast and nuclear microsatellites in Scots pine - from theory to practice. in *Analysis of*

microsatellite sequences in Scots Pine (ed. J Nowakowska), pp. 9-16. Forest Research Institute, Warsaw.

Bömcke E, Gengler N. 2009. Combining microsatellite and pedigree data to estimate relationships among Skyros ponies. *Journal of Applied Genetics* **50**: 133-143.

CABI. 2002. Pines of silvicultural importance / compiled from the forestry compendium. p. 531. CABI Pub, Oxon, UK.

Callen DF, Thompson AD, Shen Y, Phillips HA, Richards RI, Mulley JC, Sutherland GR. 1993. Incidence and origin of null alleles in the (AC)n microsatellite markers. *American Journal of Human Genetics* **52**: 922-927.

Cappa EP, Lstibůrek M, Yanchuk AD, El-Kassaby YA. 2011. Two-dimensional penalized splines via Gibbs sampling to account for spatial variability in forest genetic trials with small amount of information available. *Silvae Genetica* **60**: 25-35.

Cercueil A, Bellemain E, Manel S. 2002. PARENTE: Computer program for parentage analysis. *Journal of Heredity* **93**: 458-459.

Chagne D, Chaumeil P, Ramboer A, Collada C, Guevara A, Cervera MT, Vendramin GG, Garcia V, Frigerio JMM, Echt C et al. 2004. Cross-species transferability and mapping of genomic and cDNA SSRs in pines. *Theoretical and Applied Genetics* **109**: 1204-1214.

Chakraborty R, Smouse PE, Neel JV. 1988. Population amalgamation and genetic variation: observations on artifitially agglomerated tribal populations of Central and South America. *American Journal of Human Genetics*: 709-725.

Cheng SH, Higham NJ. 1998. A modified cholesky algorithm based on a symmetric indefinite factorization. *Siam Journal on Matrix Analysis and Applications* **19**: 1097-1110.

Cockerham C.C. 1967. Group inbreeding and coancestry. *Genetics* **56**: 89.

Dakin EE, Avise JC. 2004. Microsatellite null alleles in parentage analysis. *Heredity* **93**: 504-509.

Danzmann RG. 1997. PROBMAX: A computer program for assigning unknown parentage in pedigree analysis from known genotypic pools of parents and progeny. *Journal of Heredity* **88**: 333-333.

Davison A, Chiba S. 2003. Laboratory temperature variation is a previously unrecognized source of genotyping error during capillary electrophoresis. *Molecular Ecology Notes* **3**: 321-323.

Dellaporta SL, Wood J, Hicks JB. 1983. A Plant DNA Minipreparation: Version II. *Plant Molecular Biology Reporter* **1**: 19-21.

Devey ME, Bell JC, Uren TL, Moran GF. 2002. A set of microsatellite markers for fingerprinting and breeding applications in Pinus radiata. *Genome* **45**: 984-989.

DiGiovanni F, Kevan PG, Arnold J. 1996. Lower planetary boundary layer profiles of atmospheric conifer pollen above a seed orchard in northern Ontario, Canada. *Forest Ecology and Management* **83**: 87-97.

Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS. 1991. Touchdown PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Research* **19**.

Dong JS, Wagner DB. 1994. Paternally inherited chloroplast polymorphism in pinus - estimation of diversity and population subdivision, and tests of disequilibrium with a maternally inherited mitochondrial polymorphism. *Genetics* **136**: 1187-1194.

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. in *Phytochemical Bulletin*, pp. 11-15.

Duchesne P, Godbout MH, Bernatchez L. 2002. PAPA (package for the analysis of parental allocation): a computer program for simulated and real parental allocation. *Molecular Ecology Notes* **2**: 191-193.

Dzialuk A, Chybicki I, Burczyk J. 2005. PCR multiplexing of nuclear microsatellite loci in Quercus species. *Plant Molecular Biology Reporter* **23**: 121-128.

Echt CS, MayMarquardt P. 1997. Survey of microsatellite DNA in pine. *Genome* **40**: 9-17.

Echt CS, Vendramin GG, Nelson CD, Marquardt P. 1999. Microsatellite DNA as shared genetic markers among conifer species. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere* **29**: 365-371.

El-Kassaby YA, Cappa EP, Liewlaksaneeyanawin C, Klápště J, Lstibůrek M. 2011. Breeding without Breeding: Is a Complete Pedigree Necessary for Efficient Breeding? *PloS one* **6**: e25737.

El-Kassaby YA, Lstibůrek M. 2009. Breeding without breeding. *Genetics Research* **91**: 111-120.

El-Kassaby YA, Lstibůrek M, Liewlaksaneeyanawin C, Slavov GT, Howe GT. 2006. Breeding without Breeding: Approach, Example, and Proof of Concept. in *IUFRO Division 2 Joint Conference: Low imput breeding and genetic conservation of forest tree species*, pp. 1-235, Antalya, Turkey.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *Plos One* **6**.

Elsik CG, Williams CG. 2001. Low-copy microsatellite recovery from a conifer genome. *Theoretical and Applied Genetics* **103**: 1189-1195.

Eriksson G, Ekberg I, Clapman D. 2006. *An Introduction to Forest Genetics*, Uppsala, Sweden.

Ewen KR, Bahlo M, Treloar SA, Levinson DF, Mowry B, Barlow JW, Foote SJ. 2000. Identification and analysis of error types in high-throughput genotyping. *American Journal of Human Genetics* **67**: 727-736.

Falconer DS, Mackay TFC. 1996. *Introduction to Quantitative Genetics*. Longman, London.

Farjon A. 1998. *World Checklist and Bibliography of Conifers*. Royal Botanical Gardens at Kew, Richmond, UK.

Friesen N, Brandes A, Heslop-Harrison JS. 2001. Diversity, origin, and distribution of retrotransposons (gypsy and copia) in conifers. *Molecular Biology and Evolution* **18**: 1176-1188.

Funda T, Lstibůrek M, Lachout P, Klápště J, El-Kassaby YA. 2009. Optimization of combined genetic gain and diversity for collection and deployment of seed orchard crops. *Tree Genetics & Genomes* **5**: 583-593.

Gagneux P, Boesch C, Woodruff DS. 1997. Microsatellite scoring errors associated with noninvasive genotyping based on nuclear DNA amplified from shed hair. *Molecular Ecology* **6**: 861-868.

Gerber S, Chabrier P, Kremer A. 2003. FAMOZ: a software for parentage analysis using dominant, codominant and uniparentally inherited markers. *Molecular Ecology Notes* **3**: 479-481.

Gerber S, Mariette S, Streiff R, Bodenes C, Kremer A. 2000. Comparison of microsatellites and amplified fragment length polymorphism markers for parentage analysis. *Molecular Ecology* **9**: 1037-1048.

Gernandt DS, Lopez GG, Garcia SO, Liston A. 2005. Phylogeny and classification of Pinus. *Taxon* **54**: 29-42.

Giertych M. 1975. Seed orchard designs. in *Forestry Commission Bulletin*, pp. 25-37, London, UK.

Gilmour AR, Cullis BR, Welham SI, Thompson R. 2002. *ASReml reference manual*. NSW Agriculture, Australia.

Gilmour AR, Gogel BJ, Cullis BR, Thompson R. 2006. *ASReml User Guide Release 2.0*. International Ltd., Hemel Hempstead.

Gilmour AR, Thompson R, Cullis BR. 1995. Average information REML: An efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics* **51**: 1440-1450.

Gomes I, Collins A, Thomas NS, Wilkinson J, Watson M, Morton N. 1999. Hardy-Weinberg quality control. *Annals of Human Genetics*: 535-538.

Goodnight KF, Queller DC. 1999. Computer software for performing likelihood tests of pedigree relationship using genetic markers. *Molecular Ecology* **8**: 1231-1234.

Grattapaglia D, Ribeiro V, Rezende G. 2004. Retrospektive selection of elite parent trees using paternity testing with microsatellite markers: an alternative short term breeding tactic for Eucalyptus. *Theoretical and Applied Genetics*: 192-199.

GRIN. 2010. National Genetic Resources Program Germplasm Resources Information Network. National Germplasm Resources Laboratory, Beltsville, Maryland.

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

Guevara MA, Soto A, Collada C, Plomion C, Savolainen O, Neale DB, Gonzáles-Martínez SC, Cervera MT. 2005. Genomics applied to the study of adaptation in pine species. *Invest Agrar: Sist Recur For* **14**: 292-306.

Guo XQ, Elston RC. 1999. Linkage information content of polymorphic genetic markers. *Human Heredity* **49**: 112-118.

Gömöry D, Bruchanik R, Longauer R. 2003. Fertility variation and flowering asynchrony in Pinus sylvestris: consequences for the genetic structure of progeny in seed orchards. *Forest Ecology and Management* **174**: 117-126.

Hardy OJ, Vekemans X. 2002. SPAGEDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* **2**: 618-620.

Hart JA. 1987. A cladistic analysis of conifers: preliminary results. *Journal of the Arnold Arboretum* **68**: 269-307.

Hayden MJ, Nguyen TM, Waterman A, McMichael GL, Chalmers KJ. 2008. Application of multiplex-ready PCR for fluorescence-based SSR genotyping in barley and wheat. *Molecular Breeding* **21**: 271-281.

Higham NJ. 2002. Computing the nearest correlation matrix - a problem from finance. *Ima Journal of Numerical Analysis* **22**: 329-343.

Hizume M, Shibata F, Matsusaki Y, Garajova Z. 2002. Chromosome identification and comparative karyotypic analyses of four Pinus species. *Theoretical and Applied Genetics* **105**: 491-497.

Hoffman JI, Amos W. 2005. Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. *Molecular Ecology* **14**: 599-612.

Jarne P, Lagoda PJL. 1996. Microsatellites, from molecules to populations and back. *Trends in Ecology & Evolution* **11**: 424-429.

Johansson A, Karlsson P, Gyllensten U. 2003. A novel method for automatic genotyping of microsatellite markers based on parametric pattern recognition. *Human Genetics* **113**: 316-324.

Jones AG. 2001. Gerud 1.0: a computer program for the reconstruction of parental genotypes from progeny arrays using multilocus DNA. *Molecular Ecology Notes* **1**: 215-218.

Jones AG, Ardren WR. 2003. Methods of parentage analysis in natural populations. *Molecular Ecology* **12**: 2511-2523.

Jones OR, Wang JL. 2010. COLONY: a program for parentage and sibship inference from multilocus genotype data. *Molecular Ecology Resources* **10**: 551-555.

Kalinowski ST, Taper ML, Marshall TC. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology* **16**: 1099-1106.

Karhu A, Dieterich JH, Savolainen O. 2000. Rapid expansion of microsatellite sequences in pines. *Molecular Biology and Evolution* **17**: 259-265.

Kaňák J. 2008. *Metodika zakládání semenných sadů : recenzovaná metodika*. Výzkumný ústav lesního hospodářství a myslivosti, Strnady.

Kaňák J. 2011. Návrh šlechtitelských postupů pro borovici lesní v západních a jižních Čechách (Project of breeding strategies for Scots Pine in the western and southern Bohemia). in *Department of Dendrology and Forest Tree Breeding*. CZU (CULS), Praha.

Kaňák J, Klápště J, Lstibůrek M. 2009. Úvodní genetické hodnocení semenných sadů borovice lesní v západních Čechách (Initial Evaluation of Seed Orchards of Scots Pine in the Western Czech Republic). *Zprávy lesnického výzkumu (Reports of Forestry Research)*: 189-204.

Kaňák J, Nárovcová J. 2004. Proměnlivost borovice lesní (Genetic variability of the Scots pine). *Lesnická práce* **83**: 10-11.

Kaňák K. 2001. Borovice lesní a prehistorie jejího rodu (Scots pine and the prehistory of its genus). *Lesu Zdar*: 1-4.

Klápště J, Lstibůrek M, El-Kassaby YA. 2012. Breeding without breeding: A simple pedigree-free model for estimating genetic parameters of forest tree populations. *PLoS ONE (submitted)*.

Knol DL, Tenberge JMF. 1989. Least-squares approximation of an improper correlation matrix by a proper one. *Psychometrika* **54**: 53-61.

Korecký J, Klápště J, Lstibůrek M, Kobliha J, Nelson CD, El-Kassaby YA. 2012. Genetic parameters comparison among pedigree-free, sib-ship, and combined markers-pedigree models in Scots pine multi-site open-pollinated tests. *Tree Genetics & Genomes (submitted)*.

Koski V. 1970. A study of pollen dispersal as a mechanism of gene flow in conifers. *Commun Inst Forest Fenn* **70**: 78.

Koski V. 1991. Generative reproduction and genetic processes in nature. in *Genetics of Scots Pine* (ed. M Giertych), pp. 59-72. Akadémiai Kiadó, Budapest.

Kozlowski KV, Pallardy SG. 1997. *Physiology of Wood Plants*. Academic Press, USA.

Krutovsky KV, Elsik CG, Matvienko M, Kozik A, Neale DB. 2007. Conserved ortholog sets in forest trees. *Tree Genetics & Genomes* **3**: 61-70.

Kruuk LEB. 2004. Estimating genetic parameters in natural populations using the 'animal model'. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **359**: 873-890.

Lai BS, Funda T, Liewlaksaneeyanawin C, Klápště J, Van Niejenhuis A, Cook C, Stoehr MU, Woods J, El-Kassaby YA. 2010. Pollination dynamics in a

Douglas-fir seed orchard as revealed by pedigree reconstruction. *Annals of Forest Science* **67**.

Lambeth C, Lee BC, O'Malley D, Wheeler N. 2001. Polymix breeding with parental analysis of progeny: an alternative to full-sib breeding and testing. *Theoretical and Applied Genetics* **103**: 930-943.

Lefort F, Echt C, Streiff R, Vendramin GG. 1999. Microsatellite sequences: a new generation of molekular markers for forest genetics. *International Journal of Forest Genetics* **6**: 15-20.

Li CC, Weeks DE, Chakravarti A. 1993. Similarity of DNA fingerprints due to chance and relatedness. *Human Heredity* **43**: 45-52.

LI-COR. 2002. Saga (user guide).

Liewlaksaneeyanawin C, Ritland CE, El-Kassaby YA, Ritland K. 2004. Single-copy, species-transferable microsatellite markers developed from loblolly pine ESTs. *Theoretical and Applied Genetics* **109**: 361-369.

Lindgren D. 1994. Effect of tree cover on Scots pine pollination and seeds. *Forest Genetics* **1**: 73-80.

Lindgren D, Cui J, Son SG, Sonesson J. 2004. Balancing seed yield and breeding value in clonal seed orchards. 11-22.

Lindgren D, Gea LD, Jefferson PA. 1997. Status number for measuring genetic diversity. *International Journal of Forest Genetics* **v. 4(2):** 69-76.

Lindgren D, Mullin TJ. 1997. Balancing gain and relatedness in selection. *Silvae Genetica* **46**: 124-129.

Lindgren D, Mullin TJ. 1998. Relatedness and status number in seed orchard crops. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere* **28**: 276-283.

Liston A, Gernandt DS, Vining TF, Campbell CS, Piñero D. 2003. Molecular phylogeny of *pinaceae* and pinus. *Proceedings of the International Conifer Conference Acta Horticulturae*: 107-114.

Liston A, Robinson WA, Pinero D, Alvarez-Buylla ER. 1999. Phylogenetics of Pinus (Pinaceae) based on nuclear ribosomal DNA internal transcribed spacer region sequences. *Molecular Phylogenetics and Evolution* **11**: 95-109.

Lstibůrek M, Ivanková K, Kadlec J, Kobliha J, Klápště J, El-Kassaby YA. 2011. Breeding without breeding: minimum fingerprinting effort with respect to the effective population size. *Tree Genetics & Genomes* **7**.

Lstibůrek M, Klápště J, Kobliha J, El-Kassaby YA. 2012. Breeding without Breeding Effect of gene flow on fingerprinting effort. *Tree Genetics & Genomes* **8**.

Lstibůrek M, El-Kassaby YA. 2010. Minimum-Inbreeding Seed Orchard Design. *Forest Science* **56**: 603-608.

Lynch M, Ritland K. 1999. Estimation of pairwise relatedness with molecular markers. *Genetics* **152**: 1753-1766.

Malécot G. 1948. *Les mathematiques de I´heredite*. Masson & Cie, Paris.
Marshall TC, Slate J, Kruuk LEB, Pemberton JM. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology* **7**: 639-655.
Masi P, Zeuli PLS, Donini P. 2003. Development and analysis of multiplex microsatellite markers sets in common bean (Phaseolus vulgaris L.). *Molecular Breeding* **11**: 303-313.
Metzenberg S. 2007. *Working with DNA*. Taylor and Francis Group, Abingdon, UK.
Mikeska M, Vacek S, Prausová R, Simon J, Minx T, Podrázský V, Malík V, Kobliha J, Anděl P, Matějka K. 2008. *Lesnicko-typologické vymezení, struktura a management přirozených borů a borových doubrav v ČR (Forestry-typological determination, structure and management of natural pine and pine-oak forests in the Czech Republic)*. Lesnická práce, Kostelec nad Černými lesy.
Miller J, Charles N. 1976. Early evolution in the Pinaceae. *Review of Palaeobotany and Palynology* **21**: 101-117.
Mirov NT. 1967. *The Genus Pinus*. The Ronald Press Company, New York.
Misevicius A. 2005. A tabu search algorithm for the quadratic assignment problem. *Computational Optimization and Applications* **30**: 95-111.
Molotkov PI, Patlaj IN. 1991. Systematic position within the genus *Pinus* and intraspecific taxonomy. in *Genetic of Scots Pine* (ed. M Giertych), pp. 31-40. Akadémiai Kiadó, Budapest.
Morse AM, Peterson DG, Islam-Faridi MN, Smith KE, Magbanua Z, Garcia SA, Kubisiak TL, Amerson HV, Carlson JE, Nelson CD et al. 2009. Evolution of Genome Size and Complexity in Pinus. *Plos One* **4**.
Muona O, Harju A. 1989. Effective population sizes, genetic-variability, and mating system in natural stands and seed orchards of Pinus sylvestris. *Silvae Genetica* **38**: 221-228.
Muratova EN. 1997. Cytogenetical study on Scots pine (Pinus sylvestris L.) in the Central Yakutia. in *Cytogenetic studies of forest trees and shrub species* (eds. Z Borzan, SE Schlarbaum), pp. 157-177. Hrvatske Sume and Sumarski Fakultet Sveucilista, Zagreb.
Musil I. 2003. *Lesnická dendrologie I.: jehličnaté dřeviny (Forest Dendrology I.: conifers)*. ČZU, Praha.
Musil J, Novák P, Šefl J. 2007. Semenné sady v České republice (Seed orchards in the Czech Republic). in *Aktuálne problémy lesného škôlkarstva, semenárstva a umelej obnovy lesa 2007*, p. 158. Národné lesnícke centrum, Liptovský Ján.
MZE. 2010. *Zpráva o stavu lesa a lesního hospodářství České republiky v roce 2009 (Information on Czech Forestry 2009)*. MZE, Praha

Namkoong G, Kang HC, Brouard JS. 1988. *Tree breeding: principles and strategies*. Springer-Verlag, New York.

Nielsen R, Mattila DK, Clapham PJ, Palsboll PJ. 2001. Statistical approaches to paternity analysis in natural populations and applications to the North Atlantic humpback whale. *Genetics* **157**: 1673-1682.

Nowakowska JA. 2005. Microsatellite Sequences in Scots Pine as a New Tool in Forest-Tree Genetics. in *Analysis of microsatellite sequences in Scots pine* (ed. J Nowakowska), pp. 9-16. Forest Research Institute, Warsaw.

Nowakowska JA. 2007. *Zmienność genetyczna polskich wybranych populacji sosny zwyczajnej (Pinus sylvestris L.) na podstawie analiz polimorfizmu DNA (Genetic variation of polish Scots pine (Pinus sylvestris L.) populations assessed with DNA polymorphism markers)*. Institut Badawczy Leśnictwa, Sekocin Stary.

Parantainen A, Pulkkinen P. 2003. Flowering and airborne pollen occurrence in a Pinus sylvestris seed orchard consisting of northern clones. *Scandinavian Journal of Forest Research* **18**.

Parker PG, Snow AA, Schug MD, Booton GC, Fuerst PA. 1998. What molecules can tell us about populations: Choosing and using a molecular marker. *Ecology* **79**: 361-382.

Paule L. 1992. *Genetika a šlachtenie lesných drevín (Genetics and Forest Tree Breeding)*. Príroda, Bratislava.

Plomion C, Chagne D, Pot D, Kumar S, Wilcox PL, Burdon RD, Prat D, Peterson DG, Paiva J, Chaumeil P et al. 2007. Pines. *Genome Mapping and Molecular Breeding in Plants*: 29-92.

Plíva K. 1987. *Typologický klasifikační systém ÚHÚL (Typological system of Forest Management Institut)*. ÚHÚL, Brandýs nad Labem.

Pompanon F, Bonin A, Bellemain E, Taberlet P. 2005. Genotyping errors: Causes, consequences and solutions. *Nature Reviews Genetics* **6**: 847-859.

Price R, Liston A, Strauss S. 1998. Phylogeny and systematics of *Pinus*. in *Ecology and Biogeography of Pinus* (ed. DM Richardson), pp. 49-68. Cambridge University Press, Cambridge, UK.

Price R, Olsen-Stojkovich J, Lowenstein J. 1987. Relationships among the genera of *Pinaceae:* an immunological comparison. *Systematic Botany* **12**: 91-97.

Provan J, Powell W, Hollingsworth PM. 2001. Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends in Ecology & Evolution* **16**: 142-147.

Queller DC, Goodnight KF. 1989. Estimating relatedness using genetic-markers. *Evolution* **43**: 258-275.

Rambousek J. 2003. Semenné sady lesních dřevin v České republice (The seed orchards of the forest tree species in the Czech Republic). *Lesnická práce*: 20-22.

Richardson DM. 1998. *Ecology and Biogeography of Pinus*. Cambridge University Press, Cambridge, UK.

Robinson JP, Harris SA. 1999. Microsatellite markers as a tool for the detection of intra- and interpopulational genetic structure. in *Which DNA Marker for Which Purpose? Final Compendium of the Research Project Development, optimisation and validation of molecular tools for assessment of biodiversity in forest trees in the European Union, Research Programme Molecular Tools for Biodiversity*.

Robledo-Arnuncio JJ, Gil L. 2005. Patterns of pollen dispersal in a small population of Pinus sylvestris L. revealed by total-exclusion paternity analysis. *Heredity* **94**: 13-22.

Robledo-Arnuncio JJ, Smouse PE, Gil L, Alia R. 2004. Pollen movement under alternative silvicultural practices in native populations of Scots pine (Pinus sylvestris L.) in central Spain. *Forest Ecology and Management* **197**: 245-255.

Schlötterer C. 2004. The evolution of molecular markers - just a matter of fashion? *Nature Reviews Genetics* **5**: 63-69.

Schmidt A, Doudrick RL, Heslop-Harrison JS, Schmidt T. 2000. The contribution of short repeats of low sequence complexity to large conifer genomes. *Theoretical and Applied Genetics* **101**: 7-14.

Scott KD. 2001. Microsatellites derived from ESTs, and their comparison with those derived by other methods. in *Plant Genotyping - the DNA Fingerprinting of Plants*, pp. 225-237. CABI, New York.

Scotti I, Paglia G, Magni F, Morgante M. 1999. Microsatellite markers as a tool for the detection of intra- and interpopulational genetic structure. in *Which DNA Marker for Which Purpose? Final Compendium of the Research Project Development, optimisation and validation of molecular tools for assessment of biodiversity in forest trees in the European Union, Research Programme Molecular Tools for Biodiversity*.

Sewell MM, Sherman BK, Neale DB. 1999. A consensus map for loblolly pine (Pinus taeda L.). I. Construction and integration of individual linkage maps from two outbred three-generation pedigrees. *Genetics* **151**: 321-330.

Signorovitch J, Nielsen R. 2002. PATRI - paternity inference using genetic data. *Bioinformatics* **18**: 341-342.

Slavov GT, Howe GT, Adams WT. 2005a. Pollen contamination and mating patterns in a Douglas-fir seed orchard as measured by simple sequence repeat markers. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere* **35**: 1592-1603.

Slavov GT, Howe GT, Gyaourova AV, Birkes DS, Adams WT. 2005b. Estimating pollen flow using SSR markers and paternity exclusion: accounting for mistyping. *Molecular Ecology* **14**: 3109-3121.

Slavov GT, Howe GT, Yakovlev I, Edwards KJ, Krutovskii KV, Tuskan GA, Carlson JE, Strauss SH, Adams WT. 2004. Highly variable SSR markers in Douglas-fir: Mendelian inheritance and map locations. *Theoretical and Applied Genetics* **108**: 873-880.

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

StatSoft. 2010. Statistica 9 (manual).

STRATEC Biomedical. 2011. Instruction for the Invisorb® Spin Plant Mini Kit.

Svoboda J. 2010. *Koncepce zachování a reprodukce genových zdrojů lesních dřevin u Lesů České republiky, s.p., na období 2010-2019 (The long-term conception of the maintaining and reproduction of forestry species on areas managed by LČR for the period 2010-2019)*. Lesy České republiky, Praha.

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

Taberlet P, Griffin S, Goossens B, Questiau S, Manceau V, Escaravage N, Waits LP, Bouvet J. 1996. Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research* **24**: 3189-3194.

Vandeputte M, Mauger S, Dupont-Nivet M. 2006. An evaluation of allowing for mismatches as a way to manage genotyping errors in parentage assignment by exclusion. *Molecular Ecology Notes* **6**: 265-267.

Varis S, Pakkanen A, Santanen A, Häggvik T, Pappinen A, Pulkkinen P. 2005. Paternity analysis of competition crosses using microsatellite markers - a tool to investigate pollen selection in northern Scots pine. in *Analysis of microsatellite sequences in Scots pine* (ed. J Nowakowska), pp. 89-100. Forest Research Institute, Warsaw.

Varshney RK, Graner A, Sorrells ME. 2005. Genic microsatellite markers in plants: features and applications. *Trends in Biotechnology* **23**: 48-55.

Vaughan SP, Russell K. 2004. Characterization of novel microsatellites and development of multiplex PCR for large-scale population studies in wild cherry, Prunus avium. *Molecular Ecology Notes* **4**: 429-431.

Viard F, Bremond P, Labbo R, Justy F, Delay B, Jarne P. 1996. Microsatellites and the genetics of highly selfing populations in the freshwater snail Bulinus truncatus. *Genetics* **142**: 1237-1247.

Viewegh J. 2003. *Klasifikace lesních rostlinných společenstev : (se zaměřením na Typologický systém ÚHÚL)*. Česká zemědělská univerzita, V Praze.

Viewegh J, Hitchen AV, Boulderstone PR. 2008. *Phytocenology and classification of vegetation*. Lesnická práce, Kostelec nad Černými lesy.

Vignal A, Milan D, SanCristobal M, Eggen A. 2002. A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution* **34**.

Vondrejs V, Storchová Z. 1997. *Genové inženýrství I. (Genetic engineering I.)*. Nakladatelství Univerzity Karlovy, Praha.

Wang J. 2007. Parentage and sibship exclusions: higher statistical power with more family members. *Heredity* **99**: 205-217.

Wang JL. 2002. An estimator for pairwise relatedness using molecular markers. *Genetics* **160**: 1203-1215.

Wang XQ, Tank DC, Sang T. 2000. Phylogeny and divergence times in Pinaceae: Evidence from three genomes. *Molecular Biology and Evolution* **17**: 773-781.

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

White TL, Adams WT, Neale DB. 2007. *Forest genetics*. CABI Publishers, Wallingford, Oxfordshire, UK ; Cambridge, MA.

Whittaker JC, Harbord RM, Boxall N, Mackay I, Dawson G, Sibly RM. 2003. Phylogeny and divergence times in Pinaceae: Evidence from three genomes. *Genetics* **164**: 781-787.

Wilmer JW, Allen PJ, Pomeroy PP, Twiss SD, Amos W. 1999. Where have all the fathers gone? An extensive microsatellite analysis of paternity in the grey seal (Halichoerus grypus). *Molecular Ecology* **8**: 1417-1429.

Wilson AJ, Ferguson MM. 2002. Molecular pedigree analysis in natural populations of fishes: approaches, applications, and practical considerations. *Canadian Journal of Fisheries and Aquatic Sciences* **59**: 1696-1707.

Yazdani R, Yeh FC, Rimsha J. 1995. Genomic mapping of Pinus sylvestris (L.) using random amplified polymorphic DNA markers. *International Journal of Forest Genetics* **2**: 109-116.

Zane L, Bargelloni L, Patarnello T. 2002. Strategies for microsatellite isolation: a review. *Molecular Ecology* **11**: 1-16.

ÚHÚL. 2007. *Národní inventarizace lesů v České republice (National Forestry Inventory in the Czech Republic)*. ÚHÚL, Brandýs nad Labem.

ÚHÚL. 2010a. *Evidence Reprodukčního MAteriálu (Register of Reproduction Material) erma.uhul.cz*.

ÚHÚL. 2010b. *Zpráva o stavu lesa a lesním hospodářství České republiky v roce 2009 (Report on Czech Forestry 2009)*. Ministerstvo zemědělství, Praha.

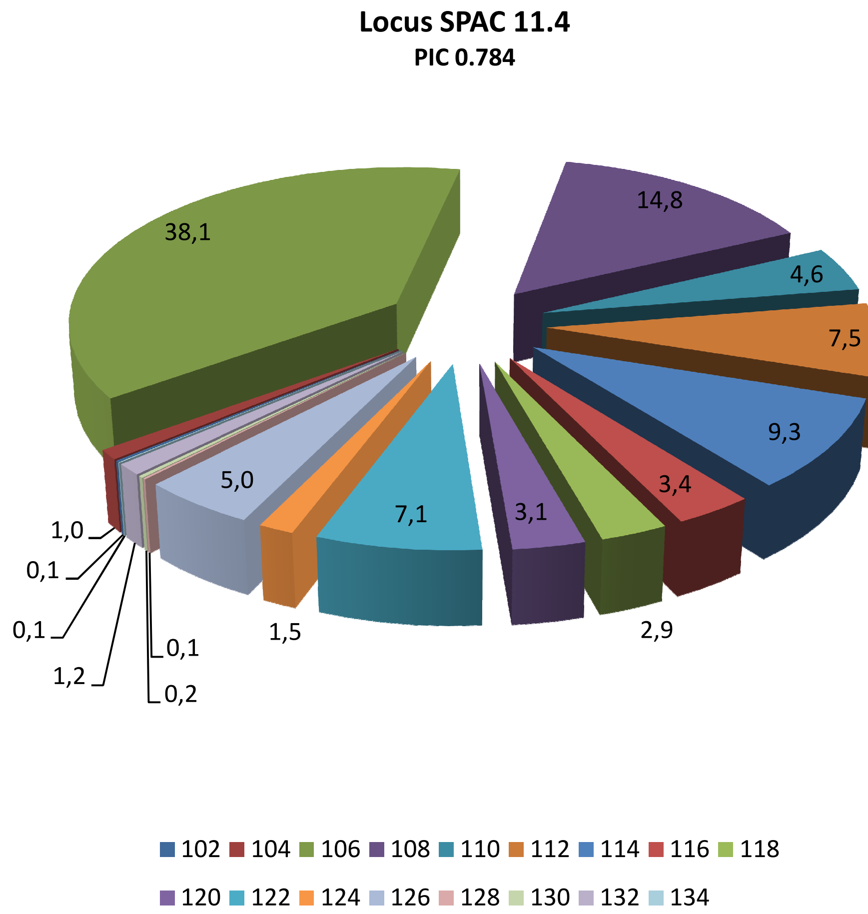
Appendices

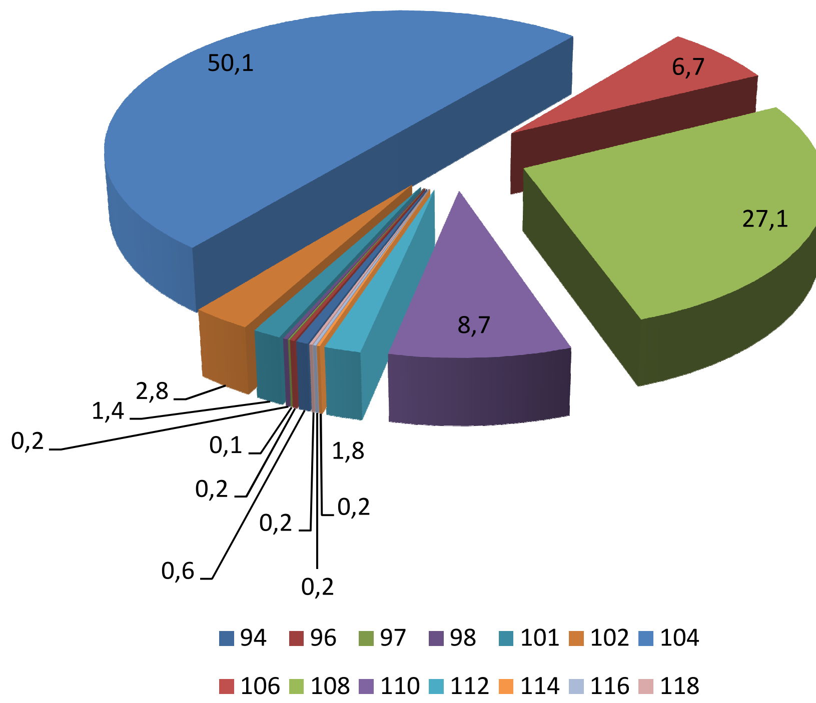
Appendix 1: Allele frequencies of analyzed loci

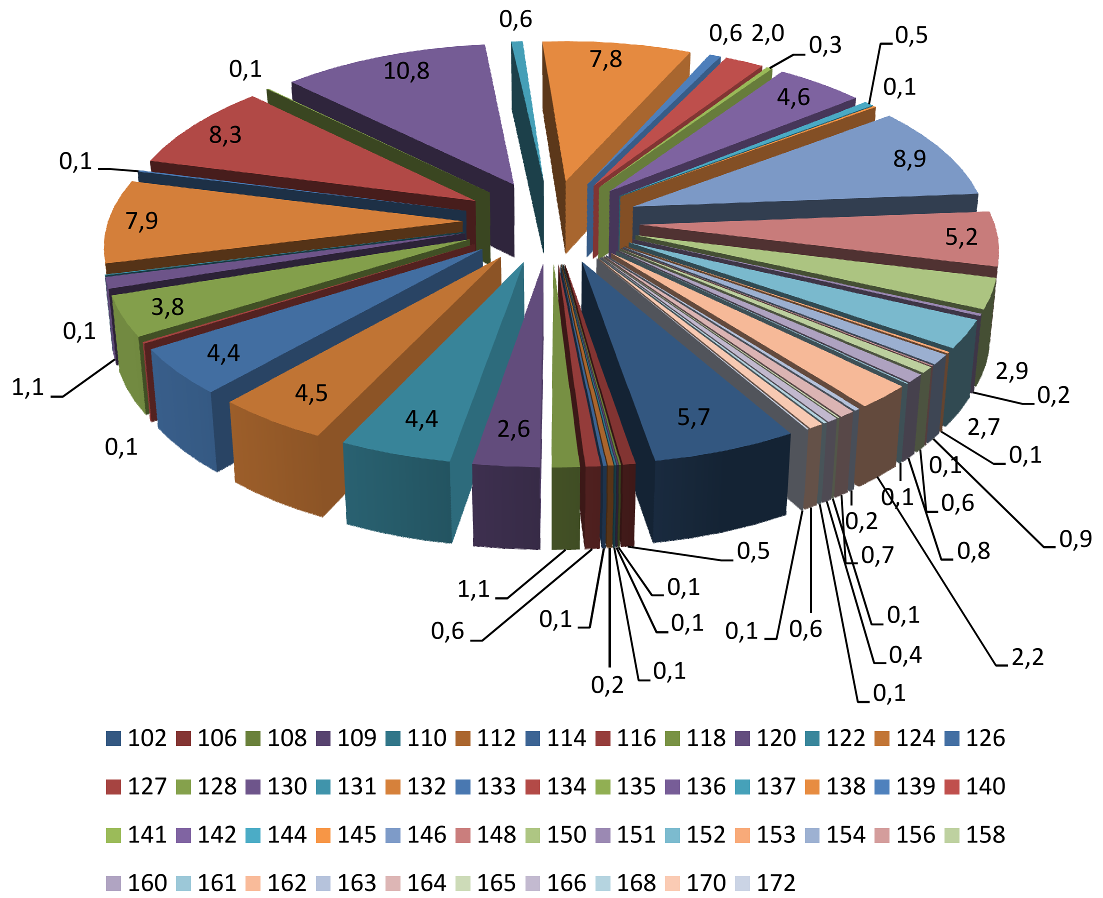
Appendix 2: Sib-ship assignment of Scots pine individuals

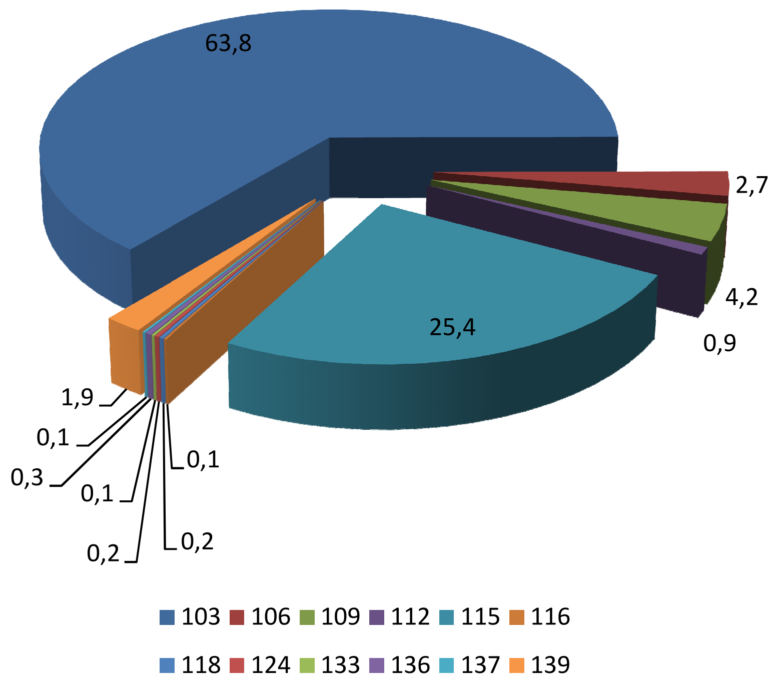
Appendix 3: Characteristics of half-sib progeny trees with confirmed mother and assigned father (pedigree reconstruction)

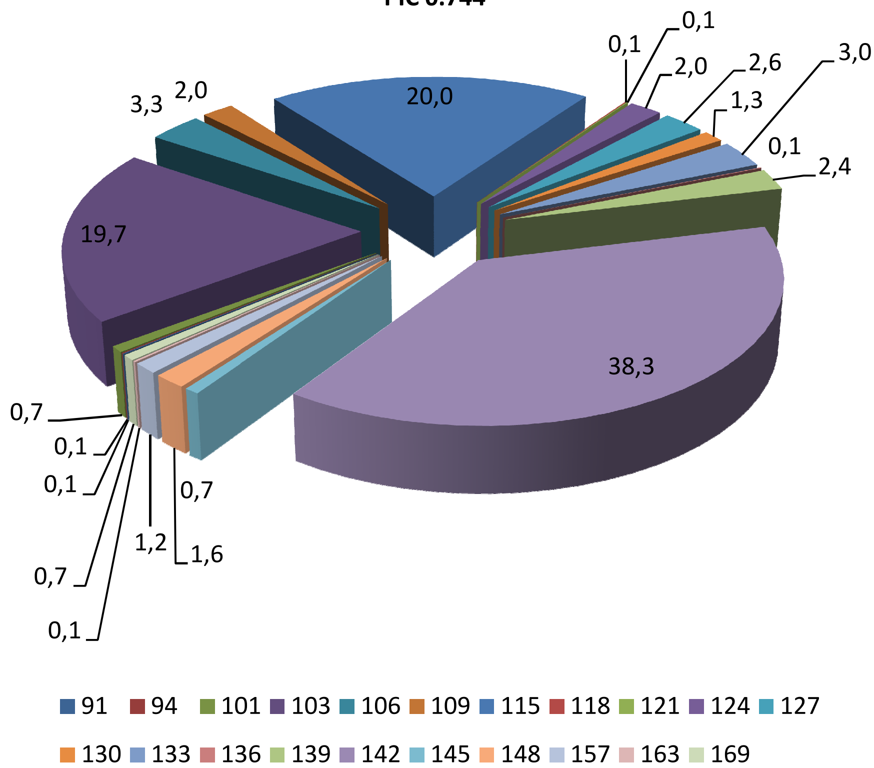
Appendix 1: Allele frequencies of analyzed loci

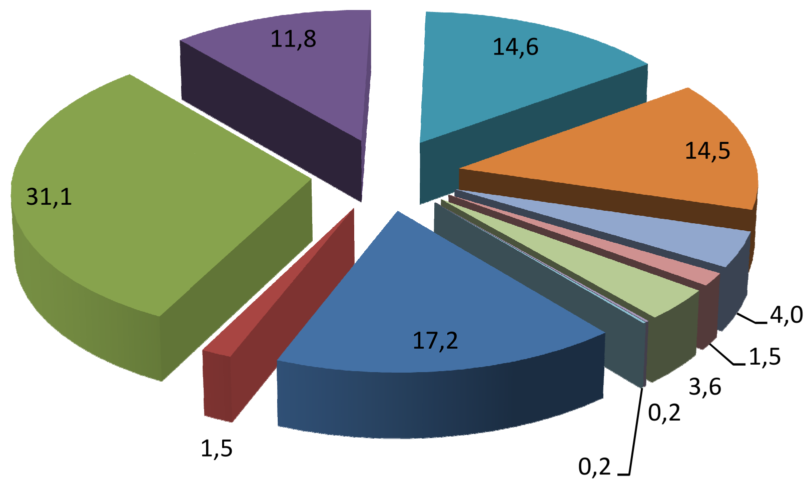



Locus LOP 1
PIC 0.616
50,1
6,7
27,1
8,7
2,8
1,4
0,2
0,1
0,2
0,2
0,6
0,2
1,8
0,2
94
96
97
98
101
102
104
106
108
110
112
114
116
118


Locus SPAC 12.5
PIC 0.940
102 106 108 109 110 112 114 116 118 120 122 124 126
127 128 130 131 132 133 134 135 136 137 138 139 140
141 142 144 145 146 148 150 151 152 153 154 156 158
160 161 162 163 164 165 166 168 170 172

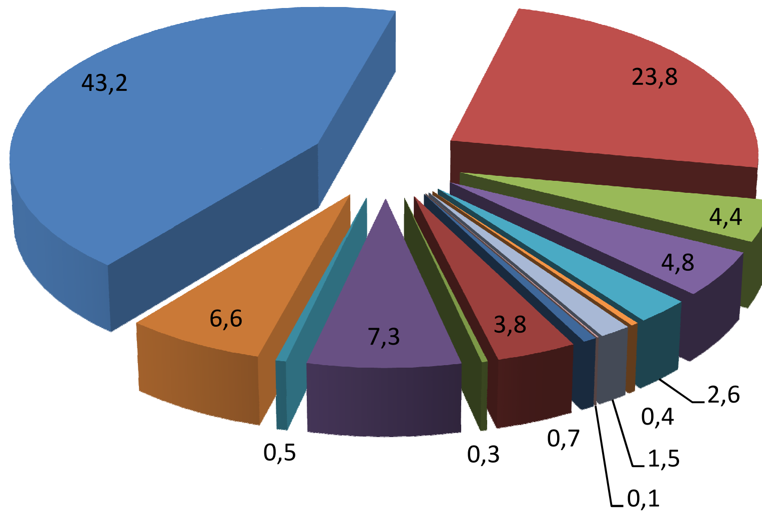

Locus PtTX 3025
PIC 0.478
63,8
2,7
4,2
0,9
25,4
1,9
0,1
0,3
0,1
0,2
0,1
0,2
103 106 109 112 115 116
118 124 133 136 137 139

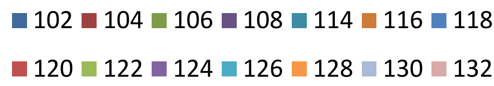

Locus PtTX 2146
PIC 0.744
0,1
0,1
2,0
2,6
3,0
1,3
0,1
2,4
3,3
2,0
20,0
19,7
38,3
0,7
0,1
0,1
0,7
0,1
1,2
1,6
0,7
91 94 101 103 106 109 115 118 121 124 127
130 133 136 139 142 145 148 157 163 169



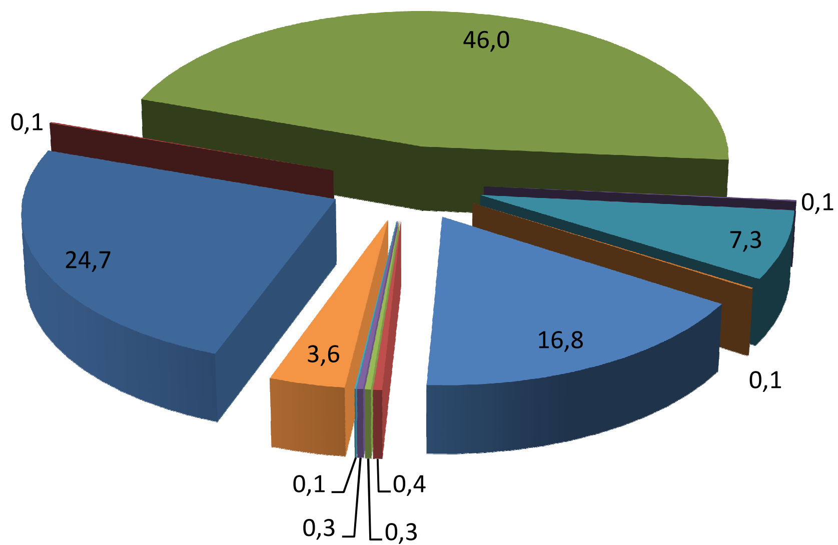
Locus SsrPt_ctg 64
PIC 0.349
11,8
14,6
14,5
31,1
17,2
4,0
1,5
3,6
0,2
0,2
1,5
103 106 109 112 115 118 121 124 127 130 133

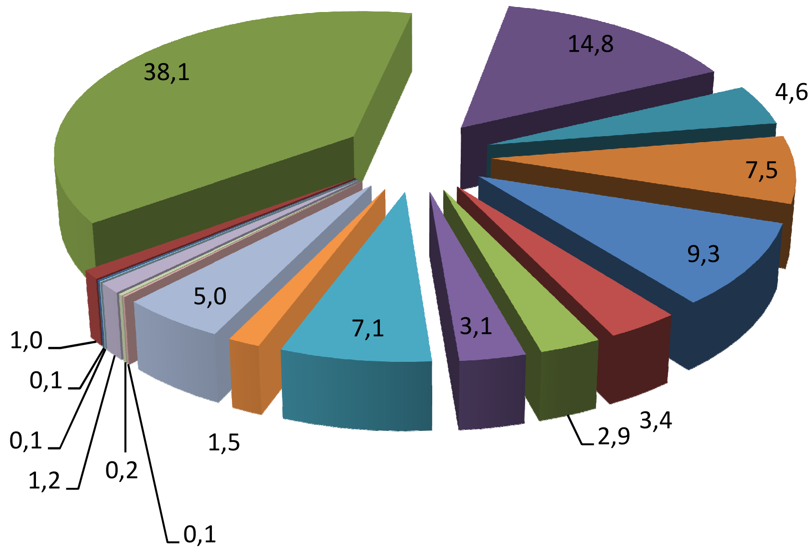
Locus PtTX 4001
PIC 0.711

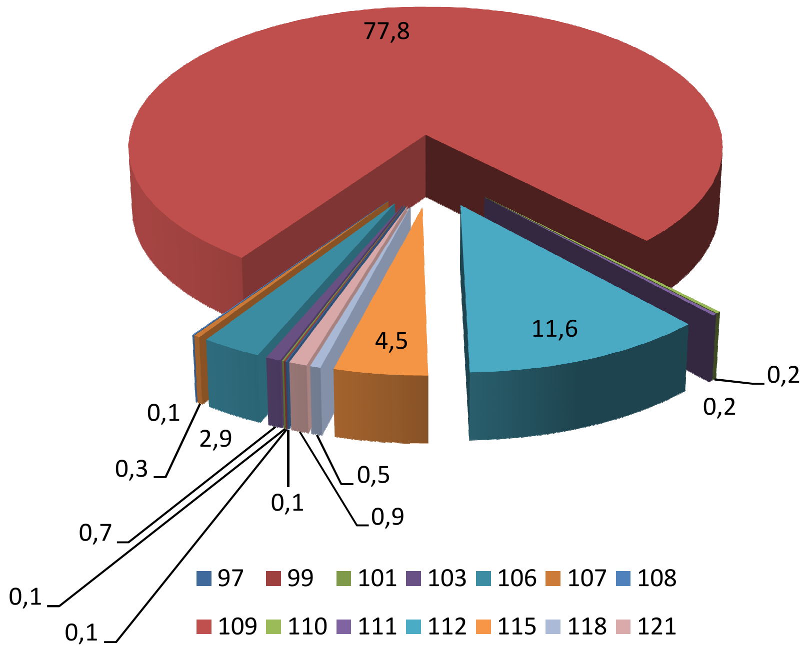




Locus PtTX 4011
PIC 0.649

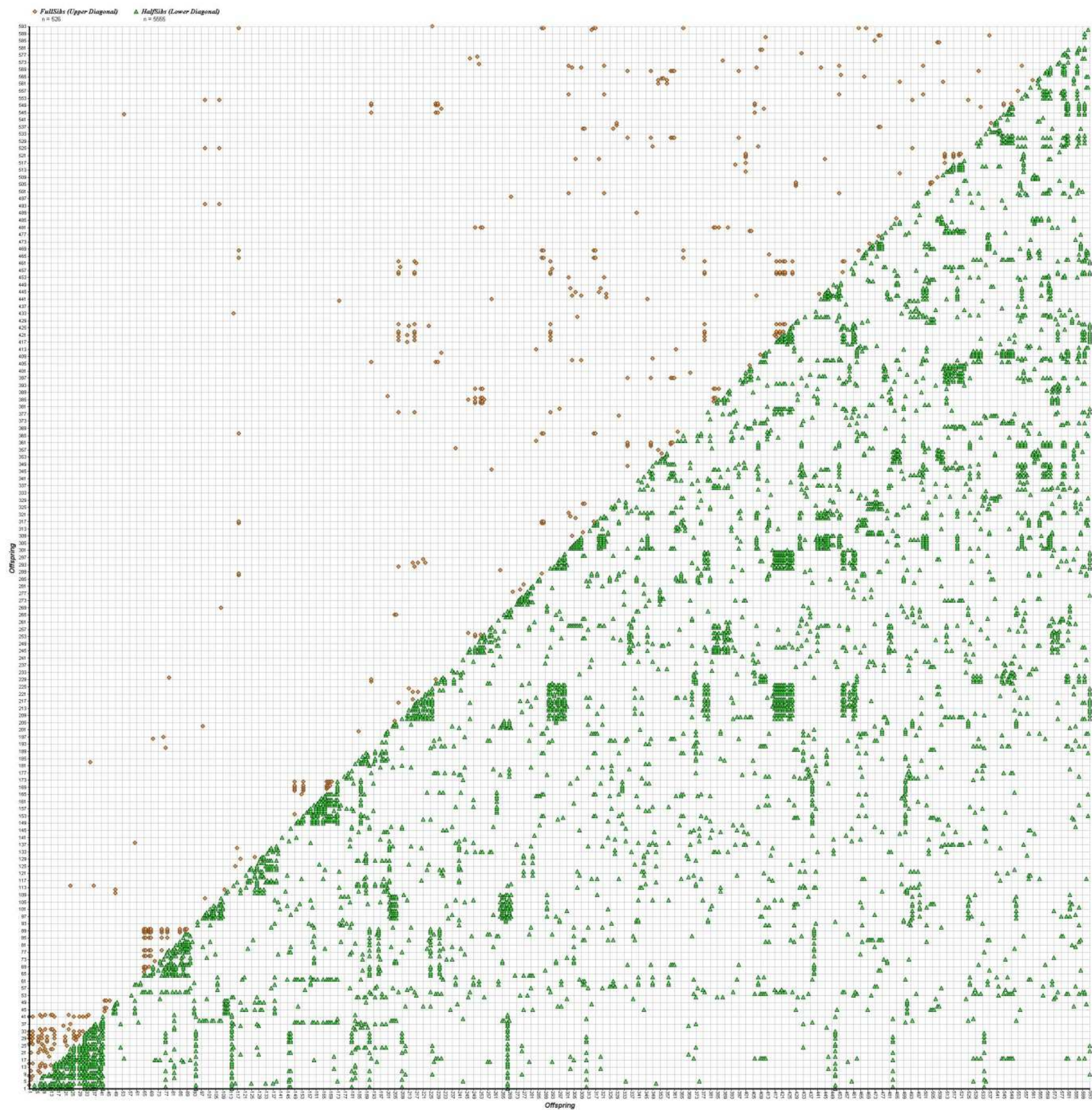



Locus SPAC 11.4
PIC 0.787
38,1
14,8
4,6
7,5
9,3
5,0
7,1
3,1
1,0
0,1
0,1
1,2
0,2
0,1
1,5
2,9
3,4
102 104 106 108 110 112 114 116 118
120 122 124 126 128 130 132 134


Locus PtTX 3107
PIC 0.784
77,8
11,6
4,5
0,2
0,2
0,1
2,9
0,3
0,7
0,1
0,1
0,1
0,5
0,9
97 99 101 103 106 107 108
109 110 111 112 115 118 121

Appendix 2: Sib-ship assignment of Scots pine individuals

Figure was generated by software COLONY (Jones and Wang 2010)



Appendix 3: Characteristics of half-sib progeny trees with confirmed mother and assigned father (pedigree reconstruction)

Skelná huť								
ID	**Plot**	**Row**	**Column**	**Clone**	**Height (cm)**	**dbh (cm)**	**Confir. mother**	**Assign. father**
1	1x21	1	21	2987	673	8	**2987**	**2369**
2	1x23	1	23	2987	608	6.4	**2987**	**2376**
3	1x29	3	29	2002			**2002**	**2658**
4	1x30	1	30	2002	721	9.6	**2002**	**2998**
5	1x81	1	81	2657	669	10.5		
6	1x21	2	21	2987	597	6.4		
7	1x25	2	25	2987	710	8.3	**2987**	**2097**
8	1x26	2	26	2002	760	10.2		
9	1x30	2	30	2002	603	6.1	**2002**	**2975**
10	1x56	2	56	2657	706	9	**2657**	**2103**
12	1x58	2	58	2657	609	6.5	**2657**	**2355**
13	1x58	3	58	2657	620	7		
14	1x24	4	24	2987	595	6.1		
15	1x25	4	25	2987	712	9.6	**2987**	**2097**
16	1x26	4	26	2002	720	8	**2002**	**2357**
17	1x56	4	56	2657	663	8	**2657**	**2985**
19	1x22	5	22	2987	702	10.2	**2987**	**2376**
20	1x25	5	25	2987	633	8.3	**2987**	**2984**
21	1x60	5	60	2657	582	6		
22	1x21	6	21	2987	703	10.2	**2987**	**2374**
23	1x29	6	29	2002	678	8.9	**2002**	**2377**
24	1x56	6	56	2657	692	7.5	**2657**	**2353**
25	1x82	6	82	2657	636	9	**2657**	**2369**
26	1x84	6	84	2657	688	7		
27	1x22	7	22	2987	658	7.3	**2987**	**2357**
28	1x85	7	85	2657	679	7.5	**2657**	**2354**
32	1x82	9	82	2657	688	11		
33	1x84	9	84	2657	494	6	**2657**	**S-2**
34	1x25	10	25	2987	664	8.6	**2987**	**2352**
35	1x29	10	29	2002	666	9.6	**2002**	**2378**
36	1x30	10	30	2002	672	6.4	**2002**	**2374**
38	1x81	10	81	2657	675	7.5		
39	2x89	11	89	2002	715	9	**2002**	**2101**
40	2x90	11	90	2002	702	7	**2002**	**2007**
41	2x107	11	107	2102	746	10	**2102**	**2097**
43	2x86	12	86	2002	754	11	**2002**	**2366**
44	2x105	12	105	2102	775	10	**2102**	**2998**
45	2x82	13	82	2987	728	9	**2987**	**2005**

46	2x105	13	105	2102	758	10	**2102**	**2988**
47	2x107	13	107	2102	616	6.5	**2102**	**2004**
48	2x117	13	117	2369	637	10	**2369**	**2097**
49	2x81	14	81	2987	703	10	**2987**	**2357**
50	2x91	14	91	2375	740	10	**2375**	**2359**
51	2x104	14	104	2102	688	8	**2102**	**2999**
52	2x106	14	106	2102	756	7.5		
53	2x108	14	108	2102	748	8.5	**2102**	**2352**
54	2x89	15	89	2002	609	11.5	**2002**	**2370**
55	2x105	15	105	2102	627	7.5		
56	2x108	15	108	2102	730	8	**2102**	**2989**
57	2x87	16	87	2002	627	6.5	**2002**	**2355**
58	2x106	16	106	2102	688	7	**2102**	**2372**
59	2x118	16	118	2369	724	9.5	**2369**	**2999**
60	2x119	16	119	2369	579	6.5	**2369**	**2355**
61	2x88	17	88	2002	673	6.5	**2002**	**2376**
62	2x106	17	106	2102	665	7.5		
63	2x83	18	83	2987	580	6.5	**2987**	**2779**
64	2x86	18	86	2002	772	13	**2002**	**2372**
65	2x87	18	87	2002	612	6	**2002**	**2987**
66	2x89	18	89	2002	589	7.5	**2002**	**2007**
67	2x90	18	90	2002	723	12.5	**2002**	**2351**
68	2x108	18	108	2102	724	9	**2102**	**2352**
69	2x116	18	116	2369	654	9	**2369**	**2377**
70	2x120	18	120	2369	565	8		
71	2x91	19	91	2375	643	9		
72	2x105	19	105	2102	675	8	**2102**	**2371**
73	2x106	19	106	2102	666	6.5	**2102**	**2377**
74	2x118	19	118	2369	690	9.5	**2369**	**2351**
75	2x90	20	90	2002	724	10	**2002**	**2373**
76	2x104	20	104	2102	624	10.5	**2102**	**2656**
77	2x105	20	105	2102	671	6.5	**2102**	**2007**
78	2x106	20	106	2102	699	9	**2102**	**2349**
79	2x120	20	120	2369	592	10		
81	3x12	21	12	2351	657	6	**2351**	**2004**
82	3x28	21	28	2975	711	8	**2975**	**2989**
83	3x29	22	29	2975			**2975**	**2099**
84	3x37	21	37	2353	694	7.5	**2353**	**2107/9**
85	3x38	21	38	2353	749	10	**2353**	**2656**
86	3x40	21	40	2353	778	7.5		
88	3x60	21	60	2980	635	9	**2980**	**2996**
89	3x61	21	61	2980	718	11	**2980**	**2360**
90	3x63	21	63	2980	651	6.5		
91	3x7	22	7	2369	566	8.5	**2369**	**2356**
92	3x14	22	14	2351	749	8.5		

93	3x31	22	31	2371	621	6.5	**2371**	**2372**
94	3x33	22	33	2371	684	7.5	**2371**	**2352**
95	3x34	22	34	2371	680	6.5	**2371**	**2002**
96	3x39	22	39	2353	688	6	**2353**	**2309**
97	3x51	22	51	2002	641	7.5	**2002**	**2010**
98	3x55	22	55	2998	667	10	**2998**	**2099**
99	3x56	22	56	2998	760	9	**2998**	**2355**
100	3x63	22	63	2980	599	6.5	**2980**	**2363**
101	3x12	23	12	2351	701	7	**2351**	**2011**
103	3x14	23	14	2351	739	7.5	**2351**	**2007**
104	3x27	23	27	2975	633	7	**2975**	**2985**
106	3x32	23	32	2371	636	8.5	**2371**	**2011**
107	3x33	23	33	2371	678	9	**2371**	**2981**
108	3x34	23	34	2371	644	6.5		
111	3x37	23	37	2353	736	9		
112	3x38	23	38	2353	662	6.5	**2353**	**2380**
113	3x61	23	61	2980	594	6		
114	3x62	23	62	2980	625	7		
115	3x6	24	6	2369	604	6.5		
116	3x7	24	7	2369	611	6	**2369**	**2378**
117	3x11	24	11	2351	643	7.5	**2351**	**2984**
118	3x12	24	12	2351	737	7.5		
119	3x14	24	14	2351	726	9	**2351**	**S-5**
120	3x15	24	15	2351	660	7.5		
121	3x29	24	29	2975	627	9	**2975**	**2986**
122	3x31	24	31	2371	647	9	**2371**	**2989**
123	3x36	24	36	2353	758	9	**2353**	**2012**
124	3x38	24	38	2353	636	7	**2353**	**2656**
125	3x55	24	55	2998	564	6	**2998**	**2990**
127	3x62	24	62	2980	680	9.5		
128	3x11	25	11	2351	652	6.5	**2351**	**2776/ 2976**
130	3x26	24	26	2975			**2975**	**2986**
131	3x27	25	27	2975	712	8	**2975**	**2359**
132	3x28	25	28	2975	553	7	**2975**	**2012**
133	3x29	25	29	2975	607	8.5	**2975**	**2659**
134	3x31	25	31	2371	680	9	**2371**	**2367**
135	3x36	25	36	2353	678	7.5	**2353**	**2377**
136	3x37	25	37	2353	594	6	**2353**	**2380**
137	3x38	25	38	2353	694	8.5	**2353**	**2975**
138	3x39	26	39	2353			**2353**	**2377**
139	3x61	25	61	2980	589	8.5	**2980**	**2657**
140	3x62	25	62	2980	651	9		
141	3x7	26	7	2369	683	10.5	**2369**	**2354**
142	3x8	26	8	2369	588	7.5	**2369**	**2349**

144	3x12	26	12	2351	796	10		
145	3x15	26	15	2351	652	6.5		
146	3x33	26	33	2371	717	10	**2371**	**2010**
147	3x38	26	38	2353	706	8.5	**2353**	**2985**
148	3x40	26	40	2353	646	7.5		
153	3x6	27	6	2369	699	8.5	**2369**	**2983**
154	3x13	27	13	2351	701	7.5		
156	3x28	27	28	2975	577	7	**2975**	**S-15**
158	3x36	27	36	2353	606	7		
159	3x38	27	38	2353	737	8.5	**2353**	**2364**
160	3x51	27	51	2002	617	6.5	**2002**	**2656**
161	3x59	27	59	2980	628	6.5	**2980**	**2657**
164	3x8	28	8	2369	661	6.5	**2369**	**2981**
165	3x12	28	12	2351	722	7.5	**2351**	**2992**
166	3x13	28	13	2351	703	10	**2351**	**2355**
167	3x37	28	37	2353	636	9	**2353**	**2004**
168	3x56	28	56	2998	720	10	**2998**	**2367**
169	3x60	28	60	2980	647	8.5	**2980**	**2360**
172	3x7	29	7	2369	669	8.5	**2369**	**2657**
173	3x11	29	11	2351	607	7.5		
175	3x8	30	8	2369	636	7.5	**2369**	**2377**
176	3x11	30	11	2351	774	9		
177	3x14	30	14	2351	654	7.5		
178	3x35	30	35	2371	650	9.5	**2371**	**2377**
179	3x36	30	36	2353	625	6.5	**2353**	**2975**
180	3x37	30	37	2353	677	9.5		
183	4x23	31	23	2657	690	10	**2657**	**2357**
184	4x54	31	54	2654	560	8	**2659**	**2004**
185	4x53	32	53	2375	580	6		
186	4x53	33	53	2375	615	6.5		
187	4x22	34	22	2657	685	10.5	**2657**	**2004**
188	4x23	34	23	2657	620	7	**2657**	**2356**
189	4x25	34	25	2657	730	10	**2657**	**2379**
191	4x23	35	23	2657	575	7		
193	4x24	36	24	2657	630	8.5	**2657**	**2657**
194	4x23	37	23	2657	666	9.5	**2657**	**2379**
195	4x51	37	51	2375	595	9.5	**2375**	**2103**
196	4x24	38	24	2657	610	6.5	**2657**	**2367**
198	4x56	38	56	2355	760	11.5		
199	4x25	39	25	2657	730	7.5		
200	4x50	39	50	2375	560	7		
201	4x56	39	56	2355	760	9.5		
203	4x23	38	23	2657				
204	4x24	40	24	2657	610	7	**2657**	**2988**
205	4x54	40	54	2654	622	8.5		

206	5x18	41	18	2102	677	9	**2102**	**2357**
207	5x25	41	25	2654	745	9	**2659**	**2099**
208	5x27	41	27	2654	710	11		
209	5x29	41	29	2659	635	9.5		
211	5x51	41	51	2002	691	10	**2002**	**2379**
212	5x52	41	52	2002	723	11	**2002**	**2007**
213	5x58	41	58	2987	672	8.5		
214	5x16	42	16	2102	688	6	**2102**	**2975**
215	5x24	42	24	2654	726	9	**2659**	**2379**
216	5x27	42	27	2654	728	9	**2659**	**2653**
217	5x29	42	29	2659	552	7	**2659**	**2104**
219	5x51	42	51	2002	489	6	**2002**	**2098**
220	5x54	42	54	2002	677	9	**2002**	**2657**
221	5x57	42	57	2987	682	7		
222	5x59	42	59	2987	685	9	**2987**	**2657**
223	5x24	43	24	2654	652	7	**2659**	**2006**
224	5x28	43	28	2659	597	6	**2659**	**2352**
225	5x34	43	34	2992	643	7		
226	5x36	43	36	2992	671	11	**2992**	**2982**
227	5x51	43	51	2002	607	6	**2002**	**2991**
228	5x52	43	52	2002	682	6	**2002**	**2104**
229	5x55	43	55	2002	733	9.5	**2002**	**2357**
230	5x59	43	59	2987	651	9	**2987**	**2351**
231	5x14	44	14	2102	668	8	**2102**	**2097**
232	5x31	44	31	2659	686	9	**2659**	**2656**
233	5x33	44	33	2992	668	7.5	**2992**	**2657**
236	5x52	44	52	2002	700	9.5	**2002**	**2981**
237	5x15	45	15	2102	612	6	**2102**	**0012**
238	5x16	45	16	2102	720	8	**2102**	**2007**
239	5x24	45	24	2654	572	7	**2659**	**2099**
240	5x26	45	26	2654	682	11	**2659**	**2991**
241	5x27	45	27	2654	642	7	**2659**	**2005**
242	5x29	45	29	2659	620	8.5	**2659**	**2984**
243	5x33	45	33	2992	616	6		
244	5x34	45	34	2992	645	8		
245	5x53	45	53	2002	642	8.5	**2002**	**2010**
246	5x54	45	54	2002	690	9.5	**2002**	**2982**
248	5x18	46	18	2102	704	10	**2102**	**2984**
249	5x25	44	25	2654			**2659**	**2377**
250	5x27	46	27	2654	685	10	**2659**	**S-6**
251	5x32	46	32	2659	631	9	**2659**	**2104**
252	5x51	46	51	2002	589	7	**2002**	**2365**
253	5x55	46	55	2002	669	9	**2002**	**2371**
254	5x17	47	17	2102	643	6	**2102**	**2377**
255	5x36	47	36	2992	676	10		

256	5x53	47	53	2002	635	9	**2002**	**2377**
257	5x56	47	56	2987	665	7.5	**2987**	**2356**
258	5x57	47	57	2987	632	6	**2987**	**2353**
259	5x58	47	58	2987	602	6.5	**2987**	**2004**
260	5x14	48	14	2102	666	8	**2102**	**2659**
261	5x32	48	32	2659	640	6	**2659**	**2984**
262	5x33	48	33	2992	630	6.5	**2992**	**2363**
264	5x51	48	51	2002	633	7	**2002**	**2366**
265	5x52	48	52	2002	689	8.5	**2002**	**2374**
266	5x55	48	55	2002	742	11.5	**2002**	**2007**
267	5x56	48	56	2987	677	7.5		
268	5x57	48	57	2987	755	9.5	**2987**	**2378**
269	5x58	48	58	2987	698	10	**2987**	**2352**
271	5x27	49	27	2654	711	10.5	**2659**	**S-15**
272	5x29	49	29	2659	665	9	**2659**	**2364**
273	5x31	49	31	2659	705	10	**2659**	**2656**
274	5x32	49	32	2659	662	7	**2659**	**2657**
275	5x33	49	33	2992	631	7		
276	5x51	49	51	2002	655	7.5	**2002**	**2101**
277	5x52	49	52	2002	707	10		
278	5x58	49	58	2987	725	11	**2987**	**2352**
279	5x14	50	14	2102	734	8	**2102**	**2377**
280	5x18	50	18	2102	671	6.5	**2102**	**2007**
281	5x27	50	27	2654	658	6.5	**2659**	**2376**
282	5x30	50	30	2659	645	7	**2659**	**2377**
283	5x31	50	31	2659	684	8.5		
284	5x53	50	53	2002	633	9.5	**2002**	**2010**
285	5x54	50	54	2002	681	9.5	**2002**	**2355**
286	5x58	50	58	2987	692	9	**2987**	**2657**
287	6x30	51	30	2355	540	6.5	**2355**	**2983**
288	6x48	51	48	2371	659	8	**2371**	**2377**
289	6x51	51	51	2975	621	8.5	**2975**	**20111**
290	6x46	52	46	2371	666	6.5	**2371**	**2102**
291	6x48	52	48	2371	666	9	**2371**	**2366**
293	6x43	53	43	2353	757	9.5	**2353**	**2992**
294	6x49	53	49	2371	648	8.5		
295	6x52	53	52	2975	770	10	**2975**	**2656**
296	6x47	54	47	2371	656	10	**2371**	**2996**
297	6x50	54	50	2371	687	9	**2371**	**2008**
299	6x54	57	54	2975	666	10.5	**2975**	**2369**
300	6x47	58	47	2371	663	11	**2371**	**2991**
301	6x49	58	49	2371	631	8.5		
302	6x42	59	42	2353	631	9		
303	6x44	59	44	2353	680	7.5		
304	6x49	59	49	2371	656	9	**2371**	**2004**

305	6x54	59	54	2975	666	9		
307	6x42	60	42	2353	614	7.5	**2353**	**2007**
308	7x3	61	3	2358	712	9		
309	7x24	62	24	2355	716	11.5	**2355**	**2981**
310	7x25	62	25	2355	666	7	**2355**	**2989**
311	7x25	63	25	2355	760	10.5	**2355**	**2379**
312	7x25	64	25	2355	790	10	**2355**	**2989**
313	7x3	65	3	2358	740	8		
314	7x31	65	31	2375	675	10.5	**2375**	**2099**
315	7x28	66	28	2375	665	8.5		
316	7x30	66	30	2375	652	8.5	**2375**	**2375**
317	7x31	66	31	2375	612	8	**2375**	**2372**
318	7x3	67	3	2358	730	9		
319	7x30	67	30	2375	626	8.5	**2375**	**2352**
320	7x25	69	25	2355	650	9		
321	7x31	69	31	2375	521	6.5		
322	7x31	70	31	2375	557	6.5	**2375**	**2656**
323		7	58					
324		41	54				**2002**	**2656**
325		51	49					
326		17	116					
327		19	108				**2102**	**2360**

Comments: Blank cells (except in columns confirmed mother and assigned father) are missing values. If parameter in columns confirmed mother and assigned father is not filled in, it indicates inconsistency between assumed mother and assigned individuals, i.e. pedigree reconstruction wasn't performed. Mothers colored in red were consistently confirmed to have distinctive genotype than stated according to the records (2654 was estimated 2659).

Nepomuk								
ID	**Plot**	**Row**	**Column**	**Clone**	**Height (cm)**	**dbh (cm)**	**Confirm. mother**	**Assign. father**
1001	1x6	1	26	2657	900	8.5	**2657**	**2918/ 2978**
1002	1x6	1	28	2657	980	11	**2657**	**2363**
1003	1x6	3	28	2657	910	8	**2657**	**2657**
1004	1x6	3	30	2657	970	11.5	**2657**	**2656**
1005	1x6	6	29	2657	930	9	**2657**	**2365**
1008	1x12	4	60	2659	950	7.5		
1010	1x12	7	58	2659	950	9.5	**2659**	**2355**
1011	1x15	1	71	2353	960	8.5		
1012	1x15	1	72	2353	1000	12	**2353**	**2380**
1013	1x15	1	74	2353	980	10	**2353**	**2379**
1014	1x15	1	75	2353	900	9	**2353**	**2376**
1015	1x15	4	73	2353	940	8		
1016	1x15	7	73	2353	970	9	**2353**	**2364**
1017	1x15	8	72	2353	940	8.5	**2353**	**2356**
1018	1x15	8	74	2353	870	7		
1020	1x15	9	75	2353	1070	12.5	**2353**	**2358**
1022	1x15	10	72	2353	970	9.5	**2353**	**2364**
1023	1x15	10	73	2353	910	9	**2353**	**2366**
1024	2x5	11	21	2369	940	8	**2373**	**2358**
1027	2x5	12	21	2369	880	7	**2373**	**2980**
1028	2x5	13	21	2369	920	8	**2373**	**2358**
1030	2x5	13	25	2369	970	8.5		
1036	2x5	17	21	2369	900	7.5	**2373**	**S-2**
1037	2x5	17	24	2369	980	8.5	**2373**	**2004**
1038	2x5	18	23	2369	1020	8		
1039	2x5	19	21	2369	910	7.5	**2373**	**2358**
1040	2x5	20	21	2369	880	7		
1041	2x10	18	50	S 5	940	8.5		
1042	2x10	19	49	S 5	970	10		
1043	2x15	13	74	2657	1000	10.5	**2657**	**2981**
1044	2x15	14	73	2657	1020	12		
1045	2x15	15	74	2657	1010	10	**2657**	**2380**
1046	2x15	15	75	2657	1010	9	**2657**	**2366**
1047	2x15	16	73	2657	1010	8.5	**2657**	**2657**
1048	2x15	17	73	2657	990	8.5	**2657**	**2097**
1049	2x15	17	75	2657	1080	9.5	**2657**	**2376**
1051	3x2	21	9	2369	940	9	**2373**	**2655**
1054	3x2	25	7	2369	1100	10	**2373**	**2351**
1055	3x2	26	9	2369	980	11	**2373**	**2376**
1056	3x2	27	7	2369	1000	8		
1057	3x2	28	7	2369	1000	9.5		

1067	3x4	27	19	S 5	1040	10		
1068	3x4	29	18	S 5	920	8		
1070	3x4	29	20	S 5	990	9		
1071	3x4	30	17	S 5	1020	8		
1072	3x4	30	20	S 5	950	9		
1076	3x5	22	22	S 6	860	7.5		
1077	3x5	25	21	S 6	960	8.5	**S-6**	**2376**
1078	3x5	26	22	S 6	1010	10	**S-6**	**2356**
1079	3x5	26	23	S 6	810	8	**S-6**	**2004**
1080	3x5	26	24	S 6	1000	10	**S-6**	**2099**
1081	3x5	28	21	S 6	820	7		
1086	3x10	23	50	2371	980	9.5	**2371**	**2358**
1087	3x10	25	48	2371	1050	11	**2371**	**2997**
1088	3x10	26	47	2371	950	10.5	**2371**	**2373**
1090	3x10	27	50	2371	1050	9		
1092	3x10	30	46	2371	970	10.5	**2371**	**2358**
1093	3x10	30	47	2371	950	10	**2371**	**2356**
1094	3x10	30	50	2371	1020	12	**2371**	**2356**
1095	3x13	21	64	2358	950	9	**2358**	**2008**
1098	3x13	23	65	2358	890	8	**2358**	**2011**
1099	3x13	25	62	2358	1070	12	**2358**	**2361**
1100	3x13	28	62	2358	870	8		
1101	3x13	28	63	2358	1060	11	**2358**	**2373**
1102	3x13	28	65	2358	1000	8.5		
1103	3x13	29	64	2358	950	9.5	**2358**	**2990**
1104	3x13	30	61	2358	880	9.5	**2358**	**2375**
1106	3x15	23	71	2359	930	8	**2359**	**2779**
1107	3x15	23	72	2359	990	10	**2359**	**2369**
1108	3x15	23	74	2359	1100	10.5	**2359**	**2376**
1109	3x15	24	73	2359	1070	10	**2359**	**2989**
1110	3x15	25	72	2359	860	8.5	**2359**	**2998**
1111	3x15	28	74	2359	1090	10	**2359**	**S-15**
1112	3x15	29	74	2359	1040	11.5		
1114	3x15	30	75	2359	940	8	**2359**	**2779**
1115	4x3	32	12	2358	870	10		
1117	4x3	34	12	2358	950	9	**2358**	**2998**
1120	4x3	37	12	2358	990	10	**2358**	**2099**
1122	4x3	38	15	2358	990	7.5	**2358**	**2097**
1125	4x5	32	24	2657	1010	9.5	**2657**	**2366**
1129	4x5	35	24	2657	1090	12	**2657**	**2107/9**
1131	4x5	38	24	2657	1000	12		
1132	4x5	38	25	2657	1120	12.5	**2657**	**2990**
1133	4x5	39	21	2657	1020	8.5	**2657**	**S-6**
1134	4x5	39	22	2657	1000	8.5	**2657**	**2975**
1138	4x6	31	30	S 2	970	9		

1139	4x6	32	28	S 2	920	10.5		
1140	4x6	36	26	S 2	1040	8	**S-2**	**2360**
1143	4x8	32	36	2351	1070	11.5	**2351**	**2371**
1144	4x8	33	37	2351	1030	9.5	**2351**	**2374**
1145	4x8	36	39	2351	1130	12.5	**2351**	**2369**
1146	4x8	38	38	2351	1030	8.5		
1149	4x9	35	44	S 6	1050	8	**S-6**	**2011**
1151	4x9	39	43	S 6	1110	11	**S-6**	**2999**
1153	5x2	42	7	2353	1030	10.5		
1154	5x2	44	7	2353	1060	10.5	**2353**	**2380**
1155	5x2	45	8	2353	1060	8.5	**2353**	**2374**
1156	5x2	48	7	2353	1000	11	**2353**	**2369**
1157	5x3	47	12	2657	1060	11.5		
1158	5x3	47	15	2657	1150	12	**2657**	**2373**
1159	5x9	41	42	2371	1010	11.5	**2371**	**2366**
1160	5x9	42	45	2371	1120	11	**2371**	**2349**
1162	5x9	43	42	2371	960	9		
1164	5x9	44	44	2371	1100	11	**2371**	**2989**
1168	5x9	46	43	2371	1070	9.5		
1169	5x9	47	44	2371	1060	10	**2371**	**2378**
1170	5x9	48	41	2371	1090	11.5	**2371**	**2370**
1171	5x9	49	42	2371	1090	12.5	**2371**	**2374**
1172	5x9	49	44	2371	1020	10.5	**2371**	**2349**
1173	5x9	50	44	2371	990	8	**2371**	**2349**
1178	5x13	48	61	2657	1140	10.5	**2657**	**2357**
1181	5x15	41	71	S 2	990	11.5	**S-2**	**2354**
1182	5x15	41	74	S 2	950	10.5	**S-2**	**2994**
1183	5x15	41	75	S 2	950	11	**S-2**	**2379**
1184	5x15	43	73	S 2	930	10	**S-2**	**2358**
1185	5x15	44	74	S 2	880	7	**S-2**	**2354**
1186	5x15	45	72	S 2	800	7	**S-2**	**2355**
1187	5x15	45	73	S 2	930	8.5	**S-2**	**2351**
1188	5x15	45	75	S 2	930	9	**S-2**	**2776/ 2976**
1189	5x15	46	71	S 2	1050	9	**S-2**	**20011**
1190	5x15	46	73	S 2	920	8	**S-2**	**20011**
1191	5x15	47	72	S 2	920	7.5	**S-2**	**S-5**
1193	5x15	49	73	S 2	990	7.5		
1194	6x1	52	5	2351	990	8	**2351**	**2376**
1195	6x1	53	1	2351	1030	10	**2351**	**2373**
1196	6x1	55	2	2351	1040	12.5		
1197	6x1	56	4	2351	980	8.5	**2351**	**2372**
1198	6x1	60	2	2351	1080	12.5	**2351**	**2370**
1199	6x1	60	3	2351	1100	9.5	**2351**	**2985**
1200	6x1	60	5	2351	1160	10	**2351**	**2374**

1201	6x2	54	7	2657	1060	9	**2657**	**2366**
1202	6x2	55	8	2657	980	8.5	**2657**	**2012**
1203	6x2	60	6	2657	1070	10.5	**2657**	**2360**
1204	6x9	51	41	2353	1040	9		
1205	6x9	51	42	2353	960	8	**2353**	**2364**
1206	6x9	51	45	2353	990	8.5	**2353**	**2369**
1207	6x9	52	43	2353	1060	11.5	**2353**	**2657**
1209	6x9	54	45	2353	1070	9.5	**2353**	**2380**
1210	6x9	55	42	2353	950	8.5	**2353**	**2004**
1211	6x9	55	44	2353	1050	11	**2353**	**2355**
1212	6x9	56	42	2353	1010	11	**2353**	**2998**
1214	6x9	57	42	2353	980	7.5	**2353**	**2999**
1215	6x9	58	42	2353	1010	10	**2353**	**2999**
1216	6x9	58	44	2353	970	10.5	**2353**	**2309**
1217	6x9	59	43	2353	1030	11	**2353**	**2354**
1218	6x9	60	41	2353	1030	9	**2353**	**2364**
1221	6x12	53	58	2375	940	10.5	**2375**	**2994**
1222	6x12	57	60	2375	880	7	**2375**	**2361**
1223	6x12	58	57	2375	850	9	**2375**	**2356**
1224	6x13	51	62	2355	1170	11.5	**2355**	**2373**
1225	6x13	51	64	2355	1050	10	**2358**	**2363**
1226	6x13	54	62	2355	1100	11		
1227	6x13	54	64	2355	1100	10	**2355**	**2657**
1228	6x13	54	65	2355	870	8.5		
1229	6x13	55	63	2355	1080	10.5	**2355**	**2012**
1230	6x13	56	64	2355	1000	11		
1231	6x13	57	65	2355	960	8.5		
1232	6x13	58	64	2355	970	7.5		
1233	6x13	60	61	2355	980	11.5	**2355**	**2377**
1234	6x13	60	63	2355	930	8.5	**2355**	**2658**
1236	6x15	51	72	2369	980	11		
1237	6x15	52	71	2369	980	8.5		
1238	6x15	53	74	2369	980	8	**2373**	**2984**
1239	6x15	54	73	2369	1070	9	**2373**	**2981**
1241	6x15	55	73	2369	1050	10	**2373**	**2005**
1242	6x15	56	75	2369	860	7.5	**2373**	**2380**
1243	6x15	57	72	2369	960	9	**2373**	**2011**
1244	6x15	59	71	2369	1120	10		
1245	6x15	59	73	2369	920	8.5	**2373**	**2352**
1246	6x15	60	71	2369	1020	8.5	**2373**	**2372**
1247	7x1	61	5	2369	1050	11.5	**2373**	**2099**
1248	7x1	62	3	2369	1050	11.5	**2373**	**2361**
1250	7x1	68	4	2369	1030	10	**2373**	**2358**
1252	7x1	69	4	2369	1040	10.5	**2373**	**2380**
1255	7x5	61	25	2353	940	8.5	**2353**	**2377**

1256	7x5	62	22	2353	880	7	**2353**	**2371**
1257	7x5	62	24	2353	920	10	**2353**	**S-5**
1258	7x5	66	21	2353	860	8	**2353**	**2656**
1259	7x5	66	23	2353	1110	13	**2353**	**2352**
1260	7x5	67	24	2353	920	8.5	**2353**	**S-6**
1262	7x5	68	21	2353	820	7	**2353**	**2097**
1263	7x5	69	25	2353	930	10	**2353**	**2355**
1264	7x11	63	51	2657	1240	11.5	**2657**	**2351**
1265	7x11	63	52	2657	1030	8.5	**2657**	**2373**
1266	7x11	64	53	2657	970	9	**2657**	**2352**
1267	7x11	66	55	2657	1000	8	**2657**	**2656**
1268	7x11	67	52	2657	1050	10	**2657**	**2358**
1269	7x11	69	51	2657	1140	10	**2657**	**2984**
1270	7x12	61	56	S 5	930	8		
1271	7x12	61	57	S 5	850	9		
1272	7x12	62	58	S 5	930	9		
1273	7x12	63	57	S 5	920	11		
1274	7x12	63	59	S 5	880	9.5		
1275	7x12	66	56	S 5	910	7.5		
1276	7x12	66	59	S 5	970	11		
1277	7x12	69	57	S 5	940	8.5		
1278	7x12	69	60	S 5	860	7		
1279	8x2	72	9	S 2	1110	12.5	**S-2**	**2351**
1281	8x2	76	7	S 2	1060	12	**S-2**	**2351**
1282	8x2	79	10	S 2	890	7.5		
1288	8x4	78	17	2371	1020	8	**2371**	**2349**
1290	8x4	79	16	2371	1010	12		
1294	8x5	75	25	2359	1070	11	**2359**	**2002**
1296	8x5	76	25	2359	950	11	**2359**	**2998**
1299	8x5	78	24	2359	960	8	**2359**	**2373**
1300	8x5	79	23	2359	1010	9	**2359**	**2991**
1301	8x5	79	25	2359	1030	11	**2359**	**2099**
1302	8x9	71	43	2355	1090	11		
1304	8x9	71	45	2355	1080	10	**2355**	**2356**
1306	8x9	74	44	2355	930	8.5		
1308	8x9	76	41	2355	1020	10	**2355**	**2380**
1312	8x9	78	43	2355	1060	11		
1313	8x9	78	44	2355	920	9		
1314	8x9	79	41	2355	1130	10.5		
1315	8x9	79	42	2355	1050	8.5	**2355**	**S-6**
1317	8x9	79	45	2355	1000	10		
1318	8x9	80	41	2355	1060	10.5	**2355**	**20012**
1319	8x9	80	42	2355	1110	11		
1320	8x9	80	43	2355	1070	10.5	**2355**	**2358**
1321	8x10	72	50	2369	950	8		

1323	8x10	76	49	2369	1210	12	**2373**	**2776**
1324	8x10	78	46	2369	1020	9.5	**2373**	**2779**
1325	8x10	79	50	2369	980	7.5	**2373**	**2372**
1326	8x11	72	52	2375	1030	9.5	**2375**	**2361**
1327	8x11	72	55	2375	1030	12	**2375**	**2361**
1328	8x11	75	52	2375	920	7	**2375**	**2361**
1329	8x11	76	51	2375	1040	12	**2375**	**2375**
1330	9x2	83	8	2359	1160	13.5	**2359**	**2366**
1331	9x2	86	10	2359	1070	9.5	**2359**	**2352**
1332	9x2	87	8	2359	1060	8		
1333	9x2	89	6	2359	940	10	**2359**	**2992**
1336	9x9	81	43	S 2	930	10	**S-2**	**S-5**
1338	9x9	82	43	S 2	890	8	**S-2**	**2376**
1339	9x9	82	45	S 2	940	8	**S-2**	**2358**
1340	9x9	83	44	S 2	1080	12.5	**S-2**	**S-6**
1342	9x9	89	41	S 2	1050	10.5	**S-2**	**2993**
1344	9x9	90	42	S 2	1080	10	**S-2**	**2004**
1347	10x5	92	23	S 2	930	10	**S-2**	**S-15**
1350	10x5	93	22	S 2	1040	10	**S-2**	**2373**
1351	10x5	93	24	S 2	970	8	**S-2**	**2372**
1352	10x5	94	25	S 2	990	12	**S-2**	**S-5**
1354	10x5	97	23	S 2	980	8	**S-2**	**2358**
1358	10x5	99	23	S 2	980	11	**S-2**	**2370**
1360	10x12	93	56	2351	1020	10.5		
1361	10x12	93	58	2351	910	8.5		
1362	10x12	94	56	2351	1020	10		
1363	10x12	94	58	2351	1030	9		
1364	10x12	95	58	2351	1040	11		
1365	10x12	95	59	2351	950	9.5	**2351**	**2991**
1366	10x12	95	60	2351	910	7.5		
1367	10x12	97	60	2351	1050	10	**2351**	**2375**
1368	10x12	98	56	2351	960	9.5		
1369	10x12	98	57	2351	1010	9.5	**2351**	**2356**
1370	10x15	94	75	S 6	1050	10	**S-6**	**2377**
1371	10x15	95	71	S 6	1020	9.5		
1372	10x15	96	71	S 6	1120	10.5	**S-6**	**2374**
1376	11x1	108	3	S 5	1040	11.5		
1377	11x1	110	5	S 5	1000	11		
1378	11x2	101	9	S 6	1130	9	**S-6**	**2366**
1379	11x2	101	10	S 6	1230	13	**S-6**	**2367**
1380	11x2	103	8	S 6	1210	10.5	**S-6**	**2367**
1381	11x2	104	7	S 6	1170	10	**S-6**	**2989**
1382	11x2	105	8	S 6	1160	10.5	**S-6**	**2994**
1383	11x2	106	6	S 6	1110	13.5	**S-6**	**2996**
1384	11x4	102	17	2351	1080	9	**2351**	**2361**

1386	11x4	103	20	2351	1130	13	**2351**	**2981**
1387	11x4	104	17	2351	1220	12	**2351**	**2356**
1389	11x4	105	18	2351	1040	9.5		
1391	11x4	108	18	2351	1150	12		
1392	11x4	108	19	2351	1150	11	**2351**	**S-4**
1393	11x4	109	16	2351	1110	11	**2351**	**2376** **2918/2**
1394	11x4	110	18	2351	1160	11	**2351**	**978**
1398	11x6	103	27	2358	1020	9		
1399	11x6	104	30	2358	1110	13.5	**2358**	**20111**
1400	11x6	105	28	2358	1110	11	**2358**	**2370**
1401	11x6	107	26	2358	1150	12	**2358**	**2371**
1403	11x6	108	29	2358	1100	9		
1405	11x6	110	28	2358	1050	9.5	**2358**	**2352**
1406	11x6	110	30	2358	1090	10	**2358**	**2658**
1407	11x8	108	36	2657	1130	12	**2657**	**2987**
1408	11x8	108	40	2657	1060	11		
1409	11x9	101	45	2359	940	8	**2359**	**2353**
1410	11x9	102	42	2359	1030	9	**2359**	**S-2**
1411	11x9	102	43	2359	1000	8		
1416	11x9	104	45	2359	980	10.5	**2359**	**2998**
1417	11x9	105	43	2359	1090	11	**2359**	**2355**
1419	11x9	110	42	2359	940	8	**2359**	**2008**
1421	11x11	102	51	2358	1080	10	**2358**	**2372**
1422	11x11	105	53	2358	1040	10.5	**2358**	**2008**
1423	11x11	105	55	2358	970	7.5	**2358**	**2099**
1424	11x11	107	54	2358	990	9.5	**2358**	**2373**
1425	11x11	110	54	2358	1040	12	**2358**	**2373**
1426	11x12	101	56	2371	1020	10.5	**2371**	**2380**
1427	11x12	101	57	2371	890	8	**2371**	**2982**
1429	11x12	103	58	2371	950	9	**2371**	**2779**
1430	11x12	103	59	2371	930	9.5	**2371**	**2991**
1431	11x12	105	60	2371	1020	11	**2371**	**2361**
1433	11x12	107	59	2371	930	10	**2371**	**2372**
1435	3x13	30	61	2358			**2358**	**2371**
1436	3x1	29	5					
1438	4x8	40	36	2351			**2351**	**2354**
1439	2x10	20	49	S 5				
1440	1x12	10	56	2659			**2659**	**2103**
1441	6x12	60	58	2375			**2375**	**2656**
1442	11x9	107	44	2359			**2359**	**2352**
1443	2x10	14	50	S 5				
1444	11x11	110	51	2358				
1445	6x1	55	5	2351			**2351**	**2374**
1446	3x4	27	18	S 5				

1447	5x2	47	10	2353			
1448	6x1	55	4	2351		**2351**	**2380**
1449	6x2	55	7	S 3			
1450	1x6	10	30	2657		**2657**	**2985**
1451	7x5	68	22	2353		**2353**	**2002**

Comments: Blank cells (except in columns confirmed mother and assigned father) are missing values. If parameter in columns confirmed mother and assigned father is not filled in, it indicates inconsistency between assumed mother and assigned individuals, i.e. pedigree reconstruction wasn't performed. Mothers colored in red were confirmed to have distinctive genotype than stated according to the records (2369 was estimated 2373).