CZECH UNIVERSITY OF LIFE SCIENCES PRAGUE
FACULTY OF ENVIRONMENTAL SCIENCE
DEPARTMENT OF ECOLOGY

Development of microsatellite markers set for population genetic structure estimation of *Peucedanum cervaria* (Apiaceae)

Diploma Thesis 2019

Author: Gabrielle Filippi
Supervisor: Prof. Mgr. Bohumil Mandák, Ph.D
Consultant: Mgr. Karol Krak, Ph.D.

CZECH UNIVERSITY OF LIFE SCIENCES PRAGUE

Faculty of Environmental Sciences

DIPLOMA THESIS ASSIGNMENT

Gabrielle Filippi

Nature Conservation

Thesis title

Development of microsatellite markers set for population genetic structure estimation of Peucedanum cervaria (Apiaceae)

Objectives of thesis

Microsatellite markers have evolved to the status of a most versatile and popular genetic marker in a ubiquity of plant systems. Due to their co-dominant, hyper-variable and multiallelic nature, they are the prominent markers of choice for conservation genetics and phylogeographic studies. Despite its development of a new set of SSR markers for a species remained time consuming and expensive. Within this study we are planning to develop a set of microsatellite markers specific for Peucedanum cervaria, a species of dry European grasslands, and try to cross-amplify developed markers on selected species from Apiaceae family. A PCR multiplex will be established for up to eight markers that can later be analyzed on a capillary sequencer by choosing different fluorescence labels and/or different fragment sizes. Using the microsatellite multiplex population genetic structure of the study species in the Czech Republic will be estimated.

Methodology

Begin with conducting DNA extraction from Plant Tissues on collected samples. Utilizing a fluorescent dye labeled PCR, unique microsatellites will be determined for Peucedanum cervaria. With up to 8 microsatellites a multiplex will be created and used to test the variability of 10 Peucedanum cervaria populations throughout Central Europe. Finally, an analysis of the primers and the population structure will be conducted.

The proposed extent of the thesis

60 pages

Keywords

microsatellites, Peucedanum cervaria, population genetic structure

---

Recommended information sources

Li, Y.-C., Korol, A. B., Fahima, T., Beiles, A., & Nevo, E. (2003). Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Molecular Ecology, 11(12), 2453–2465. https://doi.org/10.1046/j.1365-294x.2002.01643.x

Oliveira, E. J., Pádua, J. G., Zucchi, M. I., Vencovsky, R., & Vieira, M. L. C. (2006). Origin, evolution and genome distribution of microsatellites. Genetics and Molecular Biology, 29(2), 294–307. https://doi.org/10.1590/S1415-47572006000200018

Zalapa, J. E., Cuevas, H., Zhu, H., Steffan, S., Senalik, D., Zeldin, E., … Simon, P. (2011). Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. American Journal of Botany, 99(2), 193–208. https://doi.org/10.3732/ajb.1100394

Expected date of thesis defence

2018/19 SS – FES

The Diploma Thesis Supervisor

prof. Mgr. Bohumil Mandák, Ph.D.

Supervising department

Department of Ecology

Advisor of thesis

Karol Krak

Electronic approval: 22. 3. 2019

doc. Ing. Jiří Vojar, Ph.D.

Head of department

Electronic approval: 22. 3. 2019

prof. RNDr. Vladimír Bejček, CSc.

Dean

Prague on 12. 04. 2019

I hereby declare that I wrote this thesis independently, under the direction of prof. Mgr. Bohumil Mandák, Ph.D. and Mgr. Karol Krak, Ph.D. I have listed all literature and publications used to acquire the information included in this thesis.

In Prague, 18.04.2019 Gabrielle Filippi

Acknowledgements

I would like to thank both my advisor, Prof. Mgr. Bohumil Mandák, Ph.D as well as my Consultant, Mgr. Karol Krak, Ph.D., for their guidance with the development of this thesis. I would also like to thank the women who assisted me during my lab work, specifically Alena Hrdličkova and Kateřina Machynková, as well as the others in this lab for their patience and energy. Finally I would like to thank my family and friends for their moral and technical support during this process.

Abstract

Phylogeography of dry grassland species remains a largely unexplored area of study. Yet these postglacial migration patterns are imperative to understanding the distribution of these highly valuable and under protected areas. The study species, *Peucedanum cervaria* (Apiaceae), is a characteristic species of Central European dry grassland ecosystems. In order to determine population genetic variability and history of *P. cervaria* in Central and Eastern Europe, six microsatellite markers specific to this species were developed. It has been determined, that there are several independent genetic lineages present in Central and Eastern Europe. On one side, there is a lineage specific to northern Central Europe that has a relationship to populations occurring in the Bohemian Massif. On the other side, there are lineages occurring exclusively in Carpathians, unique from all delimited lineages, including Czech ones. Finally, lineages occurring in the Pannonian Lowland are also present in the Czech Republic. Therefore, all analyzed Czech populations are derived from populations occurring in Pannonia, but those not from northern populations in fact represent an extension of the Pontic area.

Key words:

Microsatellites, *Peucedanum cervaria,* population genetic structure, Carpathian Mountains, cryptic refugia

Abstrakt

Fylogeografie druhů xerotermních trávníků nebyla doposud podrobněji studována a to i přesto, že jejich posglaciální historie je zásadní pro pochopení historie těchto silně ohrožených společenstev. Studovaný druh *Peucedanum cervaria* (Apiaceae) je charakteristickým druhem xerotermních trávníků střední Evropy. Za účelem stanovení populačně genetické variability a historie druhu *P. cervaria* bylo vyvinuto šest variabilních mikrosatelitových markerů. Analýzou reprezentativního počtu populací bylo ukázáno, že na území střední a východní Evropy se nachází několik nezávislých genetických linií. Na jedné straně jsou zde linie vyskytující se na severu střední Evropy, které nemají žádný vztah k populacím České kotliny. Na druhou stranu jsou zde linie karpatské, velmi odlišné od všech vylišených linií, včetně českých. Poslední poměrně početnou skupinou jsou linie vyskytující se v panonské nížině, zasahující i do Čech. V tomto případě můžeme předpokládat, že všechny české populace jsou odvozeny z populací panonských, a ne z populací vyskytujících se na severu střední Evropy, představující extenzi Pontické oblasti.

Table of Contents

1 Introduction 1
2 Aims 3
3 Literature Review 4
3.1 Grassland ecosystems 4
3.2 Phylogeography of Dry Grasslands 7
3.3 Risks 10
3.4 Characterization of *Peucedanum cervaria* 12
3.5 Microsatellites 13
4 Methodology 16
4.1 Study Sites and Sample Collection 16
4.2 DNA extraction, PCR, and Fragment Analysis 18
4.3 Multiplex Creation & Optimization 20
4.4 Data Analysis 21
4.4.1 Population Genetic Parameter Estimates 21
4.4.2 Population Structure Analysis 21
5 Results 22
5.1 Microsatellite Development 22
5.2 Population Structure 30
6 Discussion 33
6.1 Microsatellite Data Analysis 33
6.2 Population Structure 35
7 Conclusion 39
8 References 41
9 Appendix 45

1 Introduction

The distribution of taxa across Europe and the globe are currently only understood on a basic level. Furthermore, these current distributions cannot fully be understood until historical influences are taken into account. Throughout Europe the influence of glaciations is an important factor to consider when discussing the phylogeographic pattern of a species. The commonly accepted hypothesis on survival of temperate species during the Late Pleniglacial (LPG), 24 to 15 kyr BP, is the Southern Refugia Paradigm (Hewitt, 2000). Although this hypothesis is accepted for most temperate species, for boreal, continental, and arcto-alpine species their distributions were likely greater during the LPG than during the Holocene (Kajtoch et al., 2016; Horreo et al., 2018; Mráz et al., 2007). For this reason, the Holocene is more likely their time period of survival. The lack of knowledge and research on this situation is the motivation for this thesis.

The research on European steppes and dry grasslands up to this point although considerable has not been comprehensive. With their recently recognized conservation value, developing an understanding of their phylogeographic patterns has become a necessity (Wesche et al., 2016). The current data on steppe distributions throughout Europe suggest that these systems were more prevalent during glacial times (Binney et al., 2017; Janská et al., 2017). Kajtoch et al. 2016 describes phylogeographic patterns of dry grassland species across Central Europe for different groups of organisms. Common phylogeographic patterns were determined for steppe species and distinctiveness of populations was established, the results support the theory of survival during the LPG (Kajtoch et al., 2016). One of the downfalls of this analysis was that different molecular markers were used with different sampling methods (Kajtoch et al., 2016). Some of the analyzed molecular techniques, specifically AFLP, do not provide suitable data for analyzing demographic histories. A study utilizing one type of molecular marker with consistent sampling and analytical techniques is crucial to determining the true distributions of both species and communities.

Due to this necessity, the project: *What is the origin of dry grasslands in Central Europe? A synthesis of comparative phylogeography and palaeodistribution modeling*

was undertaken by prof. Mgr. Bohumil Mandák, Ph.D. Under this project, characteristic dry grassland plant species and communities are to be analyzed while routes of expansion will be evaluated across time and space (B. Mandák, personal communication). In order to achieve this overall goal three main aims were identified: nine dry grassland plant species phylogeographic history should be established; these established phylogeographic histories will be compared and analyzed to determine if recent immigration or long-term persistence occurred among these species; and refugia habitats and niche differentiation will be determine across phylogeographic lineages. The nine species represent three biogeographical groups with three species each, in order to provide a thorough understanding of each community. Each group represents a main vegetation type existing in European dry grasslands: *Festucetalia valesiaceae, Brometalia erecti* and *Stipo-Festucetalia pallentis*, classified respectively as species of dry steppes, meadow steppes, or submediterranean dry grassland species (Willner et al., 2017).

This thesis addresses one component of the first aim of the previously mention project, the Reconstruction of the late Pleistocene and Holocene phylogeographic histories of individual species. *Peucedanum cervaria*, the target species, is a characteristic species of submediterranean dry grasslands, thus an analysis of its population genetic structure is relevant to this project.

Currently, *Peucedanum cervaria* has no molecular markers available for genetic studies. The six unique microsatellite markers completed by this project are species specific and highly variability, fulfilling this resource gap. The population genetic structure completed in this thesis, utilizing the developed markers, will allow a detailed analysis of the species current distribution and the historical spread throughout Central Europe to be determined. Determining how many populations, or genetic clusters, exist within Central Europe and comparing their similarity will allow future studies on the evolution of dry grassland species to be performed.

2 Aims:

The aims of this thesis can be separated into 3 main sections:

1. To write a comprehensive literature review on the factors effecting dry grasslands species in Central Europe.
    - Central European dry grasslands are ecosystems of high conservation value, yet few genetic analyses have been conducted on their characteristic species. Gathering the current information on these systems and the possible mechanisms of genetic analysis on their species is an important step to determining their phylogeography.
2. To develop a set of variable microsatellite markers for *Peucedanum cervaria,* a characteristic species of dry grasslands in Central Europe.
    - Since microsatellite markers are species specific and highly variable it is important to develop a unique set for each target species. Utilizing microsatellite markers allows for a modern and consistent methodology to be used in comparative studies. Other characteristic species of dry grasslands also have unique microsatellite markers, thus allowing a comparative study to be conducted after development.
3. To analyze the population structure of *Peucedanum cervaria* across Central Europe.
    - Very little information is available on the species *Peucedanum cervaria.* Thus, the analysis of population genetic structure of the species will provide insight on the phylogeny of *Peucedanum cervaria.*

This thesis will focus on the population genetic structure of *Peucedanum cervaria.* In order to understand the post-glacial history on dry grasslands throughout Central Europe, microsatellite markers must be developed and analyzed. Developing an understanding of the historical influences on *Peucedanum cervaria* will allow comparative studies to be conducted in the future.

3 Literature Review

3.1 Grassland ecosystems

Grasslands are among the Earths most species-rich systems yet are one of the least protected biomes (Janišová et al., 2011). Around the globe, they are known by different names; prairies in North America, Pampas in South America, African Savannas, and Eurasian steppes. Despite their different names all grasslands share common features mainly that the dominant vegetation is a grass species and the system is limited by precipitation levels. A further separation of grasslands is into two categories, tropical and temperate. The temperate grasslands, such as those within Eurasian steppes, are characterized by 25-75 cm of rainfall per year, shorter grasses, as well as two seasons: dormant and growth (Wilson, 1988). Recent research has distinguished a more specific definition of *steppes* to include areas whose climates are exceptionally dry and which are dominated by herbs, in particular grasses, with the presence of chamaephytes (Wesche et al. 2016). Dry grasslands within the Eurasian steppes typically have nutrient poor soils, are on the lower end of the precipitation level of temperate grasslands, and can have up to 80 plant species per square meter (Silva, 2008). With grassland species containing the second largest group of vascular plants endemic to Europe, almost twice the amount of forest endemics, this system has a high biological value (Janišová et al., 2011). Some Central Europe plots hold world records for richness of vascular plant species making them important areas for conservationists (Palpurina et al., 2015).

Several factors are known to affect the diversity within a grassland, including social influences, connectivity, and glaciation. Social, or human, influences include events from the fall of the Iron Curtain in 1989 to agricultural land use (Kizeková et al., 2017). The fall of the Iron Curtain not only reestablished connectivity between habitats but also influenced agricultural impacts and societal views on grasslands (Kizeková et al., 2017). In the 1990s, an increase in grassland abandonment and more extensive use of grasslands became prevalent due to the changing political situations (Kizeková et al., 2017).

The connectivity of ecosystems in Europe has changed drastically throughout history, especially during the last century as political changes including both World Wars, are becoming more visible in the landscape. As grasslands are restored, they will increase the connectivity throughout Europe. Studies have shown that as habitats are restored more specialist species appear (Waldén et al., 2017). This is an important feature of ecosystems as it not only encourages natural relationships to form but also allows communities to develop into stable forms.

Climate change has also influenced the distribution as well as the characteristics of European grasslands. As temperatures are expected to continue to change in the future, looking to the characteristics of specific ecosystems can provide us with insight on how each region will be effected (Wesche et al., 2016). Glaciations are considered heavily influential on species distribution on a larger time scale. As large areas of Europe were historically covered by glaciers, it is necessary to consider their impacts on ecosystems and species development. Due to the historical conditions, species with particular adaptations were able to thrive and/or survive throughout Central Europe. The Pannonian region and its dry grasslands are a specialized example, impacted by many of these factors, thus making it a unique system to study. An example of a site is shown in Figure 1.

Figure 1: Grassland Sample Site located in Beroun, Czech Republic; this location is a possible site to collect characteristic dry grassland species such as *Peucedanum cervaria*.

3.2 Phylogeography of Dry Grasslands

To have a full understanding of a species, evolutionary history is necessary to determine. One of the major principles in evolution is phylogeny, the study of the changes of species or organism groups. With the constant development of evolutionary studies, one focus that has recently developed is the analysis of impacts by historical processes on current geographic distributions of species (Lowe et al., 2009). This field of study is phylogeography and it allows genetic genealogy to be determined and analyzed. As phylogeography has developed as a field, the common tools used for studies are microsatellites, specifically mtDNA in animals and cpDNA for plants (Lowe et al., 2009).

Within the Pannonian region, steppe ecosystems have persisted during the last glaciation and throughout the Holocene (Magyari et al., 2010). The phylogeographic pattern of temperate species, including those of dry grasslands, has typically been assumed to follow the "southern refugia paradigm" (Hewitt, 2000). However, there are two current hypotheses about the position of refugee areas where species survived the Late Pleniglacial Period (LPG; 24-15 kyr BP). These two hypotheses are the "southern refugia" and the "cryptic refugia" (Hewitt, 2000; Steward et al., 2009).

The "southern refugia" assumes many species survived in the Iberian, Italian, and Balkan Peninsulas. The southern refugia paradigm has three main patterns associated with different species, the grasshopper pattern, bear pattern, and the hedgehog pattern. Each is associated with a unique expansion pattern. The grasshopper is a migration from the Balkan region, the bear is from Iberia and eastern refuge, and the hedgehog is when all three regions contribute to the post-glacial spread of a species (Sommer, 2009).

As the leading hypothesis on survival during glaciation of temperate species, there have been numerous studies demonstrating the "southern refugia" post-glacial recolonization (Demesure et al., 1996; Dumolin-Lapègue et al., 1997; King & Ferris, 1998). Many of these studies have typically been conducted on temperate tree species, while grasses and flowering plants have been neglected. Yet until recently the southern

refugia hypothesis has been widely accepted as the most likely condition for all temperate plant species.

The “cryptic refugia” hypothesis supports many species surviving in numerous stable micro-refugia in the north (Steward et al., 2010). Recently fossil pollen records and mammal studies on survival during the LPG have shown taxa not considered able to survive the glaciation, did indeed survive in higher latitudes than expected in pockets of suitable habitats.

While these hypotheses attempt to explain how species adapted to interglacial time periods survived during glacial times, arctic-alpine or boreal taxa more likely followed an interglacial refugia system. This “interglacial refugia” attempts to explain how species adapted to glacial times are able to survive during interglacial periods.

Due to the persistence of dry grasslands since the Pleistocene, the eastern Balkan Peninsula has been considered an important migration corridor (Palpurina et al., 2015; Pokorný et al. 2014). However, due to the characteristics of many species to be more similar to arctic-alpine or boreal taxa there is a possibility they could follow an “interglacial refugia” pattern instead of one of these “glacial refugia” systems (Kajtoch et al., 2016).

With the modern change in understanding of the glacial landscapes, the inclusion of continental steppe-tundra-woodland habitats create a more accurate description of the structure during glacial times (Hodková et al., 2019; Janská et al., 2017). Thus, it is likely that continental steppes and grasslands were more widespread than previously believed. Current studies have shown that the climate and vegetation distributions throughout Central and Eastern Europe during the LPG are likely similar to southern Siberia currently, thus allowing distribution models to be created (Janská et al., 2017).

The grasslands of interest for this study belong to the Pannonian Region (Willner et al., 2017). Kajtoch et al. (2016) provides regions of the distribution of steppe species throughout Central Europe (see Figure 2).

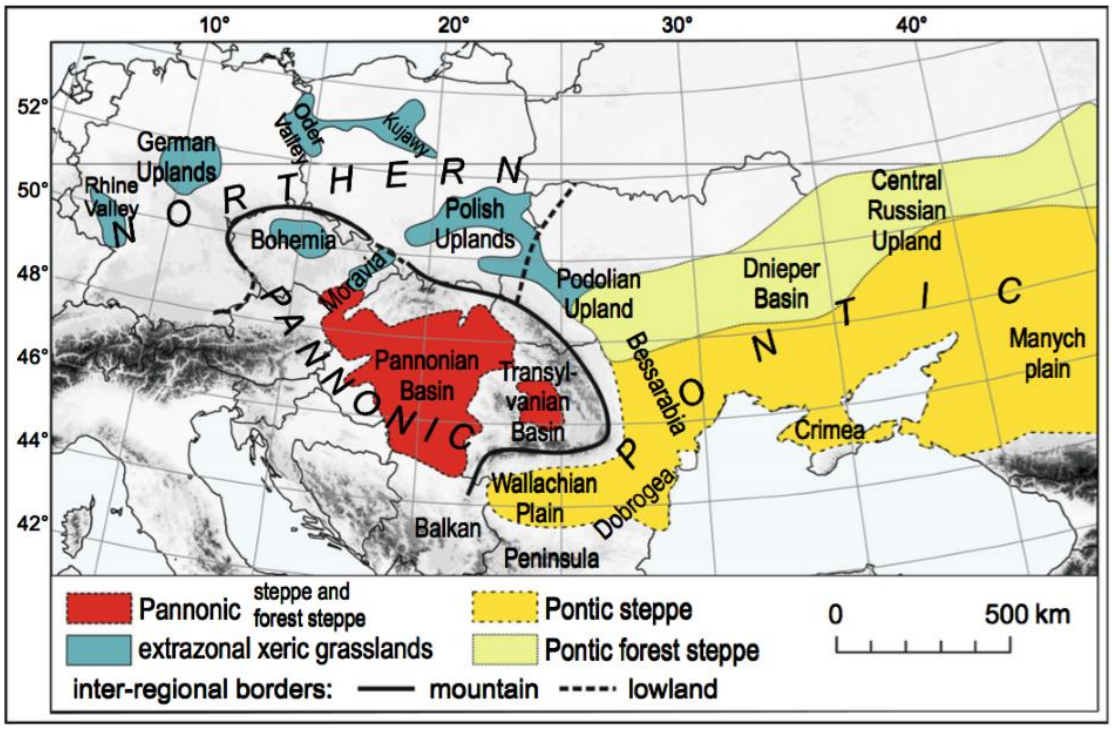


Figure 2: Simplified map of steppe and forest-steppes distribution in Pontic and Pannonian Regions as well as xeric grasslands in Eastern Central Europe. This figure was adopted from Kajtoch et al. (2016).

3.3 Risks

Human influenced threats to grasslands in Europe include changes in land use, afforestation, changes in livestock densities, and increased intensification of management (Silva, 2008; Janišová et al., 2011). These habitats are important to protect as they contain seed beds for many crops, which are genetically important for cultivations (Silva, 2008). Afforestation is another historical threat on European grasslands. Ecosystems are currently dealing with the effects of historical afforestation. *Robinia pseudoacacia* (Black Locust) is one example of a threatening species to grasslands, due to its ability to colonize open areas, especially areas with dry soils (Vítková et al., 2017). Dry and semi-dry grasslands, dry forests, and abandoned agricultural areas are some of the most often invaded habitats by Black Locusts in Europe (Vítková et al., 2017). The changes in the seed bank of these modified system is an important characteristic to consider, when addressing threats and protection. A study by Szabó & Ruprecht (2018) on afforested dry grasslands, determined that changes in the seed bank can be significant. Their suggestion for restoration is that although grassland species are able to regenerate after the removal of trees, it is necessary to provide some weed management at the start of regeneration in order to prevent nonnatives and better competitors from establishing.

Grassland ecosystems offer habitat and energy for many biological levels, specifically livestock and wild herbivores will be greatly impacted by the destruction of grasslands. Climate change is another threat to grassland ecosystems. As the seasons, temperatures, and moisture levels change, both on the global level as well as in respect to seasonality, shifts in ecosystems ranges can be seen worldwide.

To maintain dry grasslands in Europe it is necessary to understand the factors that naturally sustain their existence. The intermediate-disturbance hypothesis is typically followed by dry grasslands, demonstrating that light grazing yields the greatest richness (Škornik et al. 2010; Janišová, 2011). Some mechanisms to prevent succession from continuing in these systems is the grazing of herbivores and fire regimes. Light grazing, typical for a natural population of herbivores, has been adopted as a conservation mechanism. This allows farmers to utilize the land while maintaining the natural

landscapes. Intensive grazing, such as is typical in modern agriculture is not a viable system for conserving these grasslands. The greater level of disturbance limits the species with the ability to survive. Fire regimes are natural forces in many systems and determining their role in dry grassland requires research. Furthermore, as secondary succession begins, mechanisms to prevent development should be favored over mechanisms assisting development. Additionally, restoration techniques are possible to increase the likelihood of natural dry grassland species to establish (Pitz et al., 2019).

3.4 Characterization of *Peucedanum cervaria*

The Apiaceae family is one of the best-known families of flowering plants. The major division of this family is into three subfamilies (Hydrocotyloideae, Saniculoideae, and Apioideae) and 12 tribes common (Downie & Katz-Downie, 1996). The "umbellifer" subfamily of Apioideae display characteristic compound umbels. The fruit, without scales, contains two one-seeded mericarps suspended from a common bifurcate carpophore. The terminals arise from stylopodium common (Downie & Katz-Downie, 1996). Stipules are absent and chemical compounds of flavones, furanocoumarins, and phenylpropenes are common (Downie & Katz-Downie, 1996).

Peucedaneae are likely polyphyletic, characterized by distinct dorsal flattening of mature fruit with lateral ribs expanded into wing-like appendages, is second largest tribe, including 60 genera and 550 species (Downie & Katz-Downie, 1996).

Peucedanum cervaria is a polycarpic perennial diploid herb with reaching heights between 30-150 cm (Pladias.cz, 2019). It has a straight, oblong, stem slightly fluted at the top, with blue- green leaves, 2-3 times pinnatifid (Podesva, 2009). The small white flowers form in umbels, with 9-30 parts flowering between July and September. It is a hemicryptophyte, with ribbed elliptical to round fruits (Grulich, 1997).

Figure 3: A sample of *Peucedanum cervaria* in Czech Republic. A: A top down view of *Peucedanum cervaria*; B: Leaves of *Peucedanum cervaria* Photo by Dijana Cortan, CZ

3.5 Microsatellites

A genetic analysis will allow a deeper understanding of *Peucedanum cervaria* phylogenetically. In the age of technology, there are many techniques for studying genetic variability. Each technique has their respective advantages and disadvantages.

One such technique applies to the analysis of microsatellite markers. Microsatellites are also known as short tandem repeats or simple sequence repeats. Microsatellites contain DNA sequences sets, also referred to as motifs. These motifs are short, typically between one and five bases long, and they repeat in tandem to variable lengths. These microsatellites can be classified by the type of repeat: perfect, imperfect, interrupted, or composite (Oliveira et al., 2006). Perfect repeats are uninterrupted, while imperfect sequences are. Interrupted sequences have a pair of bases between the repeat motifs, while composite types are two different repeated sequences adjoined to each other (Oliveira et al., 2006). Some features of microsatellites that make them so intriguing and suitable for plant research are that they are co-dominant markers, their high mutation rates, and their prevalent existence within genomes.

Since microsatellites can be highly revealing about population genetic structure, they have become increasingly common in plant research. They began as a mechanism to study humans and diseases; however, they are now utilized for much more in many different fields (Oliveira et al., 2006). Microsatellites are both abundant and versatile markers within plant genomes. Typically considered neutral markers, the evolution of these markers and the forces that act on them are important aspects in understanding their significance (Li et al., 2002 & Weising et al., 2015).

Although microsatellites exist in both coding and noncoding regions of the genome, their behavior is different in each. In coding regions, microsatellites are less common than in noncoding regions and studies suggest that negative selection acts to minimize them in these regions (Li et al., 2002). A connection with triplet repeats was made to neurological diseases in humans. These triplet and hexanucleotide repeats occur more frequently than other size repeats in coding regions, likely due to mechanisms

preventing frame-shift mutations (Li et al., 2002). In noncoding regions of the genome, mutations are allowed on a greater level.

In plant species, dinucleotide microsatellites are the most common type utilized for studies (Li et al., 2002). Studies have shown that a large overall size of microsatellites are not often tolerated (Li et al., 2002). Thus, the common dinucleotides tend to have more repeats than tetra- or larger sequences. However, dinucleotide repeats tend to be more sensitive to stuttering, thus showing greater genotypic variability (Zalapa et al., 2012). Microsatellites also tend to appear in greater quantities within centromere regions (Li et al., 2002).

Microsatellites are also able to form a variety of secondary structures. This feature may be important to DNA structure, allowing the stabilization during denaturation or renaturation, or simply during transcription (Li et al., 2002). The chromosome organization may also be influenced by microsatellites, specifically through their maintenance of genomic features (Li et al., 2002). Greater allelic variation can be found in longer, perfect microsatellites (Zalapa et al., 2012). Recombination rates have been shown to be greater in microsatellites with more repeats (Li et al., 2002).

Due to the non-randomness and selectivity of microsatellites, they are excellent tools for population genetic studies (Hartl, 2000). Specific current uses for microsatellites in plant genomes include: linkage map development, parentage analysis, cultivar fingerprinting, genetic diversity studies, and gene flow and evolutionary studies (Zalapa et al., 2012).

Due to microsatellites ability to detect multiple alleles per locus they offer a specific advantage over Single Nucleotide Polymorphisms (SNPs). Transferability to related species is often considered an advantage of microsatellites, however, this is only impactful when studying particular organismal groups. Although there is a possibility of using a known microsatellite between closely related species in animals, specifically among mammals, in plants the likelihood of cross-transferability among different plant species is significantly lower (Barbara et al., 2007). The exception to this cross-transferability in plants is among many cultivar and landrace studies, which utilize the

same species, thus the microsatellites should transfer 100% of the time (Rossetto, 2007). Additional advantages microsatellites offer include: consistency, requirement for small amounts of DNA per reaction, and their ability to utilize fluorescent genotyping and multiplexing (Weising et al., 2015 & Zalapa et al., 2012).

There are some disadvantages to microsatellite usage, such as: the process time for development, the cost is not always low, and prior characterization of sequences containing microsatellites is important for primer design (Zalapa et al., 2012).In addition to the high cost and time required to develop microsatellites SNPs are typically more prevalent within a genome. In recent studies, both SNPs and microsatellites are found for target species and their prevalence and performance are compared. Many of these studies show that SNPs are more common, while both are acceptable for detecting genetic structure (Tsykun et al., 2017).

One focus of this study will be the development of microsatellite markers for *Peucedanum cervaria*. The analysis of characteristics such as heterozygosity, frequency of null alleles, and deviation from Hardy-Weinberg Equilibrium allow primers to be compared and the population structure to be determined.

4 Methodology

4.1 Study Sites and Sample Collection

The samples collected for this project came from Central and Eastern Europe: Czech Republic, Germany, Slovakia, Romania, Hungary, Austria, and Poland. Localities are shown on Figure 4 and details are provided in Appendix 1-1. All of the sites are located on natural or semi-natural dry grassland stands. Each locality has about 20 individuals collected and each individual is collected from at least 5m from each other. Exact totals of individuals collected are shown in Table 1 below. A total of ten populations were selected for this study. Among these ten populations, a total of 191 individuals were collected and analyzed.

Table 1: Total number of individuals sampled from each sample site

Population	Individuals
11	20
169	20
2	20
54	19
64	18
10	19
15	20
31	20
36	20
9	15

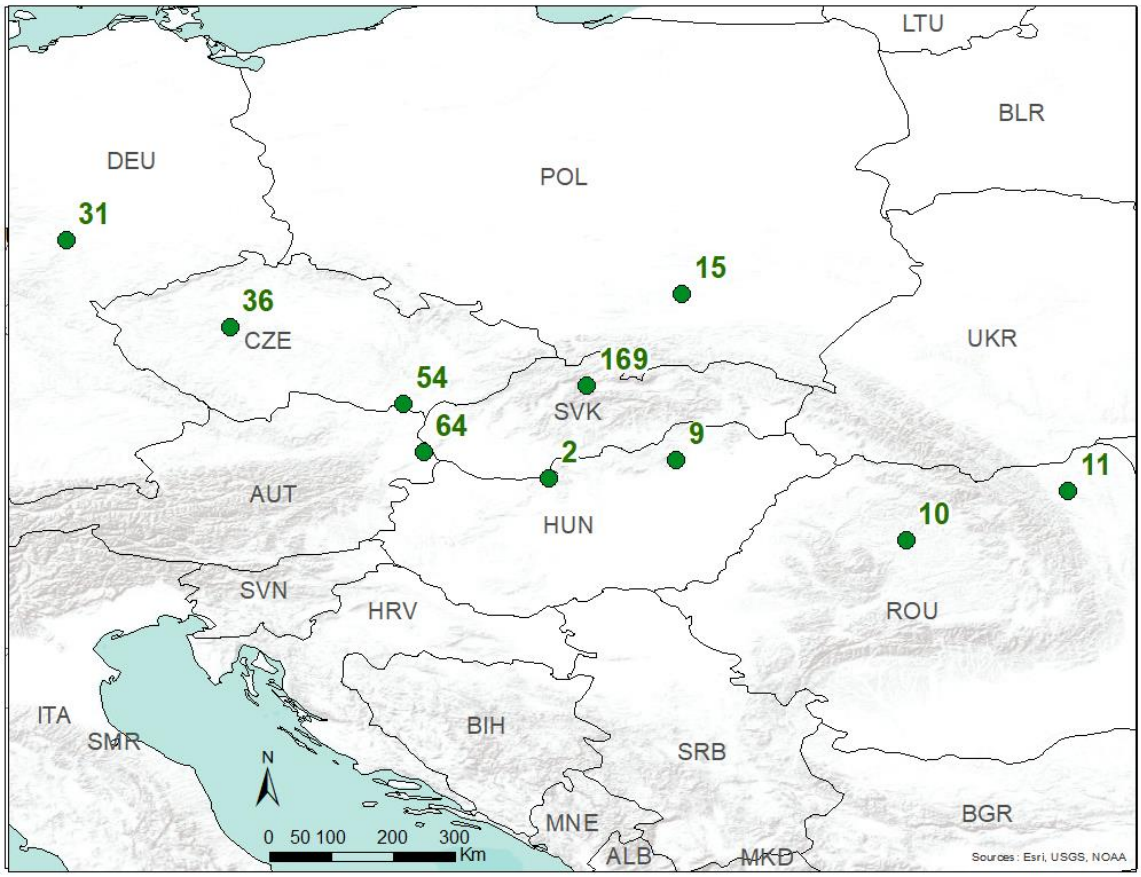


Figure 4: Map of ten localities of *Peucedanum cervaria.* Each dot indicates a population collected for this study and is labelled by the locality number.

4.2 DNA extraction, PCR, and Fragment Analysis

In order to select suitable candidate primers, three benchmarks are completed. They are: gel electrophoresis analysis, fragment analysis, and a variability test. In order to complete each step DNA extraction and a PCR, polymerase chain reaction, must be completed.

Collected leaves are dried and stored in silica gel. Total genomic DNA was extracted using the DNeasy 96 Plant Kit protocol with the adjustment in the last step the AE Buffer, which was originally 2x 100 µL was adjusted to 2x 50 µL as well as these two elutions were completed into the same eppendorf (Qiagen, 2018). The primers used for the PCR were designed for this steppe project, thus are being tested for extension on *Peucedanum cervaria. F*or the complete table of primers tested see Appendix 1-2. Using the extracted DNA, a PCR reaction was conducted using the adjustments to the procedure outlined by Schuelke in 2000. The volumes were optimized for this study to volumes of 5µL of QIAGEN PCR Master Mix (Qiagen), 0.01µL of forward primer, 0.4 µL of reverse primer, 0.4 µL of fluorescent dyed primers (6-FAM, VIC, PET, NED; Applied Biosystems), 3.19 µL of distilled water, and 1 µL of DNA per well. The PCR plate was set up for seven samples and one negative control to be tested, from three sample populations. The reaction cycle was: 95 ºC for 15 minutes, followed by 25 cycles of 95 ºC for 30 seconds, 55 ºC for 30 seconds, and 72 ºC for 2 minutes, which was followed by another 10 cycles of 95 ºC for 30 seconds, 50 ºC for 30 seconds, and 72 ºC for 2 minutes, and finished with an elongation step of 72 ºC for 10 minutes.

The PCR products were inspected on a 2% agrose gel for quality and size of bands. In order to be considered a successful primer pair at least five of the seven individuals tested must have shown one solid band on the gel. If no bands appeared in any individual, the primer was considered eliminated due to no amplification. If two or more of the individuals presented two or more bands, the primer was considered eliminated due to multiple alleles. If four or less individuals presented one band, the primer was considered eliminated due to inconsistent amplification.

Those primers considered successful during the gel electrophoresis stage were able to move into the fragment analysis stage. During this stage, the previously outlined process for DNA extraction and PCR were completed and fragment analysis was run following the protocol outlined by Mandák et al., 2016. One μL of PCR product was mixed with 0.2 μL of GeneTrace 500 LIZ (Carolina Biosystems) and 12μL of FORMAMID (Applied Biosystems); which is then denatured at 95º C for 10 minutes in the thermocycler and capillary electrophoresis is run on ABI3500 genomic analyzer. GeneMarker version 2.4.0 (SoftGenetics, USA) was then used to determine allele sizes. Allele sizes are then used to determine success of the primers at this stage. Seven samples representing the three sample populations were analyzed for quantity and size of peaks. During this stage if more than two samples showed a lack of peaks, meaning no peaks present, the primer was eliminated. Only those primers which were considered successful during this stage were tested on the full set of individuals from all three populations.

Testing the full twenty individuals from all three sample populations used the same protocol outlined above for the representative samples. Success was determined by the requirement for each individual should have a maximum of two true peaks. True peaks were determined by the removal of stutter bands as well as -A alleles. The strength of a peak should also be higher than 100 to be considered a true peak. The final condition of true peaks used during this analysis was the allele size must be greater than 100 base pairs long, in order to eliminate primer dimers. For a primer to be considered successful at least fifteen of the twenty individuals per population test should meet the two true peak requirement. Only those primers that showed consistent amplification and a degree of variability were used within the Multiplex, created in the following steps.

4.3 Multiplex Creation & Optimization

After an analysis of the size ranges and variability for each primer a multiplex was created consisting of six primers. The six primers chosen were those which showed consistent extension and a degree of variability during the previously completed primer testing phase (shown on Table 2). Their fluorescent dyes were chosen based on their size range to ensure clear peak recognition and seven samples were used to optimize the concentrations of each primer. The optimized concentration of 0.2 μM was used for each forward and reverse primer in the Multiplex and the PCR reaction conditions were adjusted to: 95 ºC for 15 minutes, followed by 35 cycles of 95 ºC for 30 seconds, 55 ºC for 30 seconds, and 72 ºC for 2 minutes, and finished with an elongation step of 72 ºC for 10 minutes.

All 10 population samples were processed with the optimized PCR reaction and the sequencing protocol outlined in Section 4.2. GeneMarker version 2.4.0 was used to determine allele sizes for each sample and the multiplex analysis followed.

4.4 Data Analysis

4.4.1 Population Genetic Parameter Estimates

The six chosen primers were analyzed using the PopGenReport and the diverRsity packages in R. Data from three test populations, each with 20 samples, were used. These R packages were used to determine their expected and observed heterozygosity (H_e and H_O), the frequency of null alleles present for primer (B), and determine if there was a significant deviation from Hardy-Weinberg Equilibrium (HWE), using a Fischer Exact Test.

Our full sample set, ten populations with about 20 individuals from each collected from Central Europe, were tested with the Multiplex for polymorphism. The allele sizes were determined with GeneMarker version 2.4.0 (SoftGenetics, USA). Expected and observed heterozygosity, frequency of null alleles, and the deviation from Hardy–Weinberg Equilibrium were calculated using the PopGenReport and the diverRsity packages in R. For the specific R code refer to Appendix 1-3.

4.4.2 Population Structure Analysis

Bayesian model-based clustering was determined with the STRUCTURE version 2.3.3 for the microsatellite data (Pritchard et al., 2000). Clusters numbers were set from K=1 to K=10 for the complete dataset. Each K was performed with at least 10 replicates. Using STRUCTURE HARVESTER the data was visualized and peak clusters were chosen (Earl and vonHoldt, 2012). Utilizing only the chosen clusters, CLUMPP and DISTRUCT were run to visualize the population structure. CLUMPP aligns all the analyses and DISTRUCT visualizes it. The processed STRUCTURE data was visualized on a map of Central Europe which allowed groups and patterns to be discerned.

5 Results

5.1 Microsatellite Development

Overall, 75 primers were tested for extension and variability with *Peucedanum cervaria* samples (Appendix 1-2). Of these 75 primers tested, 29 were eliminated from gel electrophoresis analysis, 15 were eliminated during the first round of sequencing, and 12 additional primers were eliminated during the second round of sequencing. Elimination during the gel electrophoresis phase was due to no amplification, a lack of consistent amplification among all populations tested, or presence of multiple alleles. A total of eleven of the 29 eliminated from this phase were eliminated due to inconsistent amplification. Seven of the 29 were eliminated due to multiple alleles shown on the gel by multiple bands. The remaining eleven, a total of 15% of all tested primers, were excluded due to amplification failure. During the first round of sequencing 15 primers were eliminated due to no peaks shown in the tested samples. During the second round of sequencing, which tested the full 20 individuals from three populations, elimination was based the presence of three or more peaks consistently in the sample set.

From the remaining 19 primer pairs tested, 6 lacked consistent amplification thus were eliminated. Of the remaining 13 primers sets, only 6 showed the required extension and variability to be utilized in a multiplex. For additional information on why each primer was eliminated refer to Appendix 1-4.

Primers chosen for the *Peucedanum cervaria* multiplex are: PC18, PC19, PC30, PC58, PC68, and PC73. Their repeats are $(AG)_{18}$, $(AG)_{18}$, $(AG)_{13}$, $(AT)_{12}$, $(AC)_{12}$, and $(AC)_{12}$, respectively. Table 2 contains specific information on each multiplex primer, including their sequences.

Table 2: Characteristics of six polymorphic microsatellites loci of *Peucedanum cervaria*

Locus	Primer sequence (5'-3')	Repeat motif	Allele size range (bp)	Fluorescent label	Concentration in multiplex reaction (μM)
PC68	F: ACCACTGACCCTAAACCTCG R: TGCGGGCTGGTAGAACTTTC	$(AC)_{12}$	166-192 (26)	6-FAM	0.2
PC58	F: GAGACGTACACTGAGATTGGG R: ATGGTCTCGTGTACTGTGGG	$(AT)_{12}$	324-332 (8)	6-FAM	0.2
PC73	F: TGATCCACCAAGAAGCAGATG R: TGCACCTCCAGTCTTCCATC	$(AC)_{12}$	195-228 (33)	VIC	0.2
PC19	F: TCTACATTCACCTCATCCTCCC R: GGCCCAGTAACTCCATGAAC	$(AG)_{18}$	262-300 (38)	NED	0.2
PC30	F: ACGATTCGATCCGGAAATCAC R: GGAGGTAAGTTAAATCGGGCTC	$(AG)_{13}$	373-386 (13)	NED	0.2
PC18	F: GTTGATCTGAGCTTCCCTGG R: GAACCATGCCACACAAGTCC	$(AG)_{18}$	245-320 (75)	PET	0.2

Microsatellite summary statistics were calculated for each locus, including number of alleles per locus (N_A), the number of null alleles per locus (B), Weir and Cockerham's parameter (f), as well as expected and observed heterozygosity (H_e & H_o) and the deviation from Hardy-Weinberg equilibrium (HWE). Hardy-Weinberg equilibrium used 9999 repetitions and F_{IS} utilized 1000 iterations for calculations. Overall averages of genetic diversity for each population are in Table 3; while the complete data on each locus per population are presented in Table 4. Overall, 121 alleles for 6 microsatellite loci were identified, with an average allelic richness of 5.85. The F_{IS} value determined all three tested populations had high average inbreeding coefficients (2: 0.3008, 10: 0.2675, 54: 0.2419). The highest F_{IS} values were shown with PC30, which were consistently over 0.39. The lowest F_{IS} values were shown with PC68, which were consistently less than 0.02. Averaged expected heterozygosity (H_e) ranged from 0.47 to 0.53; while average observed heterozygosity (H_O) ranged from 0.65 to 0.72. Of our 6 primers, null alleles were sometimes significantly present in PC18, PC58, and PC30 (see Table 4 for details).

Table 3: Average summary statistics on successful primers separated by population

Pop	N_O	N_A	A_R	HWE	F_{IS}
2	20	42	6.12	10^{-6}	0.3008
10	20	37	5.32	10^{-6}	0.2675
54	20	42	6.11	10^{-6}	0.2419
Average	20	40.33	5.85	10^{-6}	0.27

Pop - Population Number; N_o - Number of samples; N_A - Number of Alleles; A_R – Allelic Richness; HWE – p value of deviation from Hardy-Weinberg Equilibrium test; F_{IS} – Inbreeding Coefficient

Table 4: Detailed summary statistics on microsatellite loci separated by population.

Pop 2	PC68	PC73	PC19	PC18	PC58	PC30	Overall
N_O	20	20	20	20	19	19	19.67
A	7	6	9	10	4	6	42
%	100	54.55	64.29	45.45	100	60	70.72
A_R	5.9	5.45	8.46	8.19	3.87	4.88	6.12
H_o	0.75	0.45	0.85	0.25	0.37	0.37	0.51
H_e	0.75	0.69	0.76	0.84	0.69	0.61	0.72
HWE	0.06	0.192	0.997	0	0.012	0.004	0
F_{is}	0	0.3466	-0.1221	0.7033	0.4669	0.3995	0.3008
F_{is}_Low	-0.286	0.0784	-0.264	0.4697	0.1304	-0.0017	0.1602
F_{is}_High	0.2835	0.6258	0.028	0.89336	0.7771	0.7469	0.4244
B	0	0.2096*	-0.0575	0.5423*	0.3045*	0.2496*	
Pop 10	PC8	PC73	PC19	PC18	PC58	PC30	Overall
N_O	19	20	19	20	19	17	19
A	3	6	7	11	3	7	37
%	42.86	54.55	50	50	75	70	57.07
A_R	2.88	5.76	5.79	8.62	2.97	5.88	5.32
H_o	0.47	0.75	0.63	0.6	0.26	0.12	0.47
H_e	0.48	0.78	0.59	0.75	0.47	0.8	0.65
HWE	1	0.654	0	0	0.077	0	0
F_{is}	0.0144	0.0431	-0.0729	0.2	0.4395	0.8528	0.2675
F_{is}_Low	-0.3351	-0.214	-0.3304	-0.0038	0.0138	0.6183	0.1493
F_{is}_High	0.4296	0.3183	0.2226	0.421	0.8128	1.0143	0.3958
B	0.0072	0.0220	-0.0351	0.1111*	0.2816*	0.7433*	
Pop 54	PC68	PC73	PC19	PC18	PC58	PC30	Overall
N_O	20	20	20	20	15	19	19
A	5	9	8	13	3	4	42
%	71.43	81.82	57.14	59.09	75	40	64.08
A_R	4.61	7.96	6.63	11.02	2.95	3.48	6.11
H_o	0.7	0.85	0.6	0.6	0.27	0.16	0.53
H_e	0.61	0.83	0.73	0.9	0.58	0.54	0.7
HWE	1	0.007	0.202	0.868	0.009	0	0
F_{is}	-0.1429	-0.0287	0.1809	0.3333	0.5402	0.7054	0.2419
F_{is}_Low	-0.3118	-0.1908	-0.0375	0.0888	0.122	0.3664	0.141
F_{is}_High	0.0446	0.1638	0.401	0.5673	0.9144	1.02	0.3369
B	-0.0666	-0.0141	0.0994	0.2*	0.3700*	0.5449*	

Summary statistics, including heterozygosity's, null alleles, allele range, and deviation from Hardy-Weinberg equilibrium, were calculated for each population. Hardy-Weinberg equilibrium used 9999 repetitions and F_{IS} utilized 1000 iterations for calculations. Overall calculations per population are shown in Table 5 and null allele frequencies are shown in Appendix 1-5. No significant amounts of null alleles were identified among any of the populations. A total of 415 alleles at the 6 microsatellite loci were identified, averaging 41.5 alleles per locus. While the allelic richness averaged 5.58. Only one population average did not vary significantly from Hardy-Weinberg equilibrium with a value of 0.211. Appendix 1-5 shows each populations Hardy-Weinberg deviation by locus, of which most did not vary significantly from Hardy-Weinberg equilibrium. A majority populations with significant deviations were typically only not in Hardy-Weinberg equilibrium for PC30, PC18, or PC19, or a combination of these loci. Almost all populations showed a high inbreeding coefficient in their average. However, some markers, PC68 and PC19, consistently showed low inbreeding, while others markers, PC58 and PC30, showed higher inbreeding coefficients. Population 169, located in the Tatras Mountains of Slovakia shows the lowest allelic richness of all populations as well as a higher than average inbreeding coefficient compared to the tested populations. The Czech Republic population 54 shows the highest allelic richness, although its overall inbreeding coefficient is on the larger scale. Population 10 of Romania has the highest inbreeding coefficient as well as a relatively low allelic richness. Population 15 of Poland has the smallest inbreeding coefficient and a relatively low number of allelic richness. Population 31 located in Germany shows average allelic richness and inbreeding.

Table 5: Average summary statistics separated by population

Country	Pop	N_O	N_A	A_R	H_o	H_e	HWE	F_{IS}
Romania	11	20	41	5.81	0.68	0.73	0.211	0.0695
Slovakia	169	18	32	4.15	0.46	0.55	10^{-6}	0.163
Hungary	2	19.5	51	6.76	0.64	0.73	10^{-6}	0.1332
Czech	54	17.83	53	7.05	0.63	0.75	10^{-6}	0.1678
Austria	64	17.67	50	6.33	0.57	0.69	10^{-6}	0.1704
Romania	10	17.17	34	4.65	0.48	0.63	10^{-6}	0.2366
Poland	15	19.67	31	4.35	0.63	0.61	10^{-6}	-0.0311
Germany	31	19.17	40	5.34	0.57	0.67	0.001	0.1563
Czech	36	17.83	42	5.57	0.52	0.64	0.007	0.1846
Hungary	9	13.83	41	5.79	0.54	0.62	10^{-6}	0.1258
Average		14.41	41.5	5.58	0.572	0.662	0.022	0.137

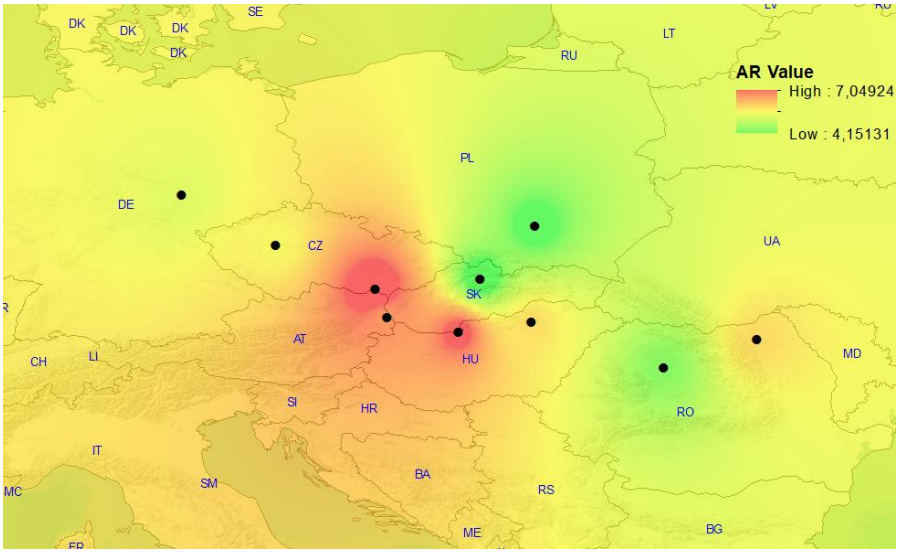
Pop - Population Number; N_o - Number of samples; N_A - Number of Alleles; A_R – Allelic Richness; HWE – p value of deviation from Hardy-Weinberg Equilibrium test; F_{IS} – Inbreeding Coefficient

By visualizing the calculated population genetic parameters with the geographic area of focus, a clearer understanding of the *Peucedanum cervaria*'s distribution is determined. A map of interpolated allelic richness (A_R) (Figure 5A) and a map of the interpolated inbreeding coefficient (Figure 5B) were created using the inverse distance weighting interpolation tool in ArcMap 10.6.1 (ESRI, Redlands, CA, USA).

On both parameters show a clear separation between the populations located within the Pannonian Basin and the Bohemia Massif and those surrounding these regions. This clear separation shows highest allelic richness within the Pannonian Basin as well as a very low inbreeding coefficient on the northeastern side of Sudetes.

The populations within the Pannonian Basin show a high level of allelic richness and a relatively high inbreeding coefficient. Population 10 in Romania and population 169 of Slovakia both show a very high inbreeding coefficient while containing low allelic richness.

A



B

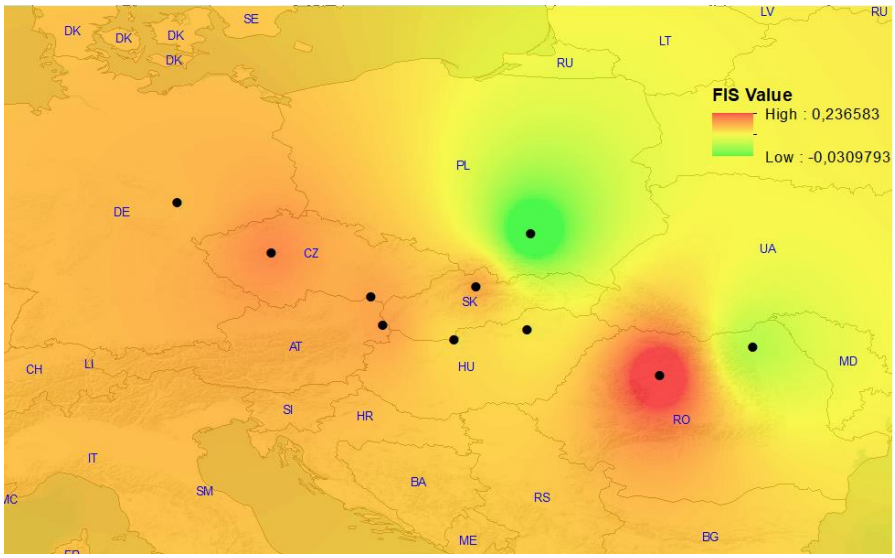


Figure 5: Map of overall A: allelic richness and B: inbreeding coefficients for *Peucedanum cervaria* across Central and Eastern Europe; visualized with the inverse distance weighting interpolation method in ArcGIS (ESRI, Redlands, CA, USA).

5.2 Population Structure

By using STRUCTURE version 2.3.3, the optimal separation of clusters was determined to be K = 5 and K = 8, due to their high Delta K values and their low ranges of L(K) (see Appendix 1-6). Both K groupings were visualized with their geographical distribution as well as a bar plot representing each individual for all 10 populations tested (see Figure 6 and 7).

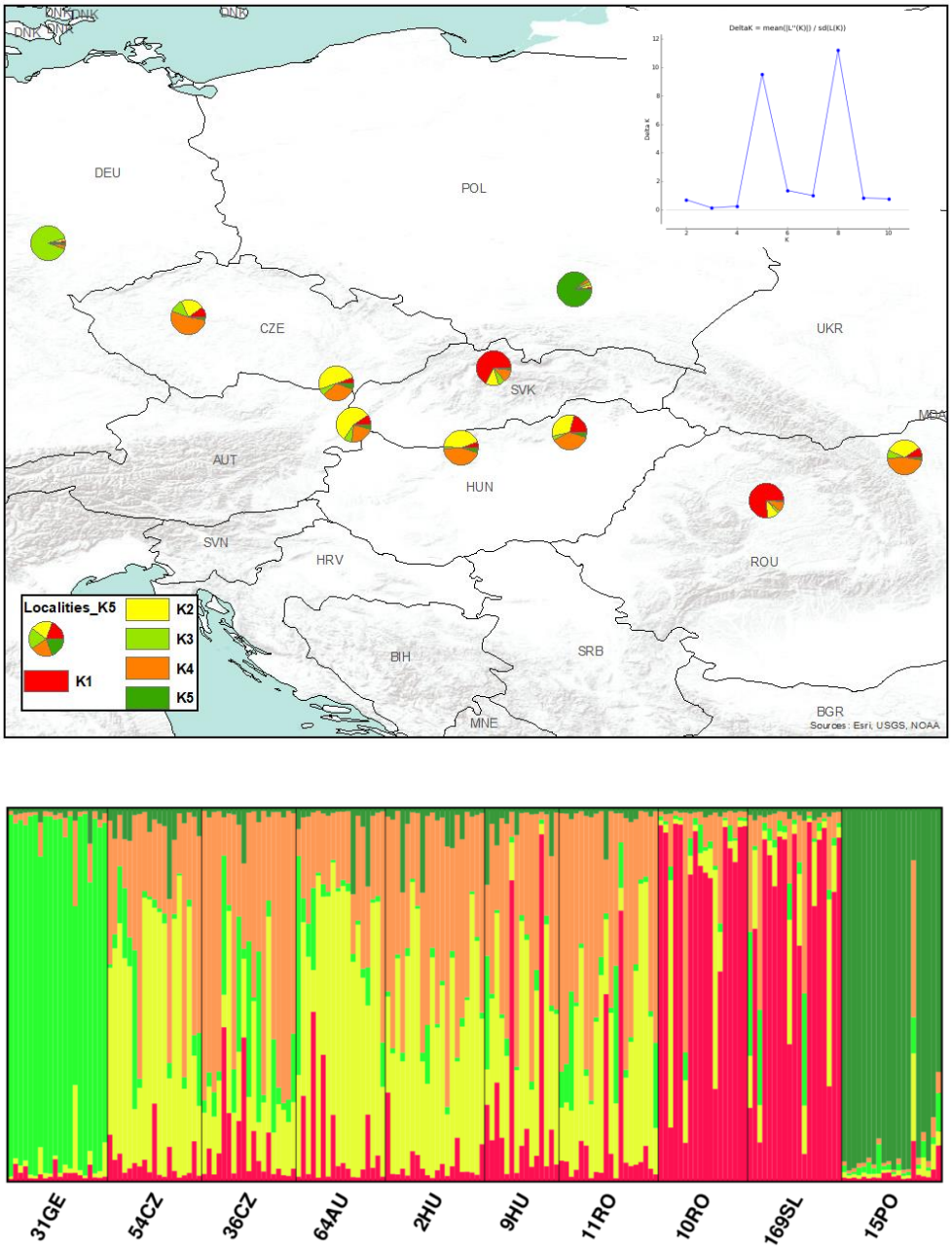


Figure 6: A: Geographical distribution of *Peucedanum cervaria* populations into 5 genetic clusters across Central Europe (pie chart contains proportions of individuals from each population assigned to each cluster by STRUCTURE) The Delta K graph in the top right corner shows 5 and 8 clusters provide the most information about the structure of the data. B: Bar chart representing the proportion of each cluster which each individual in a population belongs to.

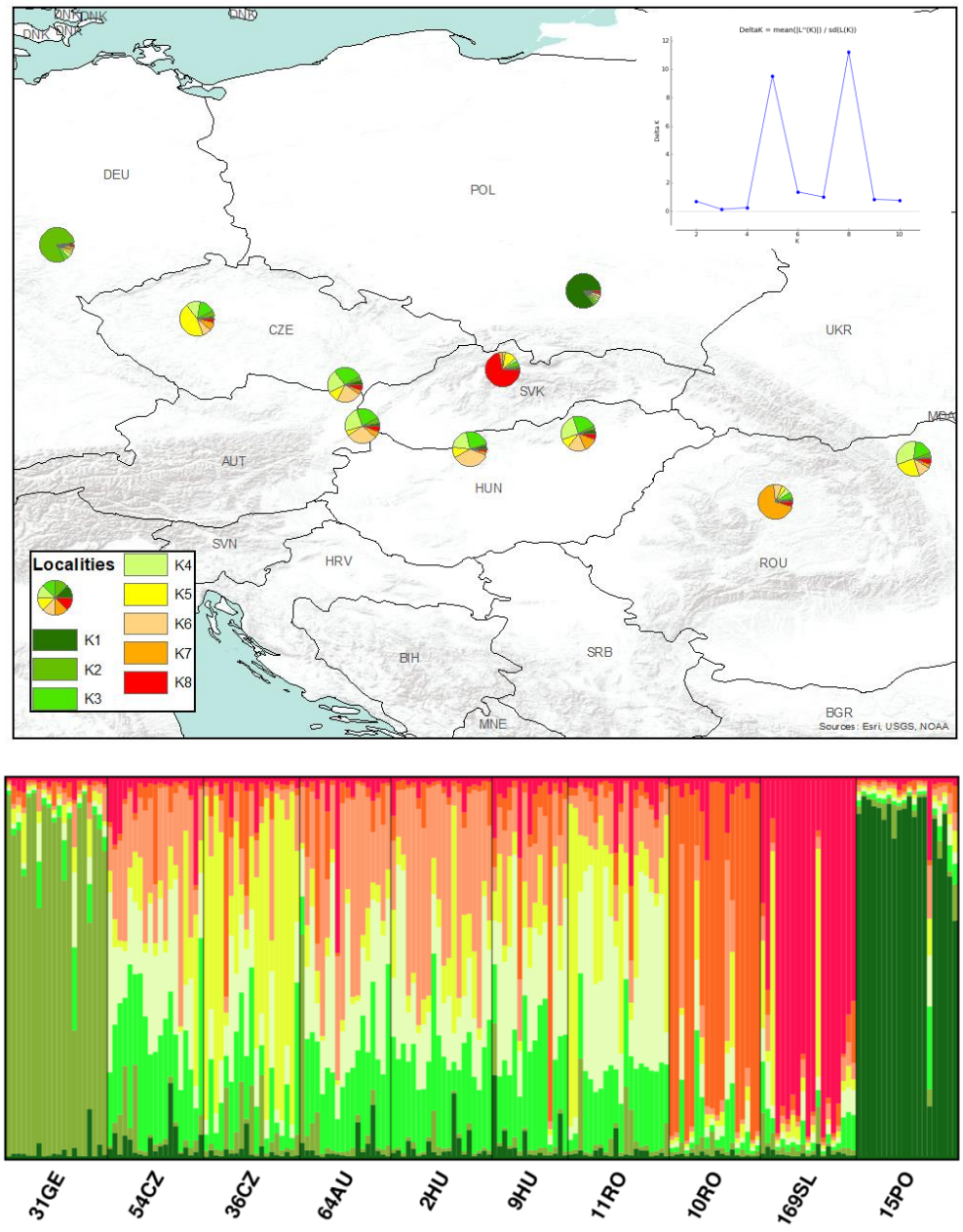


Figure 7: A: Geographical distribution of *Peucedanum cervaria* populations into 8 genetic clusters across Central Europe (pie chart contains proportions of individuals from each population assigned to each cluster by STRUCTURE) The Delta K graph in the top right corner shows 5 and 8 clusters provide the most information about the structure of the data. B: Bar chart representing the proportion of each cluster which each individual in a population belongs to.

6 Discussion

6.1 Microsatellite Data Analysis:

Genetic diversity within and across all populations are expressed with null allele frequencies, inbreeding, heterozygosity (expected and observed), number of alleles as well as allelic richness. Frequencies calculated with R showed that primers PC18, PC58, and PC30 have high null allele frequencies (shown in Table 3). However, when determined for all populations no excess presence of null alleles was found for any locus (shown in Appendix 1-5). Thus, the high frequency of null alleles found in primers PC18, PC58, and PC30 during initial microsatellite testing is not impactful on our tested populations. All population averages showed significant deviation ($p < 0.05$) from Hardy-Weinberg Equilibrium during our analysis of primers, which indicates at least one assumption is violated. However, during the analysis of each population by microsatellite loci significant deviation from Hardy-Weinberg Equilibrium was not consistently found. Population 11 was the only population to show on average no deviation from Hardy-Weinberg Equilibrium during the analysis of all populations. As well as all populations showing a majority of individual levels of deviation as non-significant. Due to H_o consistently being lower than H_e (the only exception being population 15 from Poland) in our populations, the difference is likely due to inbreeding from self-pollination. Additional indicators of genetic diversity, H_e and A_R ranged from 0.55-0.75, with the average of 0.662 and from 4.15 to 7.05, with an average of 5.58, respectively.

When interpreting the microsatellite analysis in connection with geographic distribution it becomes clear that the populations in the Pannonian Basin and the Bohemian Massif are connected. Due to the high level of allelic richness distributed throughout this region it is clear that the population sizes are considerably large and long-surviving. If the relatively high inbreeding coefficients of this region are also considered it becomes clear that self-pollination is likely a pollination strategy present within these populations.

An additional factor to consider for the Slovakia population with the highest allelic richness, as well as a high inbreeding coefficient is that this population is separated from the other populations of the Pannonian region due to a geographic barrier. This is likely the case based on the topography of the sample location.

The highest inbreeding coefficient is seen in the Romania population 10, which is located within the Carpathians. This population has a high inbreeding coefficient as well as a low allelic richness which implies it is a relatively isolated population and due to its geographic position it is likely a refugia site from glacial times, which is not fully connected with the other populations of the Pannonian region.

6.2 Population Structure:

By comparing the proportions each population has in each cluster it is possible to determine similarities between populations as well as infer information about the post-glaciation migration. Before making inferences about migrations it is necessary to compare the analyzed population locations with the locations provided by the current species distribution map in order to know if each gene flow is possible between the populations (see Figure 8).

The geographic distribution of the populations represented with five clusters show four unique characteristics. First, population 31 located in Germany is genetically unique from all other populations. The same is true of Poland's population 15. Additionally, Slovakia's population 169 on the Čebrat Hill and Romania's population 10 from Cluj-Napoca are less similar to the remaining populations of Central Europe than they are to each other. The majority of the populations throughout Central Europe show inclusion to multiple genetic clusters and this ratio seems to spread from Southeastern Europe into Central Europe beginning at the Carpathian Mountain Range, encompassing the entire Pannonian Region.

When analyzing the data with more clusters, further divisions between the populations became visible. Germany remains a unique genetic cluster, which when comparing this information with *Peucedanum cervaria*'s species distribution map is determined to be reasonable. The species distribution map (Figure 8) shows that Germany's population of *Peucedanum cervaria* is not connected to the major distribution across Europe. Additionally, it is located north of the Ore Mountains, the presumed geographic barrier for temperate species.

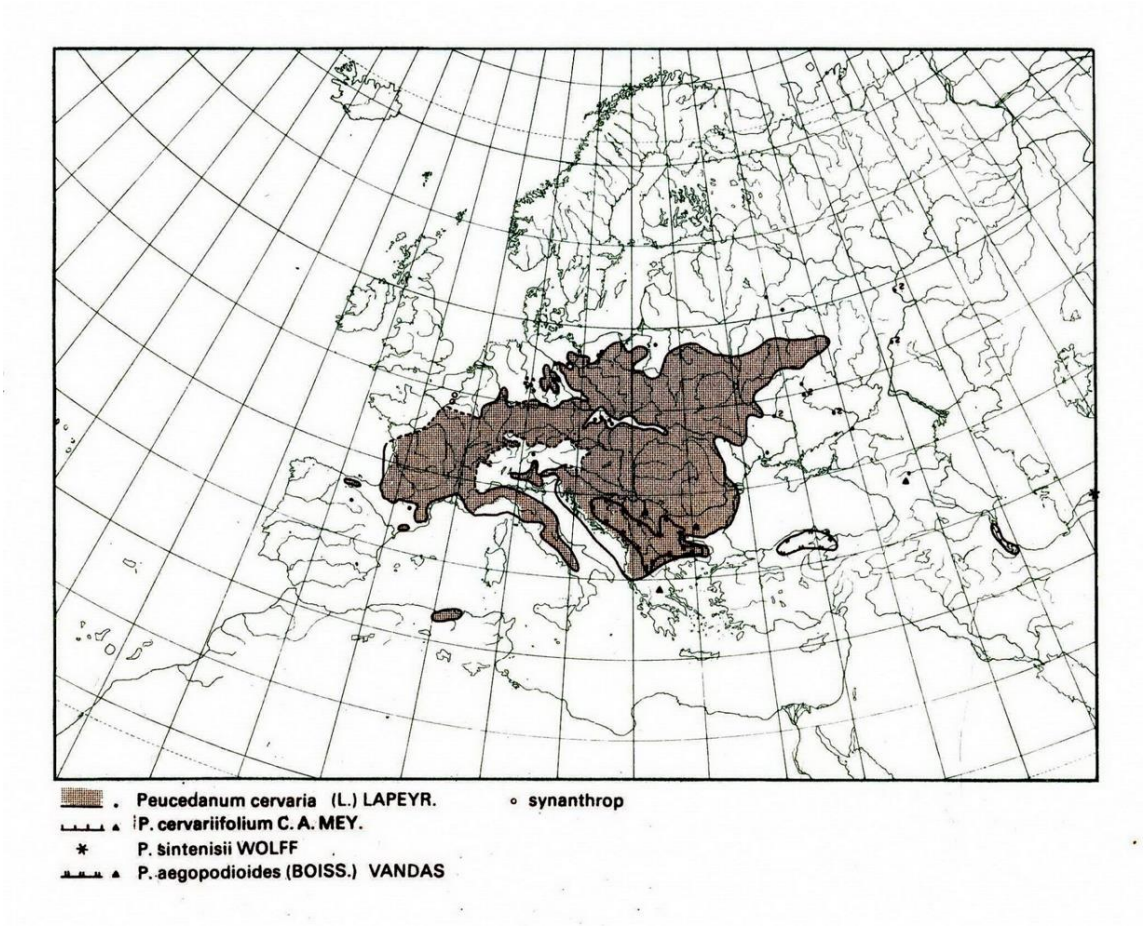


Figure 8: Peucedanum cervaria Current Distribution Map adopted from Meusel, H and Jäger, E http://chorologie.biologie.unihalle.de//choro/map.php?SearchString=peucedanum+cervaria&Spec=Peuced anum+cervaria&Text=322&Map=319a&Volume=II&Sort=1&Lang=E

Poland's population 15 is unique from all other tested populations, however, it represents the North Eastern genetic group when compared to the species distribution map (Figure 8). Thus it is also likely effected by the geographic barrier outlined on Figure 2 by Kajtoch et al. (2016), which extends around the Sudetes and southeastern around the Carpathians. The clusters shown on Figure 6 and Figure 7, show that the Carpathians are not a geographic barrier for *Peucedanum cervaria* but the Sudetes are likely a limiting feature. With a more comprehensive collection of populations this geographic barrier can be definitively determined for *Peucedanum cervaria*.

With 8 clusters representing our Central European genetic groups a further differentiation of Slovakia's population 169 on the Čebrat Hill and Romania's population 10 from Cluj-Napoca is visible. Not only are both populations separated from the remaining populations in Central Europe but they are also separated from each other in this more detailed view. Both Slovakia's population 169 and Romania's population 10 are located in the foothills within the Pannonian Region, thus they were likely connected by the Carpathians during the LPG and represent relic populations.

This differentiation into a unique cluster of population 10 during this eight cluster analysis did not show complete separation. Population 10 still contains greater than 10% of the Central European cluster classifications. Therefore, this population is likely connected to the Central European distribution; however, shows minimal influence on the migration.

The Central European cluster refers to the populations which are most similar to each other, populations 36, 54, 64, 2, 9 and 11. These populations spread from North Romania into and across Hungary, along the Austria-Slovakia border and throughout Czech Republic. Due to the consistent genetic similarity of these populations, it is clear that the cryptic refugia hypothesis was followed post-glacially. The cluster analysis shows that there similar proportions of four main clusters, K3, K4, K5, and K6. These clusters vary in which has a majority but all four of these clusters combined in each population represents a majority. In populations located in Western Czech Republic, the current northern and western limit of our Central European tested populations, more individuals are within cluster K5. This distribution is most similar to the proportions exhibited by Romania's eastern population, showing that this Romanian population is a part of the connected distribution of *Peucedanum cervaria* into Central Europe. Since this Romanian population is located on the northern side of the Carpathians it is possible to conclude that the Carpathian Mountains were not a barrier for *Peucedanum cervaria*.

Although temperate tree species have been the focus of phylogeographic for over two decades, non-tree species are still seldom studied (Daneck et al., 2016). Most recent non-tree studies have found a unique pattern of expansion, some suggest that the

southeastern connections with Central Europe are influential to allowing migration post-glacially while others support this cryptic refugia survival pattern (Daneck et al., 2011; Dvořáková et al., 2010; Vrancken, et al., 2009). The Carpathians and the Bohemian Massif have been established as important refugia sites for glacial survival (Daneck et al.). Connections within the Carpathian Region allowed for survival during glacial times and is currently not a barrier for migration of *Peucedanum cervaria*. The Carpathians, although a unique geographic feature of Central Europe, have variable levels of impacts on species distributions.

7 Conclusion

The microsatellite analysis showed consistent genetic variability among populations. With the analysis of loci and population structure of *Peucedanum cervaria* it is clear that an interglacial refugia paradigm is not followed. Thus determining the post-glacial migration route is paramount to understanding *Peucedanum cervaria'*s distribution.

Six microsatellite loci where able to be developed for *Peucedanum cervaria* during this study. From these loci, eight distinct genetic groups were able to be identified and their geographic distribution mapped. With this distribution, inferences on the post-glacial recolonization of *Peucedanum cervaria* has been determined to follow the cryptic refugia paradigm. More specifically, the conclusion that *Peucedanum cervaria* populations in Central Europe persisted in relic populations which are still present in Central and Eastern Europe and are expanded northward until the Bohemian Massif is possible. A northern corridor of expansion is rejected as a hypothesis due to the clear differentiation of German and Polish populations. The southern corridor hypothesis is also rejected due to the genetic similarities within the Pannonian Basin and a lack of information on the Balkan Region.

The lack of data representing the Balkan region means entirely excluding the possibility that the lineage of Central Europe was impacted by a migration from this southern route is not possible. With further studies including a greater geographic region, it will be possible to determine if the Balkan region played any role in the current distribution of *Peucedanum cervaria.* However, the distinct similarities between the Central Europe genetic cluster, as well as the unique populations of Slovakia (population 169) and Romania's (population 10) located in the foothills make the cryptic refugia hypothesis the most likely survival pattern of *Peucedanum cervaria.* Additionally, the Carpathian Mountains are clearly not a migration barrier population 11 located in eastern Romania shares a similar genetic structure to the Central European cluster.

The comprehensive study will be completed as a part of the ongoing Project: What is the origin of dry grasslands in Central Europe? A synthesis of comparative phylogeography and palaeodistribution modeling (B. Mandák, personal communication). To further this analysis a second multiplex, with up to 8 microsatellite markers could be developed and utilized to expand the detailed genetic analysis of *Peucedanum cervaria*. It could also be used in collaboration with this multiplex to conduct the comprehensive study on Europe in order to create a clearer image of how this species recolonization Central Europe post-glacially.

8 References

Barbara, T., Palma-Silva, C., Paggi, G.M., Bered, F., Fay, M.F., Lexer, C. (2007). Cross-species transfer of nuclear microsatellite markers: potential and limitations. *Molecular Ecology,* 16, 3759-3767. https://doi.org/j.1365-294X.2207.03439.x

Binney, H., Edwards, M., Macias-Fauria, M., Lozhkin, A., Anderson, P., Kaplan, J. O., Andreev, A, Bezrukova, E, Blyakharchuk, T, Jankovska, V, Khazina, I, Krivonogov, S, Kremenetski, K, Novenko, E, Ryabogina, N, Solovieva, N, Willis, K, Zernitskaya, V. (2017). Vegetation of Eurasia from the last glacial maximum to present: Key biogeographic patterns. *Quaternary Science Reviews*, 157, 80–97. https://doi.org/10.1016/j.quascirev.2016.11.022

Daneck, H., Fér, T., Marhold, K. (2016). Glacial survival in northern refugia? Phylogeography of the temperate shrub *Rosa pendulina* L. (Rosaceae): AFLP versus chloroplast DNA variation. *Biological Journal of the Linnnean Society,* 119, 704-718.

Demesure, B., Comps, B. and Petit, R. J. (1996), Chloroplast DNA Phylogeography of the Common Beech (*Fagus sylvatica* l.) in Europe. *Evolution*, 50, 2515-2520. doi:10.1111/j.1558-5646.1996.tb03638.x

Downie, S. R., and Katz-Downie, D. S. (2012). A Molecular Phylogeny of Apiaceae Subfamily Apioideae : Evidence from Nuclear Ribosomal DNA Internal Transcribed Spacer Sequences. *American Journal of Botany*, 83(2), 234–251.

Dumolin-Lapègue, S., Demesure, B., Fineschi, S., Le Come, V. and Petit, R. J. (1997). Phylogeographic Structure of White Oaks Throughout the European Continent. *Genetics,* 146, 1475-1487

Dvořáková, H., Fér, T. and Marhold, K. (2010). Phylogeographic pattern of the European forest grass species *Hordelymus europaeus*: cpDNA evidence, *Flora - Morphology, Distribution, Functional Ecology of Plants*, 205, 418-423.

Earl, Dent A. and vonHoldt, Bridgett M. (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4(2), 359-361. doi: 10.1007/s12686-011-9548-7

Grulich, V. (1997). Peucedanum. In B. Slavik, J. Chrtek, and P. Tomsovic (Eds.), *Květena České republiky,* 5th ed., pp. 406–418. Praha: Academia.

Hartl, D. L. (2000). *A Primer of Population Genetics*, 3rd ed. Sinauer Associates, Inc.

Hewitt GM. 2000. The genetic legacy of the Quaternary ice ages. *Nature,* 405: 907–913.

Hodková, E., Doudová, J., Douda, J., Krak, K., and Mandák, B. (2019). On the road: Postglacial history and recent expansion of the annual *Atriplex tatarica* in Europe. *Journal of Biogeography*.

Horreo, J.L., Pelaez, M. L., Suarez, T., Breedveld, M.C., Heulin, B., Surget-Groba, Y., Oksanen, T.A. and Fitze, P. S. (2018). Phylogeography, evolutionary hisotry and effects of glaciations in a species (*Zootoca vivipara*) inhabiting multiple biogeographic regions. *Journal of Biogeography*, 45, 1616–1627. https://doi.org/10.1111/jbi.13349

Janišová, M., Bartha, S., Kiehl, K., and Dengler, J. (2011). Advances in the conservation of dry grasslands: Introduction to contributions from the seventh european dry grassland meeting. *Plant Biosystems*, *145*(3), 507–513. https://doi.org/10.1080/11263504.2011.603895

Janská, V., Jiménez-Alfaro, B., Chytrý, M., Divíšek, J., Anenkhonov, O., Korolyuk, A., Lashchinskyi, N, and Culek, M. (2017). Palaeodistribution modelling of European vegetation types at the Last Glacial Maximum using modern analogues from Siberia: Prospects and limitations. *Quaternary Science Reviews*, *159*, 103–115. https://doi.org/10.1016/j.quascirev.2017.01.011

Kajtoch, Ł., Cieślak, E., Varga, Z., Wojciech, P., Mazur, M. A., Sramkó, G., and Kubisz, D. (2016). Phylogeographic patterns of steppe species in Eastern Central Europe: a review and the implications for conservation. *Biodiversity and Conservation*, *25*(12), 2309–2339. https://doi.org/10.1007/s10531-016-1065-2

King, A. R. and Ferris, C. (1998), Chloroplast DNA phylogeography of *Alnus glutinosa* (L.) Gaertn.. Molecular Ecology, 7, 1151-1161. doi:10.1046/j.1365-294x.1998.00432.x

Kizeková, M., Hopkins, A., Kanianska, R., Makovnikova, J., Pollak, S., and Palka, B. (2017). Changes in the area of permanent grassland and its implication for the provision of bioenergy: Slovakia as a case study. *Grass and Forage Science*, 73(1), 218–232. https://doi.org/10.1111/gfs.12333

Li, Y.-C., Korol, A. B., Fahima, T., Beiles, A., and Nevo, E. (2003). Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Molecular Ecology*, 11(12), 2453–2465. https://doi.org/10.1046/j.1365-294x.2002.01643.x

Lowe, A., Harris, S., and Ashton, R. (2009). Ecological Genetics: Design, Analysis, and Application. Blackwell Science Ltd.

Magyari, E. K., Chapman, J. C., Passmore, D. G., Allen, J. R. M., Huntley, J. P., and Huntley, B. (2010). Holocene persistence of wooded steppe in the Great Hungarian Plain. *Journal of Biogeography*, 37(5), 915–935. https://doi.org/10.1111/j.1365-2699.2009.02261.x

Mandák, B. (2019). *What is the origin of dry grasslands in Central Europe? A synthesis of comparative phylogeography and palaeodistribution modelling*. Personal communication

Mandák, B., Vít, P., Krak, K., TráníCek, P., Havrdova, A., Hadincová, V., Zákravský P., Jarolímová, V., Bacles, C.F.E., Douda, J. (2016). Flow cytometry, microsatellites and niche models reveal the origins and geographical structure of Alnus glutinosa populations in Europe *Annals of Botany*, 117, 107-120

Meusel, H and Jäger, E. 2010. Comparative Chorology of the Central European Flora. 2. Map 319a. http://chorologie.biologie.unihalle.de//choro/map.php?SearchString=peucedanum+cervaria&Spec=P euced anum+cervaria&Text=322&Map=319a&Volume=II&Sort=1&Lang=E

Mráz, P., Gaudeul, M., Rioux, D., Gielly, L., Choler, P., Taberlet, P., and the IntraBioDiv Consortium. (2007). Genetic structure of *Hypochaeris uniflora* (Asteraceae) suggests vicariance in the Carpathians and rapid post-glacial colonization of the Alps from an eastern Alpine refugium. *Journal of Biogeography*, 34, 2100–2114. https://doi.org/10.1111/j.1365-2699.2007.01765.x

Oliveira, E. J., Pádua, J. G., Zucchi, M. I., Vencovsky, R., and Vieira, M. L. C. (2006). Origin, evolution and genome distribution of microsatellites. *Genetics and Molecular Biology*, 29(2), 294–307. https://doi.org/10.1590/S1415-47572006000200018

Palpurina, S., Chytrý, M., Tzonev, R., Danihelka, J., Axmanová, I., Merunková, K., Karakiev, T. (2015). Patterns of fine-scale plant species richness in dry grasslands across the eastern Balkan Peninsula. *Acta Oecologica*, 63, 36–46. https://doi.org/10.1016/j.actao.2015.02.001

Pitz, C., Mahy, G., Harzé, M., Uyttenbroeck, R., and Monty, A. (2019). Comparison of mining spoils to determine the best substrate for rehabilitating limestone quarries by favoring native grassland species over invasive plants. *Ecological Engineering*, 127, 510–518. https://doi.org/10.1016/j.ecoleng.2018.10.004

Pladias.cz. (2019). *Peucedanum cervaria*.

Podesva, Z. (2009). *Peucedanum cervaria* (L.) Lapeyr. – smldník jelení. Retrieved December 12, 2018, from https://botany.cz/cs/peucedanum-cervaria/

Pokorný, P., Chytrý, M., Juřičková, L., Sádlo, J., Novák, J., and Ložek, V. (2015). Mid-Holocene bottleneck for central European dry grasslands: Did steppe survive the forest optimum in northern Bohemia, Czech Republic? *Holocene*, 25(4), 716–726. https://doi.org/10.1177/0959683614566218

Pritchard, JK Stephens M Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics,* 155, 945 959.

QIAGEN. (2012). DNeasy® Plant Handbook: DNeasy Plant Mini Kit, (October), 22–25.

Rossetto, M. (2001). Chapter 14: Sourcing SSR markers from related species. Plant genotyping: the DNA fingerprinting of plants, 221-224.

Schuelke, M. (2000). An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology*, 18(2), 233–234. https://doi.org/10.1038/72708

Silva, P. J., Toland, J., Jones, W., Eldrigde, J., Thorpe, E., and O'Hara, E. (2008). *LIFE and Europe's grasslands: Restoring a forgotten habitat*. European Communities. https://doi.org/10.2779/23028

Škornik, S. Vidrih, M. and Kaligarič, M. (2010). The effect of grazing pressure on species rich composition and productivity in North Adriatic Karst pastures. *Plant Biosystems*, 144: 355–364

Sommer, R.S. and Zachos, F.F. (2009). Fossil evidence and phylogeography of temperate species: 'glacial refugia' and post-glacial recolonization. *Journal of Biogeography,* 36, 2013-2020.

Stewart, J.R., Lister, A.M., Barnes, I., and Danén, L. (2010). Refugia revisited: individualistic responses of species in space and time. *Proceedings of The Royal Society,* 277: 661-671. https://doi.org/10.1098/rspb.2009.1272.

Szabo, A., and Ruprecht, E. (2018). Restoration possibilities of dry grasslands afforested by pine: the role of seed bank and remnant vegetation. *Tuexenia*, (38), 405–418. https://doi.org/10.14471/2018.38.013

Tyskun, T., Rellstab, C., Dutech, C. Sipo, G., and Prospero, S. (2017). Comparative assessment of SSR and SNP markers for inferring the population genetic structure of the common fungus *Armillaria cepistipes*. *Heredity,* 119, 371-380.

Vrancken, J., Brochmann, C., and Wesselingh. (2009). How did an annual plant react to Pleistocene glaciations? Postglacial history of *Rhinanthus angustifolius* in Europe. *Biological Journal of the Linnnean Society,* 98, 1-13.

Vítková, M., Müllerová, J., Sádlo, J., Pergl, J., and Pyšek, P. (2016). Black locust (Robinia pseudoacacia) beloved and despised: A story of an invasive tree in Central Europe. *Forest Ecology and Management*, 384, 287–302. https://doi.org/10.1016/j.foreco.2016.10.057

Waldén, E., Öckinger, E., Winsa, M., and Lindborg, R. (2017). Effects of landscape composition, species pool and time on grassland specialists in restored semi-natural grasslands. *Biological Conservation*, 214, 176–183. https://doi.org/10.1016/j.biocon.2017.07.037

Weising, K., Wöhrmann, T., and Huettel, B. (2015). The use of high-throughput DNA sequencing for microsatellite discovery in plants. In *Next-Generation Sequencing in Plant Systematics.* 1–26. https://doi.org/10.14630/000002

Wesche, K., Ambarlı, D., Kamp, J., Török, P., Treiber, J., and Dengler, J. (2016). The Palaearctic steppe biome: a new synthesis. *Biodiversity and Conservation*, 25(12), 2197–2231. https://doi.org/10.1007/s10531-016-1214-7

Willner, W., Kuzemko, A,. Dengler, J., Chytrý, M., Bauer, N., Becker, T., Biţă-Nicolae, C., Botta-Dukát, Z., Čarni, A., Csiky, J., Igić, R., Kącki, Z., Korotchenko, I., Kropf, M., Krstivojević-Ćuk, M., Krstonošić, D., Rédei, T., Ruprecht, E., Schratt-Ehrendorfer, L., Semenishchenkov, Y., Stančić, Z., Vashenyak, Y., Vynokurov, D., and Janišová, M. (2016). A higher-level classification of the Pannonian and western Pontic steppe grasslands (Central and Eastern Europe). *Applied Vegetation Science*, 20(1), 143–158. https://doi.org/10.1111/avsc.12265

Wilson, E. O. (1988). *Biodiversity*. National Academies Press.

Zalapa, J. E., Cuevas, H., Zhu, H., Steffan, S., Senalik, D., Zeldin, E., McCown, B, Harbut, R, Simon, P. (2011). Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. *American Journal of Botany*, 99(2), 193–208. https://doi.org/10.3732/ajb.1100394

9 Appendix 1

Appendix 1-1: Population Localities

Population ID	Locality	Latitude	Longitude
10	Romania, Cluj-Napoca, Bărăi	46.85796	23.90922
11	Romania, Suceava, Moara Carp	47.57491	26.25542
2	Hungary, Dorog, Kis-Strázsa-hegy, Nagy-Strázsa-hegy	47.750617	18.741983
64	Austria, Hundsheim (distr. Bruck an der Leitha): Hundsheimer Berg 1.0-1.2 km NW-N of the village	48.1281389	16.9333611
15	Poland, Skorocice	50.41889	20.66948
54	Czech Republic, Pavlovské vrchy	48.834861	16.639778
36	Czech Republic, Srbsko	49.946837	14.133322
169	Slovakia, Ružomberok Distr., Ružomberok: Dry grassland developed on limestone slopes of the Čebrat Hill, NW edge of the town	49.09425	19.29616667
31	Germany, Nissmitz	51.195998	11.763920
9	Hungary, Bükk, Eger	48.012401	20.578951

Appendix 1-2: All Primer Sequence Data

Primer	Motif	Left Sequence (Tag)	Right Sequence
PC-di-01	$(AG)_{19}$	GGAAACAGCTATGACCATCTCTCACGTAAACTCGCCG	CAGAGCCTCAAATCGAAGTCTC
PC-di-02	$(AG)_{19}$	GGAAACAGCTATGACCATGAACCACCACCGCTTCAC	GGACAGCGGTAGAGAGGC
PC-di-03	$(AG)_{19}$	GGAAACAGCTATGACCATCCACATACACTTGTCTCGTCC	CAACAGATGAGGACGTACTGAC
PC-di-04	$(AG)_{19}$	GGAAACAGCTATGACCATCACCGGAAACCAACCCATC	GTGATGGCCGTGAAGTGAAG
PC-di-05	$(AG)_{19}$	GGAAACAGCTATGACCATTTCCGGCGAGTCAGTGG	CCCGAAACTCCTCTTTGATCC
PC-di-06	$(AG)_{19}$	GGAAACAGCTATGACCATAAACCCTAAGTGAGACTGCG	ATTTCAATTTCCCGGGCCAG
PC-di-07	$(AG)_{19}$	GGAAACAGCTATGACCATGCTAAGACAAGTACAGCGCC	CCATTCCTAGCCTGAGTCGG
PC-di-08	$(AC)_{19}$	GGAAACAGCTATGACCATCTCTCTCGGCCCTTTACC	GTTCGGGTGTGGGTGTATCC
PC-di-09	$(AG)_{19}$	GGAAACAGCTATGACCATCTCTCTCCAACTCACCCAGC	TTTACGTGCTGGGTGGGTTG
PC-di-10	$(AG)_{19}$	GGAAACAGCTATGACCATAAATCGGGAAGGAGGAGGG	ATCCGCAAACAACCCATCAC
PC-di-11	$(AG)_{19}$	GGAAACAGCTATGACCATACCTTTCCGATGACCACCG	AGGAGAGATTGTTGGCGGAG
PC-di-12	$(AT)_{19}$	GGAAACAGCTATGACCATGCACAATCGGCGAAGATGG	TTGGGTGTAAACGCGAGATC
PC-di-13	$(AG)_{19}$	GGAAACAGCTATGACCATCCGCCACTGATTTCAACCC	GCTTCCCGCCTATATTCACC
PC-di-14	$(AT)_{19}$	GGAAACAGCTATGACCATCCTCCCATGTCCTGCAC	TGCTCCTGGTCATACGATCC
PC-di-15	$(AG)_{19}$	GGAAACAGCTATGACCATACACACGCAGAGAAGATAAAGG	CACCTGCAGTTCCAACACAC

PC-di-16	$(AG)_{19}$	GGAAACAGCTATGACCATGAAACAAGACCGGATCGC	TCTTTCTCTCTCGAACTAGCCG
PC-di-17	$(AG)_{18}$	GGAAACAGCTATGACCATGTAGAGGCGGAAAGGGAG	TGTCTCGACCTCCCTCCC
PC-di-18	$(AG)_{18}$	GGAAACAGCTATGACCATGTTGATCTGAGCTTCCCTGG	GAACCATGCCACACAAGTCC
PC-di-19	$(AG)_{18}$	GGAAACAGCTATGACCATCTACATTCACCTCATCCTCCC	GGCCCAGTAACTCCATGAAC
PC-di-20	$(AG)_{18}$	GGAAACAGCTATGACCATCCTTCTGCACTCTCATGTAAGC	GTCGAAACCTAGTGCCAATGG
PC-di-21	$(AG)_{18}$	GGAAACAGCTATGACCATGCTCGGAAAGTCAAGCCATC	GGAAGGTTGTGTTGTGGAGG
PC-di-22	$(AG)_{18}$	GGAAACAGCTATGACCATCGGAAGCAAGTAAAGGTGTG	AACTTCTCCTCACCTGGCTG
PC-di-23	$(AT)_{18}$	GGAAACAGCTATGACCATGCACAGCCTCACTCTCTTAG	TCTGTGAAGGTGTATGTAGCTG
PC-di-24	$(AG)_{18}$	GGAAACAGCTATGACCATCGATTTCAATTCGACCCACC	GAGGTAGGAAGGCAGGTGG
PC-di-25	$(AT)_{18}$	GGAAACAGCTATGACCATGCGCGCCCAGGTATTATAC	GGAACTCTCAAGGTACGTTCTC
PC-di-26	$(AG)_{13}$	GGAAACAGCTATGACCATACCCAACACCAATGAACTCC	TGGCTCCAAACCAGATTAGAG
PC-di-27	$(AG)_{13}$	GGAAACAGCTATGACCATCTTTGCCAGCTTGTACTTCC	CGTGTGGGAACTGAAACGAC
PC-di-28	$(AT)_{13}$	GGAAACAGCTATGACCATACATCAGGAAGACACCAACC	TGCAGCACCACTTAATCTCG
PC-di-29	$(AG)_{13}$	GGAAACAGCTATGACCATACAATCGGGCCACAACG	ATGGGTTGGAGGTGGAATGG
PC-di-30	$(AG)_{13}$	GGAAACAGCTATGACCATACGATTCGATCCGGAAATCAC	GGAGGTAAGTTAAATCGGGCTC
PC-di-31	$(AT)_{13}$	GGAAACAGCTATGACCATACTTGCCAGAACCACCATTG	TGGCTCCCATTGATTGACTC
PC-di-32	$(AG)_{13}$	GGAAACAGCTATGACCATCACAACACATGAATCCACATCC	CCTCATCGCTGAAAGATCTGG

PC-di-33	$(AT)_{13}$	GGAAACAGCTATGACCATGCATTGTGAATAGTCCCGC	AACGTGCAAGTGTTGGGTAC
PC-di-34	$(AT)_{13}$	GGAAACAGCTATGACCATGACCAGTATTTCAAACACAGGC	TAGAACACTCCCAGACGACG
PC-di-35	$(AG)_{13}$	GGAAACAGCTATGACCATGGTTCGAAAGTCAAGCCACC	GAGTTCGATTTGGGCCAAGG
PC-di-36	$(AG)_{13}$	GGAAACAGCTATGACCATGAGCAGCAACCAAGAAGATG	GCAGGCATAATCACCCAGAG
PC-di-37	$(AG)_{13}$	GGAAACAGCTATGACCATCGACCCGGAGATAAGCTTC	GGTGTTGCCTTAGAAACCCG
PC-di-38	$(AT)_{13}$	GGAAACAGCTATGACCATGAACACAAGACTCTGGCAC	GGCGGAATCCAAGGTGTACC
PC-di-39	$(AT)_{13}$	GGAAACAGCTATGACCATTCTCCACGGTTCTCCACG	CTATTTGCCCACGTCAGCAC
PC-di-40	$(AG)_{13}$	GGAAACAGCTATGACCATACCCACCGGAAGAATCTCTC	GGTGGAGGTTGTGAAGAGGG
PC-di-41	$(AG)_{13}$	GGAAACAGCTATGACCATTTGCCAGGTCCACTCTCC	AGCAATTGAGATTTGGAGCGG
PC-di-42	$(AG)_{13}$	GGAAACAGCTATGACCATGGACGAAGGGACAACTTATC	ACACTTGACACGACATTCTCAC
PC-di-43	$(AT)_{13}$	GGAAACAGCTATGACCATGACAAGGAGTGGGAGCAAAC	CAGCTGCTGGTTACACGG
PC-di-44	$(AG)_{13}$	GGAAACAGCTATGACCATACCAAGTTCGCCACAAACG	TTTGTACTTGGCGAATATTGGC
PC-di-45	$(AT)_{13}$	GGAAACAGCTATGACCATCCAACCAACCAAATATCCCATG	ACTGTTGGTGTCTCGGTTTG
PC-di-46	$(AC)_{13}$	GGAAACAGCTATGACCATAACCAAGCAAGCAACAACTC	TGAGAAATGGTGCTTTGATGTG
PC-di-47	$(AC)_{13}$	GGAAACAGCTATGACCATGCAACTGGAAGTGAGTGTC	TGCCCATGTAAGACGTAGGG
PC-di-48	$(AT)_{13}$	GGAAACAGCTATGACCATGACCAATAGCGTCCGTTG	ATCCAAACCCGATCCGAACC
PC-di-49	$(AT)_{13}$	GGAAACAGCTATGACCATAACATTCCACCGAGATCTGC	TCATCAGTGACGTGGCAGTG

PC-di-50	$(AG)_{13}$	GGAAACAGCTATGACCATCACACGAACGTCCGATCATC	GCTGCTATACATCTTGAGAGGC
PC-di-51	$(AC)_{13}$	GGAAACAGCTATGACCATAAGCGCGCAAACATCAC	TCGTTAGCTGTACATCTTGAGG
PC-di-52	$(AT)_{13}$	GGAAACAGCTATGACCATCACCTTCTACCCGGACG	CACGGTTTAGGTTGCTCTCG
PC-di-53	$(AG)_{13}$	GGAAACAGCTATGACCATGAACCCGAACAGCCTGTAG	CAGGCGGGCTCAATCTTTG
PC-di-54	$(AT)_{13}$	GGAAACAGCTATGACCATGCCATCTACACAAAGGACC	GGATACCCGACCCTTGGATC
PC-di-55	$(AT)_{13}$	GGAAACAGCTATGACCATGGTGCCTGTGGAAATAAAGC	CGCCCATCCGATTCACTATG
PC-di-56	$(AT)_{12}$	GGAAACAGCTATGACCATTCCGCCAAATTCACCTCTC	GTGTTCTCCATCACCTGCAC
PC-di-57	$(AC)_{12}$	GGAAACAGCTATGACCATACTCAGCACCCATTCATCCG	ATCGTCAAGGTCTCCAATGC
PC-di-58	$(AT)_{12}$	GGAAACAGCTATGACCATGAGACGTACACTGAGATTGGG	ATGGTCTCGTGTACTGTGGG
PC-di-59	$(AC)_{12}$	GGAAACAGCTATGACCATCACACAGAGGAGAGCACTG	GGTAGTCTAGAGGGCCAAGG
PC-di-60	$(AT)_{12}$	GGAAACAGCTATGACCATAAAGCCAAGGAAGCATCTCG	CTCAGCAATAGAGGTTCTCCC
PC-di-61	$(AT)_{12}$	GGAAACAGCTATGACCATGTAACAACCCAACTGCCG	TGTGGGCCTCAAATTTGGTC
PC-di-62	$(AG)_{12}$	GGAAACAGCTATGACCATGCACCCACTACACGTTACC	GGTTAGGGTGTGCATAAATCAC
PC-di-63	$(AG)_{12}$	GGAAACAGCTATGACCATAAATTGCTTGGCCACGTGTC	GTGCGGTTAGAGAGGTGAGG
PC-di-64	$(AC)_{12}$	GGAAACAGCTATGACCATGTATTTGTTGCAGCGTGTGG	AAATTCGAACCCACCTGCAC
PC-di-65	$(AC)_{12}$	GGAAACAGCTATGACCATACACTCCAAATATGCCTGCAC	GAGTGTGGAGGTTTCTGGTG
PC-di-66	$(AG)_{12}$	GGAAACAGCTATGACCATCCTCTCTTCCACAACGC	TGGCTGGTCGTGATGAGAG

PC-di-67	$(AT)_{12}$	GGAAACAGCTATGACCATCAAGACAACATTCAACCCAAC	TCTGAGAACCTGAGTCCTTGTG
PC-di-68	$(AC)_{12}$	GGAAACAGCTATGACCATACCACTGACCCTAAACCTCG	TGCGGGCTGGTAGAACTTTC
PC-di-69	$(AG)_{12}$	GGAAACAGCTATGACCATCACACGAACGTCCGATTGTC	GCTGTTCCAACTGTTGAGAGG
PC-di-70	$(AG)_{12}$	GGAAACAGCTATGACCATAATAAACTCCGGGCCTGAAG	TTGTGTTCCGCCATTGAGAG
PC-di-71	$(AT)_{12}$	GGAAACAGCTATGACCATCCGATCCCTGCAGTTAATCC	GCTGAGCAATTCTTTCGCTC
PC-di-72	$(AG)_{12}$	GGAAACAGCTATGACCATCAGACACATCACACAAGTAGG	TGGATATGGAACCAGGCATTG
PC-di-73	$(AC)_{12}$	GGAAACAGCTATGACCATGATCCACCAAGAAGCAGATG	TGCACCTCCAGTCTTCCATC
PC-di-74	$(AT)_{12}$	GGAAACAGCTATGACCATCCCAGTAGTATGACCCTTGG	CAGTGAAGTGTGGTCAATAGGG
PC-di-75	$(AG)_{12}$	GGAAACAGCTATGACCATACCCTCCGAATCAAATCTCG	TGGTGGTGATGTTTGGGTCC

Appendix 1-3: R code for Genetic Analysis

```
install.packages("adegenet")
install.packages("hierfstat")
install.packages("PopGenReport")
install.packages("diveRsity")
install.packages("HWxtest")
install.packages ("tidyverse")
install.packages ("tibble")

library(adegenet)
library(diveRsity)
library(hierfstat)
library(PopGenReport)
library(tidyverse)

setwd("E:/2019/PrimerAnalysis/PopAnalysis")
MIKRODATA <-read.csv("PC_AllPopData.csv", head=TRUE,colClasses='character')
colnames(MIKRODATA)[1]<-"POPULACE"
MIKRODATA

## data conversion from table to genind
MIKRODATA_tibble <- as_tibble(MIKRODATA)
tt <-  unique(MIKRODATA_tibble[,1])
pop_data <- list()
for (i in tt$POPULACE) {pop_data [[i]] <- filter(MIKRODATA_tibble, POPULACE == i)
}
## creates geneind objects from each df in a pop_data list
data_gen_pop <- list()
for (i in tt$POPULACE) { data_gen_pop [[i]] <- assign(paste0("DATA_GEN_pop", i), df2genind(pop_data[[i]][-1],
pop=pop_data[[i]]$POPULACE, ploidy=2, type="codom",ncode=10))
}
## replace missing data (NA) by zero
nn<-colnames(DATA_GEN@tab)
DATA_GEN@tab[is.na(DATA_GEN@tab)] <- 0
DATA_GEN@tab<-matrix(as.integer(DATA_GEN@tab), nrow=nrow(DATA_GEN@tab),
ncol=ncol(DATA_GEN@tab))
colnames(DATA_GEN@tab)<-nn

library(HWxtest)
source("GenePopUnflatten.R") # konverze dat z dataframe do genepopu
source("transformdata.R") # uprava genepop formatu
source("transformdata2.R") # genpopformat adjustment, similar to transformdata, for missing data uses"000000"
source("HW_table.R") # uprava tabulky vysledku

Hw_test2 <- divBasic(infile = tabledata2, outfile = "results_divbasic2", HWEexact = TRUE, bootstraps = 1000, mcRep =
9999) # calculation in package diveRsity
str(Hw_test2)

nullalleles_pop_summaries <- list()
for (i in tt$POPULACE){nullalleles_pop_summaries [[i]] <- assign(paste0("null_alleles_pop_sum", i),
nullalleles_pops[[i]]$null.allele.freq$`summary1`)}
str(nullalleles_pop_summaries)
```

Appendix 1-4: Elimination of Primers Table

Primers	Gel	Sample Seq	Full Pop Seq	Amplification	Variability	Multiplex
PC1		X				
PC2			X			
PC3					P	
PC4	I					
PC5			X			
PC6			X			
PC7	0					
PC8			X			
PC9	I					
PC10	M					
PC11		X				
PC12	0					
PC13	I					
PC14	I					
PC15					P	
PC16	0					
PC17	0					
PC18						ACCEPTED
PC19						ACCEPTED
PC20	0					
PC21	I					
PC22				X		

PC23		X				
PC24	0					
PC25		X				
PC26	M					
PC27	I					
PC28				X		
PC29		X				
PC30						ACCEPTED
PC31				X		
PC32			X			
PC33			X			
PC34	M					
PC35	I					
PC36			X			
PC37	M					
PC38				X		
PC39	0					
PC40	I					
PC41	I					
PC42	0					
PC43		X				
PC44		X				
PC45					NV	
PC46		X				
PC47	I					
PC48	M					
PC49		X				
PC50		X				

PC51		X				
PC52					P	
PC53			X			
PC54	0					
PC55			X			
PC56	I					
PC57		X				
PC58						ACCEPTED
PC59					NV	
PC60	0					
PC61	0					
PC62		X				
PC63	M					
PC64			X			
PC65				X		
PC66			X			
PC67	M					
PC68						ACCEPTED
PC69		X				
PC70					P	
PC71				X		
PC72			X			
PC73						ACCEPTED
PC74		X				
PC75					NV	

X – Eliminated; 0 – Eliminated due to amplification failure; I – Eliminated due to Inconsistent amplification; M – Eliminated due to multiple alleles

Appendix 1-5: Population Summary Statistics

11	X68	X58	X73	X19	X30	X18	Overall
N	20	20	20	20	20	20	20
A	7	2	6	9	5	12	41
%	58.33	40	46.15	52.94	38.46	26.67	43.76
Ar	6.17	2	5.41	7.3	4.45	9.54	5.81
Ho	0.75	0.35	0.65	0.9	0.45	0.95	0.68
He	0.78	0.47	0.77	0.76	0.69	0.88	0.73
HWE	0.381	0.325	0.543	0.738	0.088	0.095	0.211
Fis	0.0385	0.2533	0.1586	-0.1765	0.3466	-0.0826	0.0695
Fis_Low	-0.1819	-0.193	-0.0864	-0.3396	0.0282	-0.172	-0.0009
Fis_High	0.2753	0.6772	0.4179	-0.018	0.6556	0.0301	0.1292
B	-0.07816	-0.2232	-0.0610	-0.0752	-0.0869	-0.0335	
169	X68	X58	X73	X19	X30	X18	Overall
N	19	15	19	19	17	19	18
A	3	2	8	5	4	10	32
%	25	40	61.54	29.41	30.77	22.22	34.82
Ar	2.5	2	6.37	3.99	3.33	6.7	4.15
Ho	0.21	0.33	0.89	0.58	0.12	0.63	0.46
He	0.27	0.49	0.76	0.54	0.6	0.63	0.55
HWE	0.218	0.282	0	0.148	0	0.014	0
Fis	0.2284	0.3213	-0.1745	-0.0636	0.8052	0	0.163
Fis_Low	-0.1259	-0.1499	-0.3431	-0.2893	0.478	-0.1894	0.0308
Fis_High	0.7589	0.8491	0.0131	0.1708	1.0153	0.1876	0.2908
B	-0.4019	-0.2064	-0.0760	-0.1626	-0.1917	-0.0989	
2	X68	X58	X73	X19	X30	X18	Overall
N	20	19	20	20	18	20	19.5
A	7	4	6	9	8	17	51
%	58.33	80	46.15	52.94	61.54	37.78	56.12
Ar	5.62	3.8	5.28	8.02	5.52	12.35	6.76

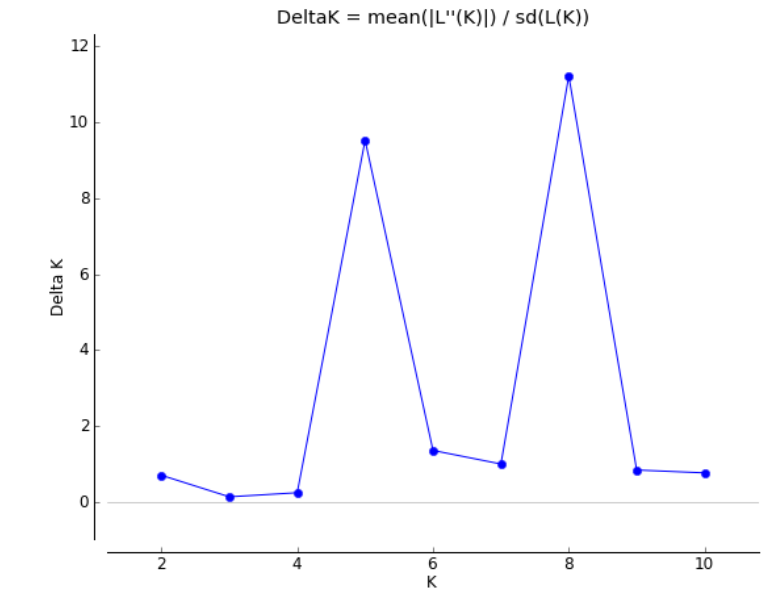
Ho	0.75	0.37	0.5	0.85	0.44	0.9	0.64
He	0.75	0.69	0.66	0.76	0.62	0.91	0.73
HWE	0.03	0.01	0.752	0.997	0.002	0	0
Fis	0.005	0.4669	0.2481	-0.1221	0.2836	0.0123	0.1332
Fis_Low	-0.2965	0.1395	0.0097	-0.2645	-0.0944	-0.1158	0.0311
Fis_High	0.3032	0.7798	0.53	0.0226	0.661	0.1663	0.2387
B	-0.0899	-0.0989	-0.1080	-0.0471	-0.125	-0.0282	
54	X68	X58	X73	X19	X30	X18	Overall
N	19	12	19	19	19	19	17.83
A	5	4	10	9	6	19	53
%	41.67	80	76.92	52.94	46.15	42.22	56.65
Ar	4.47	3.6	8.15	6.93	5.12	14.01	7.05
Ho	0.68	0.33	0.84	0.74	0.26	0.89	0.63
He	0.58	0.71	0.83	0.78	0.67	0.93	0.75
HWE	1	0.002	0.022	0.294	0	0	0
Fis	-0.1706	0.5294	-0.0116	0.0584	0.6074	0.0415	0.1678
Fis_Low	-0.3494	0.0452	-0.1783	-0.1284	0.3497	-0.0886	0.0629
Fis_High	0.0248	0.9159	0.177	0.2704	0.863	0.2	0.2499
B	-0.1159	-0.1338	-0.0478	-0.0540	-0.1589	-0.0299	
64	X68	X58	X73	X19	X30	X18	Overall
N	18	17	18	18	17	18	17.67
A	6	4	6	9	7	18	50
%	50	80	46.15	52.94	53.85	40	53.82
Ar	4.3	3.41	5.25	7.11	5	12.9	6.33
Ho	0.44	0.35	0.72	0.83	0.35	0.72	0.57
He	0.6	0.64	0.74	0.79	0.45	0.92	0.69
HWE	0.135	0.001	0.82	0	0.078	0	0
Fis	0.2615	0.4457	0.0209	-0.0485	0.2093	0.2108	0.1704
Fis_Low	-0.1064	-0.0271	-0.244	-0.2486	-0.1384	0.0117	0.0601
Fis_High	0.5995	0.8363	0.3071	0.1616	0.65	0.4246	0.2658
B	-0.125	-0.1491	-0.0657	-0.0657	-0.2324	-0.0384	

10	X68	X58	X73	X19	X30	X18	Overall
N	19	17	18	19	13	17	17.17
A	3	3	5	8	5	10	34
%	25	60	38.46	47.06	38.46	22.22	38.53
Ar	2.83	2.77	4.55	5.96	4.26	7.53	4.65
Ho	0.47	0.24	0.67	0.58	0.15	0.76	0.48
He	0.48	0.38	0.75	0.63	0.74	0.78	0.63
HWE	1	0.252	0.793	0	0	0.002	0
Fis	0.0144	0.3818	0.1148	0.0752	0.792	0.0243	0.2366
Fis_Low	-0.3384	-0.1278	-0.1837	-0.2409	0.4979	-0.178	0.0792
Fis_High	0.408	0.8718	0.4147	0.3957	1.014	0.2492	0.3825
B	-0.2013	-0.2988	-0.0657	-0.1056	-0.1229	-0.0528	
15	X68	X58	X73	X19	X30	X18	Overall
N	20	20	20	20	18	20	19.67
A	6	3	5	8	3	6	31
%	50	60	38.46	47.06	23.08	13.33	38.66
Ar	4.64	2.98	4.7	6.16	2.7	4.9	4.35
Ho	0.55	0.5	0.95	0.9	0.11	0.75	0.63
He	0.51	0.59	0.72	0.78	0.37	0.69	0.61
HWE	0.001	0.748	0.263	0	0.001	0.038	0
Fis	-0.0837	0.1471	-0.3287	-0.1502	0.6987	-0.0909	-0.0311
Fis_Low	-0.3294	-0.1511	-0.4996	-0.3042	-0.0204	-0.3663	-0.1326
Fis_High	0.1534	0.491	-0.1618	0.0321	1.0357	0.2207	0.0629
B	-0.1834	-0.1363	-0.1428	-0.0695	-0.3728	-0.1204	
31	X68	X58	X73	X19	X30	X18	Overall
N	20	20	20	20	15	20	19.17
A	7	3	3	9	5	13	40
%	58.33	60	23.08	52.94	38.46	28.89	43.62
Ar	5.62	2.53	2.56	7.33	4.06	9.96	5.34
Ho	0.8	0.45	0.25	0.85	0.27	0.8	0.57
He	0.73	0.52	0.52	0.83	0.64	0.8	0.67

HWE	0.09	0.355	0.004	0.964	0.001	0.244	0.001
Fis	-0.094	0.1283	0.5227	-0.0256	0.5862	0.0062	0.1563
Fis_Low	-0.3056	-0.2913	0.1037	-0.19	0.2216	-0.1297	0.0304
Fis_High	0.1172	0.54	0.8304	0.1821	0.8981	0.1431	0.2789
B	-0.0810	-0.1940	-0.1940	-0.0362	-0.1508	-0.0443	
36	X68	X58	X73	X19	X30	X18	Overall
N	19	18	19	19	13	19	17.83
A	6	2	6	5	6	17	42
%	50	40	46.15	29.41	46.15	37.78	41.58
Ar	5.69	1.99	5.45	3.78	4.62	11.89	5.57
Ho	0.63	0.22	0.74	0.37	0.31	0.84	0.52
He	0.79	0.28	0.77	0.32	0.74	0.9	0.64
HWE	1	1	0.856	1	0.015	0	0.007
Fis	0.2028	0.2	0.0466	-0.1368	0.5857	0.0675	0.1846
Fis_Low	-0.055	-0.1874	-0.1697	-0.2464	0.1928	-0.0971	0.0452
Fis_High	0.4813	0.7314	0.2869	-0.046	0.9204	0.2576	0.3018
B	-0.0448	-0.4025	-0.0571	-0.2824	-0.0868	-0.0358	
9	X68	X58	X73	X19	X30	X18	Overall
N	15	14	15	15	9	15	13.83
A	7	3	9	7	2	13	41
%	58.33	60	69.23	41.18	15.38	28.89	45.5
Ar	5.44	2.67	8.23	6.32	1.86	10.24	5.79
Ho	0.47	0.14	0.87	0.8	0.11	0.87	0.54
He	0.53	0.48	0.85	0.74	0.28	0.84	0.62
HWE	0.002	0.001	1	0.043	0.125	0.29	0
Fis	0.1176	0.7037	-0.0209	-0.0778	0.6	-0.029	0.1258
Fis_Low	-0.1287	0.2585	-0.2031	-0.3097	-0.086	-0.1656	-0.0004
Fis_High	0.3559	1.0187	0.1801	0.1384	1.1117	0.1279	0.2432
B	-0.1936	-0.2645	-0.0440	-0.1056	-0.4594	-0.0392	

Appendix 1-6: Delta K and Likelihood Graphs

A:



B: