Final protocols to monitor genetic diversity of Fagus sylvatica, Quercus spp, Picea abies and P. pinaster at pan-European scale

Deliverable D2.1

The research that led to these results was funded by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 289119.

Document Properties

Document number	FP7-FORGER 289119 / D2.1
Document title	Final protocols to monitor genetic diversity of Fagus sylvatica, Quercus spp, Picea abies and P. pinaster at pan-European scale
Deliverable number	D2.1
Work package nr.	WP 2
Work package title	
Responsible	Bernd Degen
Due time (month)	42
Dissemination level	PU
Dependencies	related to WP 4
Author(s)	Bernd Degen, Celine Blanc-Jolivet, Jaroslaw Burczyk, Igor Chybicki, Giuseppe Vendramin, Andrea Piotti, Koen Kramer
Version	V1.2
Date of last revision	27.11.2015
Status	final

Abstract:

We conducted a pilot study on genetic monitoring of four European forest species. In each plot, adults, seedlings and seeds were sampled and genetically analysed at two types of molecular markers. Spatial configuration of stands, demographic parameters and geographic position were analysed together with genetic parameters to find correlates of genetic diversity. Results were discussed and recommendations could be provided to improve the existing protocols in a cost-effective way.

Keywords:

Monitoring, genetic diversity, molecular marker, forest management

Contents

Executive summary

Genetic diversity is important to ensure adaptability of forest trees to changing biotic and abiotic factors. The concept of genetic monitoring has been developed to estimate changes in genetic diversity, which are monitored with genetic and demographic indicators. Several studies have been conducted on this purpose in Europe, but there is still discussion on the protocols which should be applied. In particular, the decreasing costs of genetic methods make such protocols attractive, as cost-effectiveness is an important factor.

We conducted a pilot study on *Fagus sylvatica*, *Quercus robur*, *Picea abies* and *Pinus pinaster* at the European scale. Four monitoring plots were selected per species The plots contrasted either in geographical position, management intensity or degree of isolation. In each plot, adults, seedlings and seeds were sampled and genetically analysed for diversity in microsatellite (SSR) and SNP nuclear loci. Spatial configuration of stands (spatial position of adults and seedlings), demographic parameters (diameter at breast height) and geographic position were analysed together with genetic parameters to find correlates of genetic diversity.

We found that plots located in eastern Europe exhibited higher genetic diversity, suggesting that biogeographic history and actual climate do influence levels of diversity. Within a plot family structure had a negative impact on genetic diversity, indicating an effect of management regimes.

Based on our results, we recommend: 1) to sample only adults and seedlings 2) to use only one type of molecular marker (SNP), and 3) to analyse a minimum of 100 adults and 100 seedlings well-distributed over the stand at a minimum of 150 loci. To provide further insights on the dynamics of genetic diversity, the analysis should be repeated every 10 years.

1 Introduction

Background

Genetic diversity is elementary for the adaptability of tree populations (EEA 2010d). This calls for a system to monitor the dynamics of genetic diversity of trees in a network of plots to detect changes in genetic diversity. An early concept for the monitoring of impact of forest management on genetic diversity was developed by Namkoong et al. (1996). This concept intended to use genetic and demographic indicators to evaluate the efficiency of driving demographic processes, such as genetic drift, migration, mating system, to maintain existing levels of genetic diversity. Following this concept, a pilot study of genetic monitoring for beech (Fagus sylvatica) and Wild Cherry (Prunus avium) was applied in Germany in 2006 to 2008 (Gregorius & Degen 2006, Konnert et al. 2010, Jolivet et al. 2011).

There is currently a strong advance in genomics and statistical tools that make genetic monitoring more efficient and economical (Schwartz 2006). Genetic diversity and driving processes can be measured better than ever before. Neutral genetic markers such as microsatellites have become an efficient tool to study genetic variation and to infer demographic processes (Chybicki and Burczyk 2010). Bevor the project started evidence of association between single nucleotide polymorphism (SNP) and adaptive traits have been found in forest trees (Holliday et al. 2010, Derory et al. 2010). But before the project SNPs in adaptive genes have been never applied for monitoring purposes. In short, we assumed that the rapid developments of new molecular genetic techniques and the fast decrease of costs involved, allow genetics to become an important tool for monitoring changes in diversity, both at the neutral and adaptive level.

Tasks within ForGer

The work on genetic monitoring in ForGer is subdivided in three sub-tasks:

- 2.1.1. Selection of sites and determination of monitoring intensity of the pilot study on Fagus sylvatica, Quercus spp., Picea abies and Pinus pinaster
- 2.1.2. Pilot study genetic monitoring on Quercus robur, Fagus sylvatica, Picea abies and Pinus pinaster at the European scale
- 2.1.3. Formulating improved protocols for the monitoring of genetic diversity of trees in Europe

We aimed to run a the pilot study for genetic monitoring on a European scale. We wanted to use the results of the pilot study to formulate an improved protocols for monitoring genetic diversity in Fagus sylvatica, Quercus spp., Picea abies and Pinus pinaster. The protocol should be applicable throughout Europe and should include:

- the criteria for the selection of monitoring plots
- a sampling design for the efficient and effective monitoring of genetic diversity
- the criteria to select gene markers to apply in the monitoring
- a detailed description of the methodology to screen genetic and demographic
- providing reference values for genetic and demographic parameters representing
- a detailed description of ecological alternatives for economic, rapid and sound

The protocols on genetic monitoring should be developed in close collaboration of with FORGER's stakeholders.

2 Method

2.1. Selection of monitoring plots

The analyses of the pilot study for genetic monitoring were done based on the set of monitoring plots established for four species: *Fagus sylvatica*, *Quercus* spp., *Picea abies* and *Pinus pinaster*. For each species four sample plots have been selected, in different countries. We selected isolated and non-isolated *Quercus robur* stands, and different management regimes in *Fagus sylvatica* (Table 1).

Table 1: *Stands selected for the study*

Species	Stand Name	Country	Coordinates	Selection criteria	Type
Quercus robur	Lubartów	Poland	N 22°77'82.64" E 51°46'81.28"	not isolated	Gene Conservation Unit
Quercus robur	Grosser Bruch	Germany	N 52°2'44.10" E 8°34'23.35"	not isolated	
Quercus robur	Lame del Sesia	Italy	N 45°25'36'' E 8°23'45''	isolated	
Quercus robur	Lot-et-Garonne	France	N 44°12'21" E 0°09'38"	isolated	
Fagus sylvatica	Behlendorf	Germany	N 53°42'41.5" E 10°39'51.91"	species diversity/ low management	
Fagus sylvatica	Pradaccio	Italy	N 44°23'46'' E 10°00'53''	pure stand/ coppice	
Fagus sylvatica	Lesko	Poland	N 49°23'58.53" E 22°14'57.08"	almost pure stand/ low management	Gene Conservation Unit
Fagus sylvatica	Solling	Germany	N 51°44' E 009°39'	species diversity/ intensive management	ISS
Pinus pinaster	Montefalcone	Italy	N 43°44'44'' E 10°42'58''	geographical distribution	
Pinus pinaster	Ain	Spain	N 39° 53' 38'' W 0° 21' 09''	geographical distribution	Gene Conservation Unit
Pinus pinaster	Coca	Spain	N 41° 14' 19'' W 4° 31' 10''	geographical distribution	Gene Conservation Unit
Pinus pinaster	Lacanau	France	N 44°57'37'' W 001°09'47''	geographical distribution	
Picea abies	Paneveggio1	Italy	N 46°18' E 11°45'	geographical distribution	
Picea abies	Struga Żytliejmska	Poland	N 54°20'44.52'' E	geographical distribution	Gene Conservation Unit

			22°38'34.16''		
Picea abies	Nowy Targ (Stańcowa)	Poland	N 49°32'50.47" E 19°01'53.84"	geographical distribution	Gene Conservation Unit
Picea abies	Punkaharju	Finland	N 61°48' E 029°19'	geographical distribution	ISS

2.2. Sampling design for monitoring of genetic diversity

In each stand, plots of 100 neighbour adult trees were sampled. Individuals with DBH > 10 cm (*Fagus sylvatica*, *Quercus robur*, *Picea abies*) and DBH > 3 cm (*Pinus pinaster*) were considered as adults. Among those, 15 trees well-dispersed along the stand were selected for seed harvesting. When possible, 20 seeds per seed-tree were sampled. In most *Quercus robur* and *Fagus sylvatica* stands, seeds were sampled from the ground. Further, 100 seedlings were selected within the plot. For each adult and seedling, spatial coordinates were measured as well as diameter at breast height in adults. DNA from cambium, leaf, needle or embryo material was extracted from all individuals. A subset of 96 adults and seedlings was defined for each stand for further analysis.

2.3. Selection of gene markers to be applied in monitoring of genetic diversity.

Two types of genetic markers have been selected for this study. Firstly, all sampled material was genotyped using microsatellites as genetic markers. Specific sets of loci have been used for each species:

Fagus sylvatica

Altogether, 16 loci (*csolfagus_05, csolfagus_06, csolfagus_19, csolfagus_29, csolfagus_31, FS105, sfc_0036, sfc_1143, concat14, DE576, DUCKT, DZ447, EEU75, EJV8T, Emily, ERHBI;* Lefevre et al. 2012) were used for genotyping. The loci were designed into two validated multiplexes according to Lefevre et al. (2012). The two multiplexes have been finally used in genotyping all sampled material. PCR reactions were conducted with PTC200 thermal cycler (MJ Research). PCR products were sized using a capillary sequencer ABI PRISM 3130XL (Applied Biosystems, Foster City, USA).

Picea abies

Eleven nuclear microsatellite markers have been selected for this study (*SS52,* A'Hara & Cottrell 2007, *SpAGC1,SpAG2, SpAGG3*, Pfeiffer et al. 1997, *PAAC23,* Rungis 2004, *PAAC3,* Scotti et al. 2000, *Eatc1D02A, Eatc3H03, Eatc1B02,* Scotti et al. 2002, *Eatc2B02, Eatc2G05*, Tollefsrud et al. 2009). Out of 11 loci, two PCR multiplexes were optimized including six (marker set I, *SpAGC1,SpAG2, SpAGG3, Eatc1D02A, Eatc2B02, Eatc2G05*) five (marker set II, *SS52, PAAC23, PAAC3, Eatc3H03, Eatc1B02*). PCR reactions were conducted with PTC200 thermal cycler (MJ Research). PCR products were sized using a capillary sequencer ABI PRISM 3130XL (Applied Biosystems, Foster City, USA).

Pinus pinaster

DNA from dried needle material was extracted according to Dumolin *et al.* (1995). Genetic diversity was screened at five nuclear SSR markers optimized in a single multiplex (epi3-FW (Sebastiani & Vendramin), gPp14-FW (Pinzauti *et al.* 2012), A6F03-04 (Guevara *et al.* 2005), epi5-FW (Sebastiani & Vendramin), and NZPR1078-5 (Chagne *et al.* 2004). Fragment separation occurred

on a ABI 3730 capillary sequencer and scoring was performed with GeneMarker v 2.4.0 (© Softgenetics).

Quercus robur

DNA from dried leaf or cambium material was extracted according to Dumolin *et al.* (1995). Genetic diversity was screened at a 8-plex set of nuclear SSR markers described in Guichoux *et al.* (2011). Fragment separation occurred on a ABI 3730 capillary sequencer and scoring was performed with GeneMarker v 2.4.0 (© Softgenetics).

Secondly, samples of adults and seedlings (complete data sets: 96 adults + 96 seedlings x 4 stands x 4 species), have been subjected to SNP genotyping.

2.4. Demographic parameters describing monitoring plots

All stands (adult populations) were characterized with a number ecological/demography parameters. These included: area of sample plot, tree density and mean distance to nearest neighbour. The R statistics was used to describe pattern of spatial structure of the stand (R=1 – random distribution; R<1 – clustering; R>1 – overdispersion, i.e. uniform distribution), which was tested (based on Z-test) for departure from the null hypothesis of R=1.

Each adult tree was measured for DBH (diameter at 1.30m above ground). Standard descriptive statistics of DBH were calculated. Additionally, the distribution of DBH within the stand was tested to see any pattern of spatial autocorrelation of DBH. Here, the overall significance of Moran's I auto-correlogram, and the value of autocorrelation in the first distance class were provided.

2.5. Parameters describing genetic diversity of populations

We used GDA_NT (Degen, unpublished) to estimate the effective number of alleles (Ae), the unordered number of single-locus genotypes (NG), the intrapopulation fixation index (Fis), and the genetic distance from Gregorius among ontogenetic stages within and among populations.

2.6. Parameters describing the processes affecting genetic structure of populations

2.6.1. Mating system

Based on microsatellite genotypes of seeds assigned to maternal families, the following parameters of mating system were estimated: i) single-locus outcrossing rate (t_s), multi-locus outcrossing rate (t_m), iii) single-locus correlation of allele frequencies in pollen pool (r_{ps}), iv) multi-locus correlation of allele frequencies in pollen pool (r_{pm}), correlation of outcrossing between loci (r_{tl}), correlation of outcrossing among progeny (r_{tp}). In addition, differences $t_m - t_s$ and $r_{ps} - r_{pm}$ were of interest because they are informative about biparental inbreeding and genetic differentiation between maternal neighbourhoods. The parameters were estimated with MLTR 3.4 program (Ritland 2002). Standard errors of estimates were approximated with 1000 bootstrap samples (family taken as a re-sampling unit).

2.6.2. Spatial genetic structure

The extent and intensity of the spatial genetic structure characterizing each population at each life stage were investigated through spatial autocorrelation analysis for both marker types. Spatial

autocorrelation analysis was performed calculating the Nason's kinship coefficient F_{ij} (Loiselle et al. 1995) for each pair of individuals. Tests for statistical significance of average F_{ij} values for each distance class were conducted by *i*) permutation of individual locations, and *ii*) jackknifying over loci. Analyses were performed using both the even distance classes option (using 5 m, 10 m and 20 m wide distance classes) and the even sample size options (distributing all possible pairs in ten distance classes with similar numbers of pairs per class). The intensity of SGS was measured by the *Sp* statistic (Vekemans & Hardy 2004), computed as $Sp = b_F/(F_1-1)$, where b_F is the regression slope of the kinship estimator F_{ij} computed among all pairs of individuals against the natural logarithm of their geographical distances, and F_1 is the average kinship coefficient between individuals of the first distance class (0–20 m). All analyses were performed using SPAGeDi 1.5 (Hardy & Vekemans 2002).

2.6.3. Gene dispersal

The extent of gene dispersal through seeds and pollen was investigated based on parentage/paternity modelling. For this purpose, genotypes of seeds/seedlings were taken as a sample from progeny generation while genotypes of trees were used as a sample of parental generation (or candidate parents to the sampled progeny). The parentage was modelled using the spatially-explicit mating model (the *neighborhood* model; Adams and Birkes 1989; Burczyk et al. 2006), which allows to characterize so-called dispersal kernels (probability functions of seed or pollen dispersal distance) through their means (mean dispersal distances).
In addition, self-fertilization (s) and immigration of pollen (m_p) or seeds (m_s) from outside the plot as well as the effects (selection gradients) of trunk diameter on male (b_m) and female (b_f) reproductive success were estimated. In this study we assumed that seed and pollen dispersal follows the exponential kernel, which represents intermediate probability distribution between thin- and fat-tailed dispersal kernels. Standard errors of estimates were computed using the inversed Hessian. The estimates were obtained using NM+ 1.1 software (Chybicki and Burczyk 2010), with a possibility of genotyping errors taken into account.

2.6.4. Effective population size

We estimated the effective population size based on linkage disequilibrium as implemented in NeEstimator V2.01 (Do *et al.* 2014). This approach was first proposed by Hill (1981) and further developed by Robinson and Moyer (2012) for species with overlapping generations. These studies found that the efficiency of this approach increases with sample size and with more tightly linked genes. England et al. (2006) recommend to minimally use ~100 individuals to avoid biased estimates.

3 Results

3.1. Demographic parameters describing monitoring plots

Stand structure

Sampling of adult populations was designed to collect the data on ca. 100 individuals growing together. This means that the 100 individuals formed a patch of all individuals sampled (no missing data). This sample size and density of trees within the stand determined directly the area of the study plot.

Stand size and density

Plot area and mean density varied considerably among study plots within the species. For oak, the French stand had the highest and Polish stand had the lowest density, with almost 10-fold difference (Table 2). Similarly, beech stand varied considerably with Italian stand having highest density, and Behlendorf the lowest density (Table 3). High density contrasts were also revealed for Maritime pine, where Italian stand had the highest density, but the Coca stand in Spain the lowest density (Table 4). Unlike the other species, spruce stands were quite similar in terms of area size and stand density (Table 5). The information on area of the sample plot and stand density are important for interpreting the results of genetic diversity, spatial genetic structure and patterns of seed and pollen dispersal.

Spatial structure

Distribution of adults within stands was also variable. Most of the oak stands showed random distribution of individuals, except the French stand where some overdispersion (uniform distribution) has been detected (Table 2). In case of beech, significant overdispersion was found in Prodaccio (Italy) and Solling (Germany), while another German stand (Behlendorf) revealed significant spatial clustering of adults within the stand (Table 3). Maritime pine indicated the tendency for overdispersion (noted in two stands) (Table 4). In Norway spruce, Żytkiejmy stand in Poland exhibited clustering (this is a nature reserve), while Finnish stand had significantly uniform distribution of trees (Table 5). These measures of spatial distribution of adults, although focused on nearest neighbour distance provide important information on clustering versus orverdispersion for the first distance classes and do not describe fully the dispersal patterns at larger distances, but anyway, it provides important data on stand structure useful for interpretation of the results of genetic diversity, spatial genetic structure and patterns of seed and pollen dispersal.

Table 2: *Characteristics of study sites and patterns of spatial structure of Quercus robur stands.*

Parameter	**France (Lot-et-Garonne)**	**Germany (Grosser Bruch)**	**Italy (Lame del Sesia)**	**Poland (Lubartow)**
Area [m2]	3653.7	11579.0	16331.0	23402.0
Mean density	0.0263	0.0083	0.0059	0.0041
Nearest neighbor mean distance	3.921	6.015	7.0377	7.1085
Nearest neighbor expected distance	3.0846	5.4914	6.5213	7.8067
R:	**1.2711**	1.0954	1.0792	0.91057
Z-test	5.0824	1.7873	1.4843	-1.6763
P(random)	3.73E-07	0.073893	0.13772	0.093677
Comment:	**Overdisp**.	Random	Random	Random

Note – significant parameters are indicated with **bold face**

Table 3: *Characteristics of study sites and patterns of spatial structure of Fagus sylvatica stands.*

Parameter	**Germany (Behlendorf)**	**Poland (Lesko)**	**Italy (Pradaccio)**	**Germany (Solling)**
Area [m2]	14891.0	5258.8	1107.5	9154.4
Mean density	0.0064	0.0183	0.0867	0.0105
Nearest neighbor mean distance	5.0857	3.5967	2.0988	5.5032
Nearest neighbor expected distance	6.2599	3.7006	1.6983	4.8826
R:	**0.8124**	0.9719	**1.2358**	**1.1271**
Z-test	-3.4976	-0.52666	4.4207	2.3826
P(random)	0.0005	0.5984	9.84E-06	0.0172
Comment:	**Clustering**	Random	**Overdisp**.	**Overdisp**.

Note – significant parameters are indicated with **bold face**

Table 4: *Characteristics of study sites and patterns of spatial structure of Pinus pinaster stands.*

Parameter	**Spain (Ain)**	**Spain (Coca)**	**Italy (Montefalcone)**	**France (Lacanau)**
Area [m2]	2737.0	9812.0	654.1	1664.0
Mean density	0.0351	0.0098	0.1468	0.0577
Nearest neighbor mean distance	2.9026	5.8193	1.2381	2.8694
Nearest neighbor expected distance	2.6698	5.0549	1.3051	2.0817
R:	1.0872	**1.1512**	0.94863	**1.3784**
Z-test	1.6347	2.8344	-0.96288	7.093
P(random)	0.1021	0.0046	0.3356	1.31E-12
Comment:	Random	**Overdisp**.	Random	**Overdisp**.

Note – significant parameters are indicated with **bold face**

Table 5: *Characteristics of study sites and patterns of spatial structure of Picea abies stands.*

Parameter	**Poland (Stańcowa)**	**Italy (Paneveggio)**	**Poland (Żytkiejmy)**	**Finland, (Punkaharju)**
Area [m2]	2866.8	3267.5	1796.5	2207.7
Mean density	0.0335	0.0294	0.0534	0.0435
Nearest neighbor mean distance	2.5139	3.2017	1.6879	2.9053
Nearest neighbor expected distance	2.7323	2.9171	2.163	2.3978
R:	0.9200	1.0976	**0.7804**	**1.2117**
Z-test	1.4983	1.8292	-4.1167	3.9679
P(random)	0.1340	0.0674	3.84E-05	7.25E-05
Comment:	Random	Random	**Clustering**	**Overdisp**.

Note – significant parameters are indicated with **bold face**

Tree diameter measures

Descriptive statistics on DBH measures are provided for each species and stands in tables 5-8. In general, within each species variation of DBH among stands was observed, indicating that sampled stands represent a broad variation of stand types (younger with lower DBH to older with larger DBH). Skewness and kurtosis highlighted if the frequency distribution of DBH departures from normal distribution. Also, coefficient of variation (C.V.) appeared to be variable among the stands within species. High C.V. values indicate that a stand is not uniform in respect to tree size (and likely age), which might be an additional indicator of stand composed with multiple generations, typical for naturally established (or not severely managed) stands.

Spatial autocorrelation of DBH

Detection of strong spatial autocorrelation of DBH was emphasized by statistical significance Moran's *I* correlograms but also the significance of positive Moran's *I* coefficient in the first distance class. In case of strong autocorrelation both of these parameters are expected to be significant. Spatial autocorrelation of DBH was found in all species except Maritime pine (Table 8). In oak and beech it was found in two stands (Tables 6 and 7), while in spruce it was found in just one stand (Table 9).

Table 6: *Descriptive statistics and patterns of spatial autocorrelation of DBH found in Quercus robur stands.*

Parameter	**France (Lot-et-Garonne)**	**Germany (Grosser Bruch)**	**Italy (Lame del Sesia)**	**Poland (Lubartow)**
Mean	32.14	59.13	42.00	74.09
Median	32.00	58.50	45.22	73.00
Skewness	**1.1319**	0.3456	**-0.4329**	**0.7148**
Kurtosis	**2.7318**	0.2589	**-0.7685**	1.5433
Coeff. var	19.83	13.30	38.75	16.93
Significance of Moran's I correlogram of DBH (p-value)	**0.0002**	0.5234	**0.0013**	0.3901
Moran's I in the first distance class (SE)	**0.1395 (0.0436)**	-0.0558 (0.0438)	**0.1643 (0.0462)**	0.0688 (0.0431)
Comment on autocorrelation:	**strong**	none	**strong**	none

Note – significant parameters are indicated with **bold face**

Table 7: *Descriptive statistics and patterns of spatial autocorrelation of DBH found in Fagus sylvatica stands.*

Parameter	**Germany (Behlendorf)**	**Poland (Lesko)**	**Italy (Pradaccio)**	**Germany (Solling)**
Mean	40.73	44.95	24.28	49.07
Median	43.00	44.60	22.93	49.90
Skewness	-0.1160	0.2245	**1.6883**	0.1898
Kurtosis	**-0.9586**	-0.2775	**4.4805**	0.9641
Coeff. var	38.99	18.00	29.29	22.71
Significance of Moran's I correlogram of DBH (p-value)	1	**0.0166**	**0.0002**	**0.0420**
Moran's I in the first distance class (SE)	0.0272 (0.0443)	**0.1258 (0.0438)**	**0.1619 (0.0427)**	-0.0261 (0.0439)

Comment on autocorrelation:	none	**strong**	**strong**	weak

Note – significant parameters are indicated with **bold face**

Table 8: *Descriptive statistics and patterns of spatial autocorrelation of DBH found in Pinus pinaster stands.*

Parameter	**Spain (Ain)**	**Spain (Coca)**	**Italy (Montefalcone)**	**France (Lacanau)**
Mean	24.13	40.34	7.82	27.88
Median	24.45	38.80	7.96	28.03
Skewness	-0.2235	**0.7457**	**0.5268**	-0.0111
Kurtosis	0.2305	1.0147	0.2383	**-0.4542**
Coeff. var	22.48	20.25	30.44	16.79
Significance of Moran's I correlogram of DBH (p-value)	1	1	0.2206	0.1024
Moran's I in the first distance class (SE)	-0.0062 (0.0433)	0.0482 (0.0432)	0.0233 (0.0432)	0.0452 (0.0442)
Comment on autocorrelation:	none	none	none	none

Note – significant parameters are indicated with **bold face**

Table 9: *Descriptive statistics and patterns of spatial autocorrelation of DBH found in Picea abies stands.*

Parameter	**Poland (Stańcowa)**	**Italy (Paneveggio)**	**Poland (Żytkiejmy)**	**Finland (Punkaharju)**
Mean	46.86	39.45	26.35	25.16
Median	47.10	38.50	24.80	24.60
Skewness	**0.7141**	**0.3351**	**0.5863**	0.2481
Kurtosis	1.5845	**-1.0745**	-0.1184	-0.1231
Coeff. var	15.54	55.70	29.23	22.27
Significance of Moran's I correlogram of DBH (p-value)	0.3612	**<0.0001**	1	0.4273
Moran's I in the first distance class (SE)	-0.0526 (0.0442)	**0.2213 (0.0441)**	0.0117 (0.0439)	0.0056 (0.0440)
Comment on autocorrelation:	none	**strong**	none	none

Note – significant parameters are indicated with **bold face**

3.2. Parameters describing genetic diversity of populations

Genetic diversity varied among stands, and populations located in Italy and France tended to show lower genetic variation (Table 10). Genetic distance among ontogenetic stages within stands was lower than among stands (Tables 11, 12, 13 and 14). For each species and taking into account only adults and seedlings, the maximum genetic distance among ontogenetic stages within stands (ranged from 0.102 and 0.176) was lower than the minimum genetic distance observed among stands (ranged from 0.143 and 0.247), thus showing that no substantial changes in genetic composition occurred in the seedlings compared to adults. **A genetic distance from Gregorius among ontogenetic stages within a stand higher than 0.150 might indicate ongoing genetic changes.** Indeed, in the *Quercus robur* stand located in Poland, genetic distance among adults and seedlings was 0.176. This could be explained by the low number of seedling genotypes available (33).

Table 10: *Parameters of genetic diversity (Effective number of alleles Ae and Unordered number of single-locus genotypes NG) in the studied stands.*

Species	Stand	Stage	Ae_microsat	Ae_SNP	NG_microsat	NG_SNP
Fagus	Germany (Behlendorf)	Adults	3.930	1.453	312	323
Fagus	Germany (Behlendorf)	Seedlings	3.695	1.458	310	327
Fagus	Poland (Lesko)	Adults	4.02	1.421	343	330
Fagus	Poland (Lesko)	Seedlings	4.193	1.418	330	325
Fagus	Italy (Pradaccio)	Adults	3.747	1.421	317	310
Fagus	Italy (Pradaccio)	Seedlings	3.938	1.423	308	322
Fagus	Germany (Solling)	Adults	4.029	1.433	311	327
Fagus	Germany (Solling)	Seedlings	3.62	1.457	305	328
Pinus	Spain (Ain)	Adults	2.399	1.57	51	346
Pinus	Spain (Ain)	Seedlings	2.446	1.574	46	349
Pinus	Spain (Coca)	Adults	2.699	1.567	56	351
Pinus	Spain (Coca)	Seedlings	2.563	1.559	53	348
Pinus	Italy (Montefalcone)	Adults	1.856	1.417	24	303
Pinus	Italy (Montefalcone)	Seedlings	1.737	1.418	22	298
Pinus	France (Lacanau)	Adults	2.198	1.569	42	344
Pinus	France (Lacanau)	Seedlings	2.402	1.584	40	347
Picea	Poland	Adults	5.497	-	408	-

	(Stancowa)					
Picea	Poland (Stancowa)	Seedlings	5.704	-	406	-
Picea	Italy (Paneveggio)	Adults	5.036	-	368	-
Picea	Italy (Paneveggio)	Seedlings	5.301	-	354	-
Picea	Poland (Zytkiejmy)	Adults	6.011	-	425	-
Picea	Poland (Zytkiejmy)	Seedlings	5.935	-	416	-
Picea	Finland (Punkaharju)	Adults	6.363	-	400	-
Picea	Finland (Punkaharju)	Seedlings	6.397	-	414	-
Quercus	France (Lot-et-Garonne)	Adults	4.454	-	230	-
Quercus	France (Lot-et-Garonne)	Seedlings	4.459	-	208	-
Quercus	Germany (Grosser Bruch)	Adults	6.181	-	278	-
Quercus	Germany (Grosser Bruch)	Seedlings	6.008	-	283	-
Quercus	Italy (Lame del Sesia)	Adults	4.973	-	235	-
Quercus	Italy (Lame del Sesia)	Seedlings	4.984	-	215	-
Quercus	Poland (Lubartow)	Adults	5.866	-	269	-
Quercus	Poland (Lubartow)	Seedlings	5.511	-	147	-

Table 11: *Genetic distance (Gregorius) among and within Fagus sylvatica stands*

		Germany (Behlendorf)		Poland (Lesko)			Italy (Pradaccio)			Germany (Solling)		
		Seedlings	Seeds	Adults	Seedlings	Seeds	Adults	Seedlings	Seeds	Adults	Seedlings	Seeds
Germany (Behlendorf)	Adults	0.107	0.086	0.239	0.234	0.239	0.221	0.261	0.238	0.15	0.157	0.149
	Seedlings		0.117	0.227	0.223	0.229	0.217	0.251	0.227	0.154	0.143	0.15
	Seeds			0.238	0.23	0.238	0.224	0.27	0.236	0.147	0.144	0.14
Poland (Lesko)	Adults				0.116	0.092	0.257	0.287	0.28	0.196	0.213	0.209
	Seedlings					0.137	0.268	0.294	0.295	0.205	0.225	0.219
	Seeds						0.263	0.293	0.282	0.199	0.215	0.205
Italy (Pradaccio)	Adults							0.123	0.138	0.195	0.228	0.197
	Seedlings								0.154	0.237	0.264	0.237
	Seeds									0.235	0.238	0.223
Germany (Solling)	Adults										0.115	0.064
	Seedlings											0.112

Table 12: *Genetic distance (Gregorius) among and within Pinus pinaster stands*

		Spain (Ain)		Spain (Coca)			Italy (Montefalcone)			France (Lacanau)	
		Seedlings	Seeds	Adults	Seedlings	Seeds	Adults	Seedlings	Seeds	Adults	Seedlings
Spain (Ain)	Adults	0.052	0.077	0.216	0.215	0.194	0.235	0.254	0.235	0.302	0.329
	Seedlings		0.084	0.233	0.226	0.201	0.244	0.261	0.258	0.31	0.323
	Seeds			0.186	0.184	0.167	0.244	0.263	0.241	0.278	0.294
Spain (Coca)	Adults				0.102	0.069	0.374	0.38	0.357	0.191	0.208
	Seedlings					0.094	0.344	0.346	0.337	0.24	0.236
	Seeds						0.344	0.354	0.339	0.192	0.215
Italy (Montefalcone)	Adults							0.06	0.083	0.417	0.434
	Seedlings								0.074	0.414	0.448
	Seeds									0.398	0.419
France (Lacanau)	Adults										0.082

Table 13: *Genetic distance (Gregorius) among and within Quercus robur stands*

		Italy (Lame del Sesia)			Poland (Lubartow)			Germany (Grosser Bruch)			France (Lot-et-Garonne)
		Seedlings	Seeds	Adults	Seedlings	Seeds	Adults	Seedlings	Seeds	Adults	Seedlings
Italy (Lame del Sesia)	Adults	0.169	0.128	0.247	0.311	0.254	0.29	0.286	0.29	0.415	0.418
	Seedlings		0.22	0.325	0.36	0.325	0.328	0.346	0.341	0.47	0.486
	Seeds			0.288	0.346	0.291	0.313	0.295	0.305	0.422	0.421
Poland (Lubartow)	Adults				0.176	0.124	0.278	0.287	0.291	0.433	0.419
	Seedlings					0.176	0.337	0.361	0.348	0.49	0.479
	Seeds						0.297	0.304	0.304	0.458	0.452
Germany (Grosser Bruch)	Adults							0.141	0.095	0.356	0.348
	Seedlings								0.155	0.326	0.325
	Seeds									0.342	0.345
France (Lot-et-Garonne)	Adults										0.15

Table 14: *Genetic distance (Gregorius) among and within Picea abies stands*

		Italy (Paneveggio)		Poland (Zytkiejmy)		Poland (Stancowa)		Finland (Punkaharju)		
		Seedlings	Adults	Seedlings	Seeds	Adults	Seedlings	Adults	Seedlings	Seeds
Italy (Paneveggio)	Adults	0.098	0.24	0.265	0.278	0.222	0.232	0.304	0.287	0.32
	Seedlings		0.242	0.265	0.273	0.225	0.226	0.296	0.276	0.313
Poland (Zytkiejmy)	Adults		0.124	0.153	0.17	0.172	0.215	0.18	0.233	
	Seedlings			0.156	0.192	0.188	0.211	0.192	0.236	
	Seeds				0.195	0.207	0.232	0.216	0.237	
Poland (Stancowa)	Adults					0.136	0.265	0.253	0.286	
	Seedlings						0.257	0.246	0.275	
Finland (Punkaharju)	Adults							0.148	0.163	
	Seedlings								0.182	

Inbreeding coefficients

Inbreeding coefficient (Wright's fixation index) (Fis) was calculated for adult and seedling cohorts in each stand of four studied species using SSR markers. The estimation procedure accounted for the presence of null alleles, which are often found in microsatellites causing an upward bias of the estimates of inbreeding coefficients (Chybicki and Burczyk 2009). Results are presented in Table 15.

Table 15: *The estimates of fixation index (Fis) based on SSR markers, when null alleles were taken into account during estimation of Fis (Chybicki and Burczyk 2009)*

Species / life stage	Stand name			
Fagus sylvatica	***Germany (Behlendorf)***	***Poland (Lesko)***	***Italy (Pradaccio)***	***Germany (Solling)***
Adults	0.0075	0.0090	0.0051	0.0046
Seedlings	0.0069	0.0025	0.0032	0.0132
Picea abies	**Poland (Stancowa)**	**Italy (Paneveggio)**	**Poland (Zytkiejmy)**	**Finland (Punkaharju))**
Adults	0.0184	0.0179	0.0440	0.0098
Seedlings	0.0271	***0.0579 ****	0.0391	0.0321
Pinus pinaster	***Spain (Ain)***	***Spain (Coca)***	***Italy (Montefalcone)***	***France (Lacanau)***
Adults	0.0379	0.0472	0.0597	0.0417
Seedlings	0.0128	0.0247	0.0225	***0.0490 ****
Quercus robur	***France (Lot-et-Garonne)***	***Germany (Grosser Bruch)***	***Italy (Lame del Sesia)***	***Poland (Lubartów)***
Adults	0.0311	0.0039	0.0093	0.0085
Seedlings	***0.0267 ****	0.0217	0.0090	0.0169

Country codes – D- Germany, F – France, FIN – Finland, I – Italy, PL – Poland, S – Spain.
* - significantly different from 0 at $p<0.05$

Generally, levels of inbreeding were found to be low and insignificantly different from 0. Significant levels of inbreeding were detected in few cases in seedlings: one population of *Picea abies* (Pannevegio, Italy), one population of *Pinus pinaster* (Lacanau, France), and also one population of *Quercus robur* (Lot-et-Garonne, France). No signs of inbreeding were found in adult populations of any species. Our results suggest, that **the levels of inbreeding should not be a major concern for genetic monitoring of populations of forest trees**. It confirms many other results that populations of forest trees exhibit low levels of inbreeding, which is in line with their life characteristics (large population sizes, high outcrossing rates, extensive gene flow within and among populations, low levels of genetic relatedness within populations, weak spatial genetic structure, etc.).

3.3. Parameters describing the reproductive processes affecting genetic structure of populations

3.3.1. Mating system.

No seeds were collected from French populations of *Quercus robur* and *Pinus pinaster*, and populations of *Picea abies* in one Polish stand (Stańcowa) and Italian stand. Therefore we could not obtain any estimates on mating system and pollen dispersal patterns based on seed samples for these populations (Table 16).

Multilocus outcrossing rates were not significantly different from 1, indicating no self-fertilization. On the contrary, single-locus outcrossing rates were lower than 1 in the German populations of oak and beech as well as Finnish and one of Polish populations of spruce. However, t_s did not deviate from t_m except for the Finnish spruce population only, suggesting that both single- and multi-locus outcrossing estimates support absolute outcrossing and the absence of inbreeding within populations. In the case of the Finnish population, the difference $t_m - t_s$ suggests a presence of mating between relatives. However, because the correlation of selfing between loci (r_{ta}) was insignificant for this population, while this parameter outcompetes $t_m - t_s$ as a measure of bi-parental inbreeding (Chybicki, unpublished) we further considered neither population showed contemporary inbreeding.

Several other mating system parameters were calculated as well (correlation of t among progeny, multi- and single-locus correlation of pollen alleles, correlation of outcrossing among loci), however all those parameters appeared to be not significant. This indicates, that mating system of studied populations did not depart from random mating (given the assumptions of mixed mating model), or that departures from random mating were minor and generally beyond the power of sampling design (ca. 20 progenies from 15 mothers). However, our findings are in line with existing knowledge on mating system in forest trees. These findings confirm the results of insignificant levels of inbreeding reported earlier.

Given the above findings, it seems reasonable to consider that **mating system parameters should not be considered as important measures in genetic monitoring of populations of main European forest tree species**.

Table 16: *Estimates of multilocus (tm) and single-locus (ts) outcrossing rates calculated for each stand.*

Species / Parameter	**Study plot**			
Fagus sylvatica	Germany (Behlendorf)	Poland (Lesko)	Italy (Pradaccio)	Germany (Solling)
tm	0.993 (0.066)	1.000 (0.076)	-*	0.970 (0.024)
ts	**0.980 (0.010)**	0.983 (0.015)	0.987 (0.012)	**0.955 (0.022)**
Picea abies	Poland (Stańcowa)	Italy (Paneveggio)	Poland (Żytkiejmy)	Finland (Punkaharju)
tm	no data	no data	0.951 (0.074)	-*
ts			**0.861 (0.058)**	**0.918 (0.022)**
Pinus pinaster	Spain (Ain)	Spain (Coca)	Italy (Montefalcone)	France (Lacanau)
tm	0.958 (0.025)	0.985 (0.015)	0.969 (0.041)	no data
ts	0.967 (0.022)	0.962 (0.027)	1.033 (0.040)	
Quercus robur	France (Lot-et-Garonne)	Germany (Grosser Bruch)	Italy (Lame del Sesia)	Poland (Lubartow)
tm	no data	0.995 (0.099)	1.002 (0.096)	0.980 (0.113)
ts		**0.963 (0.016)**	0.975 (0.016)	0.968 (0.029)

* - estimation procedure could not converge to biologically realistic estimates,

3.3.2. Spatial genetics structure

The investigation of spatial genetic structure by spatial autocorrelation analysis has shown a generally higher SGS in oak > beech > spruce (Table 17, Figures 1, 2 and 3), as expected by the species-specific seed and pollen dispersal modes, although large heterogeneity among stands × life stages has been detected. In *Quercus robur*, Sp values ranged between 0.003 to 0.030 (mean 0.015) with F_1 values up to 0.057 in seedlings (Italy) and 0.048 in adults (Poland). Parameters indicating spatial clumping of genotypes are usually higher at the seedling stage, with the only exception of the Poland stand. In *Fagus sylvatica*, Sp values ranged between 0.001 to 0.012 (mean 0.006) with F_1 values up to 0.028 in seedlings (GermanySO) and 0.027 in adults (GermanyBE). Although spatial correlograms are quite similar at the adult stage among stands, they are highly different at the seedling stage, with the regression slope of the kinship estimator over distance varying from -0.001 (Poland) to -0.012 (Italy). In *Picea abies*, a nearly absence of spatial signal in the within-population genetic structure characterized all stands at both life stages, with Sp values always <0.010 and extremely low kinship coefficient in the first classes, in particular at the seedling stage. However, even in both oak and beech stands, F_n values statistically higher than zero were detected only at some stands and for distance classes extending up to 10-20 m.

Table 17: *Parameters describing within-population genetic structure in the studied plots and their standard errors (SE)*

Species	Stand	Stage	F_1	SE	b_F	SE	*Sp*
Fagus	Germany (Behlendorf)	Adults	0.0272	0.0061	-0.00806	0.00192	0.00829
Fagus	Germany (Behlendorf)	Seedlings	0.0099	0.0041	-0.00347	0.00083	0.00350
Fagus	Poland (Lesko)	Adults	0.0092	0.0045	-0.00271	0.00120	0.00274
Fagus	Poland (Lesko)	Seedlings	0	0.0008	-0.00121	0.00091	0.00121
Fagus	Italy (Pradaccio)	Adults	0.0072	0.0028	-0.00606	0.00212	0.00611
Fagus	Italy (Pradaccio)	Seedlings	0.0097	0.0027	-0.01181	0.00227	0.01192
Fagus	Germany (Solling)	Adults	0.0133	0.0044	-0.00459	0.00129	0.00465
Fagus	Germany (Solling)	Seedlings	0.0281	0.0062	-0.00740	0.00159	0.00761
Picea	Poland (Stancowa)	Adults	-0.0024	0.0027	0.00014	0.00143	-0.00014
Picea	Poland (Stancowa)	Seedlings	0.0006	0.0032	-0.00149	0.00078	0.00149
Picea	Italy (Paneveggio)	Adults	0.0105	0.0048	-0.00327	0.00237	0.00331
Picea	Italy (Paneveggio)	Seedlings	-0.0018	0.0009	0.00159	0.00112	-0.00159
Picea	Poland (Zytkiejmy)	Adults	0.0010	0.0047	-0.00097	0.00130	0.00097
Picea	Poland (Zytkiejmy)	Seedlings	0	0.0013	-0.00043	0.00096	0.00043
Picea	Finland (Punkaharju)	Adults	0.0049	0.0035	-0.00864	0.00213	0.00868
Picea	Finland (Punkaharju)	Seedlings	0.0060	0.0023	-0.00451	0.00156	0.00453
Quercus	France (Lot-et-Garonne)	Adults	0.0068	0.0091	-0.01226	0.00644	0.01234
Quercus	France (Lot-et-Garonne)	Seedlings	0.0218	0.0078	-0.02060	0.00518	0.02106
Quercus	Germany (Grosser Bruch)	Adults	0.0303	0.0067	-0.01290	0.00228	0.01330
Quercus	Germany (Grosser Bruch)	Seedlings	0.0544	0.0165	-0.01469	0.00448	0.01554
Quercus	Italy (Lame del Sesia)	Adults	0.0201	0.0091	-0.00254	0.00143	0.00260
Quercus	Italy (Lame del Sesia)	Seedlings	0.0567	0.0060	-0.02818	0.00469	0.02987
Quercus	Poland (Lubartow)	Adults	0.0477	0.0131	-0.00689	0.00208	0.00723
Quercus	Poland (Lubartow)	Seedlings	0.0083	0.0026	-0.01669	0.00563	0.01683

F_1, average kinship coefficient between individuals of the first distance class (0-10 m); b_F, regression slope of the kinship estimator F_{ij} computed among all pairs of individuals against the natural logarithm of geographical distances; *Sp*, intensity of SGS.

***Figure 1**: Correlograms from spatial autocorrelation analysis using the Nason's kinship coefficient (F_{ij}) and even samples sizes (10 distance classes) for the four beech stands (rows) and two life stages (columns). Shaded areas represent the 95% confidence interval obtained through random shuffling of individual geographic locations, black lines around mear F_{ij} values represent 95% confidence intervals around mean F_{ij} values generated by jackknifing over loci.*

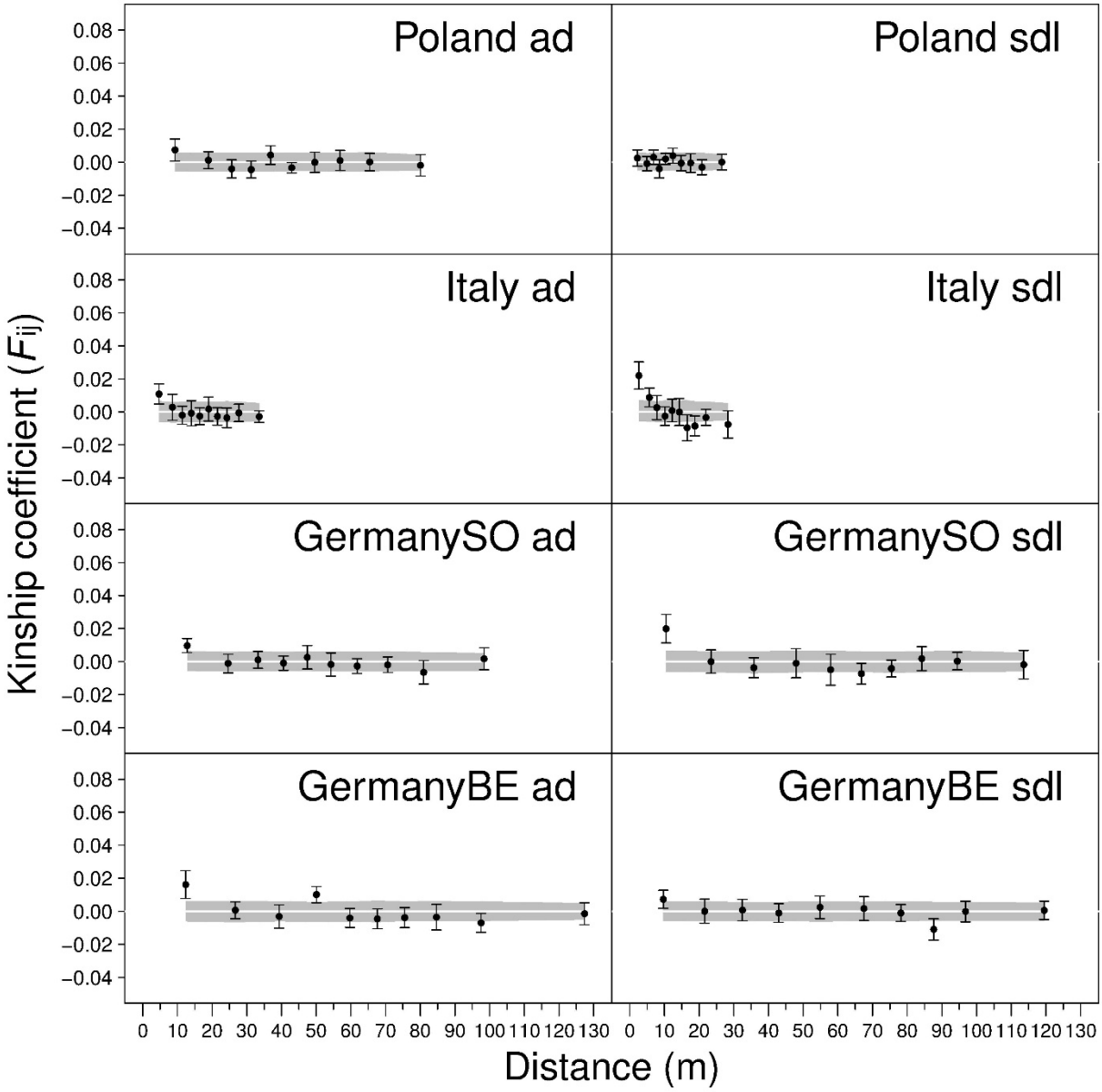


Figure 2: *Correlograms from spatial autocorrelation analysis using the Nason's kinship coefficient (F_{ij}) and even samples sizes (10 distance classes) for the four oak stands (rows) and two life stages (columns). Shaded areas represent the 95% confidence interval obtained through random shuffling of individual geographic locations, black lines around mear F_{ij} values represent 95% confidence intervals around mean F_{ij} values generated by jackknifing over loci.*

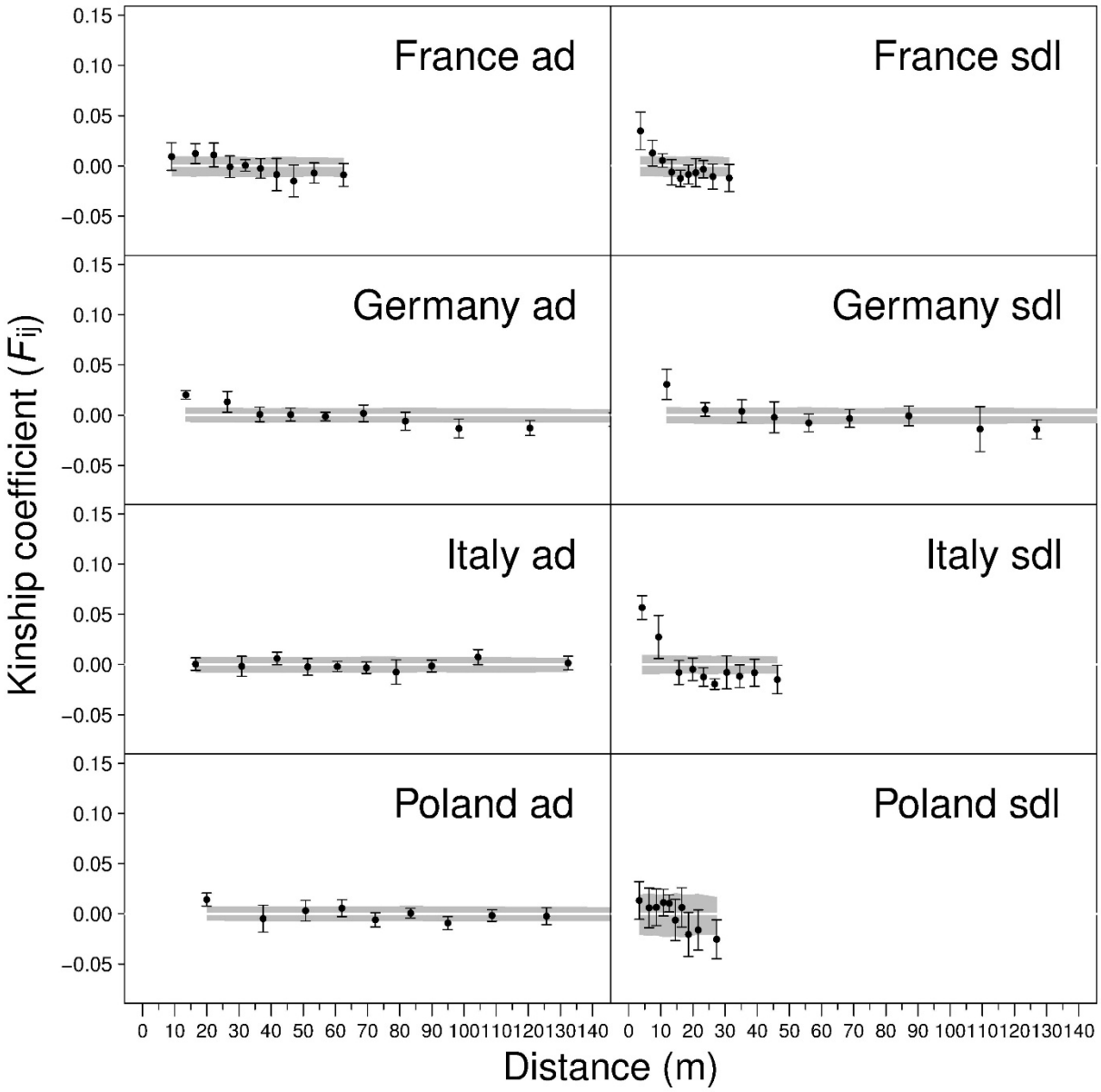
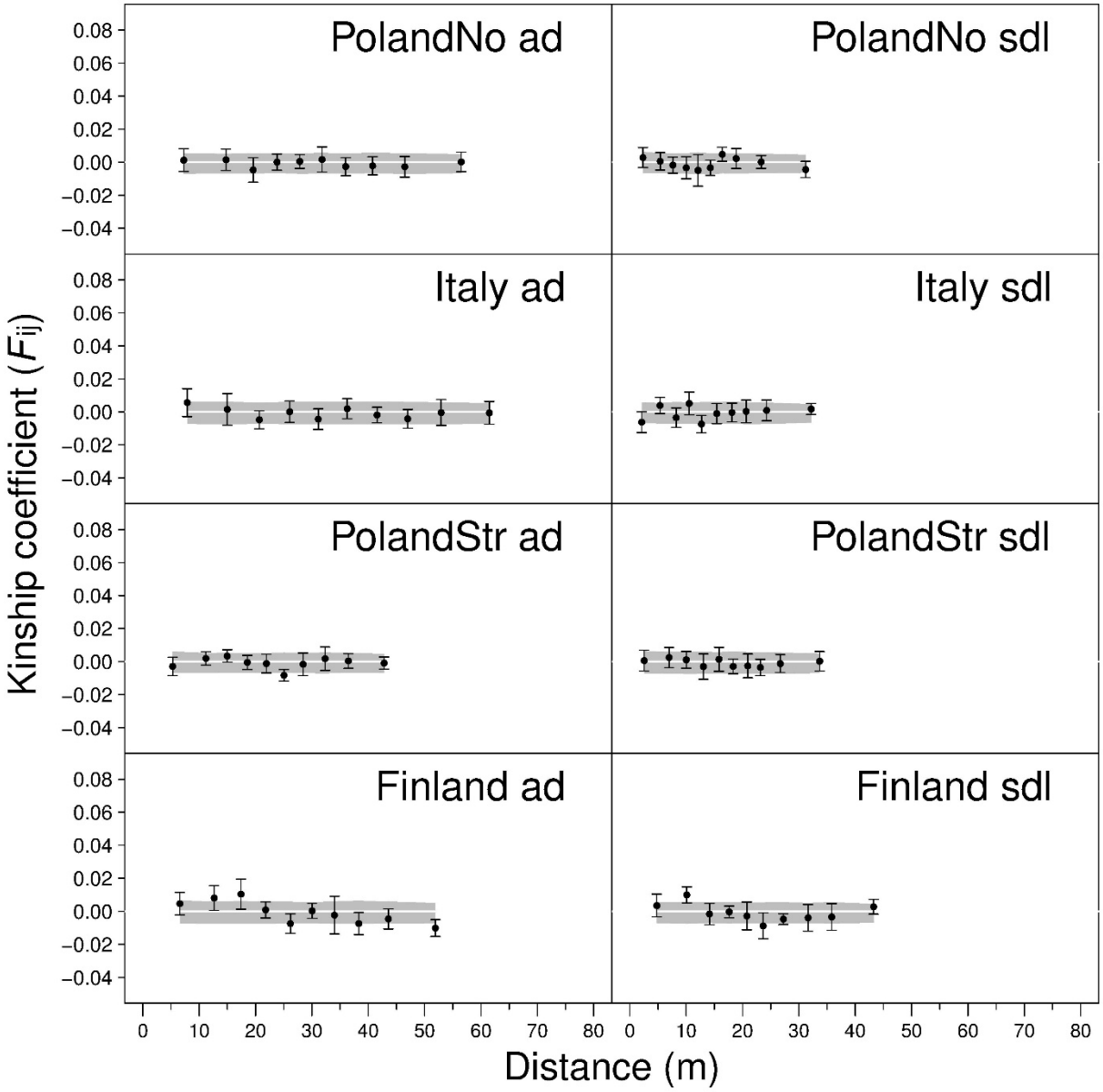


Figure 3: *Correlograms from spatial autocorrelation analysis using the Nason's kinship coefficient (F_{ij}) and even samples sizes (10 distance classes) for the four spruce stands (rows) and two life stages (columns). Shaded areas represent the 95% confidence interval obtained through random shuffling of individual geographic locations, black lines around mear F_{ij} values represent 95% confidence intervals around mean F_{ij} values generated by jackknifing over loci.*



3.3.3. Gene dispersal

Fagus sylvatica

Analyses of mating system and pollen dispersal patterns based on seed samples collected from mother trees indicated that low but significant levels of selfing are likely in beech, particularly at the stage of seeds (Table 18). However, the pollen immigration rates were consistently high (>0.65) in all study plots. This result is not surprising given the small size of the study plots. Pollen dispersal within the stand, however, was not random, and the mating success of males depended significantly on the distance to mother trees. We observed notable variation in this respect, however, which resulted in quite variable mean dispersal distances estimated for the stands. The mating success was also related to individual tree size in all populations except Lesko in Poland.

Comparing seedlings and adult data (genotypes and coordinates) allowed to model pollen and seed dispersal patterns underlying the genotypic structure of seedlings. There were no signs of selfing among the seedlings, suggesting that selfed offspring is probably effectively outcompeted during establishment of seedlings. Seed immigration rates were high and uniform in all stands except Pradaccio (Italy), where lowest immigration was detected. Pollen immigration rates observed at the seedlings stage were high (>0.67) and corresponded pretty well to the estimates obtained based on seed samples. This indicates that there were no selective forces affecting immigrants vs. local mates acting between seed and seedling stages. Patterns of pollen and seed dispersal that led to the established seedling cohorts were not random, with strong effects of seed dispersal distance and diameter on female reproductive success, and slightly weaker effects of pollen dispersal and diameter on male reproductive success. Surprisingly, pollen dispersal distance was smaller than seed dispersal distance in two cases, which deserves further attention and more detailed studies. Note, that these findings should be consider with caution due to small size of study plots and sampling of adult and seedling cohorts in close proximity of individuals. Such sampling could capture very localized patterns of seed and pollen dispersal affected by the local stand structure and local distribution of adults and seedlings.

Table 18: *Parameters of mating system and pollen and seed dispersal patterns obtained based on neighbourhood models for Fagus sylvatica.*

	Parameter	Germany (Behlendorf)	Poland (Lesko)	Italy (Pradaccio)	Germany (Solling)
Pollen dispersal based on seeds	pollen immigration (mp)	**0.702 (0.019)**	**0.844 (0.020)**	**0.646 (0.043)**	**0.794 (0.021)**
	self-fertilization (s)	0.007 (0.004)	**0.022 (0.008)**	-*	**0.044 (0.010)**
	distance effect (male success)	**-0.038 (0.003)**	**-0.047 (0.01)**	**-0.184 (0.032)**	**-0.074 (0.009)**
	diameter effect (male success)	**0.059 (0.006)**	0.037 (0.019)	**0.095 (0.015)**	**0.037 (0.010)**
	pollen dispersal distance [m]	52.1	42.3	10.9	27.1
Pollen and seed dispersal based on seedlings	seed immigration	**0.482 (0.056)**	**0.542 (0.058)**	**0.035 (0.022)**	**0.496 (0.053)**
	pollen immigration	**0.899 (0.043)**	**0.840 (0.062)**	**0.675 (0.050)**	**0.896 (0.044)**
	self-fertilization	-*	-*	-*	-*
	distance effect (female success)	**-0.050 (0.008)**	**-0.170 (0.027)**	**-0.155 (0.019)**	**-0.218 (0.029)**

	distance effect (male success)	**-0.082 (0.038)**	-0.058 (0.030)	**-0.234 (0.041)**	**-0.045 (0.021)**
	diameter effect (female success)	**0.084 (0.015)**	**0.044 (0.018)**	**0.093 (0.009)**	**0.077 (0.015)**
	diameter effect (male success)	0.061 (0.041)	**0.145 (0.067)**	0.028 (0.029)	**0.100 (0.032)**
	seed dispersal distance [m]	40.3	11.8	12.9	9.2
	pollen dispersal distance [m]	24.4	34.6	8.6	44.2

* - estimation procedure could not converge to biologically realistic estimates,

Picea abies

Analyses of mating system and pollen dispersal based on seed samples were possible only for one Polish and the Finnish populations (Table 19). They revealed that low but significant levels of self-fertilization are possible in spruce. Pollen immigration varied between the two studied stands. Pollen dispersal within stands was not random and depended on distance between males and females, but male mating success was also related to the tree size in both stands.
Analyses based on seedlings, indicated no selfing among seedlings (similar result as in beech). Seed immigration was fairly consistent among stands (≥40%), and little variation among stands was observed for pollen immigration (>60%). The pattern of pollen and seed dispersal that led to the formation of seedling cohort was not random and depended on distance to mothers (seed dispersal) and distance between males and females. Tree diameter appeared to be a significant covariate of female reproductive success in all stands, but also male reproductive success in Italian and one Polish (Żytkiejmy) stands.

Table 19: *Parameters of mating system and pollen and seed dispersal patterns obtained based on neighbourhood models for Picea abies.*

	Parameter	**Poland (Stańcowa)**	**Italy (Paneveggio)**	**Poland (Żytkiejmy)**	**Finland (Punkaharju)**
Pollen dispersal based on seeds	pollen immigration (mp)	no data	no data	**0.854 (0.032)**	**0.554 (0.025)**
	self-fertilization (s)			**0.050 (0.019)**	**0.010 (0.005)**
	distance effect (male success)			-0.036 (0.029)	**-0.100 (0.008)**
	diameter effect (male success)			**0.169 (0.040)**	**0.108 (0.013)**
	pollen dispersal distance [m]			55.7	20.0
Pollen and seed dispersal based on seedlings	seed immigration	**0.398 (0.057)**	**0.522 (0.058)**	**0.516 (0.058)**	**0.589 (0.061)**
	pollen immigration	**0.787 (0.061)**	**0.615 (0.087)**	**0.781 (0.068)**	**0.878 (0.064)**
	self-fertilization	-*	-*	-*	0.000 (0.025)
	distance effect (female success)	**-0.047 (0.013)**	**-0.133 (0.024)**	**-0.125 (0.023)**	**-0.068 (0.017)**
	distance effect (male success)	**-0.061 (0.029)**	**-0.066 (0.026)**	**-0.092 (0.041)**	-0.598 (0.295)
	diameter effect (female success)	**0.072 (0.020)**	**0.041 (0.008)**	**0.155 (0.022)**	**0.130 (0.032)**
	diameter effect (male	0.057 (0.054)	**0.059**	**0.108**	-*

	success)		**(0.016)**	**(0.046)**	
	seed dispersal distance [m]	42.3	15.1	16.0	29.2
	pollen dispersal distance [m]	32.7	30.2	21.8	-*

* - estimation procedure could not converge to biologically realistic estimates,

Pinus pinaster

Estimating of mating system and pollen dispersal (not available for the French stand) revealed that considerable but fairly uniform immigration levels are present in maritime pine. Low but significant level of selfing was observed the two Spanish stands (Table 20). Negative relationship between mating success and distance between males and females was generally noted, but it was significant only in Coca stand (Spain).

Despite availability of seedling and adult data for all stands, estimation was possible only for the two Spanish stands. The mating models with data from Italy and France could not converge to biologically realistic value causing the failure in convergence procedures. One reason could be the relatively low number of loci studied in *Pinus pinaster*. The analyses will likely be continued when number of loci will be increased or new data based on SNP markers will become available. Probably the low information content of the data (low number of loci) caused some estimation problems in Spanish stands, where most of parameter estimates were loaded with relatively high estimates of SE. Given above mentioned problem a detailed and sensible discussion of pollen and seed dispersal patterns in this species seems to be not well justified. As mentioned earlier, SNP data could be an alternative marker type for this particular data set.

Table 20: *Parameters of mating system and pollen and seed dispersal patterns obtained based on neighbourhood models for Pinus pinaster.*

	Parameter	Spain (Ain)	Spain (Coca)	Italy (Montefalcone)	France (Lacanau)
Pollen dispersal based on seeds	pollen immigration (mp)	**0.475 (0.080)**	**0.686 (0.073)**	**0.675 (0.228)**	no data
	self-fertilization (s)	**0.060 (0.022)**	**0.041 (0.017)**	0.055 (0.030)	
	distance effect (male success)	-0.018 (0.013)	**-0.102 (0.024)**	-0.307 (0.272)	
	diameter effect (male success)	**0.299 (0.038)**	0.034 (0.019)	0.166 (0.209)	
	pollen dispersal distance [m]	*113.1*	19.5	*6.5*	
Pollen and seed dispersal based on seedlings	seed immigration	0.361 (0.349)	-*	lack of convergence	lack of convergence
	pollen immigration	0.296 (0.686)	**0.643 (0.225)**		
	self-fertilization	0.000 (0.000)	0.086 (0.049)		
	distance effect (female success)	**-0.102 (0.039)**	**-0.133 (0.020)**		
	distance effect (male success)	-0.501 (0.437)	-0.026 (0.054)		
	diameter effect (female success)	0.114 (0.067)	**-0.092 (0.030)**		
	diameter effect (male success)	0.141 (0.126)	0.053 (0.070)		
	seed dispersal distance [m]	19.6	15.0		
	pollen dispersal distance [m]	4.0	77.6		

* - estimation procedure could not converge to biologically realistic estimates,

Quercus robur

Almost no selfing was detected based seed samples and neighborhood model. Only population Lubartów exhibited low but significant level of selfing (3%) (Table 21). Pollen immigration levels, effect of distance, effect of diameter on mating success, and resulting pollen dispersal distance were quite similar among different oak populations.

Analyses of pollen and seed dispersal patterns based on seedlings and respective mating models, revealed that the French population differed considerably from other populations. It shown highest levels of seed and pollen immigration, but also very high estimates of distance and diameter effects on male and female reproductive success. This in turn resulted in very low (and rather unlikely/biased) estimates of seed and pollen dispersal. The precision of these estimates was compromised probably due to high pollen and seed immigration rates, therefore the patterns of seed and pollen dispersal in French population should be considered with caution. The estimates obtained for the three other populations were similar among each other. Mean pollen dispersal distances were similar when estimated based on seeds and based on seedlings. Low distances of seed dispersal are consistent with results obtained in other studies, and are reasonable given seed dispersal based on gravity.

Table 21: *Parameters of mating system and pollen and seed dispersal patterns obtained based on neighbourhood models for Quercus robur.*

	Parameter	France (Lot-et-Garonne)	Germany (Grosser Bruch)	Italy (Lame del Sesia)	Poland (Lubartow)
Pollen dispersal based on seeds	pollen immigration (mp)	no data	**0.516 (0.037)**	**0.542 (0.038)**	**0.527 (0.038)**
	self-fertilization (s)		0.008 (0.006)	0.007 (0.007)	**0.032 (0.012)**
	distance effect (male success)		**-0.063 (0.008)**	**-0.038 (0.006)**	**-0.045 (0.005)**
	diameter effect (male success)		**0.046 (0.012)**	**0.047 (0.010)**	**0.030 (0.008)**
	pollen dispersal distance [m]		31.5	52.1	44.3
Pollen and seed dispersal based on seedlings	seed immigration	**0.336 (0.058)**	**0.228 (0.051)**	0.014 (0.014)	**0.179 (0.081)**
	pollen immigration	**0.769 (0.068)**	**0.510 (0.071)**	**0.589 (0.055)**	**0.640 (0.117)**
	self-fertilization	-*	-*	-*	0.062 (0.050)
	distance effect (female success)	**-0.657 (0.095)**	**-0.187 (0.022)**	**-0.198 (0.017)**	**-0.135 (0.026)**
	distance effect (male success)	**-0.395 (0.120)**	**-0.034 (0.008)**	**-0.044 (0.010)**	**-0.033 (0.016)**
	diameter effect (female success)	**0.160 (0.039)**	**0.025 (0.017)**	**0.052 (0.014)**	**0.045 (0.012)**
	diameter effect (male success)	0.087 (0.070)	0.019 (0.027)	**0.047 (0.016)**	0.016 (0.030)
	seed dispersal distance [m]	3.0	10.7	10.1	14.8
	pollen dispersal distance [m]	5.1	58.0	45.4	61.5

* - estimation procedure could not converge to biologically realistic estimates,

3.3.4. Effective population size

Table 22 presents the estimates of the effective population size per species, stage and marker for all the site. Overall, a sharp decline is found in effective population size from the adult, seedling and seed stages for all species. These results suggest that, to attain a sample with the same effective population size of the adults, larger samples of the both the seedling and seed stages are necessary than the sample size of the adult stage.

Table 22: *Estimates of effective population size*

Species	Marker	Site	stage	Estimated Ne	Lower 95% CI - Parametric	Upper 95% CI - Parametric	Lower 95% CI - JackKnife	Upper 95% CI JackKnife
Fagus	SNP	Germany (Behlendorf)	adults	31.2	28.5	34.2	25.1	39.2
Fagus	SNP	Germany (Behlendorf)	seedlings	32.8	29.7	36.3	25.4	43
Fagus	SNP	Poland (Lesko)	adults					
Fagus	SNP	Poland (Lesko)	seedlings	31.2	28	34.8	24	41.2
Fagus	SNP	Italy (Pradaccio)	adults	30	27.1	33.3	23.1	39.5
Fagus	SNP	Italy (Pradaccio)	seedlings	27.3	24.7	30.1	21.3	35.3
Fagus	SNP	Germany (Solling)	adults	21.8	19.8	23.9	17.8	26.7
Fagus	SNP	Germany (Solling)	seedlings	40.1	35.8	45.2	30.4	54.5
Fagus	SSR	Germany (Behlendorf)	adults	145	107.6	212.2	98.4	248.8
Fagus	SSR	Germany (Behlendorf)	seedlings	164.6	118.4	254.8	107.1	311.6
Fagus	SSR	Germany (Behlendorf)	seeds	50.5	46.6	54.5	46.1	55.2
Fagus	SSR	Poland (Lesko)	adults	345.5	107.6	1130.2	190	1315.4
Fagus	SSR	Poland (Lesko)	seedlings	132.8	118.4	184.2	93.4	212.1
Fagus	SSR	Poland (Lesko)	seeds	60.8	46.6	66.4	53.6	68.9
Fagus	SSR	Italy (Pradaccio)	adults	168.1	116.5	279.1	108	327.6
Fagus	SSR	Italy (Pradaccio)	seedlings	41.7	36.6	47.7	35.4	49.5
Fagus	SSR	Italy (Pradaccio)	seeds	22.7	20.4	25.2	20	25.7
Fagus	SSR	Germany (Solling)	adults	262	163.8	573.5	148.8	799.3
Fagus	SSR	Germany (Solling)	seedlings	113.4	86.5	157.5	79.1	182
Fagus	SSR	Germany (Solling)	seeds	61.4	56.2	67	54.8	68.8
Picea	SSR	Finland (Punkaharju)	adults	146.6	105.1	227.1	102.1	239.4
Picea	SSR	Finland (Punkaharju)	seedlings	297.6	165.1	1043.4	151	1916.6
Picea	SSR	Finland (Punkaharju)	seeds	54.2	49.5	59.2	46.4	63.1
Picea	SSR	Italy (Paneveggio)	adults	562	220.1		187.1	
Picea	SSR	Italy (Paneveggio)	seedlings	152.9	103.2	266.2	98.1	296.4

Species	Marker	Site	stage	Estimated Ne	Lower 95% CI - Parametric	Upper 95% CI - Parametric	Lower 95% CI - JackKnife	Upper 95% CI JackKnife
Picea	SSR	Poland (Stancowa)	adults	969.2	272.3		241.3	
Picea	SSR	Poland (Stancowa)	seedlings	158.5	106.4	279.4	103.6	295.6
Picea	SSR	Poland (Zytkiejmy)	adults		1221.4		740.4	
Picea	SSR	Poland (Zytkiejmy)	seedlings	112.2	82.8	164.1	80.8	170.4
Picea	SSR	Poland (Zytkiejmy)	seeds	53.8	47.9	60.4	45.4	63.7
Pinus	SSR	France (Lacanau)	adults	955.2	60.5		90.7	
Pinus	SSR	France (Lacanau)	seedlings	180.1	38.6		27.8	
Pinus	SSR	Italy (	adults	158	24.3		3.1	
Pinus	SSR	Italy	seedlings	135.4	26.3		6.6	
Pinus	SSR	Italy	seeds	29	14.8	53.7	6.5	94.1
Pinus	SSR	SpainCO	adults	58.9	29.7	162.8	18.7	3724
Pinus	SSR	SpainCO	seedlings	72.1	31.7	332	25.3	2343
Pinus	SSR	SpainCO	seeds	56.5	38.5	82.2	34.3	91.6
Pinus	SSR	SpainVA	adults	11.7	6.1	19.6	5	21.3
Pinus	SSR	SpainVA	seedlings		92.6		111.1	
Pinus	SSR	SpainVA	seeds	43.6	26.9	68.5	21.2	83.9
Quercus	SSR	France	adults	60.7	40	102.1	37.2	115.2
Quercus	SSR	France	seedlings	63.1	44	97.8	41.2	108.6
Quercus	SSR	Germany	adults	430.7	156		143.1	
Quercus	SSR	Germany	seedlings	313.6	128.2		126	
Quercus	SSR	Germany	seeds	66.2	56.1	78.4	53.6	82.2
Quercus	SSR	Italy	adults	935.6	197.4		186.4	
Quercus	SSR	Italy	seedlings	21.8	17.6	27	16.9	28.2
Quercus	SSR	Italy	seeds	37.2	30.5	45.3	28.4	48.6
Quercus	SSR	Poland	adults	67.2	45.7	108.3	46.6	105.1
Quercus	SSR	Poland	seedlings	23.3	16.4	34.7	15.1	38.8
Quercus	SSR	Poland	seeds	71.5	58	88.4	56.9	90.1

3.4. Comparison of microsatellites (SSR) and single-nucleotide polymorphism (SNP) markers to be used in genetic monitoring.

3.4.1. Genetic diversity

Estimates on genetic diversity differed mong markers. We even noticed negative correlations among estimates, for example in the effective number of alleles in *Fagus sylvatica* (Figure 4). However, if only ranks in adults are considered, a positive trend is observed (Figure 5). For *Pinus pinaster*, the correlation based on Ae values was positive.

Figure 4: *Relation between the effective number of alleles measured with microsatellite and SNP loci*

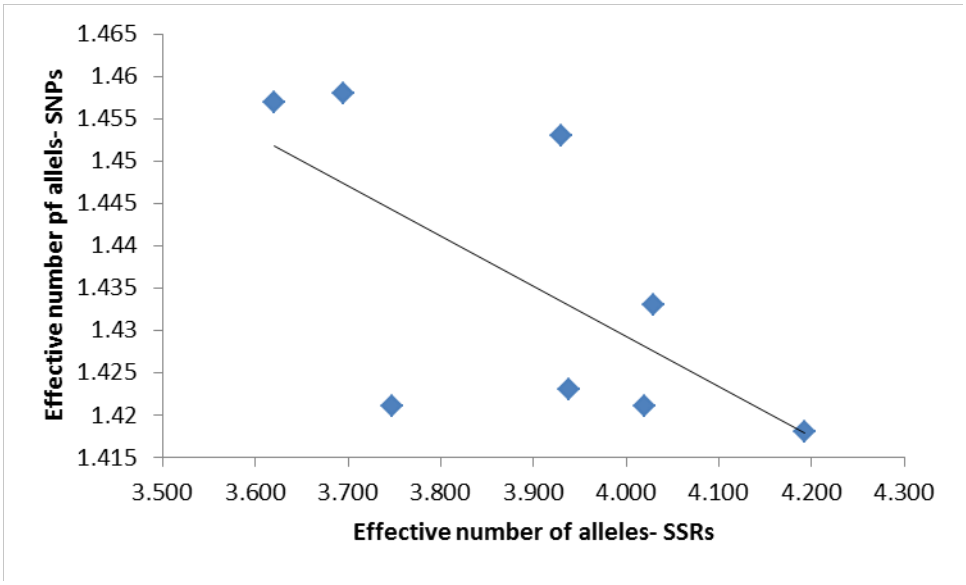
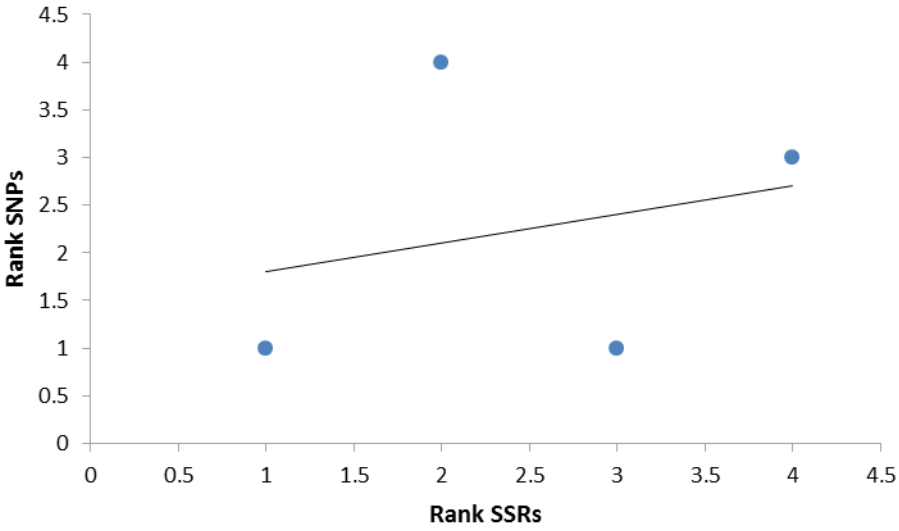


Figure 5: *Relation between the rank of the effective number of alleles measured with microsatellite and SNP loci*



3.4.2. Spatial genetic structure

SGS estimated on both SSRs and SNPs on 96 adult and juvenile individuals in each of the beech plots was assessed by producing spatial autocorrelograms based on Nason's kinship coefficient (F_{ij}) through the SpaGeDi software, using both the even distance classes (5 m-wide distance classes) and the even sample size (10 distance classes) criterions (spatial autocorrelograms produced using the 'even distance classes' criterion are reported in Figure 6, a graphic comparison of parameters describing the SGS is shown in Figure 7). The general picture emerging for the inspection of autocorrelograms and SGS parameters is a substantial concordance of the SGS with the two types

of markers. Kinship coefficients are highly correlated (Pearson's product-moment correlation: r=0.64, 95%CI=0.53-0.74; t = 9.31, df = 123, P<0.001; Figure 8).
Notwithstanding a general agreement between marker types, it is possible to highlight small scale and/or case specific differences which might reveal an effect of local adaptive patterns in shaping SGS. In all plot, at both life stages, the kinship coefficient of the 0-10 m distance class is higher (sometimes statistically higher, e.g. for adults in SOL) when calculated for SNPs, and the opposite trend is clearly visible for b-log, resulting in a slightly but generally higher intensity of SGS (as measured by the Sp parameter) at SNP loci (Figure 7). It should however be noted that linkage among SNP loci might influence this result and that the mean ΔSp is only -0.0024. Since local adaptation might involve only a few SNP markers, the possible influence of outlier loci on this general trend should be investigated.

Table 23: *Comparison between marker types (SSRs vs. SNPs) of parameters describing within-population genetic structure in the studied plots carried out on the beech dataset.*

Stand	Stage	Markers	F_1	SE	b_F	SE	*Sp*
Germany (Behlendorf)	Adults	SSRs	0.0272	0.0061	-0.00806	0.00192	0.00829
		SNPs	0.0316	0.0053	-0.00766	0.00145	0.00792
Germany (Behlendorf)	Seedlings	SSRs	0.0099	0.0041	-0.00347	0.00083	0.00350
		SNPs	0.0167	0.0050	-0.00427	0.00113	0.00435
Poland (Lesko)	Adults	SSRs	0.0092	0.0045	-0.00271	0.00120	0.00274
		SNPs	0.0103	0.0040	-0.00358	0.00137	0.00362
Poland (Lesko)	Seedlings	SSRs	0	0.0008	-0.00121	0.00091	0.00121
		SNPs	0.0016	0.0008	-0.00141	0.00089	0.00141
Italy (Pradaccio)	Adults	SSRs	0.0072	0.0028	-0.00606	0.00212	0.00611
		SNPs	0.0092	0.0018	-0.01080	0.00149	0.01091
Italy (Pradaccio)	Seedlings	SSRs	0.0097	0.0027	-0.01181	0.00227	0.01192
		SNPs	0.0139	0.0029	-0.01808	0.00277	0.01834
Germany (Solling)	Adults	SSRs	0.0133	0.0044	-0.00459	0.00129	0.00465
		SNPs	0.0324	0.0051	-0.00660	0.00126	0.00682
Germany (Solling)	Seedlings	SSRs	0.0281	0.0062	-0.00740	0.00159	0.00761
		SNPs	0.0348	0.0046	-0.01176	0.00193	0.01219

F_1, average kinship coefficient between individuals of the first distance class (0-10 m); b_F, regression slope of the kinship estimator F_{ij} computed among all pairs of individuals against the natural logarithm of geographical distances; *Sp*, intensity of SGS.

Figure 6: *Comparison between marker types (SSRs vs. SNPs, black and white dots respectively) of correlograms from spatial autocorrelation analysis using the Nason's kinship coefficient (F_{ij}) and even distance classes (5 m wide distance classes) for the four beech stands (rows) and two life stages (columns). Black lines around mear F_{ij} values represent 95% confidence intervals around mean F_{ij} values generated by jackknifing over loci. Shaded areas indicating the 95% confidence interval obtained through random shuffling of individual geographic locations were not represented.*

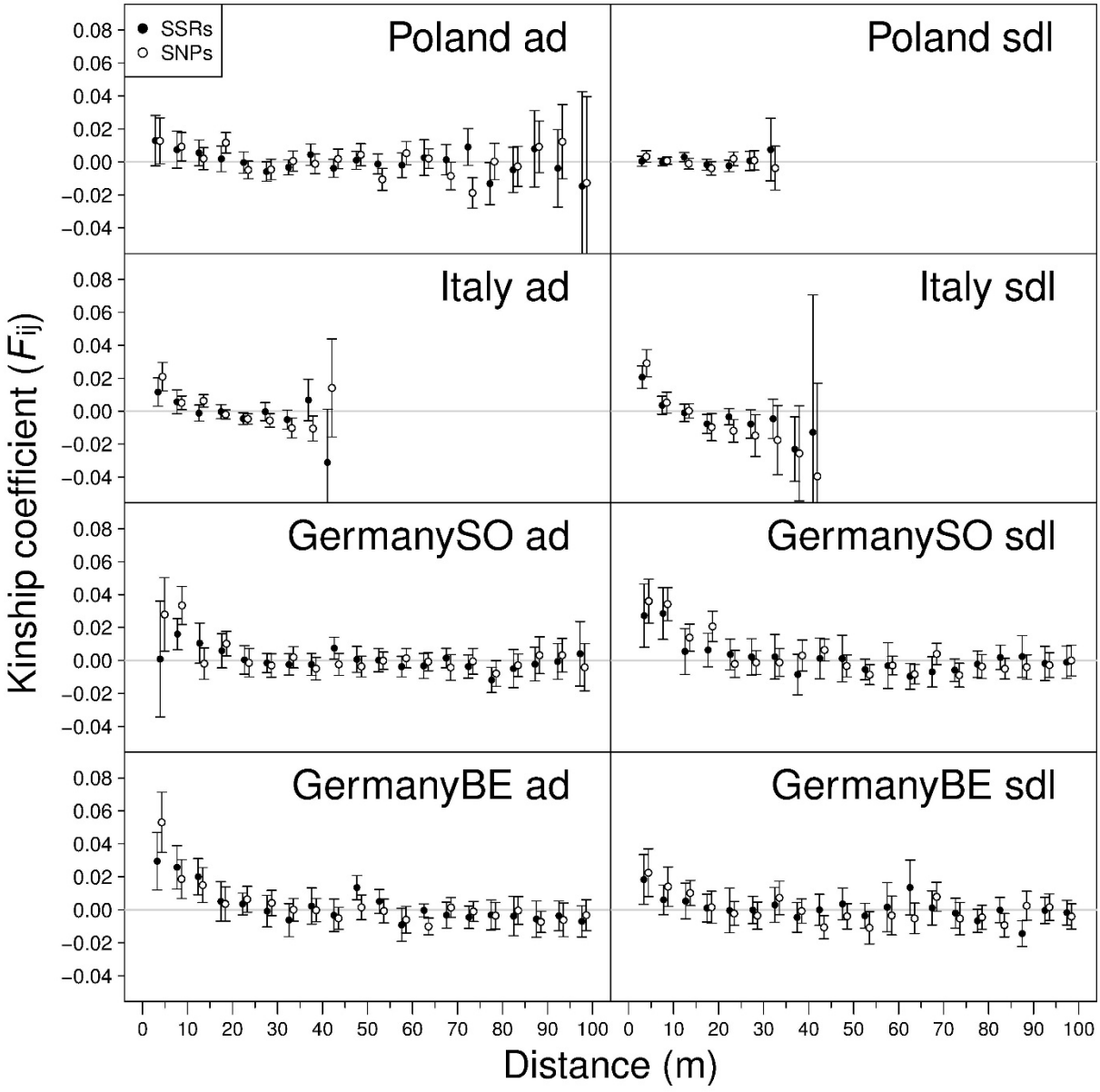


Figure 7: *CNR5 Comparison between marker types (SSRs vs. SNPs) of parameters describing the SGS for the four beech stands and two life stages.*

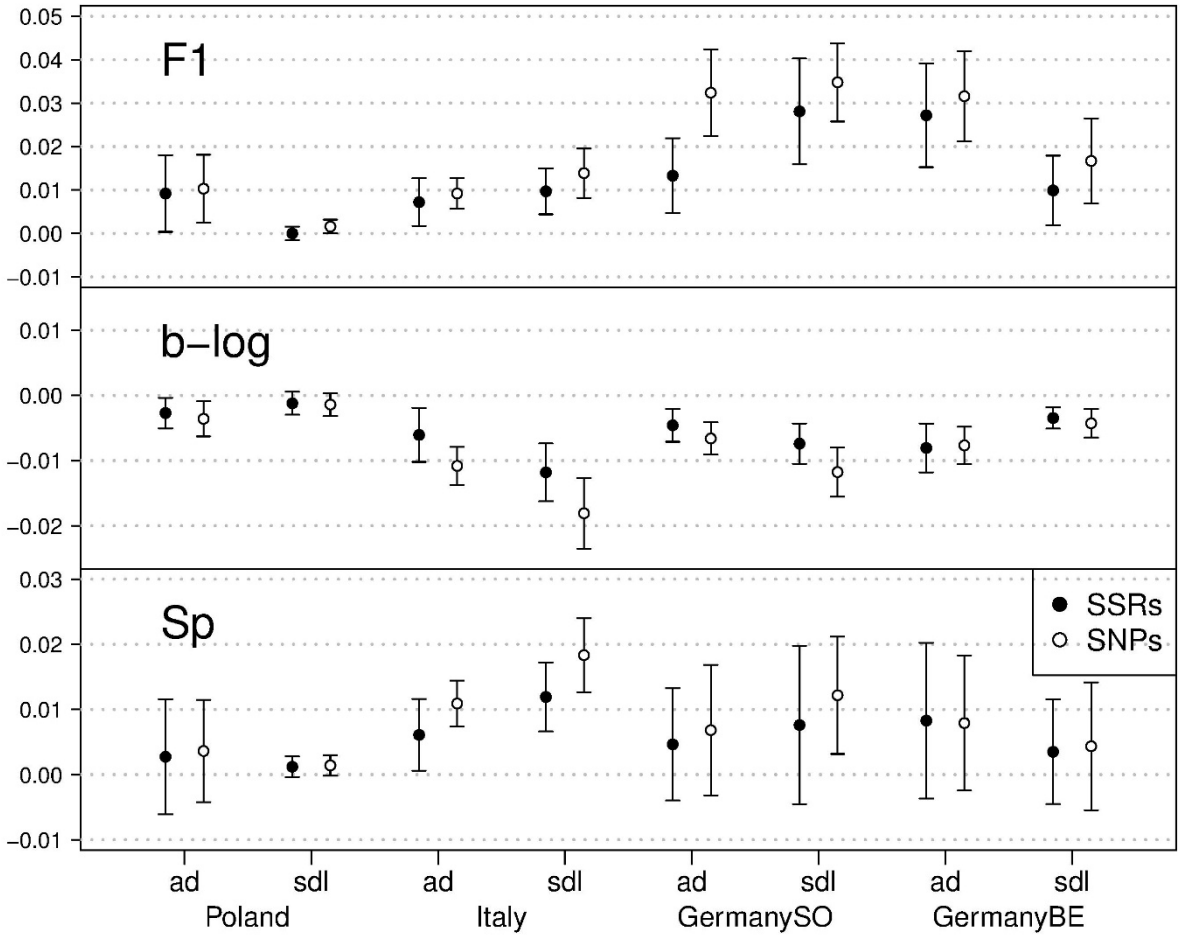
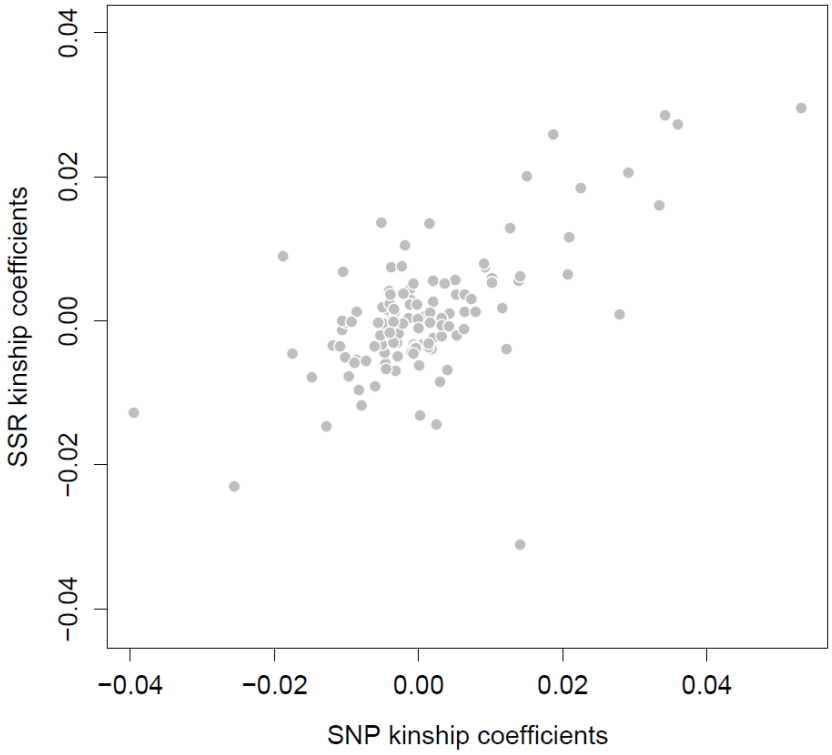


Figure 8: *CNR6 Correlation between kinship coefficients calculated on the SSR dataset (Y axis) and the SNP dataset (X axis)*



3.4.3. Gene dispersal

Comparative analysis of gene dispersal based on SSR and SNP markers

AIMS: SNP and SSR loci were compared as potential genetic markers in parentage-based modelling of gene dispersal. For this purpose, Fagus/Pradaccio (Italy) data set was chosen as an example, because based on preliminary analyses based on SSR markers it was shown to reveal minimum seed immigration and moderate pollen immigration rates as well as strong (negative) effect of distance on both mating success and maternity in respect to a given mother and seedling, respectively.

METHODS: SSR and SNP data comprised 16 and 118 polymorphic markers, respectively. Genotypic data and locations of adults and seedlings were used to infer seed and pollen dispersal patterns (seedling neighborhood model; Burczyk et al. 2006, Chybicki and Burczyk 2010a,b). In order to compare usefulness of these markers, a basic version of the neighbourhood model was used, with the following parameters enabled in the estimation: **ms** (seed immigration rate), **mp** (pollen immigration rate), **ds** (average dispersal distance of seeds) and **dp** (average dispersal distance of pollen). Because both marker types are prone to genotyping errors, two analyses per marker type were run. Firstly, data were assumed to be absolutely correct, so that typing error rates were set to zero and not accounted for in estimation procedures. Secondly, data were

assumed to contain genotyping mistakes (or mutations), and typing error rates were included in the model and set estimable along with ms, mp, ds and dp (i.e., for each locus typing errors were estimated).

RESULTS: Generally, results showed that when typing errors were not taken into account then the estimates of immigration rates were rather high (Table 24). For SNP markers, seed immigration reached 20% while pollen immigration 80%. For comparison, in the case of SSR markers, seed and pollen immigration was 10% and 82%, respectively. SNP markers showed significantly higher seed immigration compared to SSR. However, when typing errors were taken into account, immigration rates decreased substantially. Seed immigration was estimated as to be not significantly different from zero for both marker types. On the contrary, pollen immigration was still significant and appeared significantly different for the two marker types (reasons for the differences remain unclear but could result from different mutation models for SNP and SSR markers). Interestingly, when both marker types were used simultaneously (each genotype was represented by both SSR and SNP markers), estimates of immigration rates (ms and mp) were apparently closer to those for SSRs. The above mentioned results indicated that seed and pollen immigration rates are quite sensitive to genotyping errors, regardless of the type of markers. This is reasonable, because genotyping errors or mutations likely generate new multilocus genotypes which may not be compatible with any local parents.

The comparative analyses showed also that seed dispersal distance is quite similarly estimated for SNP and SSR, regardless of whether or not genotyping errors are accounted for. On the other hand, dp depended to some degree on the choice of the error treatment, but somewhat more in the case of SNP markers. This is because dispersal distance is calculated based on parentage assignment of seedlings. Typing errors (or mutations) usually generate a new genotype likely not compatible with any local parents, thus increasing the rate of pollen or seed immigration. It is rather unlikely, that typing error or mutation will generate a new genotype compatible with other local parent and thus contribute to the change of dispersal kernel and resulting dispersal distance. It is worth noting, that the SE estimates were quite comparable between SNP and SSR, and only slightly affected by the estimation model with lower SE estimates when genotyping errors were accounted for. SNP and SSR data revealed similar SE, probably because the main factor affecting SE estimates in this case is rather the sample size of seedlings and not the exclusion power of the marker set.

Estimated mistyping error rates varied much between loci. In the case of SSR, 7 out of 16 loci (44%) showed error rate significantly greater than zero. For SNP, only 9 out of 118 (8%) appeared to be significantly affected by scoring problems. However, if one note that error treatment had severe impact on the estimates of immigration rates in the case of SNP, it appears that low-intensity scoring problems can accumulate rapidly when a large number of loci is used. Therefore, it is too early to conclude that SNPs outperformed SSRs in this respect.

CONCLUSIONS: Either SNP or SSR markers may be used as genetic markers for pollen and seed dispersal studies. SNP markers are more computationally demanding due to relatively large numbers of loci as compared to SSR markers. This difference becomes more evident if genotyping errors are included in the models as estimable parameters. Both datasets provided fairly similar results of seed and pollen dispersal parameters. It has already been known that SSR markers are sensitive to genotyping errors, null alleles or mutations which complicate parentage-based analyses. However, our analyses revealed that using SNP markers may not solve this problem as some SNP loci have shown significant error rates. Not accounting for genotyping errors overestimates pollen and seed immigration rates, regardless of the marker system used. However, pollen and seed dispersal kernels appeared to be less sensitive to genotyping errors. When taking decisions on the use of SNP vs. SSR markers in monitoring pollen and seed dispersal patterns, other factors like costs of analyses might become a critical selection criterion.

Comment:

We believe, it would be very interesting to perform a broader analyses including other data sets (with beech and other species). Such comparative analyses could formulate hypotheses whether the same SNP or SSR loci show scoring problems across different populations or if this problems are site-specific. This knowledge has some fundamental importance for discussing probability of genotyping errors (or mutations) in SNP and SSR. This could provide some specific information about scoring quality for different loci (to allow decisions to retain or reject specific loci from analyses), and perhaps contribute to the ongoing discussion on the utility of SNP vs. SSR markers. This all probably could be a nice story worth of a methodological paper, which could be the side effect of our works.

Table 24: *Estimates of parameters of the neighbourhood model:* ***ms*** *(seed immigration rate),* ***mp*** *(pollen immigration rate),* ***ds*** *(average dispersal distance of seeds),* ***dp*** *(average dispersal distance of pollen). Estimate column indicates whether parameter value (Param.) or standard errer (SE) is given.*

Markers	Errors	Estimate	ms	mp	Log(ds)	Log(dp)	ds	dp
SNP	no	Param.	0.196	0.801	2.726	2.244	15.3	9.4
		SE	0.043	0.046	0.153	0.243		
	yes	Param.	0.011	0.442	2.643	2.601	14.1	13.5
		SE	0.011	0.055	0.134	0.173		
SSR	no	Param.	0.099	0.815	2.619	1.861	13.7	6.4
		SE	0.032	0.042	0.128	0.223		
	yes	Param.	0.034	0.677	2.590	2.155	13.3	8.6
		SE	0.022	0.049	0.124	0.172		
SNP+SSR	no	Param.	0.324	0.907	2.593	1.521	13.4	4.6
		SE	0.048	0.036	0.144	0.346		
	yes	Param.	0.042	0.721	2.614	1.93	13.7	6.9
		SE	0.020	0.047	0.124	0.176		

Figure 9: *Estimates of seed (ms) and pollen (mp) immigration rates and their SE calculated based on seedling neighborhood model when typing errors are ignored (off) or accounted for (on) in the estimation procedure.*

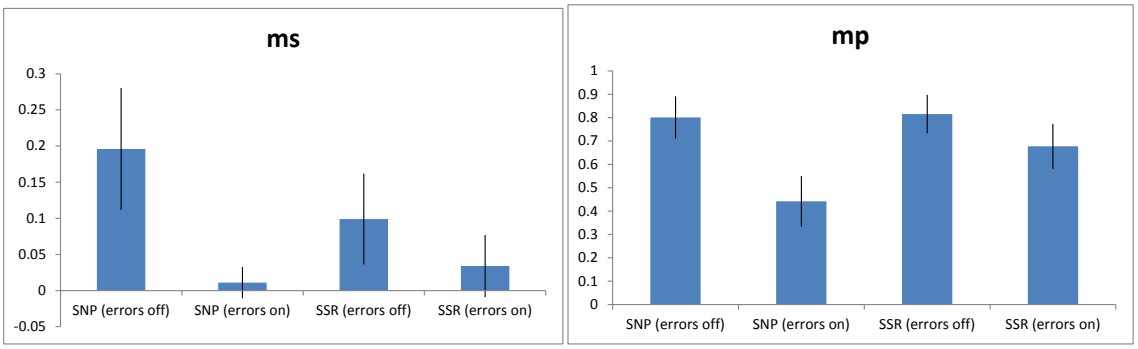


Figure 10: *Estimates of seed and pollen dispersal distances and their SE calculated based on seedling neighborhood model when typing errors are ignored (off) or accounted for (on) in the estimation procedure.*

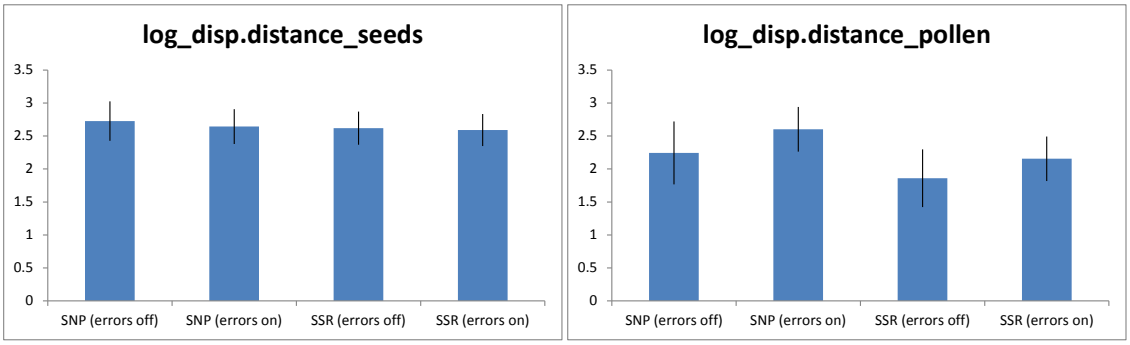
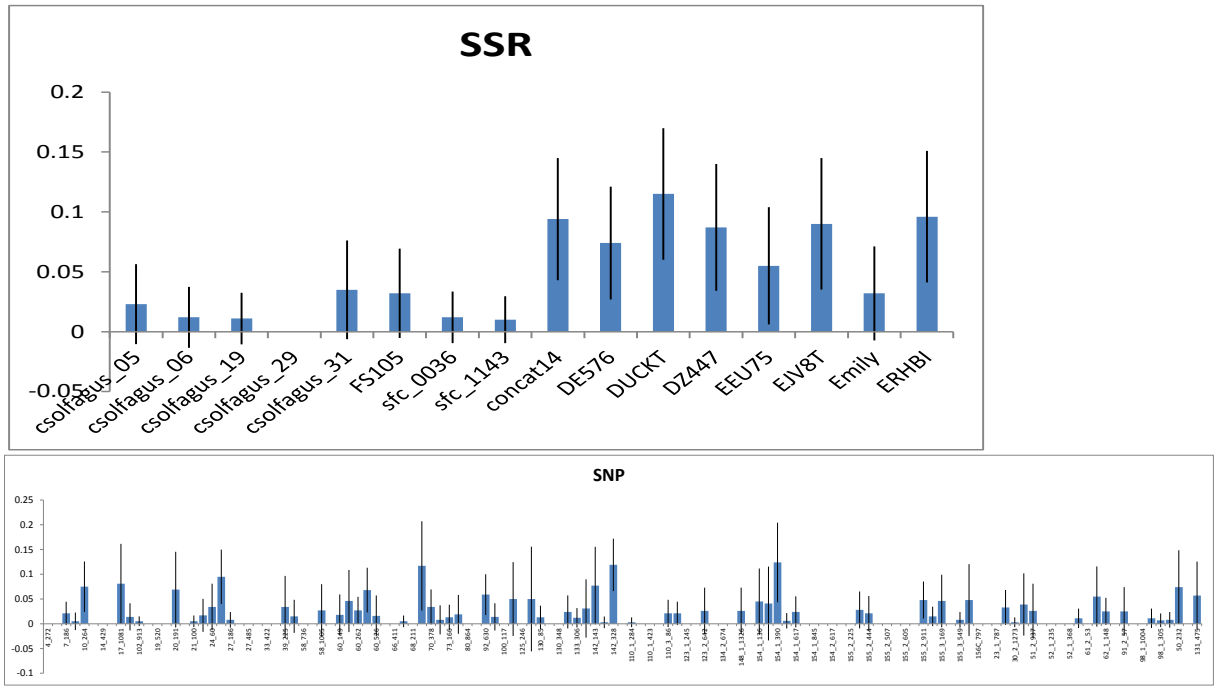


Figure 11: *The estimates of typing error rates for each locus for SSR and SNP markers:*



3.4.4. Effective population size

In case of Fagus sylvatica, the effective population size (Ne) estimated on based on the SSR was much larger than the estimated based on the SNP markers (Figure 12). Also the confidence intervals (CI), both based on the parametric method and the JackKnifing approach (See NeEstimator documentation) was much larger when based on SSR markers than based on SNP markers (Figure 13 and 14, respectively). These findings are true both for the adult and the seedling stages. In case of the SSR results, the confidence interval increased with the value of Ne. This is an undesirable feature, as it indicates that (much) larger samples are needed for populations with large Ne to attain a same confidence interval compared to populations with a relatively small Ne.

Figure 12: *Estimated of effective population size (Ne) for Fagus sylvatica based on either SSR markers or SNP markers. NB: for Lesko site estimated Ne = ∞ for adults.*

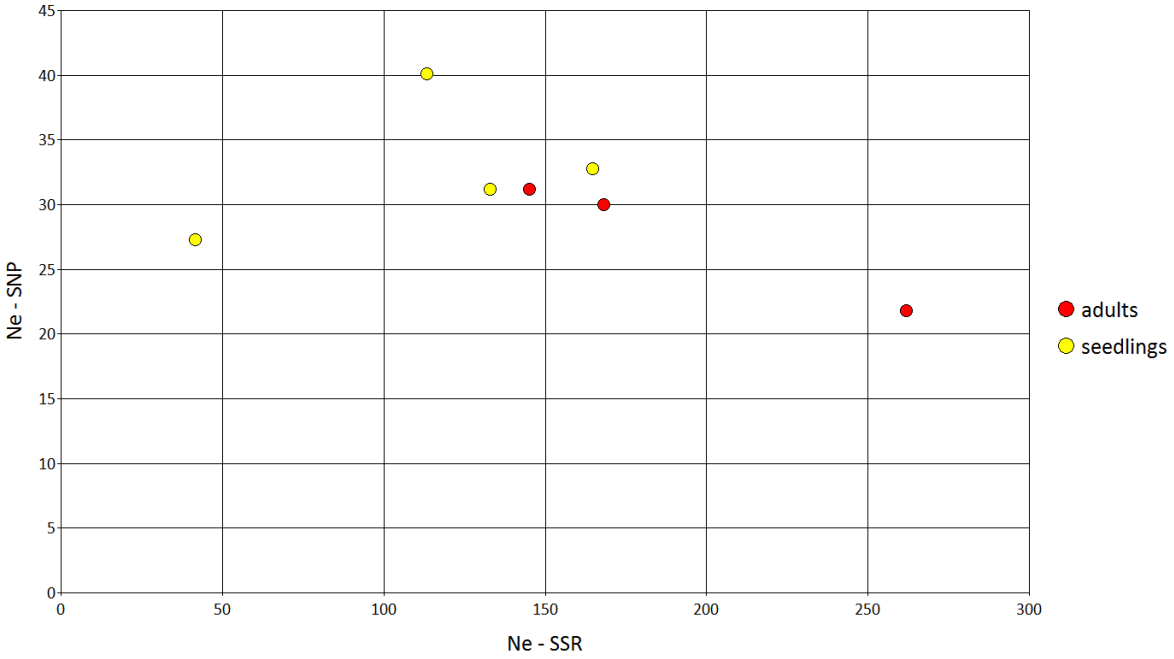


Figure 13: *Width of the confidence interval (upper – lower) of effective population size for Fagus sylvatica based on either SSR markers or SNP markers. Parametric approach used to estimate confidence interval.*

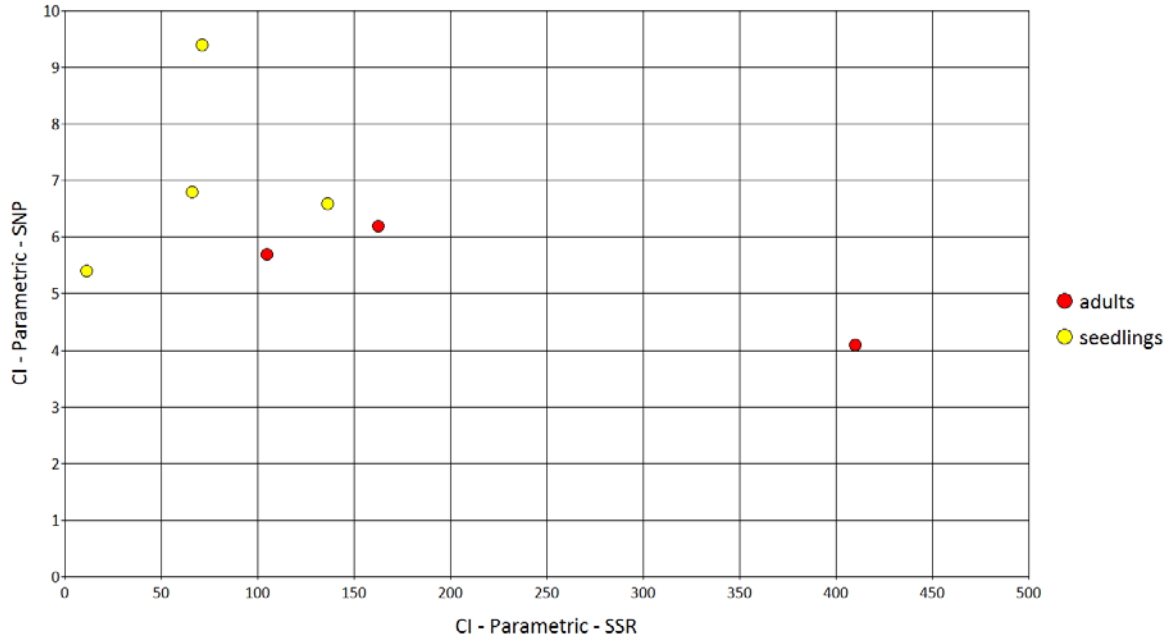
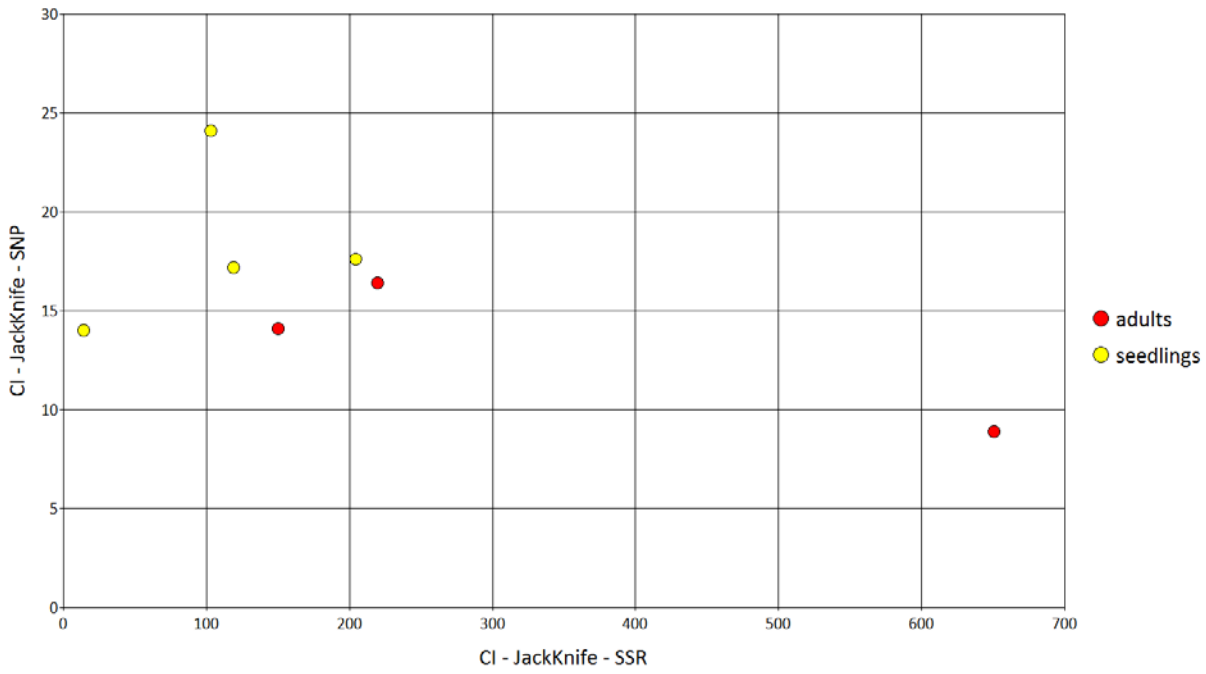


Figure 14: *Width of the confidence interval (upper – lower) of effective population size for Fagus sylvatica based on either SSR markers or SNP markers. JackKnife approach used to estimate confidence interval.*



3.5. The effect of sample size on the precision of genetic monitoring.

Objectives and methods

In order to determine the impact of sample size on the assessment of genetic structure, the simulation study was performed. Because SNP markers are biallelic, while SSR markers are multi-allelic in principle, the effective number of alleles (Ae) was used as a descriptive genetic parameter for a population. Generally, because data reduction increases stochasticity (or decreases precision) of estimates, any decrease in sample size can potentially influence conclusions regarding relative genetic values of populations or conservation units. Therefore, instead of details of the impact of data reduction on parameter values, the probability of recovery of correct genetic ranks of populations (stands) within the species was of interest. Here, genetic population ranks reflect directly relative levels of genetic variation of populations. It is likely that the reduction of sample size may generate higher stochasticity in genetic ranks when populations are weakly (insignificantly) differentiated, while it may have relatively smaller impact in the case of high (significant) differentiation among populations. Therefore, the probability of recovery of significance of the overall difference among populations in genetic variation was additionally assessed.

For each species, 8 data sets were available: 4 tree samples and 4 seedling samples. Each data set contains 96 individuals genotyped at 5-16 SSR markers. Additionally, for beech and pine, SNP data were available.

For each data set, the average Ae value was computed for both marker types. Then, Ae values were used to determine genetic population ranks (i.e., the higher Ae, the higher rank) for individual populations. Furthermore, the Friedman rank test was used to assess the significance of the difference of Ae between populations. The differences was denoted as significant if p-value < 0.05. These ranks were treated as the reference for comparisons. Subsequently, data sets were reduced to get *N* individuals (*N* = 84, 72, 60, 48, 36, 24, 12), randomly drawn (without

replacement) from the initial sample of 96 individuals. The truncated data were used to compute Ae and genetic ranks. For each *N*, the procedure was repeated 1000 times to obtain the proportion of recovery of the reference genetic ranks. In order to study how the reference ranks are affected by data reduction, the following proportions were computed: all ranks correct, the lowest rank correct, the highest rank correct, both extremes correct, intermediate ranks correct. In addition, for each truncated data set, the Friedman rank test was used to assess the significance of the difference between populations. Subsequently, the result of the test was compared to the reference. In order to compute the recovery rate for the Friedman test of differentiation, the proportion of simulations with the concordant test result was scored.

Results

Generally, reducing sample size affected population ranks both for SSR and SNP markers. In the case of SSR data, extreme ranks were the least affected while the intermediate ranks were the most affected. In the case of SNP data differences between extreme ranks and the intermediate ranks were less obvious, although still extreme ranks seemed more robust to the reduction of sample size.

Among the four species, *Quercus*, *Fagus* and *Pinus* revealed insignificant differences among populations in genetic variation at SSR markers. On the contrary, *Picea* revealed significant differences among populations. Therefore, the effect of reduction of sample size was compared between *Picea* ("differentiated") and the three genetically "uniform" species. Generally, there were slight differences between "differentiated" and "uniform" species in the proportion of recovery of both extreme and intermediate ranks when the sample size decreased from 96 down to 48 (Figure 15). However, with further reduction of sample size any differences vanished. In the case of SNP markers, the probability of recovery of both extreme and intermediate ranks decreased even more rapidly than in SSRs, allowing to conclude that ranks are generally hardly recovered with the reduced sample, regardless of marker type used.

There were some interesting patterns of recovery of differentiation between populations related to the initial (reference) significance. In the case of SSR markers, in the three species (*Quercus*, *Fagus*, *Pinus*), the difference between populations based on the reference sample was not significant. Those species showed only minor impact of the reduction of sample size on the recovery of the result of differentiation test (Figure 16). However, in the case of *Picea*, the reduction of sample size reduced the possibility of recovering of the significant differentiation down to 28% (for 12 individuals). In the case of SNP markers, both *Picea* and *Pinus* showed significant differences in Ae between populations when computed based on the complete samples (n=96) and, as compared to SSR markers (*Picea* data), they were much less affected in respect to the loss of recovery rate of differentiation test due to the reduction of sample size.

Table 25: *The impact of data reduction on the recovery of genetic ranks of populations computed based on SSR and SNP markers. The initial (reference) samples size was 96. Subsequent columns show the proportion of correctly recovered rank(s) for a given sample size.*

Species	Marker	Sample	lowest correct	highest correct	extremes correct	intermediate correct	All correct	[%] Signif. differences between pops
Fagus	SNP	96	1	1	1	1	1	1
Fagus	SNP	84	0.618	1	0.618	0.665	0.372	1
Fagus	SNP	72	0.521	0.991	0.518	0.596	0.209	0.989

Species	Marker	Sample	lowest correct	highest correct	extremes correct	intermediate correct	All correct	[%] Signif. differences between pops
Fagus	SNP	60	0.472	0.958	0.449	0.494	0.118	0.954
Fagus	SNP	48	0.406	0.909	0.361	0.419	0.072	0.880
Fagus	SNP	36	0.343	0.865	0.298	0.311	0.035	0.762
Fagus	SNP	24	0.315	0.738	0.225	0.188	0.021	0.620
Fagus	SNP	12	0.266	0.598	0.145	0.081	0.005	0.345
Pinus	SNP	96	1	1	1	1	1	1
Pinus	SNP	84	0.825	0.711	0.580	0.222	0.090	1
Pinus	SNP	72	0.725	0.568	0.411	0.176	0.050	1
Pinus	SNP	60	0.656	0.446	0.292	0.108	0.013	1
Pinus	SNP	48	0.619	0.397	0.247	0.083	0.007	1
Pinus	SNP	36	0.597	0.315	0.185	0.062	0.009	1
Pinus	SNP	24	0.593	0.263	0.142	0.056	0.005	1
Pinus	SNP	12	0.569	0.236	0.144	0.041	0.003	1
Fagus	SSR	96	1	1	1	1	1	0
Fagus	SSR	84	0.998	0.998	0.996	0.410	0.404	0.025
Fagus	SSR	72	0.932	0.97	0.905	0.317	0.261	0.048
Fagus	SSR	60	0.880	0.902	0.795	0.276	0.180	0.098
Fagus	SSR	48	0.826	0.833	0.694	0.197	0.104	0.111
Fagus	SSR	36	0.741	0.755	0.565	0.140	0.041	0.133
Fagus	SSR	24	0.663	0.649	0.424	0.068	0.016	0.128
Fagus	SSR	12	0.510	0.488	0.251	0.044	0.006	0.105
Picea	SSR	96	1	1	1	1	1	1
Picea	SSR	84	0.611	1	0.611	0.564	0.179	0.938
Picea	SSR	72	0.583	0.998	0.582	0.500	0.140	0.840
Picea	SSR	60	0.564	0.975	0.555	0.386	0.093	0.788
Picea	SSR	48	0.569	0.940	0.536	0.317	0.078	0.684
Picea	SSR	36	0.514	0.850	0.426	0.199	0.034	0.603
Picea	SSR	24	0.501	0.696	0.349	0.134	0.012	0.471
Picea	SSR	12	0.415	0.519	0.219	0.063	0.004	0.275
Pinus	SSR	96	1	1	1	1	1	0
Pinus	SSR	84	1	0.987	0.987	0.358	0.252	0
Pinus	SSR	72	1	0.893	0.893	0.217	0.125	0
Pinus	SSR	60	0.994	0.775	0.771	0.156	0.072	0
Pinus	SSR	48	0.978	0.676	0.661	0.115	0.034	0.001
Pinus	SSR	36	0.950	0.564	0.536	0.084	0.027	0.001
Pinus	SSR	24	0.865	0.458	0.400	0.062	0.008	0.006
Pinus	SSR	12	0.719	0.333	0.233	0.053	0.006	0.024
Quercus	SSR	96	1	1	1	1	1	0

Species	Marker	Sample	lowest correct	highest correct	extremes correct	intermediate correct	All correct	[%] Signif. differences between pops
Quercus	SSR	84	0.836	0.782	0.650	0.413	0.247	0
Quercus	SSR	72	0.755	0.719	0.542	0.342	0.141	0
Quercus	SSR	60	0.654	0.665	0.435	0.311	0.077	0
Quercus	SSR	48	0.585	0.637	0.368	0.258	0.052	0
Quercus	SSR	36	0.509	0.620	0.316	0.215	0.029	0.001
Quercus	SSR	24	0.441	0.561	0.257	0.213	0.015	0.005
Quercus	SSR	12	0.382	0.485	0.175	0.184	0.002	0.016

Table 26: *The impact of data reduction on the recovery of genetic ranks of populations computed based on SSR markers. The initial (reference) samples size was 96. Subsequent columns show the average proportion of correctly recovered rank(s) for a given sample size, computed across all the species.*

Sample	lowest correct	highest correct	extremes correct	intermediate correct	All correct
96	1	1	1	1	1
84	0.861	0.941	0.811	0.436	0.270
72	0.817	0.895	0.730	0.344	0.166
60	0.773	0.829	0.639	0.282	0.105
48	0.739	0.771	0.564	0.221	0.067
36	0.678	0.697	0.460	0.159	0.032
24	0.617	0.591	0.357	0.119	0.012
12	0.506	0.456	0.219	0.086	0.004

Figure 15: *The impact of reduction of sample size on the recovery rate of genetic ranks of populations.*

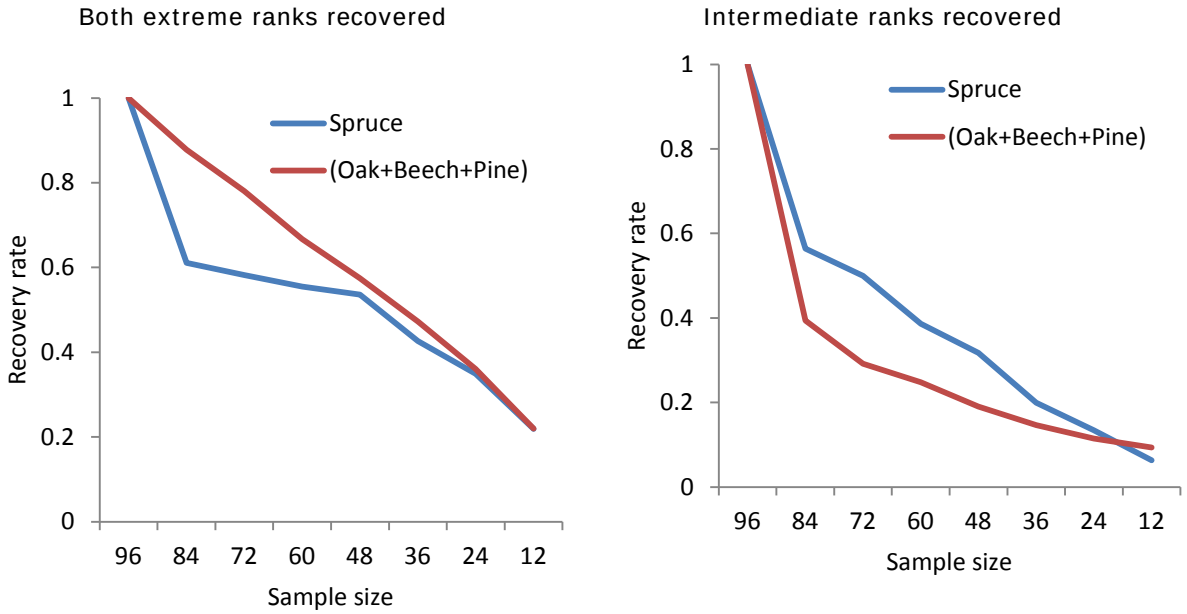
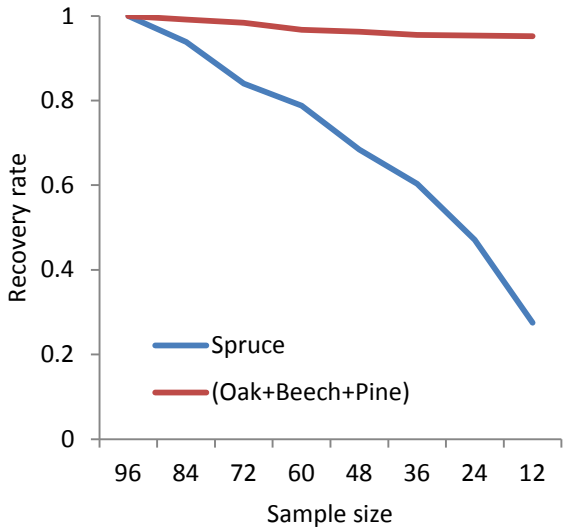


Figure 16: *The impact of reduction of sample size on the recovery rate of the overall differentiation in genetic variation levels among populations.*



Conclusions

The reduction in sample size may have strong impact on the conclusions on relative levels of genetic variation of populations. Although it seems to be extremely difficult to properly recover all ranks simultaneously (especially if many populations are compared), for the purpose of conservation of genetic resources it may be of little practical value. Instead, it is important to properly identify the extreme genetic ranks. In this respect, SSR markers seem to be less prone to

bias in recovering genetic ranks compared to SNP markers. On the other hand, SNP markers showed relatively little impact of the reduction of sample size on the chance of observing the overall difference in genetic variation levels between populations. Finally, whether populations show significant differences in genetic variation or not, has little effect on the recovery rate of extreme (low and high) genetic ranks.

3.6. The effect of number of marker loci for the precision of genetic monitoring.

Objectives and methods

In order to determine the impact of a number of markers on the assessment of genetic structure, the simulation study was performed. Because SNP markers are biallelic, while SSR markers are multi-allelic in principle, the effective number of alleles (Ae) was used as a descriptive genetic parameter for a population. Generally, because data reduction increases stochasticity (or decreases precision) of estimates, any decrease in the number of markers can potentially influence conclusions regarding relative genetic values of populations or conservation units. Therefore, instead of details of the impact of data reduction on parameter values, the probability of recovery of correct genetic ranks was of interest. Here, genetic population ranks reflect directly relative levels of genetic variation of populations. Also, the probability of recovery of significance of the overall difference among populations in genetic variation was assessed.

For each species, 8 data sets were available: 4 tree samples and 4 seedling samples. Each data set contains 96 individuals genotyped at 5-16 SSR markers. Additionally, for *Fagus* and *Pinus*, SNP data were available.

For each data set, the average Ae value was computed for all the marker. Then, Ae values were used to determine genetic population ranks (the higher Ae the higher rank) for individual populations. Furthermore, the Friedman rank test was used to assess the significance of the difference of Ae between populations. The differences were denoted as significant if p-value < 0.05. These ranks were treated as the reference for comparisons. Subsequently, data sets were reduced to get *N* markers, randomly drawn (without replacement) from the total number markers. The truncated data were used to compute Ae and genetic ranks. For each *N*, the procedure was repeated 1000 times to obtain the proportion of recovery of the reference genetic ranks. In order to study how the reference ranks are affected by data reduction, the following proportions were computed: all ranks correct, the lowest rank correct, the highest rank correct, both extremes correct, intermediate ranks correct. In addition, for each truncated data set, the Friedman rank test was used to assess the significance of the difference between populations. Subsequently, the result of the test was compared to the reference, In order to compute the recovery rate for the Friedman test of differentiation, the proportion of simulations with the concordant test result was scored.

Results

Generally, reducing the number of markers affected population ranks both for SSR and SNP markers. In the case of SNP markers, the reduction of approx. 50% led to the decrease of recovery rate for extreme ranks down to 22-29%, depending on a species. The intermediate ranks, with the recovery rate of 4-20% were even more affected. In the case of SSR markers, the reduction of the number of markers by half led to the decrease of recovery rate for extreme ranks down to 29-41%, or even 2%, when the number of markers was reduced from the total of 8 to 4 (Quercus). The reduction of the number of markers had also negative impact on the recovery of significance of differences among populations. However, this effect seemed to be strongly dependent on the species (the level of differentiation). For highly heterogeneous populations (Pinus), the reduction of SNP markers from 126 down to 46 markers only allowed to recover the significant differentiation in Ae with the chance of 88%. For comparison, the significant differentiation among beech populations was recovered with the chance of 38% only for the same number of SNP markers. In

the case of SSR markers, the effect was analogous, i.e. when there is significant heterogeneity in genetic variation (Picea), data reduction had negative impact on the recovery of this significance. On the other hand, when populations were homogeneous in respect to Ae (Fagus, Quercus), data reduction increased the chance of drawing the opposite conclusion, but only slightly compared to the case of significant heterogeneity.

Table 27: *The impact of data reduction on the recovery of genetic ranks of populations computed based on SSR and SNP markers. The reference samples are denoted with (ref.). Subsequent columns show the proportion of correctly recovered rank(s) for a given sample size.*

Species	Marker	Number of markers	lowest correct	highest correct	extremes correct	intermediate correct	All correct	[%] Signif. Differences between pops
Fagus	SNP	127 (ref.)	1	1	1	1	1	1
Fagus	SNP	110	0.484	0.989	0.482	0.573	0.254	0.971
Fagus	SNP	94	0.387	0.934	0.371	0.397	0.126	0.808
Fagus	SNP	78	0.374	0.875	0.337	0.311	0.081	0.658
Fagus	SNP	62	0.329	0.805	0.290	0.199	0.025	0.507
Fagus	SNP	46	0.302	0.677	0.226	0.151	0.033	0.378
Fagus	SNP	30	0.272	0.592	0.166	0.107	0.012	0.249
Fagus	SNP	14	0.234	0.439	0.111	0.050	0.004	0.152
Pinus	SNP	126 (ref.)	1	1	1	1	1	1
Pinus	SNP	110	0.774	0.500	0.370	0.132	0.018	1
Pinus	SNP	94	0.654	0.441	0.282	0.076	0.006	1
Pinus	SNP	78	0.631	0.409	0.257	0.043	0.002	0.999
Pinus	SNP	62	0.598	0.356	0.217	0.035	0.000	0.972
Pinus	SNP	46	0.583	0.322	0.175	0.032	0.000	0.877
Pinus	SNP	30	0.541	0.276	0.161	0.019	0.000	0.661
Pinus	SNP	14	0.478	0.230	0.116	0.025	0.000	0.356
Fagus	SSR	16 (ref.)	1	1	1	1	1	0
Fagus	SSR	14	0.994	0.776	0.77	0.173	0.121	0.152
Fagus	SSR	12	0.828	0.682	0.585	0.060	0.015	0.205
Fagus	SSR	10	0.637	0.608	0.431	0.037	0.009	0.197
Fagus	SSR	8	0.509	0.513	0.291	0.026	0.004	0.178
Fagus	SSR	6	0.376	0.445	0.197	0.023	0.002	0.12
Fagus	SSR	4	0.276	0.343	0.126	0.012	0	0.134
Picea	SSR	12 (ref.)	1	1	1	1	1	1

Picea	SSR	10	0.553	0.872	0.437	0.385	0.166	0.631
Picea	SSR	8	0.545	0.764	0.412	0.282	0.079	0.478
Picea	SSR	6	0.587	0.708	0.413	0.171	0.054	0.309
Picea	SSR	4	0.528	0.555	0.286	0.083	0.016	0.186
Quercus	SSR	8 (ref.)	1	1	1	1	1	0
Quercus	SSR	6	0.46	0.301	0.13	0.187	0	0
Quercus	SSR	4	0.191	0.165	0.02	0.167	0	0.021

Conclusions

The reduction of the number of markers has negative impact on the recovery of genetic ranks and the overall significance of heterogeneity in genetic variation among populations. Even for highly differentiated populations (species), the reduction of SNPs from 126 to 62 substantially reduces the chance for proper identification of populations with extreme levels of genetic variation. Thus, when compared to the effect of the reduction of sample size, the reduction of the number of markers seems to have more negative impact on the recovery of the genetic information.

3.7. Relationships between demographic, genetic, climatic and reproductive parameters.

Correlations between different variables

All measures were transformed into variables standardized within species x(0,1)(i.e., with a mean equal 0 and standard deviation equal 1). In this way measures obtained for different species could be compared simultaneously and differences among species are purged out. However, standardization caused that the standardized measures became incomparable among species in terms of the scale of differences within species. Note also, that number of correlation pairs is relatively low (16) making detection of significant correlations difficult. At the same time, detecting false positive correlations is also possible, so detecting significant correlations should be supported by reasonable explanation.

Below selected correlations between standardized measures of several traits are presented. All measures presented below are standardized, unless explicitly indicated that not standardized measures were used.

Geographic variables

Note that latitude and longitude of sample sites are highly correlated (0.7558; p=0.001) but standardized measures of latitude and longitude are not correlated (0.4537; p=0.078, although the same trend shown). Depending on the context, we might be interested to see relationship with real or standardized coordinates.

Latitude

- Positively correlated with effective number of alleles (Ae) in SSR loci in adults (0.5233; p=0.038) (Figure 17). Similar trend was found for real latitude but not so significant (0.4964; p=0.051) (Figure 18). The trend is determined mainly by the distribution of Ae in spruce, oak and beech. But Maritime pine mainly causes disturbance of this relationship, because it shows no trend to latitude.

Figure 17: *Relationship between Ae in microsatellites and Latitude.*

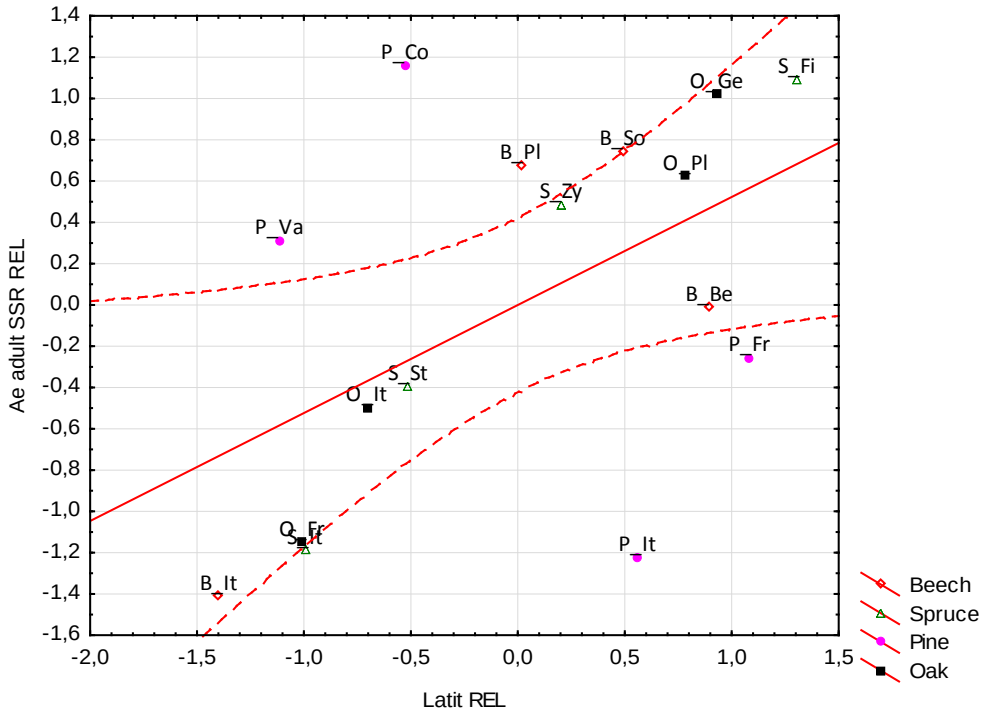
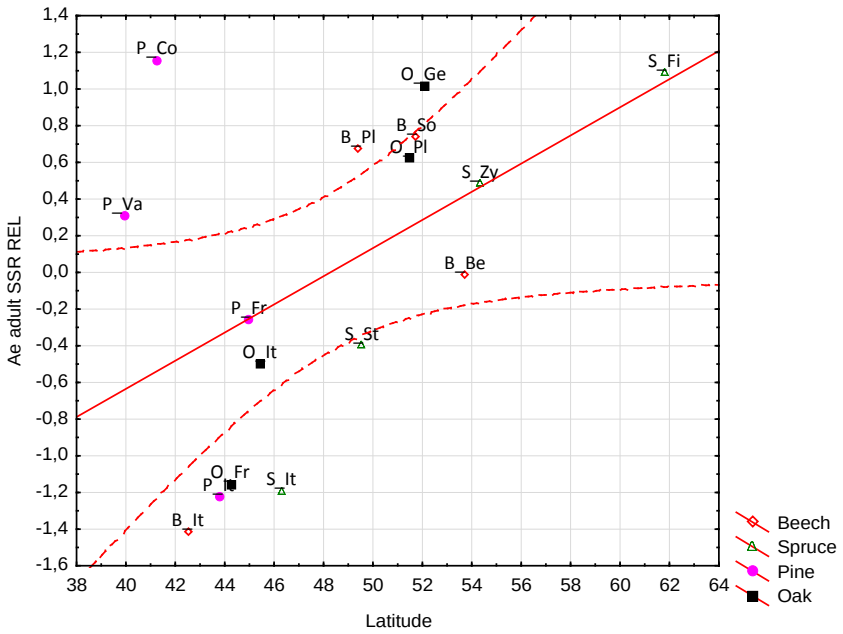


Figure 18: *Relationship between Ae (SSR) REL and Latitude (not standardized).*



Longitude

- Negatively correlated with Ae SNP adults (-0.7933; p=0.019) and Ae SNP seedlings (-0.8870; p=0.003) but sample size is small n=8. Trend determined by low Ae in Lesko (in beech) and Italian population (in pine), both in adults and seedlings.
- Negatively correlated with Fis seedlings (corrected for null alleles) (-0.5134; p=0.042). Difficult to interpret

We detected several correlations between area of sample plots, density of trees and distance to nearest neighbors, which seems obvious for this type of data.

Nearest neighbor distance

- Positively correlated with DBH (0.6964; p=0.003). Distance to nearest neighbors is increasing with increasing mean DBH within a stand.
- Positively correlated with **Ae adults SNP** (Ae increases with increasing distance among individuals (0.6026; p=0.057)
- Positively correlated with **Ae seedlings SNP** (Ae increases with increasing distance among adult individuals) (0.7312; p=0.039)
- Negatively correlated with Fis (corrected for nulls) adults SSR (Fis decreases as distance between trees increases) (-0.5311; p=0.034)
- Negatively correlated with DBH effect of male mating success (pollen) (-0.6388; p=0.025) (the importance of DBH as a covariate of male mating success decreases with increasing distance among trees)
- Positively correlated with distance effect of male mating success (seedlings) (0.5326; p=0.050) (the importance of distance between mates increases with increasing the distance to nearest neighbors) (reasonable)
- Negatively correlated with DBH effect of female reproductive success (seedlings) (-0.5874; p=0.027) (the importance of diameter as a covariate of female reproductive success decreases as distance to nearest neighbors is increasing) (if distances among trees become larger then DBH of trees is not a good predictor of their contribution to progeny).
- Positively correlated to Ne of seeds (0.6798; p=0.015)(wider spacing causes possibly mating with larger number of males increasing Ne)(reasonable)

R as a measure of clustering (<1) or overdispersion (>1)

- Negatively correlated with p-value of significance of DBH autocorrelation (-0.5451; p=0.029) –Autocorrelation of DBH is increasing (more likely) with decreasing tree clustering (i.e. DBH autocorrelation is higher in populations of more uniform tree distribution).
- Negatively correlated with pollen dispersal distance (-0.5942; p=0.042). Mean pollen dispersal distance is increasing with increasing clustering.
- Negatively correlated with the distance effect of male mating success (-0.6028; p=0.022) the importance of distance between mates (as a predictor of mating success) is increasing if trees become more clustered.

Significance of Autocorrelation of DBH (p-value)

- Negatively correlated with Moran's I in the 1st class (-0.5607; p=0.024) Moran's I in the 1st class increases with increasing significance of DBH autocorrelation (i.e. decreasing p-value)
- Positively correlated with Ae adult SSR (0,6120; p=0,012) Ae decreases with increasing significance (decreasing p-value). **Populations are more genetically diverse** when there is no DBH autocorrelation

- Positively correlated with Ae adult SNP (0,7147; p=0,046) Ae decreases with increasing significance (decreasing p-value). **Populations are more genetically diverse** when there is no DBH autocorrelation (why)

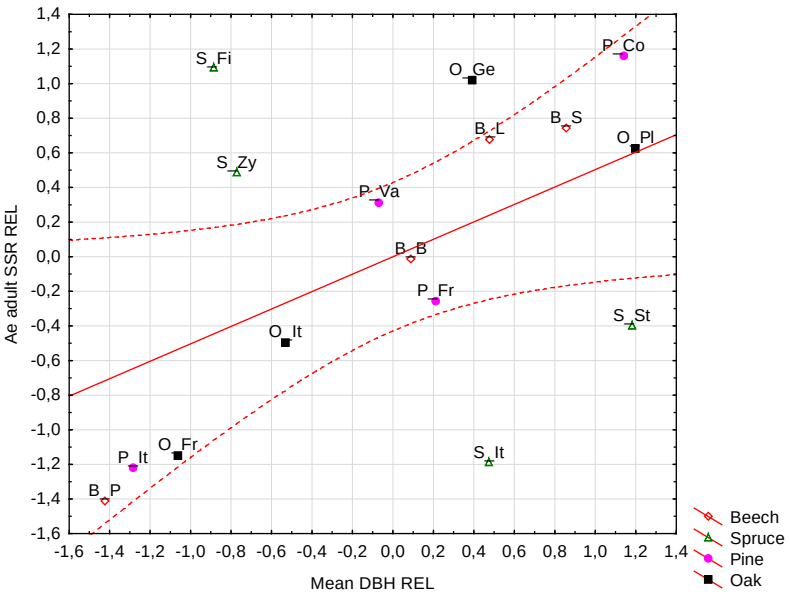
Moran's I index in the 1st class

- Negatively correlated with significance of Autocorrelation of DBH (p-value) (-0.5607; p=0.024) Moran's I increases with increasing significance of autocorrelation (i.e. decreasing p-value) (AS ABOVE)
- Negatively correlated with Ne seedlings (-0.6333; p=0.008). Ne of seedlings is decreasing with increasing DBH similarity among nearest neighbors.

DBH

- Negatively correlated with stand density (-0.8547; p<0.001). Mean DBH is increasing with decreasing stand density (kind of obvious)
- Positively correlated with nearest neighbor distance (0,6946; p=0,003) Mean DBH is increasing with increasing distance to nearest neighbors (Kind of obvious)
- **Positively correlated with Ae in adults SSR** (0.5036; p=0.047) Ae SSR is increasing with increasing mean DBH (the same trend observed for SNP but not significant at p<0.05) This trend is positive for pine beech and oak, but not for Spruce (Figure 19)

Figure 19: *Relationship between DBH and Ae in adults (SSR). The correlation is stronger without spruce.*

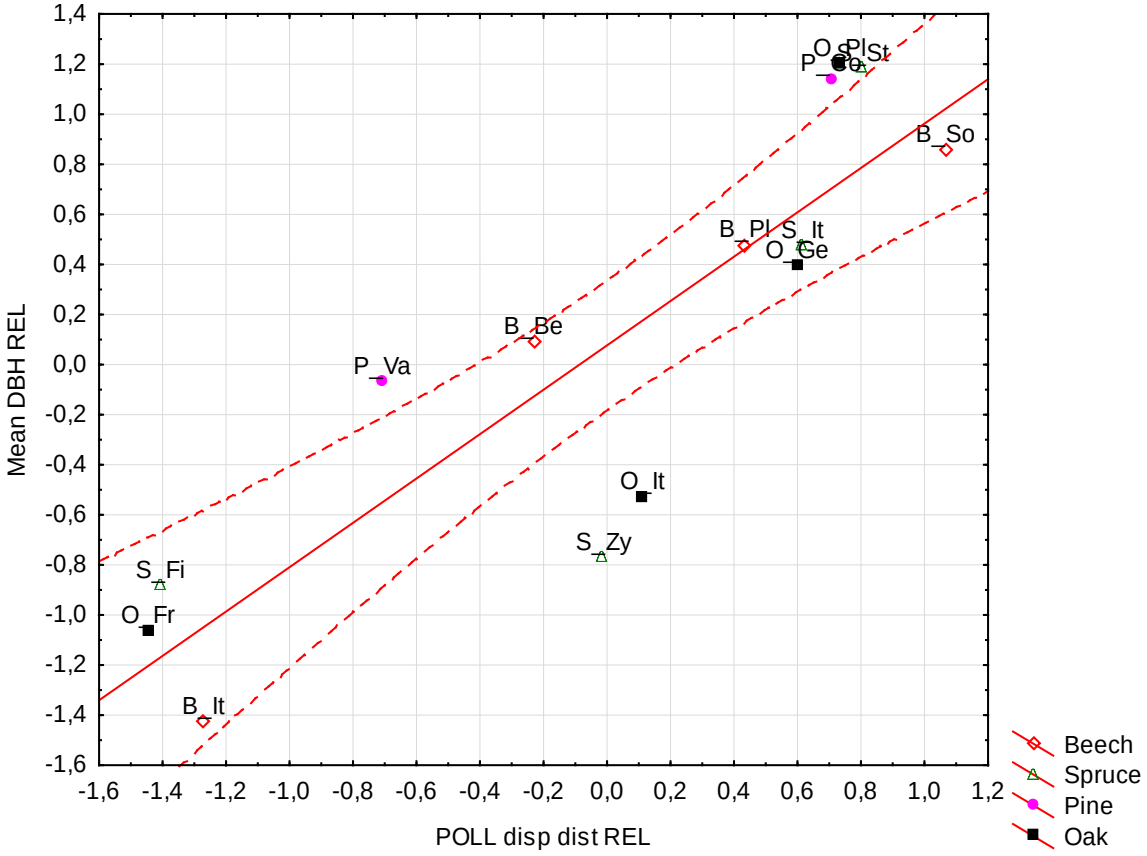


- Positively correlated with number of genotypes in SNP (0.926; p=0.001) Number of SNP genotypes is increasing with increasing DBH. The same trend is observed of NG of SSR, but not significant at p<0.05.
- Negatively correlated with diameter effect on male mating success (observed at the stage of seeds) (-0.6521; p=0.022) Importance of DBH as covariate of male mating success is decreasing with increasing stand DBH. (this may be possible because male mating success (or pollen production) per tree is not linearly related to three DBH, largest (and oldest)

trees often do not have the highest pollen production, so the effect of tree DBH of male mating success may change with the age of stand).

- Negatively correlated with diameter effect on female reproductive success (-0.7241; p=0.003) Importance of DBH as covariate of female reproductive success is decreasing with increasing stand DBH. (this may be possible because female reproductive success (or seed production) per tree is not linearly related to three DBH, largest (and oldest) trees often do not have the highest seed production, so the effect of tree DBH of female reproductive success may change with the age of stand).
- Generally positively correlated to Pollen and seed dispersal, but highly significant correlation noted for pollen dispersal distance based on seedlings (0.8760; p<0.001) (Figure 20). Mean pollen dispersal distance is increasing with increasing stand DBH. Note that DBH is negatively correlated to stand density which might have direct effect on pollen dispersal.

Figure 20: *Relationship between pollen dispersal distance and mean DBH.*



Coefficient of variation of DBH

- Negatively correlated with Ae adult SSR (-0.5019; p=0.048), Ae seedlings SSR (-0.5427; p=0.030) Ae seeds SSR (-0.8722; p<0.001). This means, that generally **Ae is decreasing with increasing variability of DBH** within the stand. Stands more uniform in DBH have higher Ae.

- Negatively correlated with Ne seeds (-0.6853; p=0.014), (adults and seedlings also had negative correlation) this indicates that **effective number is decreasing with increasing variability of DBH** within the stand.

Ae SSR (effective number of alleles) in ADULTS

- Positively correlated with Longitude REL (0.5233; p=0.038). Ae is increasing towards eastern locations (within the species) (AS ABOVE)
- Negatively correlated with mean stand density (loosely)
- Positively correlated with significance (p-value) of DBH autocorrelation (0,6120; p=0,012) Ae decreases with increasing significance (decreasing p-value). **Populations are more diverse** when there is less DBH autocorrelation
- Positively correlated with DBH (0.5036; p=0.047) Ae SSR is increasing with increasing mean DBH (Figure 19)
- Negatively correlated with CV. of DBH (-0.5019; p=0.048), Ae seedlings SSR (-0.5427; p=0.030) Ae seeds SSR (-0.8722; p<0.001). This means, that generally **Ae is decreasing with increasing variability of DBH** within the stand. Stands more uniform in DBH have higher Ae
- Positively correlated with Ae in seedlings (0.6988; p=0.003) and seeds (0.6945; p=0.012). This generally means that sampling of one cohort (eg. adults) provides good measure of genetic diversity for all cohorts. This trend is strong, but while spruce, pine and oak follow this trend the beech does not (Figure 21)

Figure 21: *Relationship between Ae in adults and Ae in seedlings (microsatellites). Note that Ae of seedlings in beech do not represent well Ae in adult populations.*
Note that comparing real (not standardized) Ae in adults and seedlings is inappropriate – shows species differences:

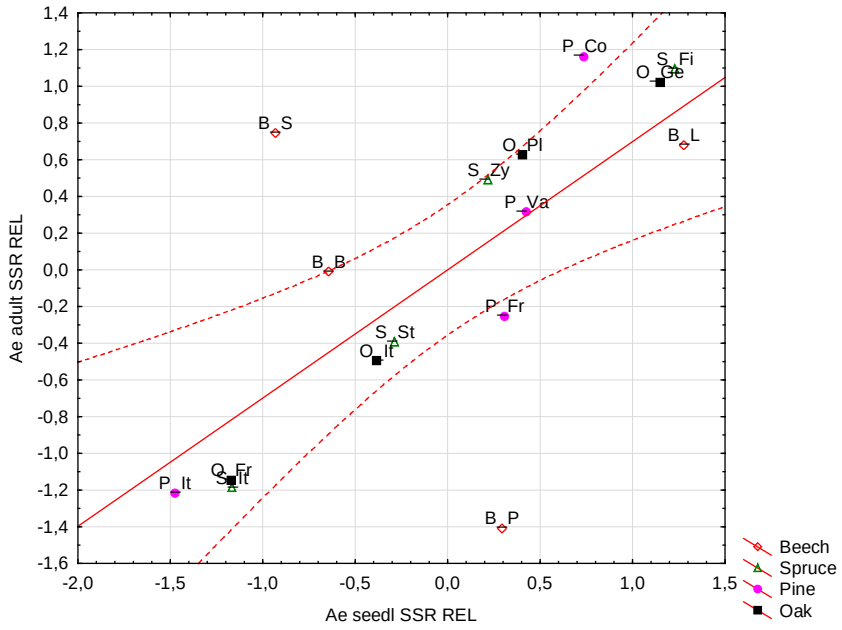
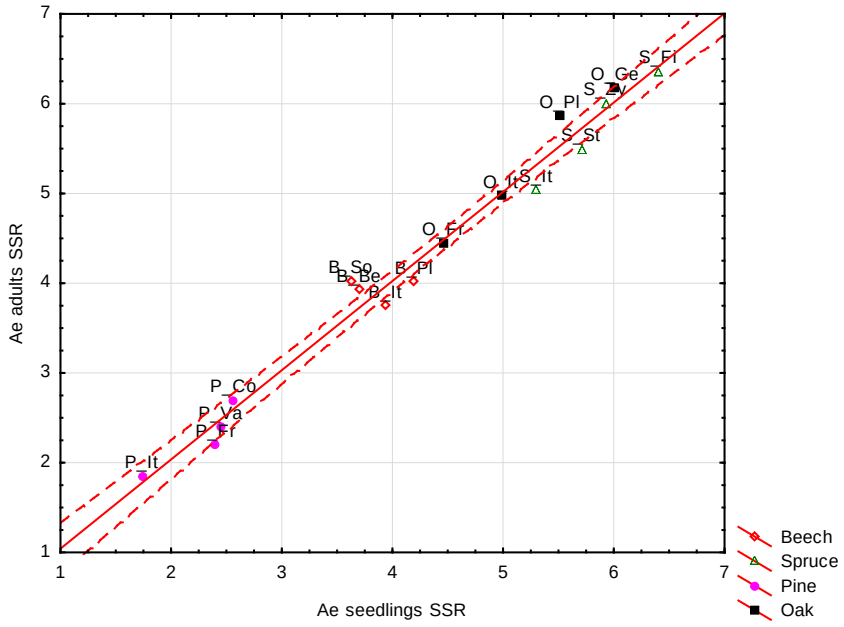


Figure 22: *Relationship between real (not standardized) Ae in adults and Ae in seedlings (microsatellites). Note that Ae of adults and seedlings show clear species structure*



- Positively correlated with Ne in seedlings (0.5071; p=0.045) and Ne in seeds (0.8913; p<0.001).
- Trends between Ae in SSR and SNP markers are positive but low significance is probably due to low number of pairs (n=8).

Ae in seedlings SSR

Negatively correlated with CV. of DBH (similarly as in adults) (-0.5427; p=0.030)
High correlations with Ae in adults (as above) and seeds (0.8329; p=0.001)

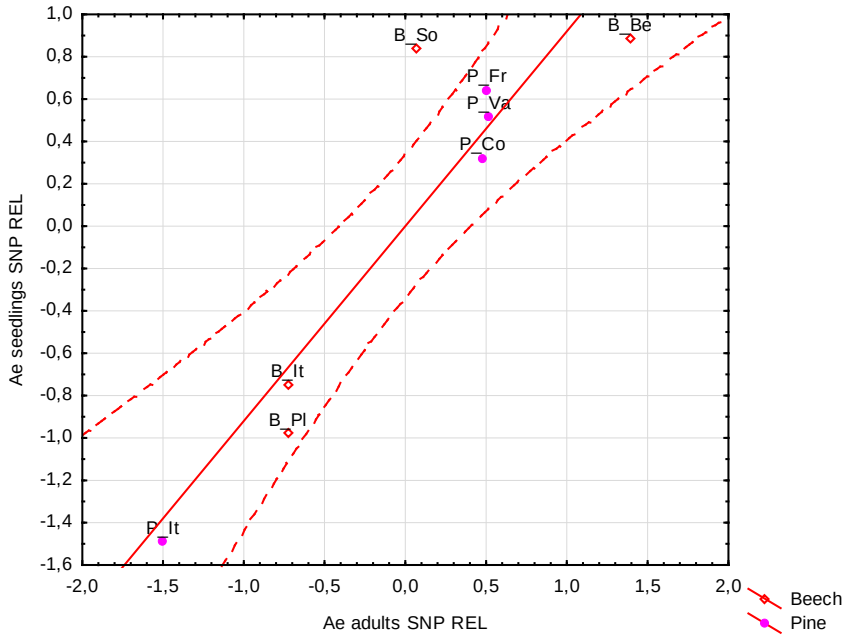
Ae in seeds SSR

- Negatively correlated with CV. of DBH (similarly as in adults and seedlings) (-0.8722; p<0.001)
- High correlations with Ae in adults and seedlings (as above)
- Positively correlated with Ne in seeds SSR (0.7185; p=0.008).

Ae adults SNP

- Negatively correlated with Longitude (-0.7933; p=0.019) (above)
- Negatively correlated with stand density (-0.7652; p=0.027).
- Positively correlated with significance (p-value) of DBH autocorrelation (0.7147; p=0.046). Ae increases with decreasing degree of DBH autocorrelation).
- Positively correlated with Ae in seedlings (0.9201; p=0.008). It is only based on 8 locations (2 species) but the signal is strong (Figure 23).

Figure 23: *Relationship between Ae in adults and Ae in seedlings within stand (SNP markers).*



Ae in seedlings SNP showed similar correlation patterns as Ae in adults SNP

Fis in adults (based on SNP markers)

- Positively correlated with pollen immigration (0.9416; p=0.002) Increase of inbreeding in adults promotes effectiveness of pollen immigration
- Positively correlated with Ne in adults SSR (0.7094; p=0.049), also positive correlation with Ne in adults in SNP, but not significant.

Fis in seedlings (based on SNP)

- Positively correlated with Ae in adults SNP

Pollen immigration (based on seeds)

- Positively correlated with distance effect of males (0.6820; p=0.021) the importance of distance as predictor of mating success is decreasing with pollen immigration – reasonable relationship given the model.
- Not much correlated with pollen dispersal distance (dispersal kernel does not depend on pollen immigration)
- Positively correlated with bF (0.8650; p=0.003), but negatively correlated with Sp (-0.8627; p=0.003). SGS increases with pollen immigration decreasing.
- Positively correlated with Ne adults SSR REL (0.7158; p=0.009).
- Positively correlated with Fis adults SSR (corrected for nulls) (0.6041; p=0.037)
- Positively correlated with Fis adults SNP (0.9416; p=0.002) (see above)
- Positively but loosely correlated with pollen immigration based on seedlings.

Self-fertilization

- Not correlated with any other trait (almost)

Distance effect on male mating success (seed stage)

- Positively correlated to dist effect in male mating success at seedlings stage (but not significant)

Pollen dispersal distance

- Pollen dispersal distance is increasing with increasing clustering (decreasing R) (-0.5942; p=0.042)
- Positively correlated with distance effect but this is obvious given computational relationship (the model) between these variables.

Seed immigration

- Positively correlated with Ne seedlings (0.6334; p=0.015). Ne of seedlings may increase with the immigration of seeds (reasonable).

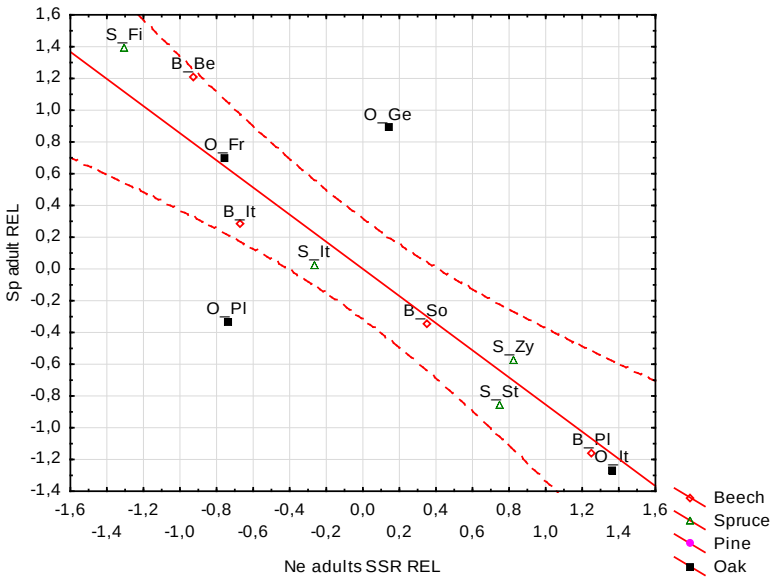
Kernel of pollen dispersal distance depends largely on the density and clustering of stands. Pollen dispersal distance is increasing with decreasing stand density and when increasing distance to nearest neighbor.

We found no correlation between SGS measures in adults and seedlings (the extend of spatial structure in seedlings and adults is not related)

Ne in adults SSR

- Negatively correlated with SGS (-0.8545; p<0.001). Ne increases with decreasing Sp (with decreasing the strength of SGS). This pattern is strong for spruce, beech and pine, but oaks loosely follow this rule.

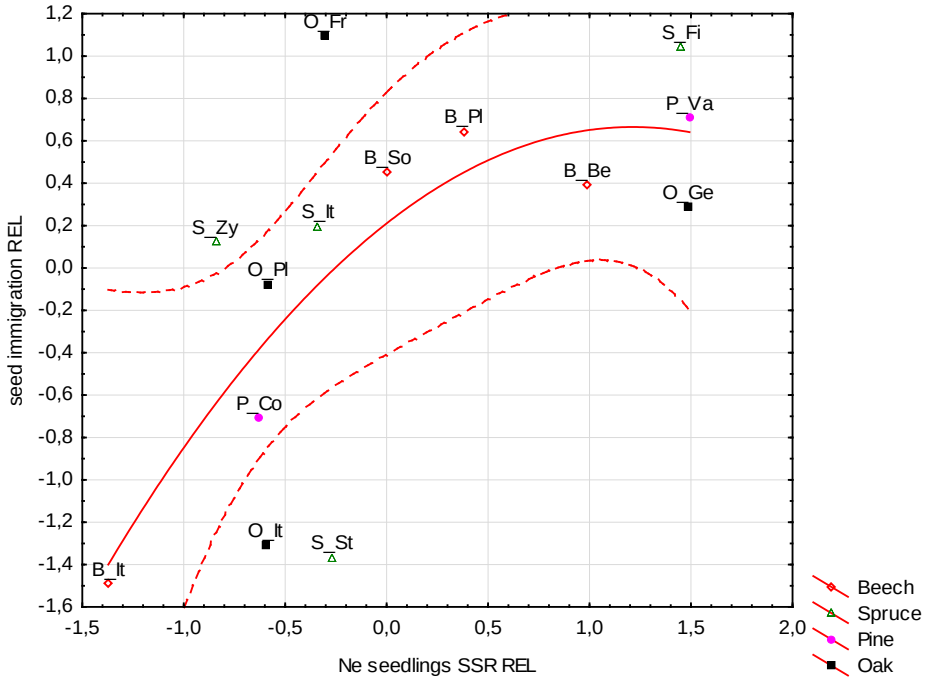
Figure 24: *Relationship between Sp and Ne (SSR) in adults.*



Ne in seedlings

- Positively correlated with level of seed immigration observed at the stage of seedlings (0.6334;p=0.015). Ne in seedling increases as seed immigration is increasing (reasonable), but the relationship may not be linear.

Figure 25: *Relationship between Ne of seedlings and seed immigration*



Ne in seeds

- Negatively correlated with stand density, but positively correlated with distance to nearest neighbor.
- Positively correlated with mean DBH
- Negatively correlated with C.V. of DBH
- Positively correlated with Ae adults SSR
- Positively correlated with Ae in seeds

Relationship between Ae and Ne:

Adults (0.0100; p=0.971)
Seedlings (0.3607; p=0.170)
Seeds (0.7185; p=0.008)

This suggests that the relationship between Ae and Ne is becoming weaker in time: the strongest at the stage of seeds, the weakest at the stage of adults.

General conclusions:

Genetic diversity (Ae) was correlated between different life stages (adult to seedlings to seeds). This suggests, that assessing genetic diversity in one cohort provides the estimates typical for the stand.

Genetic diversity (Ae) was not random in respect to geographic locations and it appeared to increase with latitude.
Genetic diversity was higher in stands with lower tree density, higher mean DBH, low variation of DBH within stands, and low spatial autocorrelation of DBH.
Genetic diversity (Ae) was positively correlated with Ne of seed and seedling cohorts.

4 Discussion

4.1. Recommendations related to sample design

Is seed and seedling sampling necessary?

Genotyping of seeds allowed estimation of mating system parameters, pollen dispersal and pollen immigration rates. No departure from random mating (insignificant inbreeding) could be observed. Further, genetic diversity was correlated among seeds and other life stages, and the effective population size in seeds was correlated with genetic diversity in adults. We can therefore conclude that data on seeds might not be useful in genetic monitoring protocols. The high costs associated with several visits to monitoring stands, seed harvesting and genotyping of many offspring could thus be spared. Another reason to leave the seed stage out is that a single seed year has only very little impact on the genetic composition of the future adult generation.

However, sampling of seedlings should be performed to detect potential changes in genetic structure. There is clear evidence for strong genetic selection in tree populations in the early ontogenetic stages (Gregorius and Degen 1994). Here a dramatic mortality occurs. In particular, a Gregorius genetic distance among adults and seedlings for nSSRs higher than 0.15 might indicate loss of genetic diversity or genetic changes which should be further characterized.

Adult and seedling sampling

Our results suggested that we should focus on the adult cohort and saplings. A representative sampling over the whole stand (with several transects) should avoid sampling of only a small part of the available genetic diversity in the stand. In order to monitor genetic changes the inventory of the sapling should be repeated every 10 years. At the same time the adults should be revisited and the mortality should be recorded and new adults (above a threshold diameter) should be sampled and genotyped.

4.2. Recommendations related to sample size

Results of the rarefaction study indicated that ranking of genetic diversity could not be recovered when a reduction of sample size and number of loci was applied. Furthermore, we cannot estimate the real ranking, and it is possible that 96 adult individuals genotyped at 120 nSNPs or at 12 nSSRs do not provide a reliable estimation on ranking of genetic diversity.
Simulations on larger datasets are therefore required to answer this question, and also to find the optimum between sample size, number of loci and sampling design. Based on our dara so far we recommend to have a sample size for adults and saplings of 150 individuals in each cohort and to use at least 120 SNPs.

4.3. Recommendations related to the type of markers

Our results indicated similar results among nSNP and nSSR markers, therefore only one marker set could be selected. Several parameters should be taken into consideration to find the best marker type. nSNP loci seem to be less affected than nSSR loci by genotyping errors, and the amount of available loci would allow selection of reliable markers. New Generation Sequencing methods also provide opportunity for development of large sets of markers.
By contrast, nSSRs genotyping often result in errors due to stuttering and large allele dropout. Further, many SSR loci are strongly affected by the presence of null alleles, which further decreases the choice of suitable loci. For this reason, the use of SNP markers might be encouraged.

Costs for SNP genotyping are strongly decreasing, which should convince the scientific community to work with SNP. Moreover the standardisation among labs would be much easier with bi-allelic SNPs instead of nSSRs with variation of fragment sizes of up to 30 different alleles.

4.4. Recommendations of parameters to be use in genetic monitoring

4.4.1. Demography

- DBH of adults
- Spatial position of adults and seedlings
- Vitality classification?

These demographic parameters allow the estimation of density, spatial structuring and age structure of the stand. This should indicate, together with geographical information, whether high or low genetic diversity is expected. At high latitudes (for beech, oak and spruce), low density, high DBH, and low variation in DBH, high genetic diversity is expected. It might also further suggest that the stand has been planted or that management regimes such as thinning occurred. We recommend to apply quick screening methods (e.g. the application of so called resistographs (http://www.iml.de/de/holzpruefsysteme/verfahren/bohrwiderstandsmessung/?gclid=CPXFuKrCqMkCFSHmwgodXcIHxA) to measure the dynamics year rings of individual trees. These profiles of year ring growth are very useful aggregated parameters of tree vitality could be correlated with time series on many environmental parameters (e.g. rainfall, temperatures).

4.4.2. Genetic diversity

- Effective number of alleles (Ae)
- Unordered number of genotypes (NG)
- Genetic distance among adults and seedlings (Gregorius)

Ranking of these parameters among populations should indicate outliers, i.e. stands with lower genetic diversity than expected and which might be at risk of genetic loss. High genetic distance among adults and seedlings might further indicate ongoing genetic erosion processes, which could be further clarified with an analysis of gene dispersal.

4.4.3. Reproductive processes

- Effective population size (Ne)

Analysis of reproductive processes did not show strong differences among the study stands and there was no evidence for inbreeding. Further, strong selection among seeds usually occurs in forest tree species, which reduces the probability to observe effects of inbreeding at the adult stage. Therefore, cost-effective protocols of genetic monitoring should not include data on seeds. The estimation of effective population size on adult and seedlings provide further estimation of genetic diversity, but also seed immigration.

5 Conclusions

Selection of plots

- Much more plots per species are needed to estimate the combined impact of different factors on genetic composition (N>50)
- More focus to main target factors (environment + forest management)
- Selection of near by pairs of plots varying in only one factor

Sample design within plots

- Genetic differentiation among different ontogenetic stages usually much smaller than genetic differentiation among plots => definition of critical thresholds possible
- Little differentiation in the mating system among plots, small impact of a single seed year on next generation of adults => leave seed studies out
- Representative sampling of adults and saplings over the whole plot (parallel transects)

Sample design within plots

- At least 100 (better 150) adult individuals and at least 100 (better 150) saplings sampled over the whole plot (parallel transects)
- Increase number of SNP gene markers including adaptive and non-adaptive ones (N>150)
- Every 10 years: Repeat the inventory of the saplings, check mortality and vitality of adults, add new adults with a diameter above threshold

Selection of gene markers

- Correlation of diversity and differentiation among nSSRs and nSNPs
- need to be checked for selective SNPs
- Similar ranking in estimated parameters of SGS, mating system and gene flow from nSSRs and nSNPs
- Genotyping and scoring errors are smaller for bi-allelic nSNPs
- Better standardisation among labs for SNPs
- Genotyping of 120 SNPs same costs as 10 nSSRs (10-30 Euro per individual)
- Information content of 120 SNPs with 2 alleles > 10 nSSRs with 10 alleles
- Useful combination of adaptive and non-adaptive SNPs
- Many SNPs are more presentative for the genome compared to a few SSRs

=> No need to apply both type of gene markers (SNPs are better!)

Observed correlations with genetic parameters

- Positive correlation among genetic diversity and latitude => increase from West to East
  - mixture of trees from different refugia in Central Europe
  - gene flow from all directions in the centre
  - more extinctions at the edges

- Negative correlation of genetic diversity with tree density, variation and spatial structure of DBH
  - negative effect of family structures

6 References

Adams, W.T., and Birkes, D.S. (1991). Estimating mating patterns in forest tree populations. In: Fineschi, S., Malvolti, M.E., Cannata, F., Hattemer, H.H. (Eds.), Biochemical Markers in the Population Genetics of Forest Trees. SPB Academic Publishing, The Hague, The Netherlands, pp. 157–172.

Burczyk, J., Adams, W.T., Birkes, D.S., Chybicki, I.J. (2006). Using genetic markers to directly estimate gene flow and reproductive success parameters in plants on the basis of naturally regenerated seedlings. *Genetics* **173***:* 363-372.

Chagne, D., P. Chaumeil, A. Ramboer, C. Collada, A. Guevara, M. T. Cervera, G. G. Vendramin, V. Garcia, J. M. M. Frigerio, C. Echt, T. Richardson, and C. Plomion (2004). Cross-species transferability and mapping of genomic and cDNA SSRs in pines. *Theoretical and Applied Genetics* **109**:1204-1214.

Chybicki, I. J. and Burczyk, J., 2010., Realized gene flow within mixed stands of Quercus robur L. and Q. petraea (Matt.) L. revealed at the stage of naturally established seedling. Molecular Ecology, 19: 2137-2151.

Chybicki, I.J., Burczyk, J. (2009). Simultaneous estimation of null alleles and inbreeding coefficients. *Journal of Heredity* **100***:* 106-113.

Chybicki, I.J., Burczyk, J. (2010). NM+: software implementing parentage-based models for estimating gene dispersal and mating patterns in plants. *Mol Ecol Resour* **10***:* 1071-1075.

Chybicki, I.J., Burczyk, J. (2010). Realized gene flow within mixed stands of *Quercus robur* L. and *Q. petraea* (Matt.) L. Revealed at the stage of naturally established seedling. *Molecular ecology* **19***:* 2137-2151.

Derory, J., Scotti-Saintagne, C., Bertocchi, E. et al. 2010. Contrasting relations between diversity of candidate genes and variation of bud burst in natural and segregating populations of European oaks. Heredity 105: 401- 411.

Do, C., Waples, R. S., Peel, D., Macbeth, G. M., Tillett, B. J. & Ovenden, J. R. (2014). NeEstimator V2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources* **14:** 209-214.

Dumolin, S., B. Demesure, and R. J. Petit (1995). Inheritance of chloroplast and mitochondrial genomes in pedunculate oak investigated with an efficient PCR method. *Theoretical and Applied Genetics* **91**:1253-1256.

EEA, 2010d. Assessing biodiversity in Europe: the 2010 report. European Environment Agency, Copenhagen.

England P.R., J.-M. Cornuet, P. Berthier, D. A. Tallmon, G.Luikart. Estimating effective population size from linkage disequilibrium: severe bias in small samples P.R. England, J.-M. Cornuet, P. Berthier, D. A. Tallmon, G.Luikart (2006). *Conservation Genetics* **7:** 303-308.

Gregorius H.-R. and Degen B. (1994): Estimating the extent of natural selection in seedlings from different Fagus sylvatica (L.) populations: Application of new measures. Journal of Heredity, 85, 183-190.

Gregorius, H.-R., Degen, B., 2006.: Monitoring genetischer Ressourcen - Prinzipien und Methoden. In: Begemann, F., Schröder, S., Wenkel, K.-O., Weigel, H.-J. (Hrsg.): Monitoring und Indikatoren der Agrobiodiversität. Schriftenreihe Agrobiodiversität des Informations- und Koordinationszentrums für Biologische Vielfalt (IBV), Band 27, 39-65, 2007.

Guevara, M. A., D. Chagne, M. H. Almeida, M. Byrne, C. Collada, J. M. Favre, L. Harvengt, S. Jeandroz, C. Orazio, C. Plomion, A. Ramboer, M. Rocheta, F. Sebastiani, A. Soto, G. G. Vendramin, and M. T. Cervera (2005). Isolation and characterization of nuclear microsatellite loci in Pinus pinaster Ait. *Molecular Ecology Notes* **5**: 57-59.

Guichoux, E., L. Lagache, S. Wagner, P. Leger, and R. J. Petit (2011). Two highly validated multiplexes (12-plex and 8-plex) for species delimitation and parentage analysis in oaks (Quercus spp.). *Molecular Ecology Resources* **11**: 578-585.

Hardy OJ, Vekemans X (2002). SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol Ecol Notes* **2**: 618-620.

Hill W.G. (1981). Estimation of effective population size from data on linkage disequilibrium. *Genetical Research* **38**: 209-216

Holliday, J. A., Ritland, K. and Aitken, S. N., 2010., Widespread, ecologically relevant genetic markers developed from association mapping of climate-related traits in Sitka spruce (Picea sitchensis). New Phytologist, 188: 501-514.

Jolivet C, Hoeltken AM, Liesebach H, Steiner W, Degen B (2011). Spatial genetic structure in wild cherry (Prunus avium L.): I. variation among natural populations of different density. Tree Genetics & Genomes 7(2): 271-283.

Konnert M, Maurer W, Degen B, Kaetzel R (2011). Genetic monitoring in forests - early warning and controlling system for ecosystemic changes. Iforest-Biogeosciences and Forestry 4: 77-81.

Loiselle B, Sork VL, Nason J, Graham C (1995). Spatial genetic structure of a tropical understory shrub, Psychotria officinalis (Rubiaceae). *American Journal of Botany* **82**: 14209–11425.

Namkoong, G.; Boyle, T.; Gregorius, H.-R.; Joly, H.; Savolainen, O.; Ratnam, W. & Young, A., 1996.: Testing criteria and indicators for assessing the sustainability of forest management: genetic criteria and indicators. CIFOR Working Paper 10, Center for International Forestry Research, Bogor (Indonesia), 15 pages; published under www.cifor.cgiar.org/Publications/DocumentDownloader.

Pinzauti, F., F. Sebastiani, K. B. Budde, B. Fady, S. C. Gonzalez-Martinez, and G. G. Vendramin. (2012). Nuclear microsatellites for Pinus pinea (Pinaceae), a genetically depauperate tree, and their transferability to p. halepensis. *American Journal of Botany* **99**: E362-E365.

Ritland, K. (2002). Extensions of models for the estimation of mating systems using n independent loci. *Heredity* **88**: 221–228.

Robinson J. D., G.R. Moyer (2012). Linkage disequilibrium and effective population size when generations overlap. *Evolutionary Applications* **6:** 290-301.

Schwartz, M.K.; Luikart, G. & Waples, R.S., 2006.: Genetic monitoring as a promising tool for conservation and management. Trends in Ecology and Evolution 22: 25-33.

Vekemans X, Hardy OJ (2004). New insights from fine-scale spatial genetic structure analyses in plant populations. *Mol Ecol* **13**: 921-935.