Origin and Genetic Variation Analysis of Austrian Northern Red Oak (*Quercus rubra*)

by

Daniel Erich Boehnke

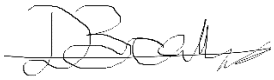
Thesis presented in fulfilment of the requirements for the degree of Master of Science in Mountain Forestry at the Institute of Silviculture, University of Natural Resources and Life Sciences (BOKU), Vienna

Supervisors: Prof Hubert Hasenauer, Dr Charalambos Neophytou

April 2020

Declaration

By submitting this thesis, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by BOKU will not infringe any third-party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.



Daniel Erich Boehnke April 2020

Abstract

The Quercus genus includes some important timber species. Northern red oak (*Quercus rubra*) belongs to the section Rubra (red oaks), confined to North America, and possesses a large natural distribution in this continent. In Europe, Northern red oak was introduced in the 17th century and is now one of the most common non-native tree species in the forest. However, very little is known about the origin and genetic variation of this species and this is also the case in Austria.

The objectives of this study are to determine the origin of ten Northern Red Oak (NRO) stands in Austria, as well as study the genetic structure among the stands in order to identify the seed sources from which they were established. The genetic diversity within the stands was also analysed and compared to native and non-native populations which may provide useful insight into making future decisions on seed transfer and breeding.

Samples were taken from 256 individuals over 10 populations in Burgenland, Austria, a region where the species has been frequently planted, and were subjected to a genetic analysis of 10 nuclear (biparentally inherited) microsatellite loci. Five chloroplast markers (maternally inherited) were used to identify the origins and ten nuclear markers were used to determine the genetic variation of the stands.

The results showed the prevalence of nine haplotypes of chloroplast DNA, namely, A, A1, A2, A3, B, C, E, F and O. Haplotype A was shown to represent a dominant three quarters of the study population and at least two lineages were revealed, possibly three. The fixation index (F_{ST}=0.034) of the nuclear SSRs was rather low, suggesting a high degree of sharing of genetic variation.

The intrapopulation haplotypic composition suggests that one of the plots was established with a seed source from the northern region of the native range. This was complimented by a cluster analysis which showed the plot was of a different cluster. These analyses placed another plot in a possibly separate cluster from the other eight.

An overwhelming frequency of haplotype A (73.7%) was found, followed by C (9.5%) and B (8.7%) and three unknown (not described in previous publication), rare haplotypes were also identified and named A1 (0.4%), A2 (2.7%) and A3 (0.4%). One population appears to be of a different seed source not found in Austria or Germany but in the native range and is of a separate lineage to the others.

Haplotype diversity within the Austrian populations was high (H_S=0.401) compared to that of the North American populations. German populations lie in between these two. Despite the low haplotypic diversity within the North American populations, the total haplotype diversity is higher than that of the Austrian populations (H_T=0.453).

In conclusion, results suggest that three different seed sources might have been used for establishment of the study stands. Two pieces of evidence (chloroplast and nuclear markers) suggest a separate seed source for at least two of the study plots. As reported by the forest owner, this population arose from native reproductive material from the northern part of the native range, which is in agreement with the results of the genetic analysis. On the other hand, eight populations displayed a similar haplotypic composition as populations from Southern Germany and low interpopulation genetic differentiation. This could be due to the use of seed sources – likely introduced (European) – which had been initially established with or originate from native material from a limited geographic area in the native range.

Acronyms

DNA	Deoxyribonucleic acid
cpDNA	Chloroplast DNA
nuDNA	Nuclear DNA
NRO	Northern Red Oak (*Quercus rubra)*
SSRs	Short sequence repeats (microsatellites)
cpSSRs	Chloroplast SSRs
nuSSRs	Nuclear SSRs
PCR	Polymerase chain reaction
MSN	Minimum spanning network
MST	Minimum spanning tree
AMOVA	Analysis of molecular variance
IBD	Identity by descent

Table of Contents

Table of Figures

Table of Tables

Chapter 1: Introduction

1.1. Background

Northern Red Oak (NRO) is a dominant hardwood species in the northeast of North America, ranging between 32° and 47° N latitude and 60° and 96° W longitude, as can be seen in Figure 1. The species is able to cope with a large variation in mean temperatures (4°C-15°C), annual rainfall (600-2000 mm) as well as soil conditions (Magni, et al., 2005).

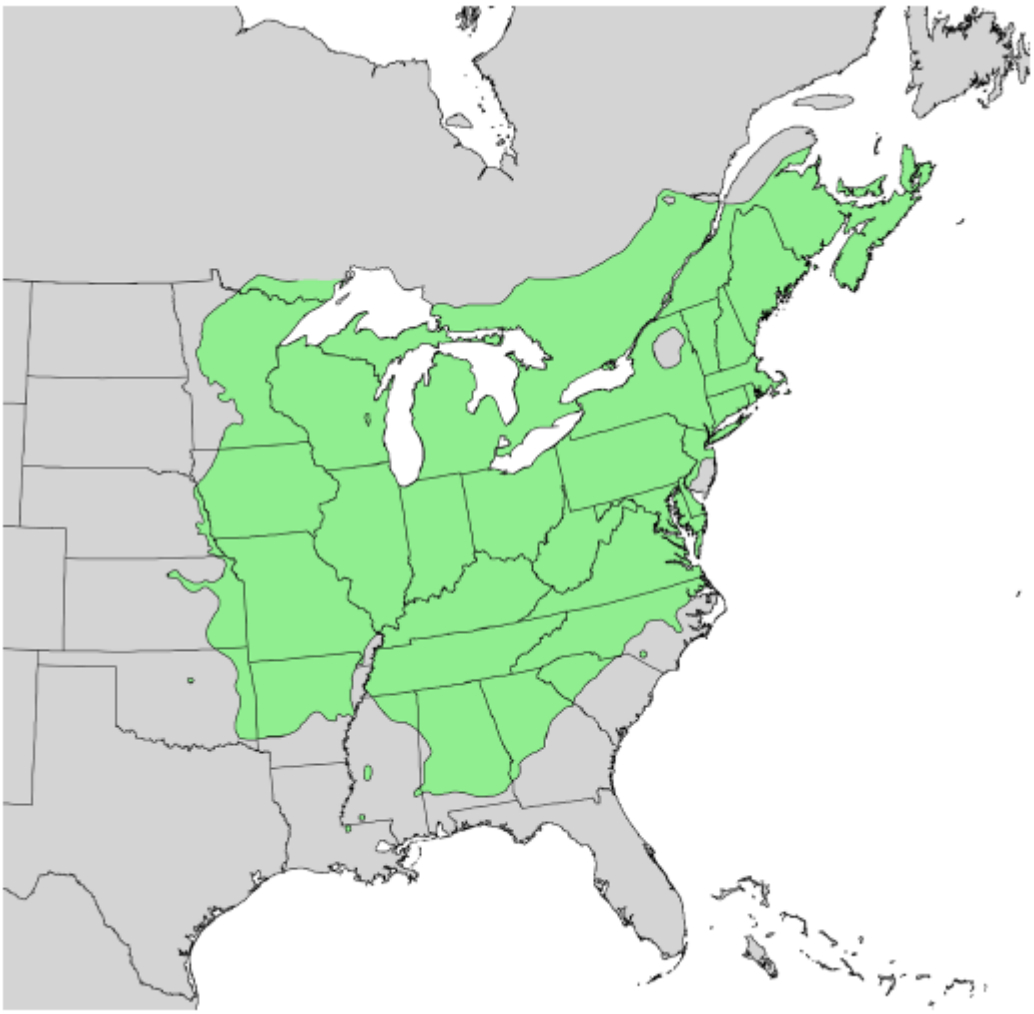


Figure 1) Native distribution range of northern red oak (United States Geological Survey)

NRO was introduced to Europe in 1691. Understandably, the entire range could not be harvested for seed and thus it is assumed that the introduced populations contained only a fraction of the genetic information relative to the native range (Barrett & Husband, 1990). Initially intended to be an ornamental tree, it eventually became an important species for wood production in Central European countries like France and Germany (Magni Diaz, 2004). This is predominantly due to faster growth, which shortens the required rotation period to 80-120 years. Compared to the native oak species' rotation periods of over 140 years, as well as lower water and nutrient requirements, NRO became a sought-after production species (Nagel, 2015).

The commercial hardwood market in western and central Europe is still contributed to significantly by NRO. The afforestation over the last 300 years would have to have been introduced from reproductive material, the origin of which is still largely unknown. As its natural regeneration abilities are relatively efficient, it is quite likely that through the 19th and 20th centuries the source of propagation was simply the few existing stands dating back two or three hundred years (Magni Diaz, 2004). Alternatively, reproductive material from native sources has been introduced on multiple occasions, resulting in a complex ancestry among NRO forests in Europe consisting of a genetic mixture of both first and advanced generation stands (Magni, et al., 2005).

In Austria, northern red oak is less common than in other Central or Western European countries. It can be mainly found in Burgenland, where it has been sporadically planted in several districts (Forstwesen, 2012). These European stands have been used as an abundant and inexpensive source of seed for afforestation throughout the 19th and 20th centuries and it is likely that the mixing of seed lots may have been practised (Magni Diaz, 2004).

Since these stands already serve as source for afforestation, it is of interest to investigate their origin and genetic variation. However, little is known about the origin of these stands.

1.2. Research objectives

1. The genetic variation among NRO stands in Austria would be determined for our sample. This could also provide further evidence about a different origin and/or admixture of different seed sources. This can be achieved by selecting common sets of loci as a standard between this and previous studies, which allows one to trace a lineage by comparing common chloroplast haplotypes with native and non-native reference samples (Pettenkorfer, et al., 2019).

2. To study genetic structure among the study stands and compare it to the distribution of chloroplast haplotypes in order to identify different seed sources used for forest stand establishment.

3. To determine the genetic diversity within the sampled stands and compare it to other, native or non-native, populations. Such data may provide useful information for future decisions on seed transfer and breeding.

Chapter 2: Literature review

The following sections address previous literature on NRO in its native range, Europe and the differences between populations.

2.1. *Quercus Rubra* L.: What we know

The genus *Quercus* falls within the Fagaceae family which includes other ecologically and economically significant taxa such as *Castanea* (chestnut) and *Fagus* (beech) (Aldrich, et al., 2003). The NRO belongs to the section Lobatae (red oaks) which is a clade confined to North America and is reproductively isolated from the section Quercus (white oaks), which includes some native European species such as *Q. robur, Q. pubescens* and *Q. patrea*, among others (Denk, 2017).

Hybridisation is possible within the section Lobatae, including but not limited to species such as *Q. palustris, Q. falcata* and *Q. ellipsoidalis,* which has frequently been observed in the native range (Aldrich *et al.*, 2003; Lind and Gailing, 2013; Zhang, Hipp and Gailing, 2015). The result of this in many cases has been introgressive hybridisation, where interspecific hybrids repeatedly backcross with one of their parent species. Also known as genetic introgression, this is sometimes an important source of genetic variation (Aldrich, et al., 2003). There is evidence of this phenomenon in our own human DNA, with strong leads pointing to the introgression of Neanderthal genes (Juric, et al., 2016). However, introgression may also have disadvantages due to negative epistatic interactions (Arnold & Bennett, 1993).

NRO does not exhibit any strong phylogeographic structure across its range, unlike the European white oaks. Molecular markers from the chloroplast DNA have been previously used to investigate phylogeographic structure in the native range (Magni, et al., 2005; Pettenkorfer, et al., 2019). Due to maternal inheritance, the distribution of chloroplast DNA takes place only through seeds. Lacking recombination, chloroplast haplotypes remain unchanged for many generations. Thus, they are suitable for tracking past migration. In the case of European white oaks, lineages of chloroplast haplotypes correspond to glacial refugia around the Mediterranean Sea, whereas their distribution reflect post-glacial migration pathways (Petit et al. 2002, Forest Ecology and Management 156: 5-26).However, refugial populations of NRO were located further to the north and were spatially less isolated which limited genetic differentiation among the various refugia (Magni, et al., 2005).

The spatial pattern of genetic differentiation is continuous and expected of a wind-pollinated species. This has been confirmed via the use of nuclear molecular markers inherited from both parents and, unlike chloroplast DNA, depend both on seed and pollen dispersal. Borkowski et al. (2017) used ten highly polymorphic microsatellites to detect a well-defined genetic cluster in the north western part of the native range, with two additional clusters mainly distributed in the north east and south. A large transition zone was observed between the clusters. Merceron et al. (2017) conducted a wide-range study using 69 single nucleotide polymorphism (SNP) markers which a gradual genetic differentiation of northern from southern populations was discovered.

However, there are some characteristics in which location within the native range differ ever so slightly. Evidence from provenance trials has been provided which shows a degree of ecotypic variation of quantitative traits:

- Leaf senescence and autumnal colouration take place earlier in northern provenances (Deneke 1974).
- A weaker correlation of bud flushing was observed with northern provenances where flushing generally occurred earlier (Kriebel et al. 1976).
- Cold hardiness is positively correlated with latitude (Kriebel et al. 1976).
- Height and diameter growth also varied geographically (Kriebel et al. 1976).
- Drought resistance increased in areas west of Mississippi exhibiting an adaptation to dry summers (Deneke 1974).

2.2. *Quercus rubra* L. in Europe

Some very extensive studies on seed origin have been performed in France and Germany. These investigations were based on chloroplast DNA (cpDNA) markers which revealed a higher than expected inter-stand haplotypic diversity. This may indicate either a larger founder population than previously thought or a higher degree of genetic variation among the native population from which the founder population material was sourced. However, the still low variation between regions suggests that intermixing has indeed occurred (Pettenkorfer, et al., 2019). Additionally, haplotypes from the southern part of the natural distribution seem to be lacking in Europe, suggesting the founder population would have been from the northern part of the natural distribution (Merceron *et al.*, 2017; Pettenkofer *et al.*, 2019).

Studies based on isozymes have indicated that European populations consist of an even higher genetic diversity due to an increase in the frequencies of rare alleles (Daubree and Kremer, 1993). This was explained by the different selection pressures in the introduced population relative to the native range. Additional research with local or regional scope showed weak or even lacking isolation by distance among populations, as well as lower genetic variation of quantitative traits (Daubree and Kremer, 1993; Merceron, 2016).

Some evidence pointing towards adaptation of NRO has been provided by studies in quantitative genetics. It has shown that European populations exhibited distinct phenological traits compared to that of the native population. Examples of these traits are superior height and/or diameter growth and earlier budburst. Interestingly, it was also shown that southern European provenances exhibit earlier budburst which is contrast to the trend observed in the native range (Merceron, 2017).

Chapter 3: Materials and Methods

3.1. Study Plots

The plots were located in Burgenland, Austria and were spread across 10 stands where NRO dominated. 20-30 individuals were sampled per plot. The plots run from Deutschkreutz approximately 80km south to Moschendorf. See Figure 2.

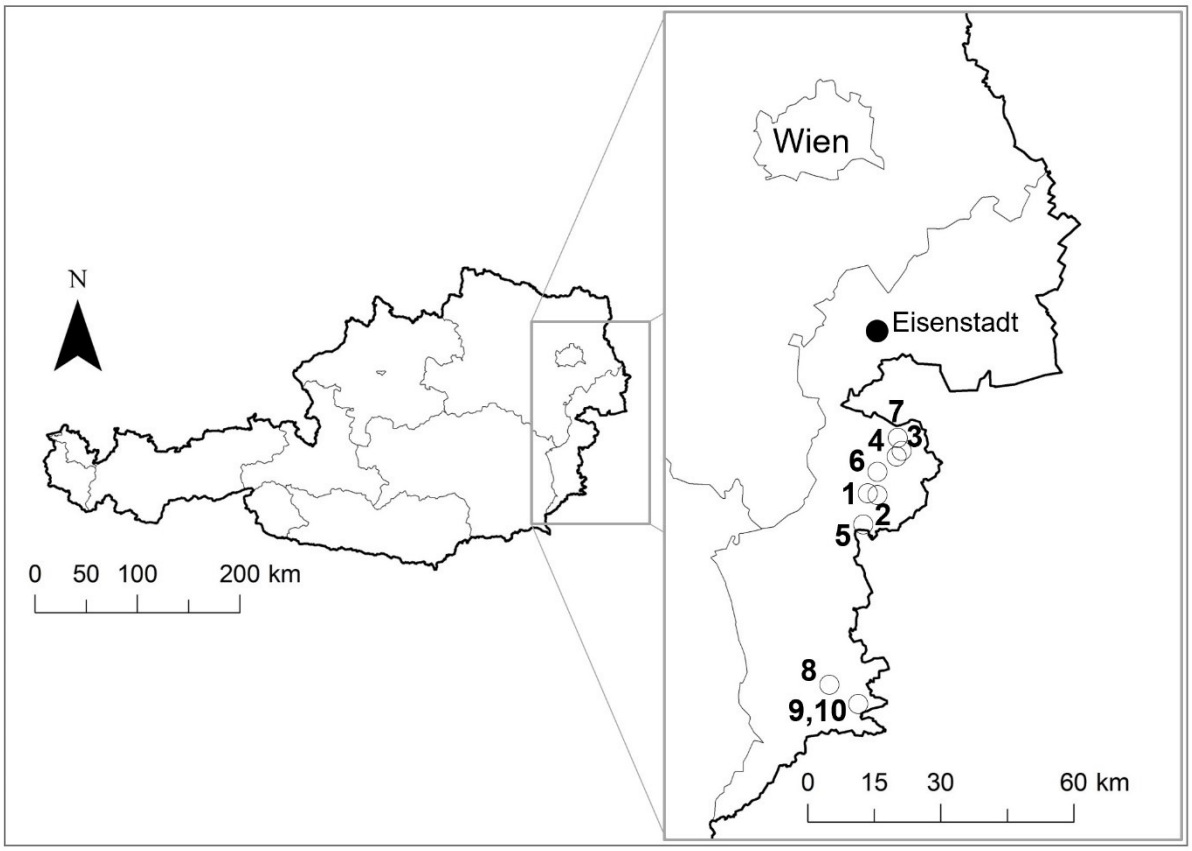


Figure 2) Plot locations. For plot names see Table 1.

The plots were selected based on our knowledge of the presence of NRO as well as that of the managing foresters whom granted us permission to sample the plots. Plots are either referred to by number or code for the rest of this document. The plot locations, code and range of individuals within them are as follows in Table 1:

Table 1) Plot summary

Plot number	Location	Location code	Sample size
1	Unterpullendorf	UN	20
2	Unterpullendorf	UN	20
3	Horitschon	HO	20
4	Horitschon	HO	20
5	Mannersdorf	MA	30
6	Langental	LA	30
7	Girm	GI	30
8	St. Michael	SM	30
9	Moschendorf	MO	30
10	Moschendorf	MO	30

3.2. Sampling Method

Once in the plot, trees were selected randomly, with the intent of achieving a representative sample of individuals covering the whole stand. NRO is rather adept at coppicing and this may lead to two distinct stems appearing to be separate individuals in close proximity. However, these individuals may be genetically identical and it is thus pointless to sample and analyse both. Therefore, sampled trees were at least 15 m apart from each other. Age was a consideration because a mother and her offspring would have near-identical chloroplast DNA and would therefore not contribute any significant data towards the origin tracing component of the study (Weising & Gardner, 1999). Each individual sampled was marked on a GPS for the purpose of creating a genotype map as well as standard record keeping for future reference.

Four populations were sampled in November 2017 and the tissue was dehydrated and stored. Twenty individuals from each population were sampled. The other 6 populations were sampled in April 2019, where 30 individuals of each were taken and similarly stored for analysis the following week.

The tissue sampled were, in most cases, buds. These buds yield a more concentrated and purer amount of DNA relative to cambial or leaf samples. A silky saw (polesaw) was used to prune live branches containing buds, which were removed and stored in plastic bags containing silica gel as a dehydration medium to prevent any mould or bacteria from growing on and contaminating the samples. The dried buds are arguably easier to work with during DNA isolation.

3.3. DNA Isolation

The protocol followed for the DNA isolation was one by QIAGEN's DNeasy® Plant Handbook. See Appendix A. DNA was extracted from 15-20 mg of bud or leaf material. The outer coating of the buds was removed and the inner tissue was then weighed and transferred to epi tubes. This amounted to 2-3 buds per sample. In the case of leaf material, 15-20 mg of tissue was cut from the intervenary space.

Changes to the official protocol included placing two tungsten beads in the epi tubes for pulverisation, as well as running the samples in the TissueLyser pulveriser for 3 minutes, swapping the plates around and running them again for another 3 minutes. The original protocol advises the use of one tungsten bead for two cycles of 30 seconds. Ideally the tissue samples should be pulverised into a fine, uniform powder. As DNA consists of very long molecular chains, too much pulverisation can destroy it. We were careful not to do so.

After the disruption step we put the samples, still in the pulverising plates, in a Hettich ROTANTA 460R centrifuge at 3000 rpm for 1 minute. This brought all the pulverised material down to the bottom of the epi tube. The rest of the centrifugations were carried out on a Eppendorf 5424 R.

Towards the end of the isolation procedure, only 50 µl of elusion buffer was used to extract the DNA from the capturing membrane in the tube, instead of 100 µl as advised by the official protocol, as the DNA concentrations of previous tests using the latter volume were rather low.

Each DNA sample was tested on a 1.5% agarose gel electrophoresis and the gels were scanned and captured on a BIO RAD Universal Hood 2. DNA concentration and purity of 100 randomly selected samples were tested on a NanoDrop machine. Once satisfied with the extract, master plates were made, diluting over-concentrated samples individually to achieve a uniform dilution of approximately 20ng/ml.

The end result was 3 plates of 96 wells each. As can be seen in Table 2, Table 3Table 4, the first plate consisted of only the first batch of samples taken in 2017, populations 1 to 4. The second plate consisted of populations 5 to 7 as well as 11 reference samples. Finally, the third plate consisted of populations 8-10 plus 5 samples from population 7. The last well (H12) was filled with deionised water as a means to prevent air from entering the capillary during sequencing. Well H12 would also serve as a negative control which is helpful in identifying problematic primers and/or sequencer runs.

Table 2) Master Plate 1

MP1	1	2	3	4	5	6	7	8	9	10	11	12
A	101	102	103	104	105	106	107	108	109	110		
B	113	114	115	116	117	18	119	120	201	202		
C	205	206	207	208	209	210	211	212	213	214		
D	217	218	219	220	301	302	303	304	305	306		
E	309	310	311	312	313	314	315	316	317	318		
F	401	402	403	404	405	406	407	408	409	410		
G	413	414	415	416	417	418	419	420	411	412		
H	111	112	203	204	215	216	307	308	319	320		

Table 3) Master Plate 2 with reference samples

MP2	1	2	3	4	5	6	7	8	9	10	11	12
A	501	509	517	525	603	611	619	627	705	713	721	D
B	502	510	518	526	604	612	620	628	706	714	722	E
C	503	511	519	527	605	613	621	629	707	715	723	F
D	504	512	520	528	606	614	622	630	708	716	724	G
E	505	513	521	529	607	615	623	701	709	717	725	H
F	506	514	522	530	608	616	624	702	710	718	A	i
G	507	515	523	601	609	617	625	703	711	719	B	M
H	508	516	524	602	610	618	626	704	712	720	C	O

Table 4) Master Plate 3

MP3	1	2	3	4	5	6	7	8	9	10	11	12
A	726	727	728	729	730	801	802	803	804	805	806	807
B	808	809	810	811	812	813	814	815	816	817	818	819
C	820	821	822	823	824	825	826	827	828	829	830	901
D	902	903	904	905	906	907	908	909	910	911	912	913
E	914	915	916	917	918	919	920	921	922	923	924	925
F	926	927	928	929	930	1001	1002	1003	1004	1005	1006	1007
G	1008	1009	1010	1011	1012	1013	1014	1015	1016	1017	1018	1019
H	1020	1021	1022	1023	1024	1025	1026	1027	1028	1029	1030	H_2O

3.4. Locus Amplification

Loci were amplified based on the polymerase chain reaction (PCR) protocol by Qiagen with the 15 SSR primers. Primers for 5 chloroplast SSR loci were used as well as 10 nuclear. The first step was to rehydrate the primers and dilute them individually to a concentration specified by the manufacturer. Then the optimal PCR programme had to be determined. In order to do this a

simplex gradient PCR was run on each primer with 4 repetitions using the same genomic DNA samples, namely, 101, 205, 308 and 401. The PCR machine used for the gradient PCR runs was a biometra Thermocycler.

PCR is a replication process which uses cycles of varying temperatures to amplify segments of DNA specified by the primer. The second step in a typical PCR run repeats many times. Each repetition creates a copy of the total pool of selected segments (x), theoretically resulting in 2^x copies of the selected sequence. A typical number of cycles is 28, which should increase the concentration of the target segments to 2^{28} relative to the rest of the rest of the DNA in the solution.

Reference samples, acquired through Goettingen University, were run with our samples (see Table 3). This served as a means to unambiguously assign already known haplotypes, identifying any shift or reading errors during allele scoring. These samples made a trustworthy reference as they formed part of a recently published study of similar nature (Pettenkorfer, et al., 2019).

3.4.1. Chloroplast loci

Chloroplast DNA (cpDNA) is rather well conserved in all plants. cpDNA is inherited maternally in angiosperms, under which the *Quercus* genus falls. Due to the lack of chromosomal recombination between mother and father, cpDNA has very little variation over generations and is only subject to slight mutations which may occur during mitosis. This sequence conservation stores invaluable information about the origin of individuals (Weising & Gardner, 1999).

The primers selected are shown in Table 5.

Table 5) Chloroplast Primers

Locus	Length min	Length max	No. of alleles	Reference 1 (primer notes)	Further references
µcd4	97	99	3	Deguilloux et al. 2003	Zhang et al. 2015, Pettenkofer et al. 2019
µdt1	84	87	4		
µdt4	145	146	2		
ccmp2	226	228	3	Weising & Garnder, 1996	
ccmp4	115	118	4		

Pettenkorfer et al., (2019) and Zhang et al., (2015) collectively used these exact primers in a similar study describing variation patterns in native and introduced populations, the former being recently published in the European Journal of Forest Research.

As mentioned above, the primers were all subject to a gradient PCR run to find their optimal annealing temperatures. Based on the annealing temperatures provided by Biomers.net, which can be seen in Appendix B, we selected the range of this gradient PCR run for 48.0-60.0°C. As the intention was to multiplex them, a high product yield in the same temperature range was ideal. It seemed as though this ideal temperature was 52.0°C.

The chloroplast primers were, at first, mixed together with each DNA sample in a multiplex and run in the PCR machine at 52.0°C. This was called combination 1 or “kombi 1” (K1), as can be seen in Table 6.

Table 6) PCR components K1

Component	Volume per reaction (µl)	Total volume + 10% (µl)
QIAGEN Master Mix	5.0	530.0
Rnase-free H_2O	3.0	318.0
Primer mix	1.0	106.0
Genomic DNA	1.0	96.0 (no reserve)

The "Primer mix" component consists of a dilution of the forward and reverse primers included in the PCR. The forward primer contained the dye. They were mixed according to the Qaigen PCR protocol, shown in Table 7.

Table 7) PCR components K1 - primer mix

Primer	Forward (µl)	Reverse (µl)	Rnase-free H_2O (µl)	Total (µl)
µcd4	4.2	4.2	64.0	106.0
µdt1	4.2	4.2		
µdt4	4.2	4.2		
ccmp2	4.2	4.2		
ccmp4	4.2	4.2		

A typical PCR run consists of three steps. The first being a denaturation of the DNA, more specifically, "melting" the double helix in order to separate it into two single strands. 95°C is enough to break the hydrogen bonds holding the strands together without destroying the information each stand holds. The second step, as mentioned above, copies the strands many times over. The third step is a final extension which is used to map the 5' ends of the strands. Our optimal PCR run for K1 can be seen in Table 8 and visually in Figure 3.

Table 8) Chloroplast primer PCR program K1

Step	Activity	Temperature (°C)	Time (min)	No. of cycles
1	Initial denaturation	95.0	5:00	1
2	Denaturation	95.0	0:30	28
	Primer annealing	52.0	1:30	
	Extension	72.0	0:30	
3	Final extension	60.0	30:00	1
	Hold	4.0	∞	1

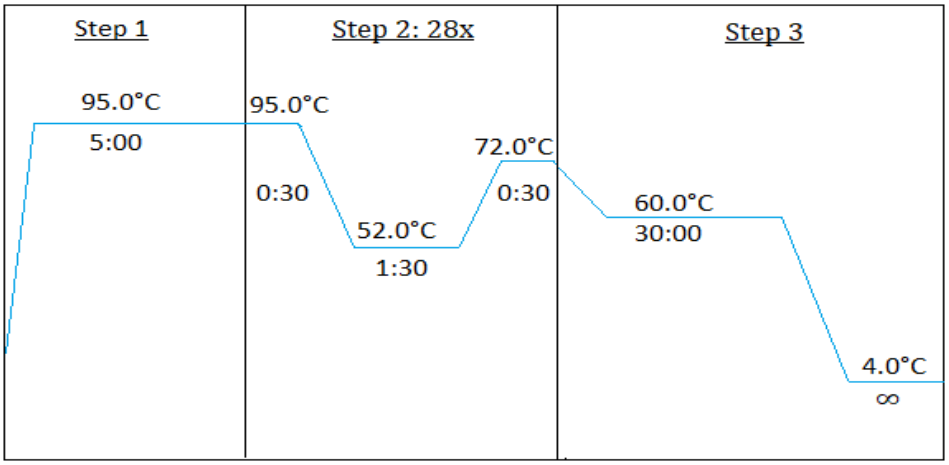


Figure 3) PCR program K1

However, there seemed to be an inhibiting interaction between the ccmp4 primer and the others. It was decided it would be best to run each primer on its own as a simplex with the same PCR program and components, only with a more diluted primer mix. Table 9 displays the simplex run five times, each time with a different chloroplast marker.

Table 9) Simplex PCR components - K1

Primer	Forward (µl)	Reverse (µl)	Rnase-free H_2O	Total (µl)
cp Primer	4.2	4.2	97.6	106.0

Once complete, the PCR products were tested in an agarose gel electrophoresis and a suitable dilution ratio was determined for the plates which would be sent for capillary electrophoresis. The PCR products were then mixed according to this dilution ratio, which can be seen in Table 10.

The gel electrophoresis was conducted at 140 volts for 20 minutes in a 1.5% agarose gel (with gel red) with a 1X TBE buffer solution.

Table 10) Chloroplast dilution plate K1

Primer	Concentration (µl per 120 µl)
µcd4	1
µdt1	3
µdt4	2
ccmp2	2
ccmp4	3

1 µl of the mixture in the dilution plate was then transferred to another plate containing 10 µl HiDi and 0.26 µl ROX standard per well. The dilution plate was sealed and stored at -80°C along with the DNA master plates.

3.4.2. Nuclear loci

Unlike cpDNA, nuDNA is indeed subject to chromosomal recombination. This results in a mixture of paternal and maternal DNA with a much higher degree of variation. The nuclear primers in Table 11 were selected based on two criteria. Their prominence in a multitude of

previous publications which may help standardise results from said publications was of utmost importance. Fragment length was the second deciding factor as primers which overlap require different colour dyes, which is explained in the description of the analysis of this paper.

Table 11) Nuclear primers

Locus	Length min	Length max	No. of alleles	Reference 1	Reference 2
QpZAG15	103	148	15	Gailing et al., 2012	Borkowski et al., 2017
PIE099	178	202	12	Konar et al., 2017	Gailing et al., 2012
GOT009	221	249	10	Collins et al., 2015	Gailing et al., 2012
WAG065	268	280	-	Durand et al., 2010	Konar et al., 2017
FIR039	111	132	5	Collins et al., 2015	Durand et al., 2010
quru-GA-1F02	166	184	18	Aldrich et al.,2003	Konar et al., 2017
quru-GA-1F07	306	348	23	Aldrich et al.,2003	Borkowski et al., 2017
GOT021	95	101	2	Collins et al., 2015	Gailing et al., 2012
FIR053	136	150	5	Collins et al., 2015	Durand et al., 2010
GOT004	264	294	6	Collins et al., 2015	Gailing et al., 2012

The nuclear primers were split into two groups based on their fragment length and annealing temperatures, as overlaps are not easily distinguished during scoring. Table 12 shows K2 with each primer and corresponding dye.

Table 12) Combination 2 (K2)

Primer	Dye
GOT009	FAM
WAG065	FAM
FIR039	HEX
quru-GA-1F02	HEX
quru-GA-1F07	HEX

The PCR components were mixed with the same ratios as the chloroplast primers but a different PCR program was run. The results of the gradient PCR suggested that a touchdown PCR of 10 and 25 cycles at 58.0°C and 52.0°C respectively would be best. During a touchdown PCR run the chance of non-specific binding can be reduced by sequentially decreasing the annealing temperature during each cycle. This reduces the frequency of unwanted amplification which can create noise when analysing the product (Korbie and Mattick, 2008). This program was used for both K2 and K3. The main components as well can the primer mix can be seen in TablesTable 13 andTable 14.

Table 13) PCR components K2

Component	Volume per reaction (µl)	Total volume + 10% (µl)
QIAGEN Master Mix	5.0	530.0
Rnase-free H_2O	3.0	318.0
Primer mix	1.0	106.0
Genomic DNA	1.0	96.0 (no reserve)

Table 14) PCR components K2 - primer mix

Primer	Forward (µl)	Reverse (µl)	Rnase-free H_2O	Total (µl)
GOT009	4.2	4.2	64.0	106.0
WAG065	4.2	4.2		
FIR039	4.2	4.2		
quru-GA-1F02	4.2	4.2		
quru-GA-1F07	4.2	4.2		

Combination 3 followed the same PCR procedure. The segment lengths were considered and dyes were appropriately requested (Table 15). The same component concentrations were used, seen in TablesTable 16 andTable 17.

Table 15) Combination 3 (K3)

Primer	Dye
QpZAG15	FAM
PIE099	FAM
GOT021	Atto550
FIR053	Atto550
GOT004	Atto550

Table 16) PCR components K3

Component	Volume per reaction (µl)	Total volume + 10% (µl)
QIAGEN Master Mix	5.0	530.0
Rnase-free H_2O	3.0	318.0
Primer mix	1.0	106.0
Genomic DNA	1.0	96.0 (no reserve)

Table 17) PCR components K3 - primer mix

Primer	Forward (µl)	Reverse (µl)	Rnase-free H_2O	Total (µl)
QpZAG15	4.2	4.2	64.0	106.0
PIE099	4.2	4.2		
GOT021	4.2	4.2		
FIR053	4.2	4.2		
GOT004	4.2	4.2		

As mentioned above, a different PCR program was used for K2 & K3. The touchdown PCR program is interpreted in table format in Table 18 and visual format in Figure 4.

Table 18) Nuclear primer touchdown PCR program

Step	Temperature (°C)	Time (min)	No. of cycles
Initial denaturation	95.0	5:00	1
Denaturation	95.0	0:30	10
Primer annealing	58.0	1:30	
Extension	72.0	0:30	
Denaturation	95.0	0:30	25
Primer annealing	52.0	1:30	
Extension	72.0	0:30	
Final extension	60.0	30:00	1
Hold	4.0	∞	1

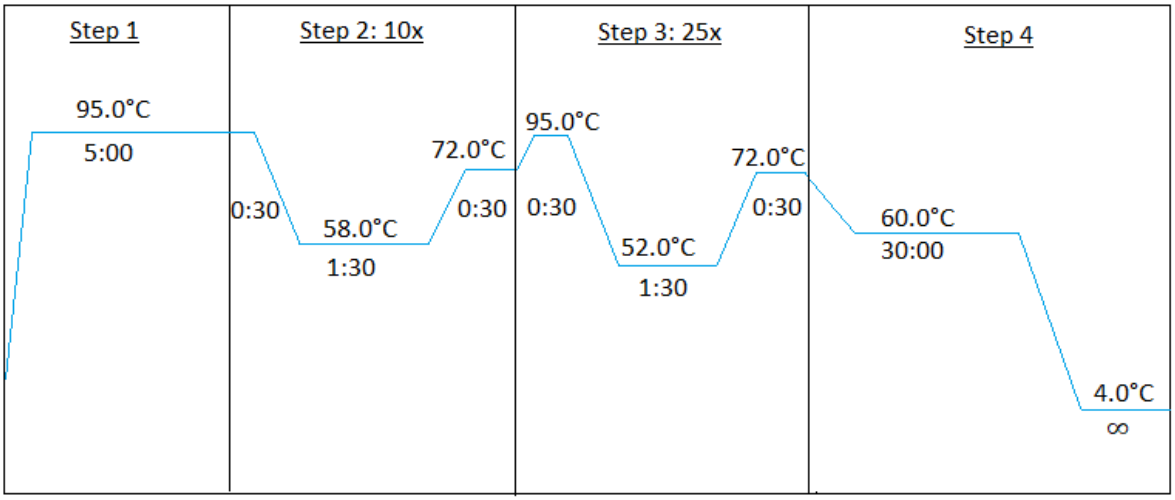


Figure 4) PCR program K2 and K3

Once complete, the PCR products were tested via gel electrophoresis and transferred to a new plate with HiDi and ROX standard and sent to be sequenced.

3.4.3. Allele scoring

Sequencer data was then imported into GeneMapper 5 software by Thermofischer. This software isolates the specified markers, which would be in abundance due to PCR replication, by portraying a graph showing base pair length on the x-axis and peak intensity on the y.

In order to distinguish between primers in a multiplex which overlap in base pair length, the forward primer of each marker contained a dye, specified by us, which is identifiable by GeneMapper.

Scoring of the allele peaks may vary in difficulty. In some cases, it may be clear enough for the GeneMapper software to correctly identify the peaks, as is the case in Figure 5. Sometimes a correction on the part of the user is required if stutter bands are misidentified as peaks, as is the case in Figure 6. On rare occasions the data may be illegible. This could be due to a contamination of the sample or a very high or low concentration of PCR product. Figure 7 is one such example and would be repeated if too many samples show such poor peaks.

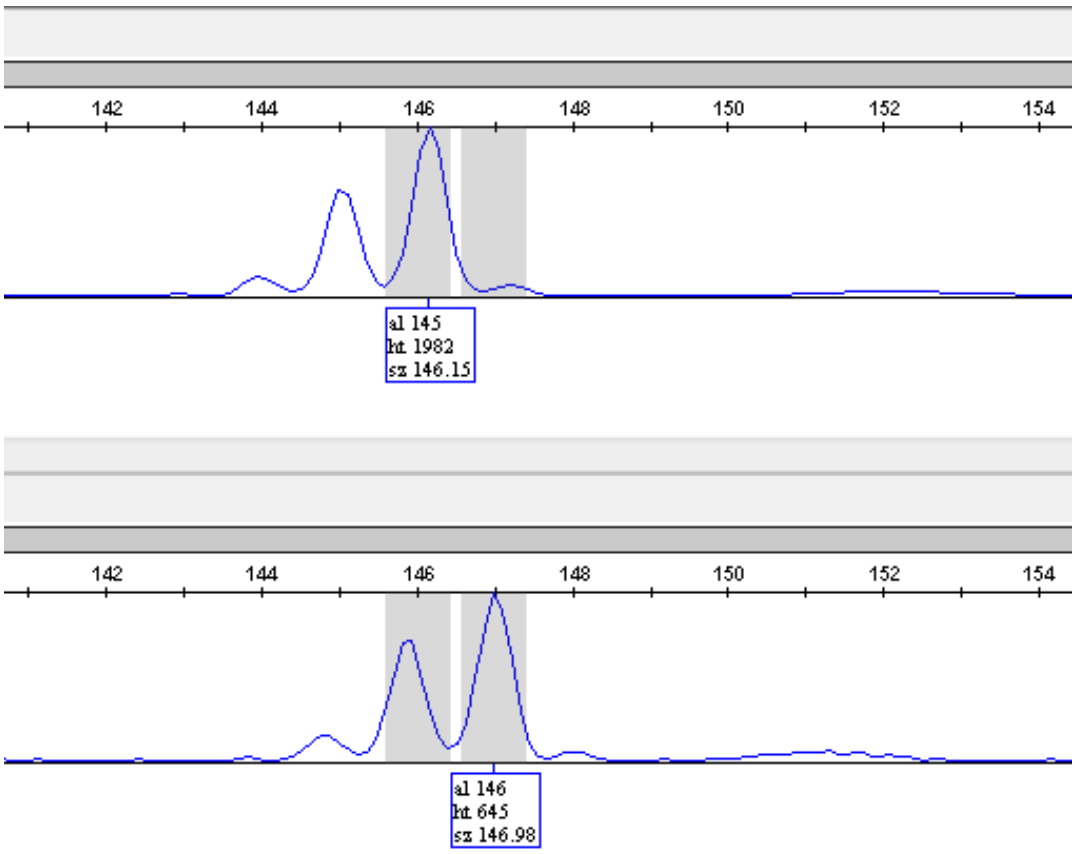


Figure 5) µcd4 - Electropherogram visualized with GeneMapper

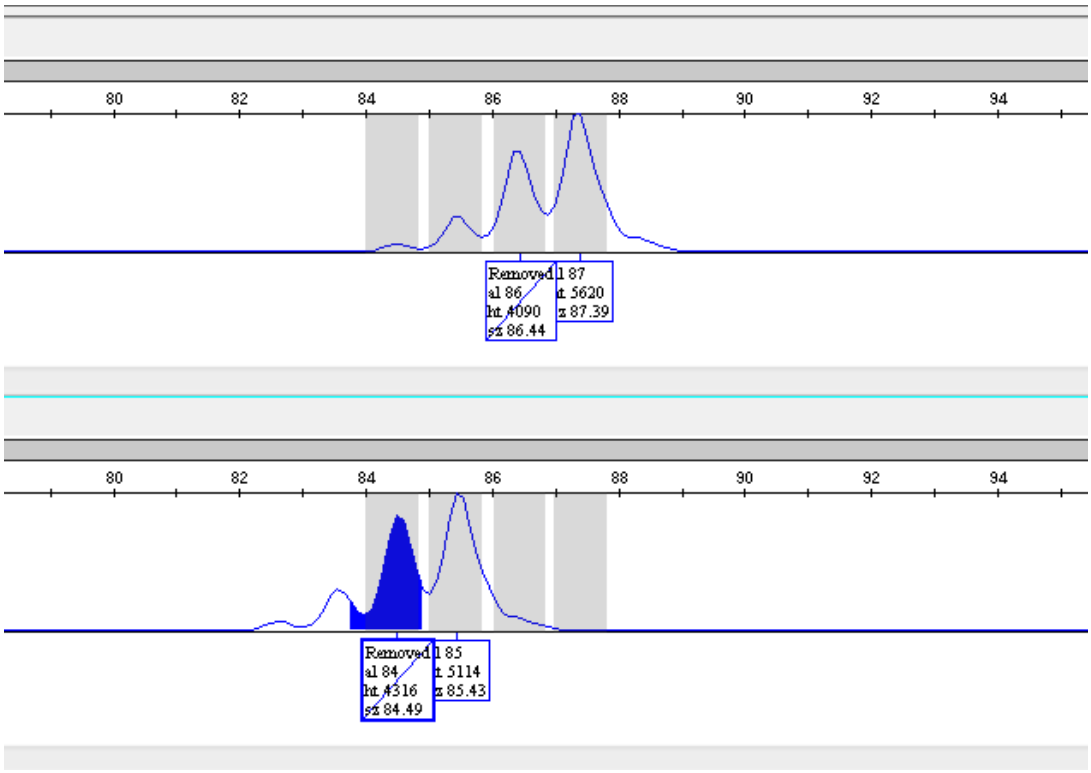


Figure 6) µcd4 - Electropherogram visualized with GeneMapper. Manual corrections were undertaken in order to remove misidentified peaks

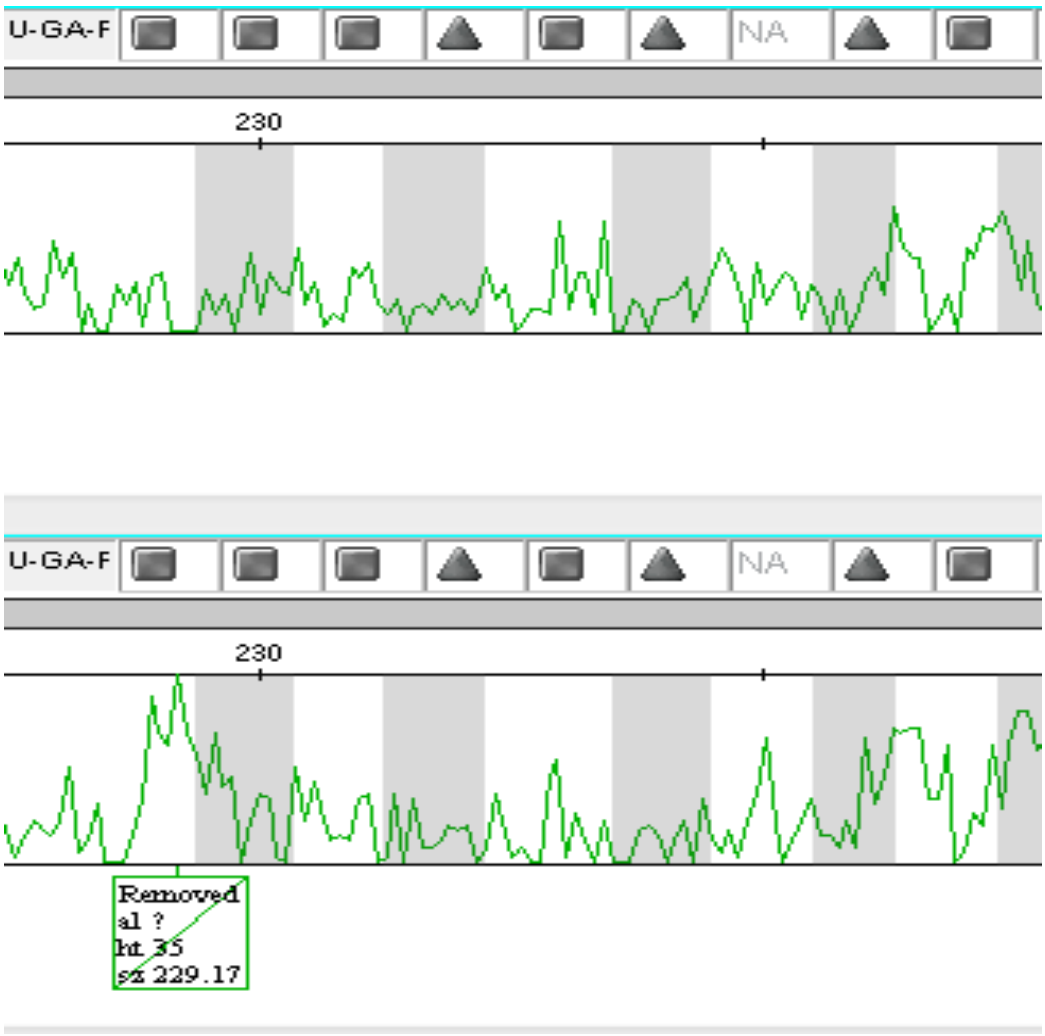


Figure 7) Nuclear microsatellite locus quru-GA-1F07 – Electropherogram visualised in GeneMapper. An example of PCR failure. Peaks were manually deleted.

As can be seen in the figures, the allele boxes are assigned an allele number, which refers to a rounded or calibrated fragment size. The actual fragment size is recorded below.

The alleles were successfully scored and the data exported into a spreadsheet format.

3.5. Population Genetic Analysis

3.5.1. Chloroplast haplotypes and phylogenetic relationships

A haplotype is a group of alleles that are inherited from a single parent. In the case of chloroplast DNA, this group of alleles are inherited together in a cluster of tightly linked genes on a chromosome which means they are likely to be conserved as a specific sequence over many reproductive generations. Therefore, relationships between haplotypes may provide valuable information on population genetics.

In order to place each sample within a haplotype the location of the allele associated with each marker is a significant identification factor. This is expressed as a calibrated fragment size in base pair length (bp). For the sake of compatibility of results between studies, the same scale may be used by different authors. In this case we used the same regime as

Pettenkorfer et al. (2019) and Zhang et al. (2015) as they collectively covered the chloroplast primers we used.

These relationships were calculated and visualised using a series of software. As mentioned before, GeneMapper produces a numerical export spreadsheet. This sheet can then be processed by Arlequin 3.5 which uses a range of methods to interpret the data (Excoffier, et al., 2005).

Phylogenetic relationships among the different haplotypes were resolved by constructing a minimum spanning network (MSN). This was based upon pairwise distances between microsatellite haplotypes. Assuming a stepwise mutation model, we used the sum of squared allele length difference between two of them using the following equation developed by Slaktin (1995) applied by the software Arlequin:

$$\hat{d}_{xy} = \sum_{i=l}^{L} (a_{xi} - a_{yi})^2$$

Besides the matrix of pairwise distances, a so-called nexus file describing the phylogenetic relationships among the haplotypes was produced and fed into the software FigTree in order to construct a Minimum Spanning Tree (MST). Finally, alternative connections provided by Arlequin were used to extend the Tree (MST) into a network (MSN) of haplotypes. This final step of visualisation was done by hand.

3.5.2. Genetic (haplotypic) variation based on chloroplast microsatellites

In order to gain an overview of the intra- and interspecific variation of chloroplast haplotypes, frequencies and pie charts per population were constructed and put on a map.

In addition, an Analysis of Molecular Variance (AMOVA) was carried out in order to partition haplotypic variation within and among populations (Excoffier, et al., 1992). The “fixation index” (F_{ST}) by Wright (1965) was calculated as a measure of genetic variation among populations (calculation according to Weir & Cockerham 1984). This refers to the variance of allele frequencies between populations as well as the probability of identity by descent, which suggest a common ancestor between two individuals. F_{ST} is a value between zero and one. Zero means that there is complete sharing of genetic material between all the populations (i.e. the same haplotypic frequencies can be observed in each one of the compared populations) whereas a value of one would indicate no sharing whatsoever (each population is fixed for a different allele). The significance of inter-population differentiation (F_{ST}) was tested by performing 1000 random permutations of haplotypes among populations.

Subsequently, measures of genetic diversity and differentiation within and among populations were computed. In particular, following measures were calculated using the software developed based on Pons and Petit (1996) and Burban et al. (1999).

H_T as a measure of the total genetic diversity across all samples (pooled populations). It is defined as:

$H_T = 1 - \sum p_i^2$ where pi is the frequency of the i-th haplotype in the sample.

H_S as a measure of the genetic diversity within populations. It is the weighted average of individual diversity values calculated by means of the aforementioned formula for H_T within each of the populations.

G_{ST}, an extension of F_{ST} for loci with multiple states of alleles. It analyses allele frequency variation among subpopulation in terms of haplotypic diversity. It serves as a measure of genetic differentiation without taking mutational steps into account.

R_{ST} is an analogue of G_{ST} and F_{ST}, assuming a stepwise mutation model (Slatkin, 1995). It is equivalent to the fraction of the total variance in allele size in terms of repeat units. That is, it does not only depend on haplotype frequency differences among populations, but also on the phylogenetic differences among them, which increase with increasing allele length differences between them.

Finally, a test for phylogeographic structure was carried out by comparing R_{ST} with G_{ST} also using the cpSSR software. As mentioned above, R_{ST} increases if phylogenetically different haplotypes prevail within different populations. On the contrary, if phylogenetic differences among populations are random, then R_{ST} with G_{ST} do not differ significantly. 1000 permutations of haplotypes between populations were performed in order to test for phylogeographic structure by comparing the observed R_{ST} with the randomly permutated G_{ST} values based on Slatkin (1995), Pons and Petit (1996) and Burban et al. (1999).

3.5.3. Genetic variation based on nuclear microsatellites (nuSSRs)

Genetic diversity and structure are determined through a series of analyses on the nuSSR genotypic data.

Pairwise F_{ST} values are calculated using GenAlEx, based on Peakall and Smouse (2006, 2012), and are used in the construction of a phylogenetic tree which illustrates the genetic relationships between populations.

An AMOVA is produced to partition the genetic variation within vs. among populations. Significant differentiation among populations indicates that different seed sources were used between populations. To test for significance, 999 random permutations were produced according to the procedure described in (Peakall and Smouse, 2012). The AMOVA and statistical tests were performed using GenAlEx.

A STRUCTURE analysis is conducted as an additional method to identify subgroups and assign individuals and populations to them (Hubisz *et al.*, 2009). In addition to providing further evidence to support whether or not populations differ from one another, this analysis can also reveal if there is an admixture of different genetic groups or seed sources within a population. In particular, the STRUCTURE algorithm models a pre-defined number of K clusters and assigns each individual a membership proportion to each one of the K modelled clusters. Twenty independent runs are carried out for each one of K-values ranging from 1 to 10. 50 000 burn-in replications and 100 000 Markov Chain Monte Carlo (MCMC) iterations are applied for each run. According to our assumptions the admixture model, correlated allele frequencies are opted. Finally, the locprior option (Hubisz *et al.*, 2009) is selected. According to this option, population information is used as a prior during the model run (instead of non-prior), increasing the method sensitivity in detecting minor differences in the genetic structure. This option is selected given that an assumed common origin of single stands.

The optimal number of clusters K, was defined based on two criteria: (i) Maximisation of the statistic ΔK (Evanno *et al*., 2005), which is based on the second order rate of change of log likelihood of the data for consecutive K values. The K value corresponding to the maximum ΔK denotes the uppermost hierarchical level (Evanno *et al*., 2005). ΔK was calculated using the program STRUCTURE HARVESTER (Earl and Von Holdt, 2012). (ii) Unimodality among runs for a particular K, for example, each one of the 20 runs for a particular K should lead to the same clustering solution. Using the online platform CLUMPAK (Kopelman *et al.*, 2015), we were able to control this by processing the multiple runs.

Subsequently, a series of genetic diversity measures are calculated using GenAlEx. These include: mean number of alleles per locus (N_a), observed (H_o) and expected heterozygosity (H_e) per locus and populations by using the software GenAlEx v6.5 (Peakall and Smouse, 2006, 2012). Finally, using the software FSTAT (Goudet, 1995) we compute inbreeding coefficient (F_{IS}) values per locus and across all loci (Weir and Cockerham, 1984), which express the deviation between H_o and H_e. A heterozygote deficit is denoted by a positive F_{IS} value. Significance of F_{IS}-values was tested by applying 1000 random permutations of alleles among individuals within populations and comparing the observed values with those produced by the permutation procedure.

Chapter 4: Results

4.1. Chloroplast Marker Results

The basic analysis of haplotypes revealed an overwhelming majority of the samples consisting of Haplotype A. Table 19 shows the haplotypes assigned and their frequencies of occurrence. Haplotypes A1, A2 and A3 were slightly different from A but seem to be so rare that no published literature has acknowledged them thus far. However, they are statistically relevant to this study as they comprise over three percent of our total population. Hence, we named these variants A1, A2 and A3.

Table 19) Haplotype description and frequencies

Haplotype	cpSSR fragment size (bp)					Frequency
	ccmp2	ccmp4	µcd4	µdt1	µdt4	
A	228	116	99	86	145	73,0%
A1	228	116	99	85	145	0,4%
A2	228	115	99	86	145	2,7%
A3	228	115	99	87	145	0,4%
B	227	115	98	85	146	8,7%
C	228	116	99	87	145	9,5%
E	228	116	97	86	146	1,1%
F	227	116	98	85	146	3,8%
O	228	115	98	85	146	0,4%

The MSN in Figure 8 shows the haplotypes in sizes relative to their frequencies as well as the linkage between them. It is clear that the haplotypes A, A1, A2, A3 and C are descended from a different lineage than F, B, O and E.

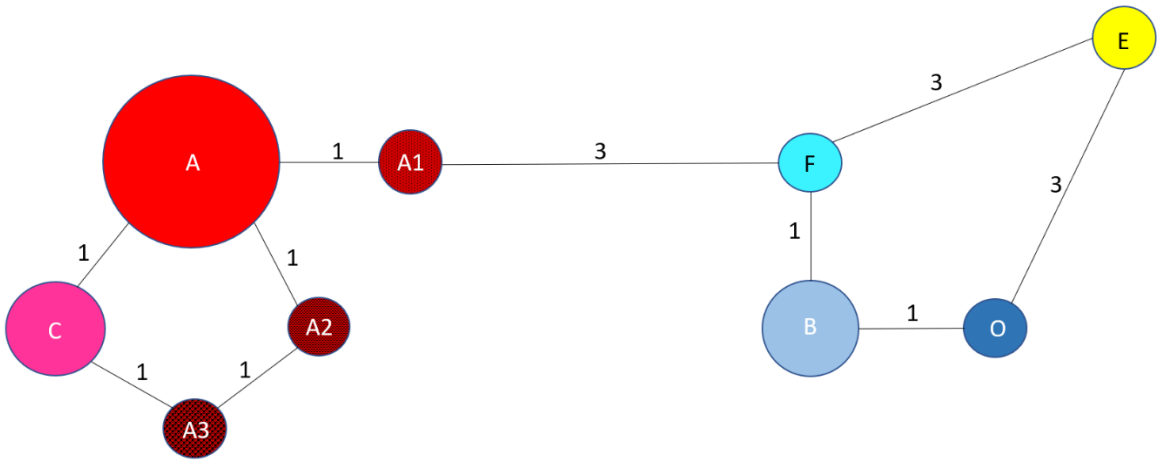


Figure 8) Minimum spanning network

The haplotypic variation within each plot is shown in a pie chart format in FiguresFigure 9Figure 18. Haplotype A is dominant in both the native and European ranges of NRO.

The first plot in Unterpullendorf consisted only of haplotype A.

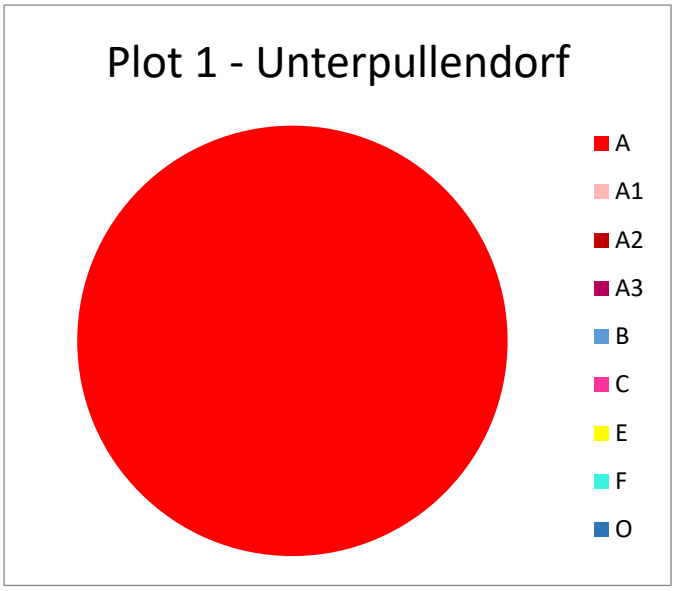


Figure 9) Plot 1 haplotype chart

The second Unterpullendorf plot is comprised of a significant amount of haplotype F, which represents 3.8% of the total population sampled in the study. The differences in haplotype proportions between the two Unterpullendorf plots are the largest observed in this study for plots so close together.

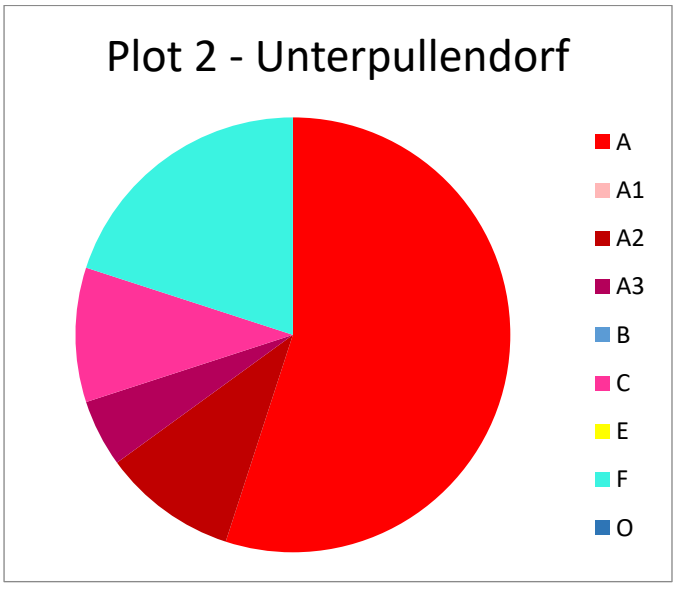


Figure 10) Plot 2 haplotype chart

The Horitschon plots also showed a considerable proportion of haplotype F, which we found to be of a different lineage to the A's and C.

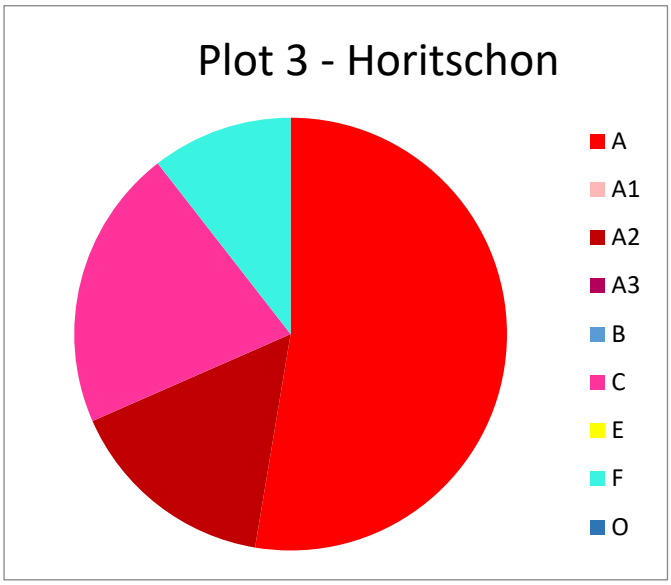


Figure 11) Plot 3 haplotype chart

Horitschon Plot 4 is very similar to Plot 3. The prevalence of haplotype C is half that of Plot 3 although they are both dominated by haplotype A.

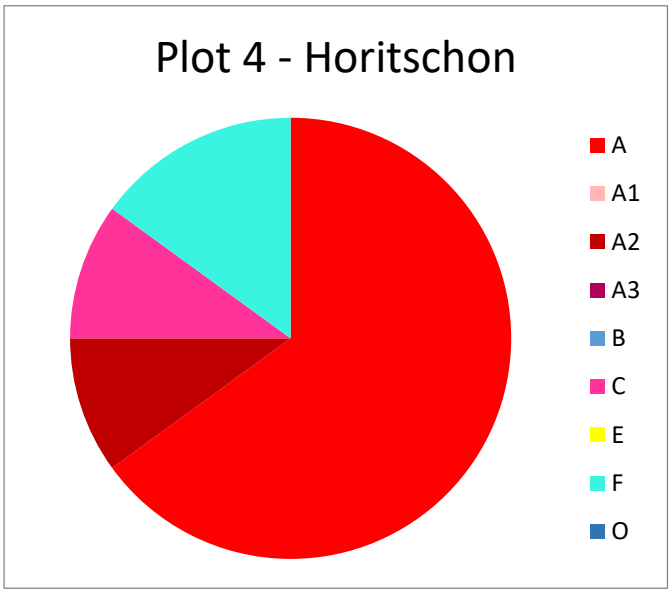


Figure 12) Plot 4 haplotype chart

The Mannersdorf plot is also dominated by haplotype A but consists of significant portions of both B and C.

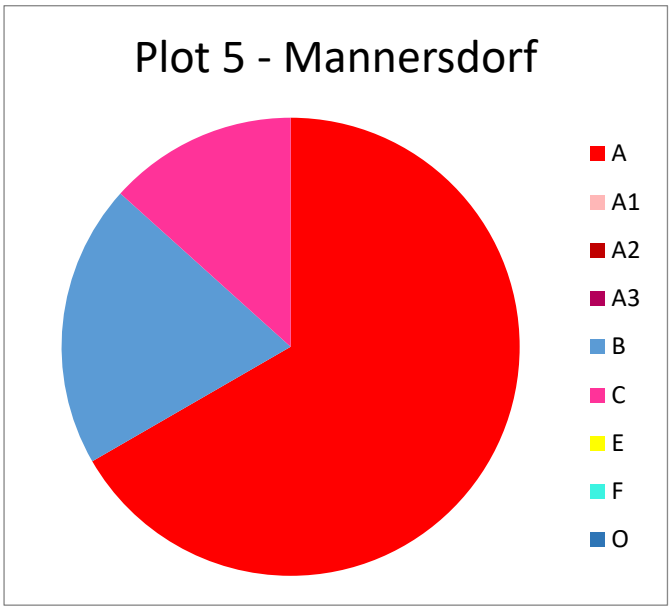


Figure 13) Plot 5 haplotype chart

The Langental plot also had an overwhelming presence of haplotype A, along with small portions of both A1 and B.

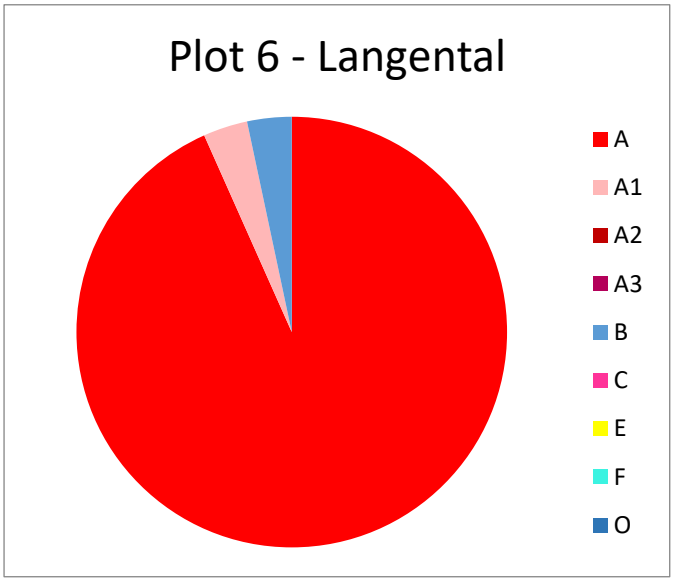


Figure 14) Plot 6 haplotype chart

The plot in Girm contains a similar genetic makeup to that of Plot 5 in Mannersdor, an over whelming presence of haplotype A with significant portions of both B and C.

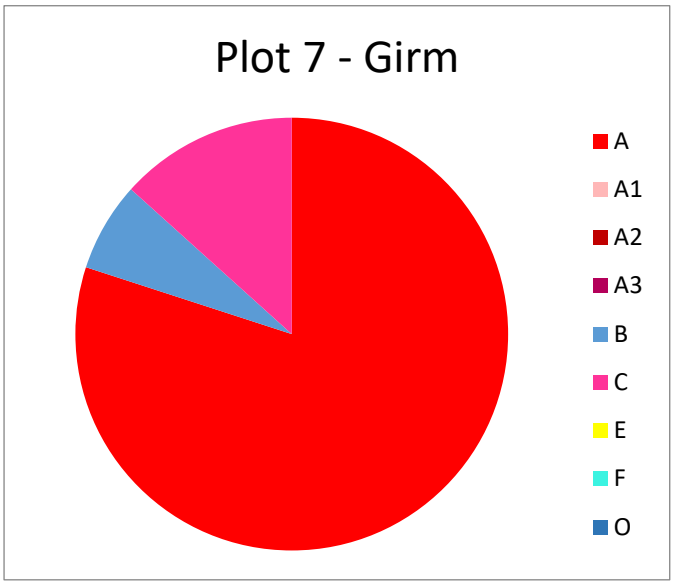


Figure 15) Plot 7 haplotype chart

The St. Michael plot is rather unique as there is a presence of haplotype O, which is rare. This haplotype makes up 0.4% of our total sample. This plot also contains the smallest portion of haplotype A.

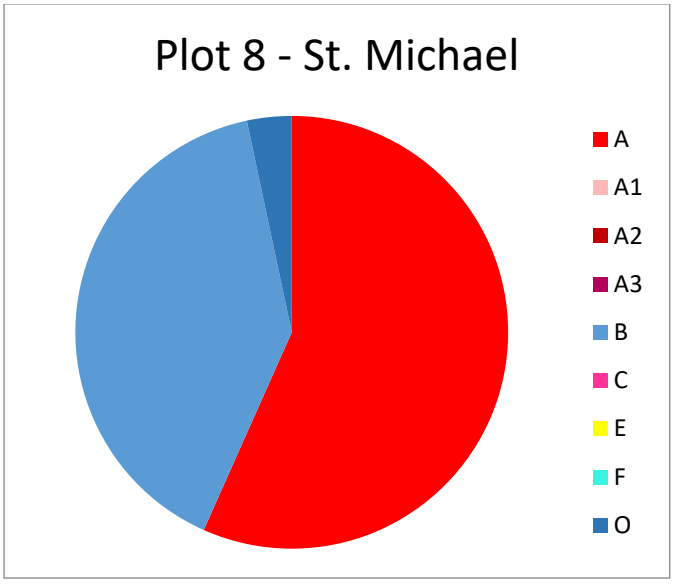


Figure 16) Plot 8 haplotype chart

The Moschendorf Plot 9 shows a large presence of haplotype A. However, the presence of haplotypes C and E together make this plot rather unique among the other plots in this study.

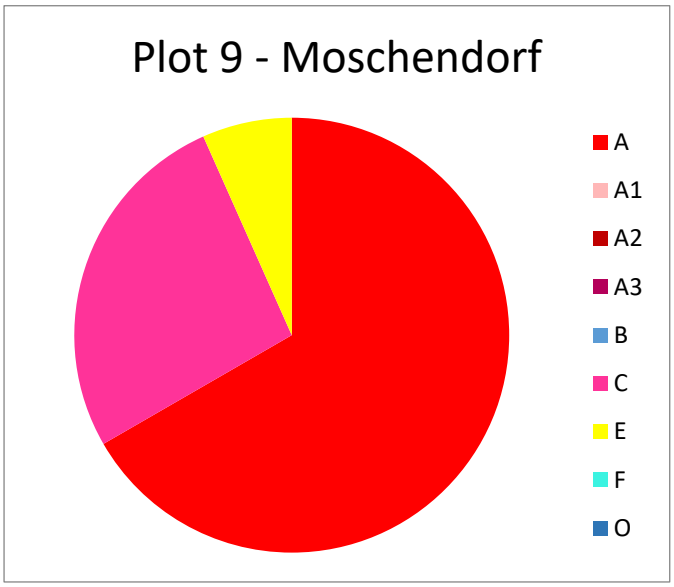


Figure 17) Plot 9 haplotype chart

The Moschendorf Plot 10 is dominated by haplotype A and contains a small portion of B. These plots are remarkably different considering their proximity to one another.

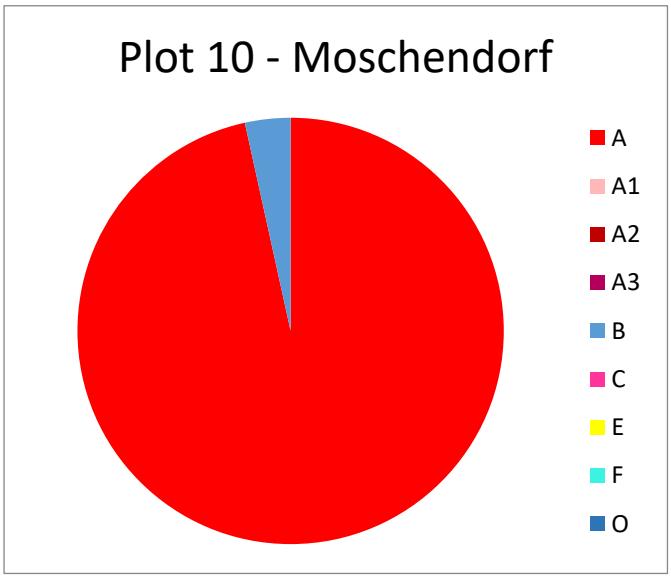


Figure 18) Plot 10 haplotype chart

A map showing the haplotypes and corresponding plots can be seen in Figure 19**Error! Reference source not found.**.

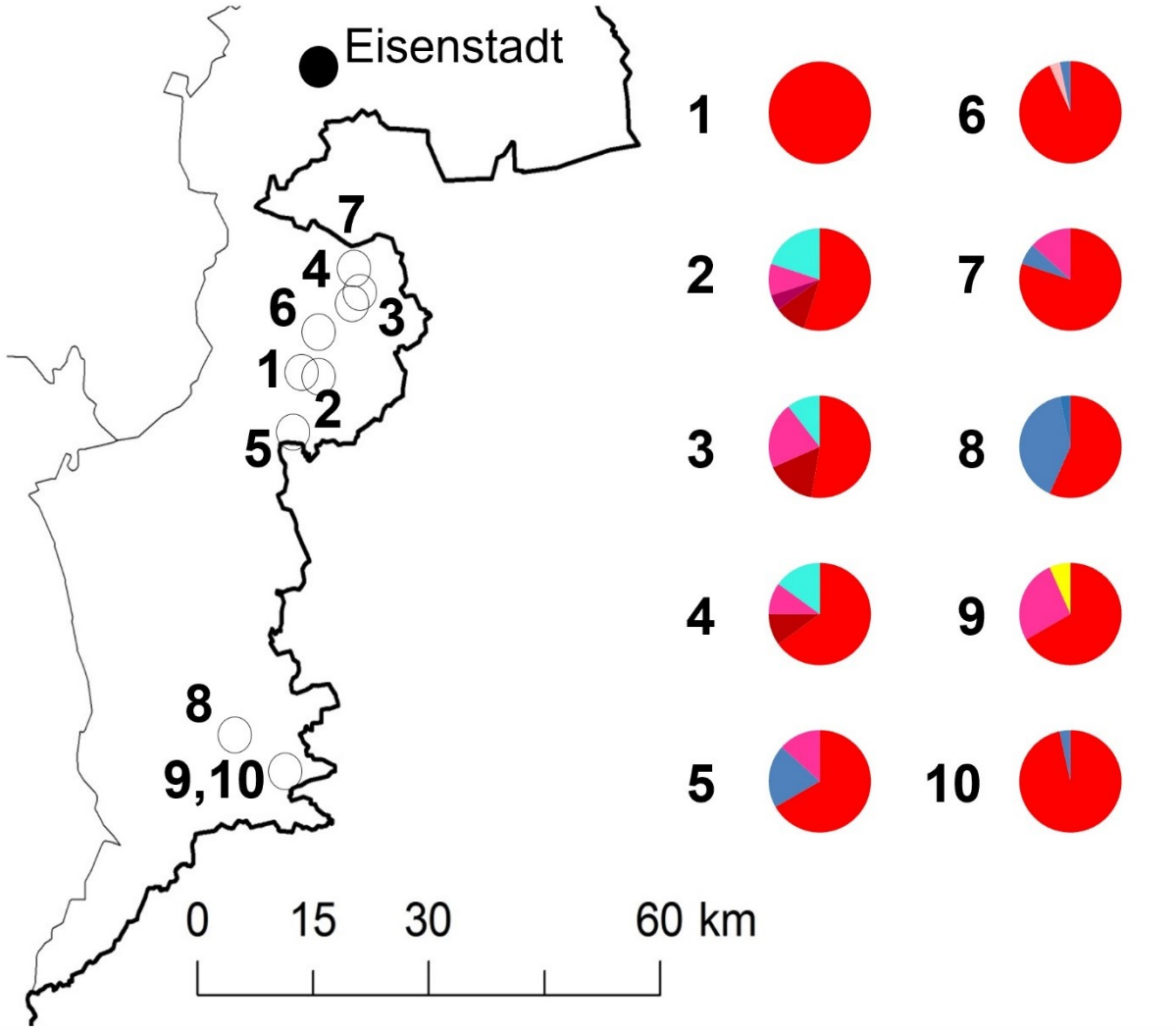


Figure 19) Haplotype/plot map

The AMOVA in Table 20 returns the F_{ST} value with a P – value of less than 0.001, pointing that the populations are, in total, significantly different in terms of their haplotypic composition.

Table 20) AMOVA of cpSSRs

Source of variation	Degrees of freedom	Sum of squares	Variance components	Percentage of variation
Among populations	9	24.787	0.08333	11.97
Within populations	248	152.012	0.61295	88.03
Total	257	176.798	0.69628	
Fixation Index (F_{ST})	0.11968			
P – value	<0.001			

Table 21) Arlequin output and measures of diversity

H_T	H_S	G_{ST}	R_{ST}
0.453	0.401	0.115	0.104

The average within-population genetic diversity (H_S), shown in Table 21 above is similar to the value within the pooled sample (H_T), which denotes that a most of the common haplotypes can be found within each plot.

Moreover, the test for phylogeographic structure was not significant, i.e. G_{ST} was not significantly different from RST (P = <0.001). In other words, this shows that the presence of phylogenetically different haplotypes across plots is rather random.

4.2. Nuclear Marker Results

4.2.1. Genetic differentiation among populations

The phylogenetic tree produced by FigTree uses pairwise F_{ST} values of each site based on nuSSRs to plot said tree in a similar way to the haplotype tree used to describe the cpSSRs. This tree, illustrated in Figure 20, shows a relatively high differentiation of plots 8 and 9 to the rest of the study plots. It also shows that plots 3 and 4, as well as 5 and 6, are almost identical in their genetic compositions, respectively.

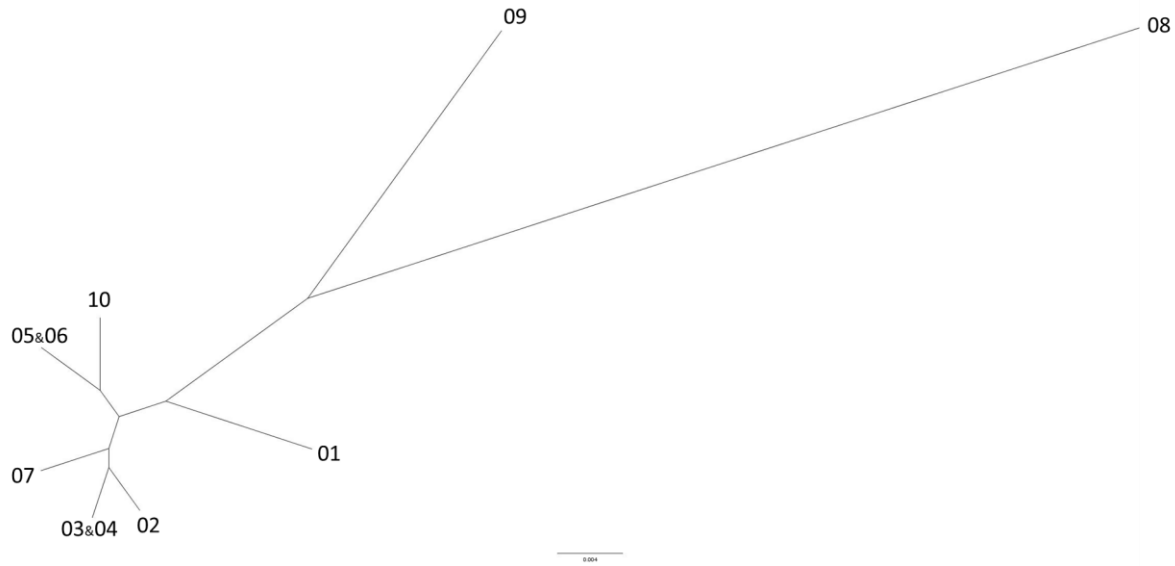


Figure 20) Genetic structure by plot

4.2.2. Genetic Diversity

The analysis of molecular variance (AMOVA) revealed the percentage of genetic variation among the entire population, as well as the fixation index, inbreeding coefficient and overall fixation index, all to a P value of less than 0.001. The AMOVA can be seen in Table 22.

Table 22) AMOVA of nuSSRs

Source of variation	Degrees of freedom	Sum of squares	Mean squared error	Estimated variance	Percentage of variation (%)
Among populations	9	95,980	10,664	0,124	3
Within populations	502	1920,676	3,826	3,826	97
Total	511	2016,656		3,960	100
Fixation Index (F_{ST})	0.034	P < 0,001			
Inbreeding Coefficient (F_{IS})	0.130	P < 0,001			
Overall Fixation Index (F_{IT})	0.157	P < 0,001			

The molecular variance measures revealed a 3% variance among populations and 97% within populations. This has been put into a pie chart format in Figure 21.

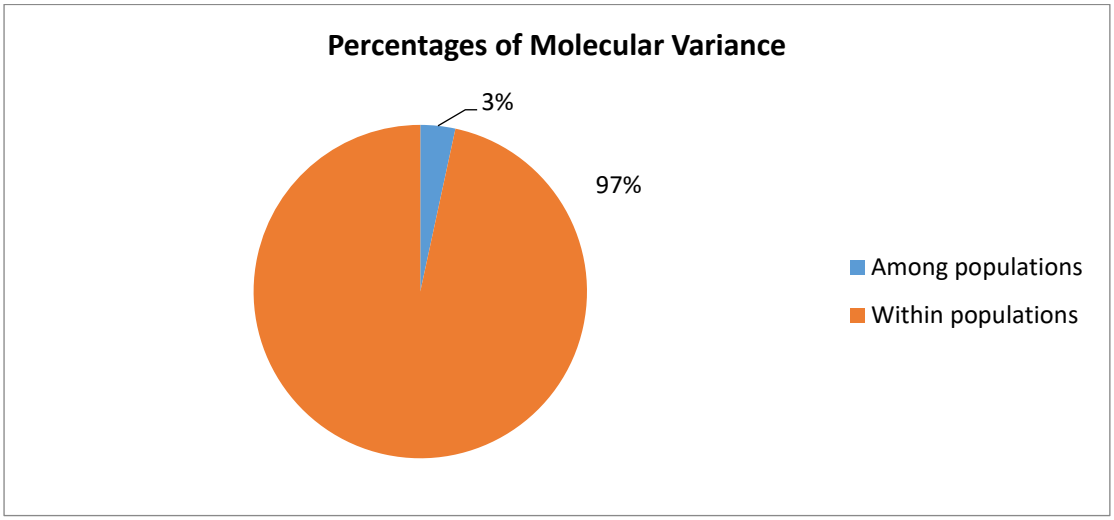


Figure 21) Chart of molecular variance

The cluster analyses in Figures 22 through 27 show the number (K) of assumed clusters and their distribution within each plot. The plots run along the x-axis while the proportion per cluster is represented on the y. Each plot is represented by a “box” and each individual within a plot by a vertical bar.

Figure 22) Cluster analysis where K=1

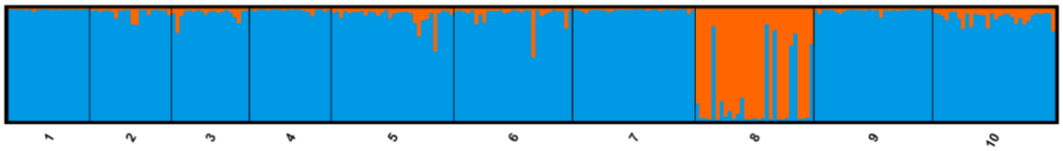


Figure 23) Cluster analysis where K=2

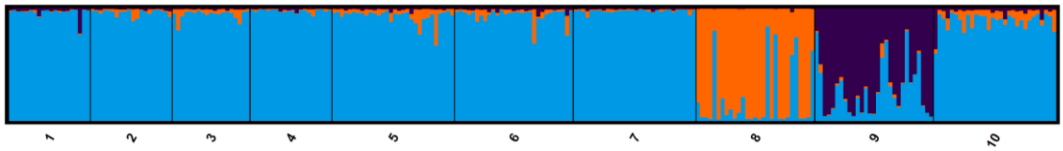


Figure 24) Cluster analysis where K=3

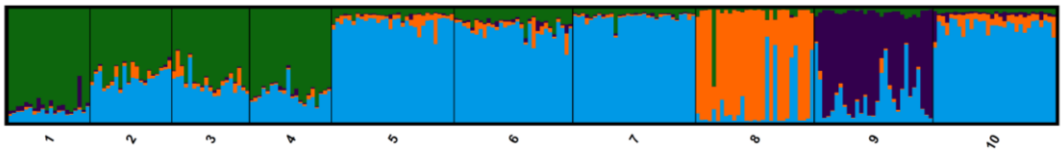


Figure 25) Cluster analysis where K=4

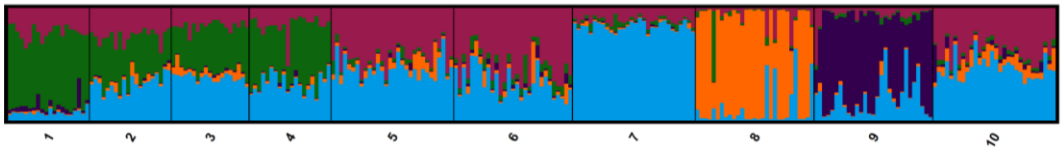


Figure 26) Cluster analysis where K=5

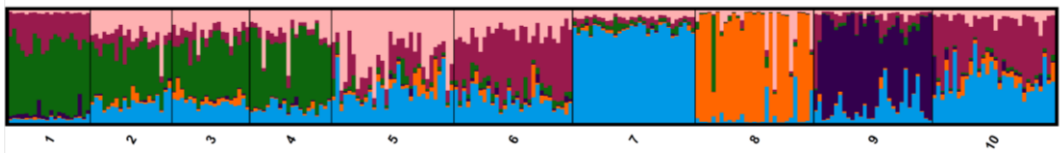


Figure 27) Cluster analysis where K=6

For K=2 (Figure 23) only one of the plots was separated from the others. The plot in question, Plot 8, is believed to be of a non-European origin by the forest owner. For K=3 (Figure 24) Plot 9, in addition to 8, is separate from the others. This corroborates the distance of plots 8 and 9 represented by Figure 20. Plots 8 and 9 are geographically separated from all the other plots with the exemption of plot 10.

Chapter 5: Discussion

5.1. Chloroplast Microsatellite Analysis Results

As chloroplast is inherited maternally in NRO, the lack of chromosomal recombination between mother and father causes very little variation over generations. This sequence conservation stores invaluable information about the origin of individuals (Weising & Gardner, 1999).

The dominance of haplotype A is no surprise, as this has been observed range-wide across the natural distribution area in North America and displays an even higher frequency in Europe (Pettenkofer *et al.*, 2019).

The rare haplotypes A1, A2 and A3 could possibly be the result of spontaneous mutations (given only one single mutational difference from haplotype A) or analysis errors in GeneMapper. However, as reference samples were included in the sequencing runs as a means of calibration, it is likely that these haplotypes do exist. By including them in this study and labelling them as such this creates a record of these haplotypes which can be used in the future.

The results also suggest that there is high degree of sharing of genetic material among the populations. However, this could also be an effect of limited genetic variation among the founder population in Europe, as well as limited introduction of new material from the native range. The lack of genetic recombination in the chloroplast could also significantly contribute to the observed measure of shared genetic material, as well as the tendency of oaks to be much greater pollen dispersers over distance than seed dispersers. As interpopulation pollination is the main means of genetic flow, as opposed to seed flow, maternal organelles remain largely unchanged genetically (Ennos, 1994).

The intrapopulation haplotypic variation (H_S) ranges from absolutely none to around fifty percent. The value of 0.401 does not fully describe the observed individual population variation but is merely an average value.

As can be seen in the minimum spanning tree there are two distinct lineages, perhaps even a third. The two main lineages are linked between haplotype A1 and F by three mutational differences. If we were to disregard A1, F would be coupled to A with four mutational differences.

There is a possibility that haplotype E belongs to a different lineage due to the distance it holds between F and O. However, to confirm such an assumption one would have to conduct an analysis with a larger sample size over the regions in which these haplotypes were found.

In comparison to the most recent such study on NRO, Pettenkofer *et al.,* (2019) found that their study plots in Germany were densely clustered with haplotypes of reference populations only found in the centre and northern regions of the native range in North America. They found thirteen chloroplast haplotypes, five of which were found only in Germany. Haplotype A was dominant in both North America and Germany although a much higher frequency of A was found in Germany.

We had similar findings to those of Pettenkofer *et al.,* (2019) in the sense that haplotype A was dominant and three of the five haplotypes only found in Germany were found in nine of the ten populations in this study. As a reference of comparison to North American and German NRO populations, Table 23 contains the chloroplast diversity measures found by Pettenkofer *et al.,* (2019) along with the measures derived from this study.

Haplotype diversity within the Austrian populations was high (H_S=0.401) compared to that of the North American populations (H_S=0.177). German populations lie in between these two (H_S=0.291). Despite the low haplotypic diversity within the North American populations, the total haplotype diversity (H_T=0.654) is higher than that of the Austrian populations (H_T=0.453).

The genetic differentiation in Austria (G_{ST}=0.115, R_{ST}=0.104) and Germany (G_{ST}=0.137, R_{ST}=0.047) is significantly less than that of the North American populations (G_{ST}=0.729, R_{ST}=0.772). The low genetic differentiation in Austria and Germany is likely due to the frequent occurrence of haplotype A and is indicative of introduced populations (Pettenkofer *et al.*, 2019) exacerbated by the loss of genetic diversity in the native range over the last glacial period in North America (Magni *et al.*, 2005). There seems to be no evidence of introduction of NRO from a source located in the southern part of the native range. However, if an initially homogenous seed source was introduced to the established population at a later stage a lower differentiation among the different gene pools is expected as there would have been genetic exchange towards a similar composition of each pool (Neophytou and Michiels, 2013). This comparison can be seen in Table 23.

Table 23) Chloroplast data measures in comparison to Pettenkorfer et al., (2019)

Data	H_T	H_S	G_{ST}	R_{ST}
Austria (this study)	0.453	0.401	0.115	0.104
Germany (Pettenkofer et al. 2019)	0.337	0.291	0.137	0.047
North America (Pettenkofer et al. 2019)	0.654	0.177	0.729	0.772

The data present a genetically more diverse population of NRO in Austria in comparison to Germany.

The owner of Plot 8 mentioned that the seed used in establishment was sourced in Canada, in the northern part of the native range. In agreement to this statement, the cpSSR results show a significant portion (40%) of haplotype B within the population, which, according to the nuSSR analyses is of a different lineage and is mostly found in the northern part of the natural distribution area. This plot also contains the smallest portion of haplotype A and according to Pettenkorfer *et al*., (2019) the native range also contains less of this haplotype. Thus, there is sufficient evidence to suggest that this population stems from a seed source different from all other stands which could have its origin in the northern part native range.

There is also reason to believe that Plot 9 may have been subject to the reintroduction of genetic material from the native range many years ago. This is evident by the prevalence of haplotype E which does not feature in any of the other plots and is also of a different lineage to the A haplotypes and possibly to the entire pool of haplotypes we found.

5.2. Nuclear Microsatellite Analysis Results

Gene flow among populations take place under two main modes. The first being dispersal of pollen from one population to another, in which the very next generation (F1) is a mixture of DNA between the two populations. The second being dispersal of seed from one population to another, resulting in a mixture among the stand and a mixed genome in the F2 generation. As

Oaks are naturally adept pollinators, being outbreeding, wind dispersed and tall, interpopulation pollen flow has been observed to be 200 times higher that seed flow (Ennos, 1994). As dispersal of acorns by birds and rodents is relatively restricted (Sork, 1984), where artificial regeneration is common, interpopulation seed flow is mostly conducted deliberately by humans, as was observed in this study.

The fixation index is low (F_{ST}=0.034), which corresponds with the cpSSR results (G_{ST}=0.115) but due to the nature of nuSSRs a large degree of sharing of this genetic material is expected due to efficient long-distance gene flow by pollen. The average fixation index (F_{ST}=0.05) found on nuSSR data in the native range by Lind and Gailing (2013) is larger than in this study. |Considering that the marker sets and study sample sizes differed , as hypothesised by Pettenforfer et al. (2013), the relatively low values shown in this study could also be due to the intermixing of seed sources with a homogenising effect (Neophytou and Michiels, 2013).

Plots 3 and 4 display no significant differentiation between each other, which is in agreement with the results from chloroplast microsatellites indicating a very similar composition of haplotypes A, A2, C and F and are also very close to one another.

According to the nuclear data plots 5 and 6 should also be closely related. However, their haplotype composition does not support this and they are approximately 15km apart. This is possibly a result of establishment with a very similar seed source.

The nuSSR results also point to evidence which suggests that Plot 8 may stem from a different seed source. The distance that the plot holds from the central node of the phylogenetic tree in Figure 20 clearly shows a separate genetic composition from the others. According to the cluster analyses, Plot 8 also shows a significant difference to the others when assuming two or more clusters.

Plot 9 tells a similar story but to a lesser extent. The nuSSR results revealed evidence to back this up as the phylogenetic tree places Plot 9 almost as far from the central node as Plot 8. In other words, plots 8 and 9 possess unique gene pools and thus are very likely of a different seed origin.

Chapter 6: Conclusion

The focus of this project, to determine the genetic origin, structure and variation of *Q. rubra* over ten plots in Burgenland, Austria, will likely contribute to future studies on the species and serve as a bridge between similar studies from which the microsatellites in question were taken. It was the goal of this study to choose and use a highly informative set of markers, allowing them to distinguish gene pools of different origin. By selecting markers from different studies, a standard can be established whereby reliable comparisons may be conducted between studies which have no markers in common, given that this study has common markers with the others.

6.1. Research Objective 1 – Genetic Variation and Origin

A total of nine haplotypes were found in our populations, six of which have been found throughout Germany (Pettenkofer *et al.*, 2019) and three of which have not been recorded at the time of this study. Due to the presence of three unique haplotypes in German and Austrian populations, at least nine of the ten populations in Austria likely originated from the same source as populations in Germany.

At least one of the plots (Plot 8) was shown to consist of "new" genetic material and the haplotypes found within it are consistent with the northernmost NRO found in the native range (Magni *et al.*, 2005). The forest owner of this plot mentioned that his grandfather sourced the seed outside of Europe (in particular, in Canada).

There seems to be no distinct trend between geographic location and haplotypes in the populations.

6.2. Research Objective 2 – Genetic Structure

The distribution of chloroplast haplotypes is skewed very much in favour of haplotype A. Much like other studies in Europe, our populations were overwhelmed by haplotypes A, B and C. These haplotypes represent a total of 91.2% and haplotype A alone representing 73.6% within our populations.

Plot 8 clearly stems from a different seed source as both the chloroplast and nuclear data show a different haplotype composition and greater linkage steps respectively.

Plot 9 seems to have been at least supplemented with genetic material of a different source as it exclusively contains haplotype E within our study sample and is shown by the nuclear data to have a greater genetic distance from the central node.

6.3. Research Objective 3 – Genetic Diversity

As expected of an introduced species, the genetic diversity measures are lower, according to the sample in this study, than those of the native populations. However, these measures showed a higher degree of genetic diversity within Austria in comparison to Germany.

As such, it would be recommended to use the populations with high haplotypic diversity for future seed transfer and breeding.

6.4. Final Remarks

The practical relevance of this study falls under a range of points which are beneficial to proficient and responsible forest practice.

There are significant differences in phenological traits among provenances (Daubree and Kremer, 1993). These traits may influence growth rates (Kriebel, 1993) and resistance to disturbances. Bud phenology is a fine example of this, as it is linked to frost hardiness and drought tolerance.

Thus, information on seed origin is relevant as it may give indications of what to expect of growth, adaptive capacity and resilience of a stand. However, seed origin is mostly untraced or lost, as was apparent when the forest owners and enterprises were asked about the origin of the study stands but no documentation could be found for nine out of ten study plots.

The tested sets of chloroplast and nuclear SSRs proved to be efficient in identifying different seed origins and meeting the objectives of this study.

References

Aldrich, P. R. et al., 2005. Genetic structure inside a declining red oak community in old-growth forest. *Journal of Heredity,* 11, 96(6), pp. 627-634.

Aldrich, P. R., Jagtap, M., Michler, C. H. & Romero-Severson, J., 2003. Amplification of North American Red Oak Microsatellite Markers in European White Oaks and Chinese Chestnut. *Silvae Genetica,* 52(3-4), pp. 176-179.

Aldrich, P. R., Michler, C. H., Sun, W. & Romero-Severson, J., 2002. Microsatellite markers for northern red oak (Fagaceae: Quercus rubra). *Molecular Ecology Notes,* 12, 2(4), pp. 472-474.

Aldrich, P. R., Parker, G. R., Michler, C. H. & Romero-Severson, J., 2003. Whole-tree silvic identifications and the microsatellite genetic structure of a red oak species complex in an Indiana old-growth forest. *Canadian Journal of Forest Research,* 13 11, 33(11), pp. 2228-2237.

Merceron_2016_Processus_ecologiques_evolutifs_Q_rubra_Europe_DISSERTATION.

Merceron, N. R. *et al.* (2017) ' Back to America: tracking the origin of European introduced populations of Quercus rubra L. ', *Genome*. Canadian Science Publishing, 60(9), pp. 778–790. doi: 10.1139/gen-2016-0187.

Arnold, M. L. & Bennett, B. D., 1993. Natural Hybridization in Louisiana irises: genetic variation and ecological determinants.. *Hybrid Zones and Evolutionary Process,* pp. 115-139.

Barrett, S. & Husband, B., 1990. The genetics of plant migration and colonization.

Borkowski, D. S., Hoban, S. M., Chatwin, W. & Romero-Severson, J., 2017. Rangewide population differentiation and population substructure in Quercus rubra L.. *Tree Genetics and Genomes,* 1 6.13(3).

Borkowski, D. S., McCleary, T., McAllister, M. & Romero-Severson, J., 2014. Primers for 52 polymorphic regions in the Quercus rubra chloroplast, 47 of which amplify across 11 tracheophyte clades. *Tree Genetics and Genomes,* 10(4), pp. 885-893.

Burban, 1999. *Molecular Ecology,* Volume 8, pp. 1593-1602.

Cockerham, C. & Weir, B., 1984. Estimating F-statistics for the analysis of population structure. *Evolution,* Volume 30, pp. 1358-1370.

Daubree, J. and Kremer, A. (1993) 'Genetic and phenological differentiation between introduced and natural populations of Quercus rubra L', *annales des sciences forestières*. Springer Nature, 50(Supplement), pp. 271s-280s. doi: 10.1051/forest:19930727.

Denk, T., 2017. An updated infrageneric classification of the oaks: review of previous adaptive evolution, and the role opf multiple introductions. *Oaks Physiological Ecology,* pp. 13-38.

Earl, D. and Vonholdt, B. (2012) 'Earl DA, VonHoldt BM.. Structure Harvester: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Cons Genet Res 4: 359-361', *Conservation Genetics Resources*, 4, pp. 1–3. doi: 10.1007/s12686-011-9548-7.

Ennos, R. A. (1994) 'Estimating the relative rates of pollen and seed migration among plant populations', *Heredity*, 72(3), pp. 250–259. doi: 10.1038/hdy.1994.35.

Excoffier, Laval & Schneider, 2005. *Arlequin ver 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1:47-50.* [Online].

Excoffier, L., Smouse, P. & Quattro, J., 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics,* Volume 131, pp. 479-491.

Forstwesen, C. f. d. g., 2012. Centralblatt für das gesamte Forstwesen. *Centralblatt für das gesamte Forstwesen ,* 77(3), pp. 169-182.

Goudet, J. (1995) 'FSTAT (Version 1.2): A Computer Program to Calculate F-Statistics', *Journal of Heredity*, 86(6), pp. 485–486. doi: 10.1093/oxfordjournals.jhered.a111627.

Hubisz, M. J. *et al.* (2009) 'Inferring weak population structure with the assistance of sample group information', *Molecular Ecology Resources*, 9(5), pp. 1322–1332. doi: doi:10.1111/j.1755-0998.2009.02591.x.

Juric, I., Aeschbacher, S. & Coop, G., 2016. The Strength of Selection against Neanderthal Introgression. *PLoS Genet,* 12(11).

Korbie, D. J. and Mattick, J. S. (2008) 'Touchdown PCR for increased specificity and sensitivity in PCR amplification', *Nature Protocols*, 3(9), pp. 1452–1456. doi: 10.1038/nprot.2008.133.

Kriebel, H. B. (1993) 'Intraspecific variation of growth and adaptive traits in North American oak species', in *annales des sciences forestières*. EDP Sciences, pp. 153s-165s.

Kopelman, N. M. *et al.* (2015) 'Clumpak: a program for identifying clustering modes and packaging population structure inferences across K', *Molecular Ecology Resources*. John Wiley & Sons, Ltd (10.1111), 15(5), pp. 1179–1191. doi: 10.1111/1755-0998.12387.

Liang, JI, Y.-J. & ZHANG, D.-X., 2015. Statistical measures of genetic differentiation of populations: Rationales, history and current states. *Current Zoology,* 61(5), pp. 886-897.

Lind, J. F. and Gailing, O. (2013) 'Genetic structure of Quercus rubra L. and Quercus ellipsoidalis E. J. Hill populations at gene-based EST-SSR and nuclear SSR markers', *Tree Genetics and Genomes*, 9(3), pp. 707–722. doi: 10.1007/s11295-012-0586-4.

Magni, C. R. et al., 2005. Chloroplast DNA variation of Quercus rubra L. in North America and comparison with other Fagaceae. *Molecular Ecology,* 2, 14(2), pp. 513-524.

Magni Diaz, C. R. (2004) 'Reconstitution de l'introduction de Quercus rubra L. en Europe et conséquences génétiques dans les populations allochtones', *Paris: ENGREF*.

Merceron, N. (2016) 'Processus écologiques et évolutifs impliqués dans le succès de l'introduction de Quercus rubra L. en Europe'. Bordeaux.

Merceron, N. R. *et al.* (2017) ' Back to America: tracking the origin of European introduced populations of Quercus rubra L. ', *Genome*. Canadian Science Publishing, 60(9), pp. 778–790. doi: 10.1139/gen-2016-0187.

Nagel, R., 2015. Roteiche (Quercus rubra L.).. pp. 219-267.

Neophytou, C., Gärtner, S. M., Vargas-Gaete, R. & Michiels, H.-G., 2015. Genetic variation of Central European oaks: shaped by evolutionary factors and human intervention?. *Tree Genetics & Genomes.*

Neophytou, C. and Michiels, H. (2013) 'Forest Ecology and Management Upper Rhine Valley : A migration crossroads of middle European oaks', *FOREST ECOLOGY AND MANAGEMENT*. Elsevier B.V., 304, pp. 89–98. doi: 10.1016/j.foreco.2013.04.020.

Peakall, R. and Smouse, P. E. (2006) 'genalex 6: genetic analysis in Excel. Population genetic software for teaching and research', *Molecular Ecology Notes*. John Wiley & Sons, Ltd (10.1111), 6(1), pp. 288–295. doi: 10.1111/j.1471-8286.2005.01155.x.

Peakall, R. and Smouse, P. (2012) 'GenAIEx V5: Genetic Analysis in Excel. Populations Genetic Software for Teaching and Research', *Bioinformatics (Oxford, England)*, 28, pp. 2537–2539. doi: 10.1093/bioinformatics/bts460.

Petit, P. a., 1996. *Genetics,* Volume 144, pp. 1237-1245.

Pettenkorfer, T. et al., 2019. Genetic diversity and differentiation of introduced red oak (Quercus rubra) in Germany in comparison with reference native North American populations. *European Journal of Forest Research.*

Slatkin, M., 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics,* Volume 139, pp. 457-462.

Sork, V. L. (1984) 'Examination of Seed Dispersal and Survival in Red Oak, Quercus Rubra (Fagaceae), Using Metal-Tagged Acorns', *Ecology*. Ecological Society of America, 65(3), pp. 1020–1022. Available at: http://www.jstor.org/stable/1938075.

Terrab, et al., 2007. Genetic Diversity at Chloroplast Microsatellites (cpSSRs) and Geographic Structure in Endangered West Mediterranean Firs (Abies spp., Pinaceae). *Taxon,* 56(2), pp. 409-416.

Weir, B. and Cockerham, C. (1984) 'Weir BS, Cockerham CC.. Estimating F-Statistics for the Analysis of Population-Structure. Evolution 38: 1358-1370', *Evolution*, 38, pp. 1358–1370. doi: 10.2307/2408641.

Weising, K. & Gardner, R., 1999. A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. *Genome,* pp. 42: 9-19.

Wright, S., 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. *Evol,* Volume 19, pp. 395-420.

Zhang, R., Hipp, A. L. and Gailing, O. (2015) 'Sharing of chloroplast haplotypes among red oak species suggests interspecific gene flow between neighboring populations', *Botany*. NRC Research Press, 93(10), pp. 691–700.

Appendix

Appendix A – DNA extraction protocol

Quick-Start Protocol March 2016 DNeasy® Plant Mini Kit

The DNeasy Plant Mini Kit (cat. nos. 69104 and 69106) can be stored at room temperature (15–25°C) for up to 1 year if not otherwise stated on label.

Further information

- DNeasy Plant Handbook: www.qiagen.com/HB-1166
- Safety Data Sheets: www.qiagen.com/safety
- Technical assistance: support.qiagen.com

Notes before starting

- Perform all centrifugation steps at room temperature (15–25°C).
- If necessary, redissolve any precipitates in Buffer AP1 and Buffer AW1 concentrates.
- Add ethanol to Buffer AW1 and Buffer AW2 concentrates.
- Preheat a water bath or heating block to 65°C.

1. Disrupt samples (≤100 mg wet weight or ≤20 mg lyophilized tissue) using the TissueRuptor®, the TissueLyser II or a mortar and pestle.

2. Add 400 µl Buffer AP1 and 4 µl RNase A. Vortex and incubate for 10 min at 65°C. Invert the tube 2–3 times during incubation. Note: Do not mix Buffer AP1 and RNase A before use.

3. Add 130 µl Buffer P3. Mix and incubate for 5 min on ice.

4. Recommended: Centrifuge the lysate for 5 min at 20,000 x g (14,000 rpm).

5. Pipet the lysate into a QIAshredder spin column placed in a 2 ml collection tube. Centrifuge for 2 min at 20,000 x g.

6. Transfer the flow-through into a new tube without disturbing the pellet if present. Add 1.5 volumes of Buffer AW1, and mix by pipetting.

7. Transfer 650 µl of the mixture into a DNeasy Mini spin column placed in a 2 ml collection tube. Centrifuge for 1 min at ≥6000 x g (≥8000 rpm). Discard the flowthrough. Repeat this step with the remaining sample.

8. Place the spin column into a new 2 ml collection tube. Add 500 µl Buffer AW2, and centrifuge for 1 min at ≥6000 x g. Discard the flow-through.

9. Add another 500 µl Buffer AW2. Centrifuge for 2 min at 20,000 x g. Note: Remove the spin column from the collection tube carefully so that the column does not come into contact with the flow-through.

10. Transfer the spin column to a new 1.5 ml or 2 ml microcentrifuge tube.

11. Add 100 µl Buffer AE for elution. Incubate for 5 min at room temperature (15–25°C). Centrifuge for 1 min at ≥6000 x g.

12. Repeat step 11.