University of Tartu
Faculty of Science and Technology
Institute of Ecology and Earth Sciences
Department of Botany

Iris Reinula

The impact of landscape change on the genetic diversity of the grassland plant *Primula veris*

Ecology and Biodiversity Conservation
Master thesis (30 EAP)

Supervisors: Tsipe Aavik
Sabrina Träger

Tartu 2018

Table of contents

Maastikumuutuste mõju niidutaimede geneetilisele mitmekesisusele nurmenuku (*Primula veris*) näitel

Hiljutised maakasutuse muutused on looduslikke ökosüsteeme oluliselt muutnud, mille tulemuseks on looduslike ja pool-looduslike elupaikade (nt loopealsed) killustumine. Käesoleva töö eesmärgiks on uurida, kas ja kuidas mõjutavad maastikumuutused niidutaimede geneetilist mitmekesisust. Selleks kasutati maastikugeneetika meetodeid. Töös analüüsiti 338 nurmenuku (*Primula veris*) indiviidi proovi, mis pärinesid 19-st populatsioonist Saare- ja Muhumaal. Geneetiliste markerite (üksiknukleotiidsed polümorfismidest e SNP-d) leidmiseks kasutasin nüüdisaegset sekveneerimismeetodit (*restriction site associated DNA sequencing, RADseq*). Saadud geneetiliste andmete abil analüüsisin ajalooliste ja tänapäevaste maastikuparameetrite mõju nurmenuku populatsioonide geneetilisele mitmekesisusele. Lisaks uurisin nurmenuku populatsioonide geneetilist struktuuri. Kasvukoha pindala kadu mõjutas geneetilist mitmekesisust negatiivselt. Tänapäevane metsa pindala avaldas geneetilisele mitmekesisusele negatiivset mõju. Inimpopulatsiooni tihedus oli samuti negatiivse mõjuga. Leidsin, et Muhu ja Ida-Saaremaa populatsioonid olid teiste populatsioonidega võrreldes üksteisele geneetiliselt sarnasemad. Kokkuvõtvalt võib öelda, et maastikumuutustel on olnud suur mõju Eesti loopealsetel kasvavatele nurmenuku populatsioonide geneetilisele mitmekesisusele.

Märksõnad: bioloogiline mitmekesisus, ddRAD, geneetiline mitmekesisus, looduskaitse, maastikugeneetika, maastikumuutused, rohumaa, pindalakaotus, SNP, üksiknukleotiidsed polümorfismid

CERCS teadusalade koodid: B225 Taimegeneetika, B270 Taimeökoloogia

The impact of landscape change on the genetic diversity of the grassland plant *Primula veris*

Recent changes in land use have substantially altered natural ecosystems resulting in fragmentation of natural and semi-natural habitats such as alvars. The purpose of this thesis is to analyse if and how landscape changes influence the genetic diversity of a grassland plant. Landscape genetic methods were used to study the genetic patterns of 338 individuals from 19 populations of *Primula veris* in semi-natural grasslands (alvars) in Saare- and Muhumaa. I applied a state of the art high-throughput sequencing method (restriction site associated DNA sequencing, RADseq) to obtain thousands of single nucleotide polymorphism (SNP) genetic markers. The obtained data were used to analyze the relationship between genetic diversity and historic and current landscape parameters, as well as to analyze the genetic structure of *P. veris* populations in the study region. The results reveal that genetic diversity decreased with increasing habitat loss. Current forest area had a negative effect on the genetic diversity of *P. veris*. Furthermore, genetic diversity decreased with increasing historic and current human population density. Populations of *P. veris* in Muhumaa and Eastern-Saaremaa were found to be genetically closer to each other than to the rest of the studied populations. In conclusion, landscape change in Estonian alvars has had a major effect on patterns of genetic diversity of *P. veris*.

Keywords: area loss, biodiversity, conservation, ddRAD, genetic diversity, landscape change, landscape genetics, grassland, single-nucleotide polymorphisms, SNP

CERCS research field codes: B225 Plant genetics, B270 Plant ecology

Glossary (based on Balkenhol *et al.* 2015 if not stated otherwise)

Allelic richness (A_r) - the average number of alleles per locus corrected for differences in sample size.

Double digest Restriction-Site Associated DNA sequencing (ddRADseq) - a method of reduced-representation genome sequencing experiments on massively parallel sequencers (RADseq) using two restriction enzymes (Peterson *et al.* 2012).

Expected heterozygosity (H_e) - the proportion of individuals that are expected to be heterozygous at a locus averaged across all loci assuming Hardy-Weinberg equilibrium.

Fixation index (F_{st}) - a measure of subpopulation level genetic differentiation relative to total population (ranging from 0 to 1) measuring allele frequency divergence among subpopulations.

Gene flow - a process resulting from migration that is moving alleles between populations and making the populations usually genetically more similar.

Genetic diversity - the differences of nucleotide sequences (alleles) in the same DNA parts within species (Frankham *et al.* 2004).

Genetic drift - random change in allelic frequencies between generations.

Hardy-Weinberg equilibrium (HWE) - a mathematical rule enabling to determine genotypic frequencies in a population, assuming random mating, based on given allelic frequencies.

Inbreeding coefficient (F_{is}) - inbreeding estimate within population.

Locus - a physical location in DNA where a gene is located.

Observed heterozygosity (H_o) - the proportion of individuals that are heterozygous at a locus averaged across all loci.

Percentage of polymorphic loci - the proportion of loci out of total loci analyzed that contain more than one allele.

Polymerase chain reaction (PCR) - a chemical process used to make millions of copies of particular target DNA region or locus.

Single-nucleotide polymorphism (SNP) - a type of genetic marker with variations of individual base pairs throughout the genome.

1. Introduction

Recent changes in environment and landscape have influenced natural ecosystems in many ways such as fragmentation of habitats (Picó & Van Groenendael 2007) and soil pollution (Bezdicek *et al.* 1996). Due to increased density of human population and intensive agriculture, land use has substantially changed towards more agricultural land and monocultures (Picó & Van Groenendael 2007; Prentice *et al.* 2006). This has caused severe fragmentation of natural and semi-natural habitats. In addition, environmental changes, such as climate change and agricultural pollution, have altered habitat conditions and not all organisms are able to cope with these changes, leading to a loss of biodiversity, including the decrease in genetic diversity of wild populations (Leimu *et al.* 2010). In return, loss of genetic diversity makes a species more susceptible to environmental changes because lower genetic diversity may reduce the adaptive potential of populations and can lead to lower fitness (Takkis *et al.* 2013). Hence, environmental changes could harm species with impoverished genetic diversity more severely (Leimu *et al.* 2010).

A comprehensive understanding about the effects of environmental and landscape changes on genetic diversity is currently missing. Nevertheless, genetic diversity provides adaptive potential of species to handle future environmental change and should therefore be one of the focal topics in conservation biology. Indeed, studies relying on species diversity only, as it has been done traditionally, could result in underestimating the threat to biological diversity (Taberlet *et al.* 2012; Whitlock 2014; Vellend *et al.* 2014). Since it has been shown that genetic diversity can react to changes in landscape faster than species diversity, the loss of genetic diversity should be a sign that species richness could decline in the future as well (Helm *et al.* 2009). Furthermore, considering that genetic diversity might not always correlate to population size, habitat size or landscape structure, nature conservation decisions should not be based only on these characteristics, but genetic diversity should be measured directly when possible (Menges *et al.* 2010).

Genetic diversity can be influenced by many different factors, such as the size of the population, habitat area and habitat connectivity. These factors affect genetic diversity mainly through influencing genetic drift, gene flow and natural selection (Leimu *et al.* 2006). The exact influence of landscape characteristics on patterns of genetic diversity depends on many different factors such as the configuration of different landscape elements, the availability of pollen and

seed vectors, life history characteristics of species etc. These factors will be more thoroughly discussed in the subsequent chapters.

1.1 Factors influencing genetic diversity

1.1.1 Landscape configuration

Decrease in the area and connectivity of habitats often imposes serious negative effects on genetic diversity (Young *et al.* 1996). Populations in habitats which have experienced strong decrease in area have often gone through a genetic bottleneck (Jacquemyn *et al.* 2010). A bottleneck is a special case of genetic drift where population size has decreased drastically (Heinaru 2012). In that case, part of genetic diversity disappears due to random fluctuations in allelic frequencies between generations. Such random fluctuations are stronger in small populations and therefore these populations have a higher chance of loss of alleles and thus loss of genetic diversity (Young *et al.* 1996).

In addition to a decrease in habitat area, isolation of habitats could also greatly influence genetic diversity. Aguilar *et al.* (2008) found in their meta-analysis that the inbreeding coefficient (F_{is}) of isolated populations was higher than that of connected populations. Furthermore, it has been found that genetic diversity is higher in grassland plant populations surrounded by other grasslands compared to isolated grassland patches (Prentice *et al.* 2006). When isolation increases, populations are unable to exchange genetic material, which leads to reduced genetic diversity and potentially increased inbreeding within isolated populations. Therefore, for restoration purposes, target areas should have as high spatial connectivity to existing habitats as possible to ensure the recovery of species as well as genetic diversity, the latter being vital for the long-term persistence of restored populations (Aavik & Helm 2018; Helm *et al.* 2009). When studying fragmentation, it should also be considered if fragmentation has been anthropogenically imposed or if the habitats are naturally fragmented since populations that are rare because of anthropogenic reasons are expected to have a more severe reaction (Honnay *et al.* 2007).

Managing habitats in a landscape in different times could also cause a type of spatio-temporal fragmentation. This could be caused by for example different mowing times. For a self- or wind-pollinating species *Bromus hordeaceus* it has been shown that in meadows that were mowed earlier plants also flowered earlier suggesting a genetically based phenological escape (Völler *et al.* 2013). If the neighbouring meadows are mowed in different times, it could

increase the genetic differentiation between populations because of reduced pollen flow and thus gene flow between populations flowering at different times (Völler *et al.* 2013).

Restoring only structural connectivity such as habitat area might not be enough since the functional connectivity of populations - the effective dispersal of propagules or pollen among habitat patches in a landscape (Auffret *et al.* 2017) - is also required (Jacquemyn *et al.* 2010). Nevertheless, structural and functional connectivity might not be correlated (Aavik *et al.* 2014). To avoid the lack of functional connectivity, assisted dispersal of seed (e.g. people involved in restoring dispersing seeds) from neighbouring populations after the restoration might be needed (Broeck *et al.* 2015). However, the genetic material introduced should not differ too much from that of local populations because there might be a threat of outbreeding resulting in less adaptation to the local environment (Picó et al. 2007). Furthermore, commercial seed mixtures should be used cautiously because plants from seed mixtures may suffer from impoverished genetic diversity (Aavik *et al.* 2012).

Landscape consists of various elements that can affect plant species and genetic connectivity differently. For instance, forest can be a favorable habitat for forest-specialist species (Smulders *et al.* 2009), but for grassland-specialist species it may act as a barrier for pollen, seed and thus gene flow since movement of pollinators and seeds is inhibited (Hahn *et al.* 2013). This was confirmed by studies of Aavik *et al.* (2014, 2017) on *Lychnis flos-cuculi* as well as *Rhinanthus osiliensis*, respectively. In contrast, Hahn *et al.* (2013) showed that the amount of forest had no influence on gene flow of *Trifolium monatnum*. Roads, on the other hand, may be barriers for both forest- and grassland-specialists since they hinder the movement of pollinators and seed dispersing animals (Hahn *et al.* 2013).

Another major factor influencing genetic diversity can be the density of human population. Helm *et al.* (2009) found that the genetic diversity of *Briza media* is negatively related to contemporary density of human settlement in the surrounding of study populations. The biggest loss of genetic diversity took place in unstable grasslands (containing just fragments of the original grassland area), whereas in more stable grasslands (which retained relatively more original grassland area) there was no negative effect. Density of human settlement 100 years ago had a positive effect on genetic diversity, probably due to traditional tillage management (Helm *et al.* 2006) ensuring the connectivity and preservation of habitat area by grazing and mowing, respectively.

1.1.2 Seed and pollen vectors

Plants are sessile organisms and thus depend on various abiotic and biotic vectors for dispersal of seed and pollen (e.g. wind, water, animals; Holderegger *et al.* 2010). Cattle, as an example for a biotic vector, has been shown to spread propagules of grassland species most of which are not specifically adapted to animal-dispersal (Holderegger *et al.* 2010). As cattle can be distribution vectors for plants, grazing could facilitate maintaining genetic diversity via increased gene flow. Thus, rotational grazing between pastures might help to increase functional connectivity between habitats (Honnay *et al.* 2006; Jacquemyn *et al.* 2010). Rico *et al.* (2014) showed that the genetic differentiation of *Dianthus carthusianorum* was lower between habitats connected by rotational cattle grazing, indicating that cattle helped to spread the genetic material of plants through increased gene flow. The same result was confirmed by DiLeo *et al.* (2017), who found a positive correlation between genetic diversity of *Pulsatilla vulgaris* and connectivity of its habitats by grazing cattle. In addition, selective grazing might increase genetic diversity within populations (Kloss *et al.* 2011; Völler *et al.* 2013).

It should also be considered if landscape elements have the same effect on both seed and pollen flow. Campagne *et al.* (2009) found that hedgerows are corridors for seed dispersal of *Primula vulgaris*, but not for pollen. Many insect-pollinated species do not have morphological structures facilitating their seed dispersal. Thus, it is likely that pollen flow is the main way of long-distance dispersal of those species. In fragmented landscapes with barriers to pollinators (such as forests; Schmitt *et al.* 2000), the reason for reproductive impairment might be pollination limitation (Aguilar *et al.* 2006). However, corridors in barriers (barrier habitats) could facilitate pollen dispersal (Tewksbury *et al.* 2002). This might be the case for insect-pollinated plant species such as *Lychnis flos-cuculi* (Aavik *et al.* 2014) and *Rhinanthus osiliensis* (Aavik *et al.* 2017), neither of which have specialized dispersal mechanism for their seeds. Furthermore, different pollinators in the same landscape could behave differently and may thus have a differential effect on pollen flow between populations (Kramer *et al.* 2011).

1.1.3 Characteristics of plant species

The response of genetic diversity of plants to landscape configuration may also depend on the life history characteristics of the particular species, such as seed dispersal strategy, dispersal distance, seed vector, life span, but also the rarity of species (Honnay *et al.* 2006, 2007). It is usually assumed that habitat specialists and rare species are more sensitive to the size and fragmentation of habitats (Honnay & Jacquemyn 2007). For example, it has been found that

inbreeding of the rare and habitat specialist plant *Rhinanthus osiliensis* has increased due to the decrease of habitat area (Aavik *et al.* 2017). However, historically rare species and species that have become rare recently need to be treated separately because the reasons of species being rare can be different and thus also their genetic structure (Aguilar *et al.* 2008). On the other hand, it has been found that common species can be as susceptible or even more susceptible to the loss of genetic diversity caused by fragmentation (Honnay & Jacquemyn 2007). This indicates that fragmentation could endanger the genetic diversity of even more species than assumed until recently.

In addition, the pollination strategy and reproduction system of the species may influence genetic diversity. The effect of habitat fragmentation is generally not as severe for self-pollinating species as it is for cross-pollinating species, because self-pollinating species do not depend on the availability of pollinating vectors whereas cross-pollinating species do (Honnay & Jacquemyn 2007; Schmidt *et al.* 2009). Sensitivity differences to landscape changes among cross-pollinating species can be caused by differences in pollination strategies. Since insect-pollinating species depend largely on their pollinators, and pollinators can be sensitive to landscape changes, modified landscape structure influences insect-pollinated species generally more than wind-pollinated species (Aguilar *et al.* 2008). Fragmentation decreases the amount of pollinating insects and thus makes it harder for cross- and insect-pollinated species to maintain their genetic diversity and successfully survive in a fragmented landscape (Honnay & Jacquemyn 2007; Schmidt *et al.* 2009). For variations between species with different reproductive systems, it is important to note that sexually reproducing species have most of their genetic diversity within populations, whereas species reproducing vegetatively have most of their genetic diversity between populations. Self-compatible species tend to have higher inbreeding as found for example for a highly selfing (reproducing vegetatively) species *Geum urbanum* (Vandepitte *et al.* 2007). Thus, a sharp decline in population size influences more the genetic diversity of sexually reproducing species (Aguilar *et al.* 2008). In a fragmented landscape small populations of sexually reproductive species could lose rare alleles more easily and thus their genetic diversity decreases (Honnay & Jacquemyn 2007).

1.1.4 Time lag

Plant populations, which have experienced severe fragmentation, may still maintain considerably high genetic diversity despite landscape changes (Hahn *et al.* 2013). This could be caused by so-called genetic extinction debt, due to which genetic diversity has not yet reacted

to the changes of landscape structure. However, genetic diversity will most probably react to the changes in future if there is nothing to counteract habitat fragmentation.

Because time lag in the response of genetic diversity to habitat fragmentation may substantially influence study conclusions and conservation recommendations, potential delayed responses need to be accounted for. The easiest way to measure lagged responses of genetic diversity is to consider historic landscape structure in addition to current one. This could reveal links between historic landscape structure and current genetic variety (Epps & Keyghobadi 2015). Münzbergová *et al.* (2013), for example, showed that the genetic diversity of *Succisa pratensis* populations was connected to historic landscape connectivity, concluding a time lag response. Similarly, Reisch *et al.* (2017) found genetic diversity of several calcareous grassland species to be related to historic landscape structure.

In addition to historic and current landscape, tissue of the particular species from different times, e.g. historic and current samples, could be studied. Furthermore, different molecular markers with different mutation speed could be used for the analysis of a time lag (Epps & Keyghobadi 2015). For instance the evolution speed of microsatellite markers is on average $5x10^{-4}$ per locus per generation (Selkoe & Toonen 2006), whereas the evolution speed of SNPs is slower. Consequently, changes in genetic diversity to landscape changes might be seen using microsatellites, but not using SNPs, influencing study results by the choice of genetic marker used (Epps & Keyghobadi 2015).

Another important aspect to consider is the age of the sampled plants. If genetic samples are taken from adult specimen who are rather old, recent changes in landscape structure might not yet be seen in the genetic patterns of this generation. Older plants can still grow in the habitat and might have high genetic diversity, but if these individuals are not able to forward the genetic diversity to the next generation, due to changes in landscape structure and accompanying decrease in effective population size, genetic diversity in the next generations will be lower as was shown with *Primula vulgaris* in Belgium (Van Geert *et al.* 2008). For wooden plants, genetic diversity of descendants in a fragmented landscape was also shown to be lower than that of adults (Vranckx *et al.* 2012).

Time lag in the response of genetic diversity to landscape changes can be also caused by hibernating seed bank, due to which plants could start to grow much later after the seed was planted (Menges *et al.* 2010). This is sometimes believed to buffer the effects of fragmentation (Plue & Cousins 2013). However, Plue *et al.* (2017), who assessed the above- and below-ground

genetic diversity of *Campanula rotundifolia* and accounted for current and historic landscape configuration, found that seedbanks may not always be able to buffer negative effects of habitat fragmentation. Thus, generation time, demographic composition and nature of seed bank of study species should be considered in order to assess time lags in the genetic response of plants to landscape changes.

However, there is not always a time lag between genetic diversity and changes of landscape structure. It was shown that the genetic diversity of *Briza media* inhabiting fragmented alvar grasslands was more related to contemporary landscape structure, but plant species diversity in the same grasslands was related to historic landscape structure, i.e. species diversity exhibited extinction debt, whereas genetic diversity did not (Helm *et al.* 2009). Extinction debt is a delayed response of species diversity (or genetic diversity) to changed environmental conditions induced by human influence and habitat fragmentation (Helm *et al.* 2009). Therefore, it can be assumed that genetic diversity is more likely to show a response to contemporary landscape structure, because genetic diversity reacts to changes in landscape structure faster than species diversity (but see the previous paragraphs about the reaction speed of genetic diversity).

It is also important which genetic measure is being used, genetic diversity or differentiation. Since a migration-drift balance is being achieved in genetic diversity and differentiation with different speed, genetic changes are more likely to be first mirrored in genetic differentiation and later in within-population genetic diversity (DiLeo & Wagner 2016). Moreover, further genetic indices react to fragmentation with different speed. For instance, allelic diversity (A_r) may react faster than expected heterozygosity, H_e (Chung *et al.* 2014). Nevertheless, common measures used for assessing genetic differentiation, such as fixation index (F_{st}), may not always be valid in landscapes which have experienced rapid recent fragmentation since fixation index assesses historic genetic differentiation (Holsinger & Weir 2009). In such landscapes, parentage analysis and assignments tests may be more appropriate methods for assessing contemporary gene flow (Holderegger *et al.* 2010). Parentage analysis detects the most likely parent or parents (Jones *et al.* 2010). Assignment test divides individuals into genetic populations they most likely belong to using genetic information (Manel *et al.* 2005). This information from an assignment test can then be compared to actual spatial locations of populations by making inferences about contemporary gene flow (Holderegger *et al.* 2010).

1.2 Main approaches to examine the impact of landscape structure on genetic diversity

Studying the impact of landscape structure on genetic diversity is often done using landscape genetic methods. Landscape genetics is a field which combines tools of population genetics and landscape ecology to study the influence of environmental heterogeneity (e.g. landscape configuration and local environmental conditions) on genetic diversity (Manel *et al.* 2003). Genetic diversity can be measured with genetic diversity measures based on different genetic markers. Most landscape genetic studies have focused on neutral genetic markers, such as microsatellites and amplified fragment length polymorphisms (AFLPs), for assessing genetic diversity and gene flow. However, genetic methods have advanced a lot in recent years. Nowadays, next generation sequencing approaches and single-nucleotide polymorphism (SNP) molecular markers are just a few examples of genetic methods enabling to focus also on those parts of the genome which are under natural selection. Using these methods has substantially facilitated studying the patterns of genetic diversity and have potential to advance making effective conservation decisions.

The influence of landscape structure on the distribution of genetic variation in the landscape can be studied in many ways, with 'link' and 'node' methods being the most common (Fig. 1; DiLeo & Wagner 2016). Link method is used to study paired relationships between populations, where the measure of gene flow, e.g. genetic differentiation between populations, is the response variable, and various parameters of landscape characteristics between populations are used to explain patterns of gene flow (DiLeo & Wagner 2016). Node method, on the contrary, focuses on within-population genetic diversity and examines the impact of landscape characteristics at and the focal population on genetic diversity (DiLeo & Wagner 2016). As plants are sessile organisms, it is easier to study their habitat and thus node method is more often used for plants. With this method, buffers with different radiuses are drawn around the habitat patches. One could for example calculate the percentage of different land use types (or suitable and unsuitable habitat area) in these buffers. This could then be correlated with genetic data (Holderegger *et al.* 2010). Link method focuses mainly on the configuration of the landscape between populations and thus on the loss of connectivity, whilst node method has been used more extensively for examining the role of the loss of habitat area (DiLeo & Wagner 2016). The importance of both components is debated and it has even been stated that only the loss of habitat area is important (Fahrig 2013). However, it has also been suggested that the configuration of the landscape enabling the movement of pollen and seed is absolutely relevant

to maintain genetic diversity within populations in many contemporary landscapes (Aavik & Helm 2018; Auffret *et al.* 2017; DiLeo & Wagner 2016).

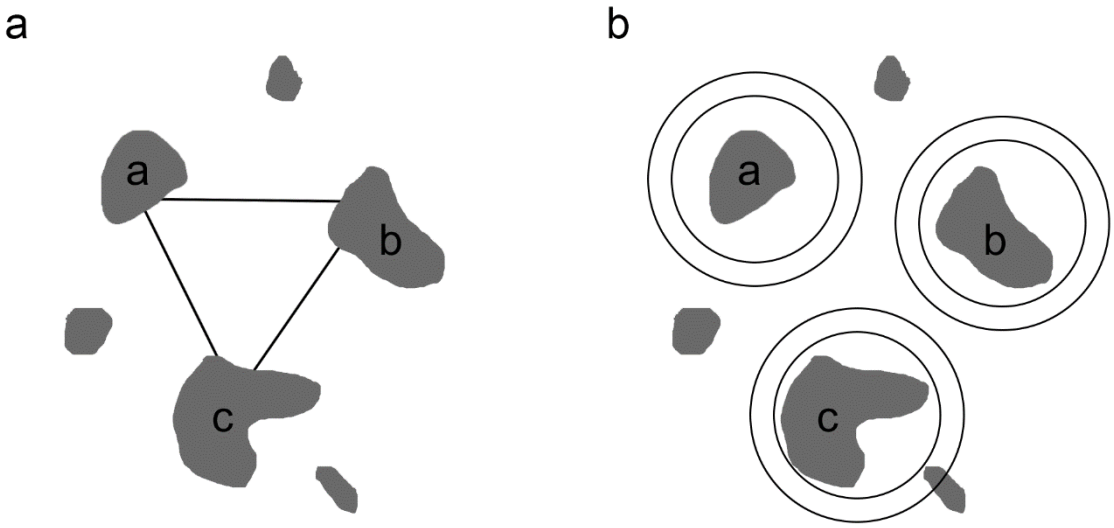

Figure 1. Two main methods for studying the influence of landscape structure on the distribution of genetic variation - 'link' (a) and 'node' (b) method. Adapted from DiLeo & Wagner (2016).

In population genetic research, the landscape is often divided in two components: the suitable habitat and the unsuitable matrix surrounding it. In classical population genetic approaches, the matrix has often been considered as uniform, but should not be treated as such, because it consists of different elements of landscape with potentially different effects on the gene flow of species. For instance, a forest and a field can influence the gene flow between isolated populations of a grassland plant differently (DiLeo & Wagner 2016) and should thus not be treated equally. With that, landscape elements can be appointed different values for degrees of resistance. For example, forest might have a higher value of resistance than a grain field since the forest is likely a stronger barrier than the field. These values could be added to a map producing a resistance landscape. If a route is then marked on this resistance landscape following the lowest resistance values, the resistance values can be added with the result being called landscape distance. This does not have to be the same as geographical distance and as a result the shortest landscape distance might not be the same as geographical distance. This data can be correlated with measures of gene flow between populations to examine how the structure of the landscape influences gene flow and genetic diversity (Holderegger *et al.* 2010). It is also

possible to not look at the whole landscape, but just parts of it, like transect or corridor approaches do. With this approach a straight corridor with a buffer zone is being used between two sites. The structure of the landscape is then assessed only in the corridor (Balkenhol *et al.* 2015). One easier method is to use overlays. In that case, the genetic structure of populations is visualized on a map, which also depicts the distribution of landscape elements one is interested in. Then the overlays of the locations of the genetic groups and the landscape elements are studied (Holderegger *et al.* 2010). However, overlay studies are rather subjective as they are only studied visually.

1.3 Objectives of the thesis

During the last century, the area of semi-natural grasslands has drastically decreased throughout Europe. Because of changes in land management strategies, grasslands are nowadays grazed and mown substantially less frequently than a century ago. Many former pastures are being forested, became overgrown, or have been turned into intensively managed agricultural fields, resulting in a decrease and fragmentation of grassland area. Therefore, many typical grassland species went locally extinct or have become scarce in these areas (Hahn *et al.* 2013; Honnay *et al.* 2007). Having experienced severe fragmentation, semi-natural grasslands serve as a suitable study system for examining the consequences of recent human-induced landscape changes for biodiversity.

In Estonia, alvars have lost most of their historic area and have become more isolated during the last 100 years. Lack of mowing and grazing is the main reason for area loss and increased isolation because alvars, being semi-natural grasslands, require moderate management to persist. Analyzing the effects of such a vast land use change on genetic diversity of plant species in alvars is thus of great interest. In this thesis, *Primula veris* was used as a study species because it is characteristic to semi-natural grasslands, such as alvars. *P. veris* is a cross-pollinating and insect-pollinated plant, making it very susceptible to landscape changes. In addition, knowledge about genetic diversity of plants (e.g. *P. veris*) in fragmented alvars could help to optimize conservation activities to account for maintaining genetic diversity.

Nineteen populations of *Primula veris* occurring in alvar grasslands of Saaremaa and Muhumaa were selected for genetic analysis. In addition, I recorded current and historic landscape configuration to examine the response of genetic diversity to landscape changes. I asked the following questions:

1) Do historic landscape characteristics (habitat area, forest area, human population density) have a stronger effect on current genetic diversity than current landscape features, i.e. do patterns of genetic diversity exhibit a lagged response to landscape change?
2) Does overgrowing of habitats (habitat loss, increase in forest area) influence the genetic diversity of grassland plant populations?
3) What influence does historic and current human population density have on genetic diversity?

2. Materials and methods

2.1 Study sites, study species, sampling

Study sites were located in calcareous grasslands – alvars – in western Estonia on the islands of Saaremaa and Muhumaa. Annual mean temperature in the area is 6°C and precipitation 585 mm (Ilmateenistus). Alvars have characteristic shallow calcareous soils occurring on limestone bedrock. They offer a habitat for many plant species adapted to the specific soil conditions at local as well as at community scale and thus have high conservation value (Pärtel & Zobel 1999). In Estonia, alvars are typically located near coastal areas in northern and western Estonia, making the climate of Estonian alvars more humid and with smaller temperature range. The study sites belong to the framework of a large-scale restoration project LIFE to Alvars. The aim of this project is to restore 2500 hectares of the most valuable, but currently overgrown alvar grasslands in Estonia. This is achieved by cutting the trees and shrubs on the sites and consequent grazing of the area (Project LIFE to Alvars).

Primula veris (*Primulaceae*), the study species, is an herbaceous perennial plant flowering in May in Estonia. *P. veris* is mostly found in well-drained, herb-rich meadows and grasslands, in shrub or woodland ridges and edges, and on calcareous cliffs. It is a shade-intolerant and drought-tolerant species (Brys & Jacquemyn 2009). *P. veris* was chosen as a study species, because it is a common plant in semi-natural grasslands and characteristic to alvar grasslands (Helm 2003). The study species is cross- and insect-pollinated. It is mostly pollinated by different species of *Hymenoptera* (e.g. bees), but also some species of *Coleoptera* (beetles) and *Lepidoptera* (butterflies). Pollen flow of *P. veris* is generally limited to a few meters from parental plants and seed dispersal is restricted to a few centimeters from maternal plants (Brys & Jacquemyn 2009).

Samples were collected from 19 populations of *Primula veris* occurring on restoration project sites scattered over Saaremaa and Muhumaa (Fig. 2). Population sizes ranged from approximately 20 to 5000 individuals. Leaves (1-3 from each individual) of *Primula veris* were collected in the summer of 2015 and 2016. From each population, leaves of 30 randomly sampled individuals (or as much as possible) were collected and stored in silica gel. The distance between sampled individuals within a population was at least one meter. 20 individuals (or as much as possible) from each population were used for genetic analysis totaling 338 individuals.

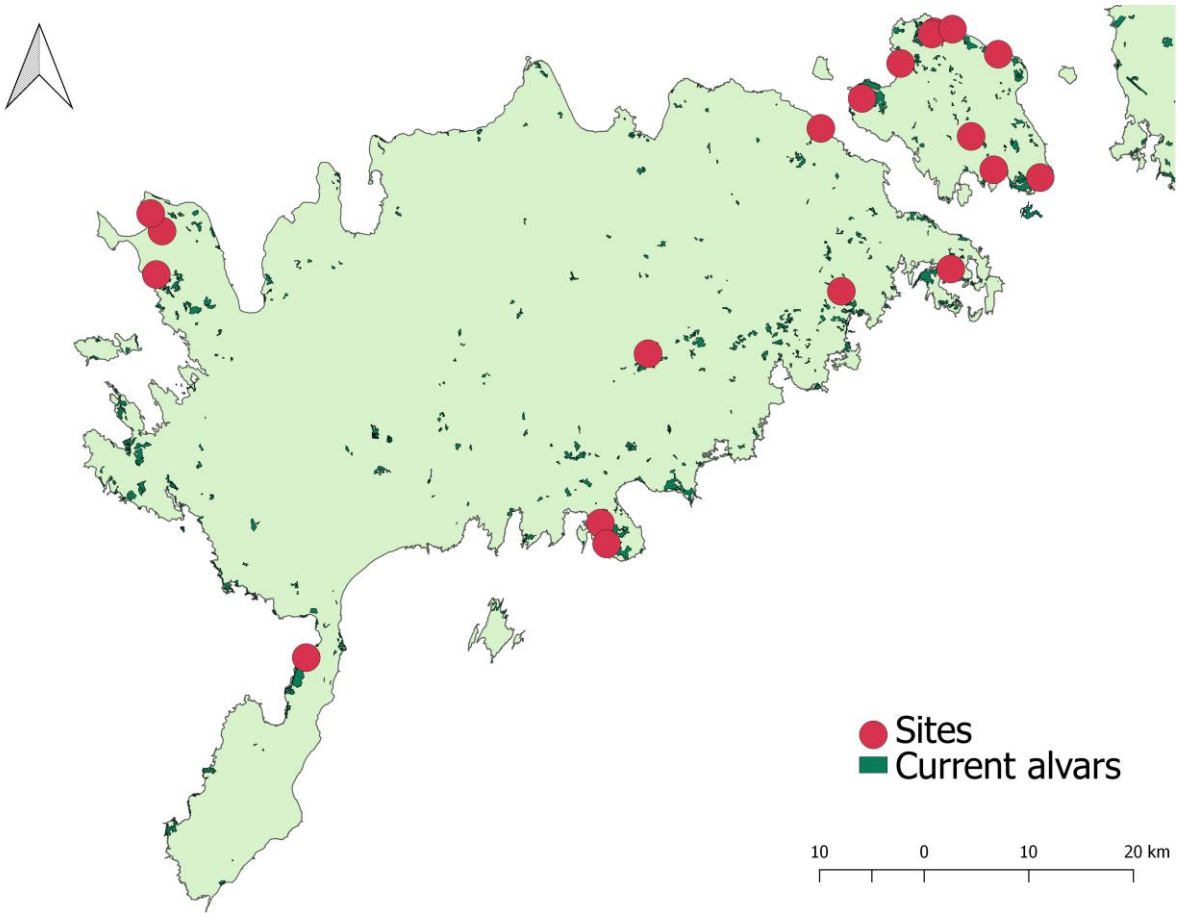

Figure 2. Study sites in Saaremaa and Muhumaa and current alvar distribution in the area.

2.2 Labwork

2.2.1 DNA extraction

Samples were randomized and c. 25 µg of leaf material per sample was weighed. Infected and damaged parts of leaves were avoided if possible. Leaves were ground for 2 minutes using two 2.3-mm metal beads in a Mixer Mill 301 (Retsch GmbH, Haan, Germany). DNA was extracted using the LGC speadex plant maxi kit (LGC, Berlin, Germany) with the modification of using 400 µl lysis buffer mix (containing 4 µl RNAse, 0.8 µl proteinase K and 395.2 µl PN lysis buffer), 420 µl binding buffer and 10 µl bead solution, two washing steps with 400 µl PN 1 washing buffer, one washing step with 400 µl PN 2 washing buffer, and eluting extracted DNA in a final volume of 50 µl elution buffer. Extraction steps of binding, washing and elution were

done on a KingFisher Flex Purification System (Thermo Fisher Scientific, Waltham, MA, USA). The concentration of extracted DNA was measured using Spark M10 Multimode Microplate Reader (Tecan Trading AG, Switzerland).

2.2.2 Sequencing library preparation

Extracted DNA was prepared for sequencing using double digest restriction-site associated DNA sequencing (ddRADseq; Peterson *et al.* 2012). This method was chosen because it is suitable for large sample sizes and for obtaining single-nucleotide polymorphism (SNP) data that are being used for further analysis. All laboratory work was done in collaboration with the Genetic Diversity Centre (GDC), ETH Zurich, Switzerland.

For adapter annealing, corresponding top and bottom P1.1 and P1.2 oligonucleotides (each 100 µM) were combined (1:1 ratio) with annealing buffer of a final concentration of 1x (AB; 10x AB = 500 mM NaCl, 100 mM Tris/Cl pH 7.5). Annealing was done in 98°C for 2.5 min followed by cooling down to room temperature at a rate of 2°C/min in a Labcycler (SensoQuest, Göttingen, Germany). The result of annealing was a solution of 40 µM per Eco- and Taq-site adapters with final working concentration for Eco-site adapters being 0.5 µM, and for Taq-site adapters 5 µM, using 1x AB. Taq-site top P2.2 adapters were biotinized at the 3' end. 48 different Eco-adapters and 2 different Taq-adapters were used (Peterson *et al.* 2012). To distinguish polymerase chain reaction (PCR) duplicates in bioinformatics analysis, I used four degenerated bases (equal mixture of A, C, G, T nucleotides at each nucleotide position) within the Taq-site (Tin *et al.* 2015).

All DNA samples were standardized to a concentration of about 5.77 ng DNA/µl. DNA was digested (i.e. cutting DNA at certain sites) in a two-step process using EcoRI and TaqI enzymes. In the first step, 0.5 µl EcoRI-HF NEB enzyme and 3 µl 10x Smartcut buffer NEB were added to 26 µl standardized DNA (per sample), and DNA was digested for 45 minutes at 37°C. In the second step, 0.5µl TaqI NEB and 0.5ul 1x Smartcut buffer were added to the previous mix and digested for another 45 minutes in 65°C.

Digested DNA was purified using custom-made SPRI bead solution (OpenWetWare) in a 1:1 ratio. After an incubation time of 15 minutes at room temperature, samples were placed on a magnetic stand to separate beads with attached DNA and the supernatant, which was discarded. Beads with attached DNA were washed two times using 70% ethanol and eluted in a final volume of 20 µl using PCR clean H_2O.

For ligation, a ligation mix containing 2 µl P2-biotin Taq-Adapter (5uM), 3 µl T4 Ligase Buffer 10x, 1 µl T4 Ligase NEB (400U/µl) and 2 µl H_2O per sample was prepared. 20 µl purified DNA and 2 µl P1 Eco-Adapter (0.5uM) were added to 8 µl ligation mix, and incubated for 25 minutes at 23°C, 10 minutes at 65°C, and then cooled to 4°C at a Labcycler (SensoQuest, witec ag, Göttingen, Germany).

Samples containing the same Taq-P2-biotin adapter were pooled together. For size selection, 300 µl ligated DNA pool and 210 µl AMPure BECKMAN beads (Beckman Coulter, Indianapolis, USA) were mixed, incubated for 10 minutes at room temperature and separated on a magnetic stand. Supernatant, containing size-selected DNA fragments, was saved, mixed with 0.12x undiluted AMPure BECKMAN beads, and incubated for 10 minutes at room temperature. The mix was then placed on a magnetic stand and supernatant was discarded. After two consequent washing steps with 70% ethanol, size-selected DNA fragments were dried for 10 minutes and eluted in 30 µl of (PCR clean) H_2O. The mix was incubated for 2 minutes at room temperature, again separated on a magnetic stand, and supernatant was saved. The concentration of purified size-selected DNA fragments was measured using Qubit fluorometer (ThermoFisher Scientific, Waltham, MA, USA).

For selecting for fragments with P2-biotin labeled adaptors, Dynabeads M-270 Streptavidin (Dynal, Invitrogen, ThermoFisher Scientific, Waltham, MA, USA) were used. First, 15 µl of beads were washed three times with 1x B&W buffer (100 µl) and resuspended in 2x B&W buffer in 2x the original volume (= 30 µl). 30 µl of size selected DNA was added to 30 µl of Dynabead solution. The mix was incubated for 15 minutes at room temperature, spinning the tube every 5 minutes. The mix was placed on a magnetic stand and supernatant was discarded. Beads with attached DNA fragments were washed three times with 1x B&W buffer (100 µl each) and resuspended in 45 µl (PCR clean) H_2O.

For polymerase chain reaction (PCR) 45 µl bead suspension from previous step, 3 µl primer 1 (10 µM), 3 µl primer 2 (10 µM), and 50 µl KAPA HiFi Hotstart Ready mix (KAPA Biosystems, Wilmington, MA, USA) were mixed. The mix was divided into four equal volumes to avoid PCR bias and PCR reactions were performed on a Labcycler (SensoQuest, Witec ag, Göttingen, Germany) with the following conditions: 95°C for 2 min (pre-heated), 9 cycles of 98°C for 20s, 65°C for 30s, and 72°C for 30s, and final cooling down to 4°C. PCR reactions were combined, cleaned as above with custom-made SPRI bead solution (0.6x; OpenWetWare), and eluted in a final volume of 20 µl using (PCR clean) H_2O.

Final DNA solution was analyzed for its concentration and fragment size using Qubit fluorometer (ThermoFisher Scientific, Waltham, MA, USA) and 2200 Tapestation (Agilent Technologies, Santa Clara, CA, USA), respectively.

2.2.3 Preparation for sequencing

For the final steps prior to sequencing, the molarity of the final ddRAD libraries was calculated according to their mean fragment size:

$$M[nM] = \frac{C}{F \cdot M_{bp}} \cdot 6, \quad (1)$$

where M is the molarity of the final library (in nM) with a mean fragment size F (in bp; i.e. 450 bp), calculated with the overall library concentration c (in ng/µl) after final purification and $M_{bp} = 660\ g/mol$.

Libraries with distinct multiplexing indices were combined resulting in a final library of at least 5 nM consisting of 96 individuals (2x 48 uniquely barcoded individuals with two multiplex indices). A reference library accounting for 15% of final library volumes was used (consisting of *Spinacea oleracea* and *S. tetrandra*, Claudia Michel, ETH Zurich, Switzerland). The reference library contained multiplex indices 4, 5, 7, 9 to increase variability of indices within our samples per Illumina sequencing lane.

Pooled libraries were prepared according to guidelines of the sequencing facility and sequenced on an Illumina HighSeq2500 (Illumina, Inc, San Diego, CA, USA) at the Functional Genomics Center Zurich (Switzerland), using one lane per library with 125 cycles in single-end read (125 bp), high-output mode.

2.3 Bioinformatics

The bioinformatic analysis was done in collaboration with the GDC, ETH Zurich. The analysis was carried out with UNIX-based software. Reads (sequenced DNA fragments) were demultiplexed using the process_radtags program of *Stacks* version 1.47 (Catchen *et al.* 2011; Catchen *et al.* 2013) separating individuals that were pooled together for sequencing. PCR duplicates were removed using the clone_filter program of *Stacks*. Sequences were filtered using trimmomatic v0.36 (Bolger *et al.* 2014) removing low quality bases from sequences and filtering sequences depending of their length (minimum 50 bases). Sequences were aligned and mapped against a reference genome (Nowak *et al.* 2015) using Burrows-Wheeler Aligner v0.7.17 (BWA; Li 2013). The final step was SNP calling, i.e. finding SNPs in the produced

reads using the reference genome. SNP calling was done using freebayes v1.1.0-54-g49413aa (Garrison & Marth 2012).

To exclude SNPs with low quality and individuals with too much missing data, SNP filtering was done using vcftools v0.1.12b (Danecek *et al.* 2011) following dDocent SNP Filtering Tutorial (Puritz *et al.* 2014a; 2014b). Genotypes called below 80% (across all individuals) and SNPs that had a minor allele count less than three were filtered out. Further, genotypes with more than three reads and individuals with more than 20% missing information were filtered out. Data were restricted to variants called in high percentage of individuals. Filtering by mean depth of genotypes with a threshold of 20 was also done. In addition, filtering by a population specific call rate was applied by estimating missing information for loci in each population using a threshold of 20%. Only loci with an allele balance between 0.25 and 0.75 as well as close to 0 (indicating almost fixed alleles) were kept. Following, a filter looking at the ratio of mapping qualities between reference and alternate alleles was applied. RADseq loci and alleles should start from the same genomic location which is why there should not be a large discrepancy between mapping qualities of them.

Li (2014) found with whole genome samples that high coverage can lead to inflated locus quality scores. Consequently, Li proposed that for read depths greater than the mean depth plus 2-3 times the square root of mean depth the quality score will be twice as large as the depth in real variants and below that value for false variants. Because this might be too conservative for RADseq data, two filters were used in the present study. First, loci that had a quality score below 1/4 of the depth were removed. Second, loci with the mean depth plus 2-3 times the square root of mean depth that did not have the quality score twice as large as the depth were removed. The final filtering was done using vcfallelicprimitives from *vcflib* (Garrison & Marth 2012) and *vcftools* where indels (i.e. insertions or deletions of bases in genome) were removed.

Genotype information was extracted from the resulting VCF file using *vcftools*. The file was transformed to a genind object in R version 3.4.2 (R Core Team 2017) using the package *adegenet* and to a GenAlEx input file using the package *poppr*. Population based genetic diversity indices (unbiased expected and observed heterozygosity, uH_e and H_o, respectively, percentage of polymorphic loci, %P) were calculated using GenAlEx version 6.503 (Peakall & Smouse 2005; 2012). Unbiased expected heterozygosity accounts for differences in population sizes. Inbreeding coefficient (F_{is}) was calculated using the package *genepop*.

2.4 Landscape data

Landscape data was obtained in the frames of a biodiversity inventory of LIFE to Alvars sites. Landscape data was calculated using historic maps from 1930s that were done during historic vegetation survey (Laasimer 1965) and contemporary maps from 2010 obtained from Seminatural Community Conservation map layer for semi-natural grasslands. Map analyses were done in ArcGIS version 10.4 (ESRI 2016). Both historic and contemporary area of alvars were calculated for each site. In addition, I assessed the absolute and proportional area loss of alvar grasslands. Forest area was calculated for each site in a 5 km buffer for both historic and contemporary data using Estonian Basic Map (Estonian Land Board) for contemporary data. 5 km radius was chosen because with shorter radiuses there was often no forest area in the buffer. Historic population density in 5 km buffers around each site was calculated from a map based on the official population census in 1922 (Tammekann 1929). Contemporary population density was obtained from the official population census in 2011 from Statistics Estonia (Statistikaamet). Historic and current human population density was calculated as human population in 5 km buffer divided by the area of the buffer (people/ha). These specific variables in addition to the area of alvars were chosen because they have previously been shown to be significant for genetic diversity of grassland plants (Aavik *et al.* 2017; Helm *et al.* 2009).

2.5 Data analysis

All statistical analyses were done in R version 3.4.2 (R Core Team 2017). Linear mixed-effects models were done with genetic diversity measures as dependent variables and population size in 2017, current and historic grassland area, loss of habitat area, proportional habitat loss, current and historic forest area in 5 km buffers and current and historic human population density as independent variables (package *lmer*). As the percentage of polymorphic loci represented frequency, generalized linear mixed-effects models (package *lmerTest*) were applied. All models had region (Saaremaa and Muhumaa) as a random variable, because it may have had an effect on dependent variables but the region effect itself was not of main interest. In all models, I included only those explanatory variables, which were not strongly correlated to each other ($r \leq 0.6$; Table 1). Non-significant variables were removed step by step. All models with significant variables were ranked according to Akaike information coefficient (AIC; Akaike 1974). Akaike weights were also calculated to show the relative likelihood of the models.

To test if populations geographically closer to each other are also genetically more similar to each other, an isolation by distance (IBD) analysis with 1000 permutations was done using the package *vegan*. Fixation index (F_{st}) was used as a measure for genetic distance.

Table 1. Correlations between dependent variables. Significant correlations ($p < 0.05$) are marked with *.

	Population size	Historic habitat area	Current habitat area	Area loss	Proportional area loss	Historic forest area (5km buffer)	Current forest area (5km buffer)	Current human popula-tion density	Historic human popula-tion density
Population size		-0.423	-0.226	-0.377	-0.155	-0.120	-0.177	-0.043	0.089
Historic habitat area	-0.423		0.086	**0.979***	**0.537***	-0.098	0.028	0.221	0.126
Current habitat area	-0.226	0.086		-0.118	-0.283	**-0.567***	-0.407	0.386	0.125
Area change	-0.377	**0.979***	-0.118		**0.592***	0.018	0.111	0.142	0.100
Proportional area loss	-0.155	**0.537***	-0.283	**0.592***		-0.092	-0.106	0.165	0.166
Historic forest area (5 km buffer)	-0.120	-0.098	**-0.567***	0.018	-0.092		**0.808***	-0.428	-0.312
Current forest area (5 km buffer)	-0.177	0.028	-0.407	0.111	-0.106	0.808		-0.409	-0.192
Current human population density	-0.043	0.221	0.386	0.142	0.165	-0.428	-0.409		**0.814***
Historic human population density	0.089	0.126	0.125	0.100	0.166	-0.312	-0.192	**0.814***	

To analyze the genetic composition of examined populations, Discriminant Analysis of Principal Components (DAPC; Jombart *et al.* 2010) was done using the package *adegenet*. DAPC is a multivariate method designed to identify and describe clusters of genetically related individuals (Jombart *et al.* 2010). It was chosen because it does not rely on a particular population genetic model and is thus free of assumptions about Hardy-Weinberg equilibrium (HWE), for example, which is a prerequisite for most of the other methods for analyzing genetic structure. However, HWE of genetic data is difficult to meet, when populations suffer from the negative consequences of severe habitat fragmentation. The genind object previously used for making GenAlEx file was used as input file for DAPC. The optimal number of principal components was found using cross-validation. The optimal number of clusters was found using

k-means clustering. Finally, data was transformed using PCA (Principal Component Analysis) and Discriminant Analysis was performed.

3. Results

3.1 Genetic diversity and landscape data of the study populations

SNP-data of 4588 loci were obtained for subsequent data analysis after bioinformatic analysis. Unbiased expected heterozygosity ranged from 0.244 to 0.313, observed heterozygosity from 0.291 to 0.368, inbreeding coefficients from -0.280 to -0.108 and percentage of polymorphic loci from 60.6% to 92.0% (Table 2). Information about landscape data used for statistical analyses is shown in Table 3.

Table 2. General genetic information about the study populations of *Primula veris* in Muhu- and Saaremaa. N - number of samples obtained for the analysis, H_o - observed heterozygosity, uH_e - unbiased expected heterozygosity, F_{is} - inbreeding coefficient and %P - percentage of polymorphic loci. Average values for Muhu- and Saaremaa genetic diversity estimates are also provided.

Region	Site	Longitude	Latitude	N	Population size 2017	uH_e	H_o	F_{is}	%P
Muhu	Koguva	23.091	58.6108	19	100	0.303	0.360	-0.193	87.6%
Muhu	Lõetsa1	23.3141	58.65	19	500	0.312	0.359	-0.152	89.5%
Muhu	Mäla	23.2709	58.5794	15	100	0.288	0.337	-0.175	80.2%
Muhu	Nõmmküla	23.2085	58.6686	19	1500	0.313	0.350	-0.121	91.7%
Muhu	Nõmmküla	23.2042	58.6668	20	100	0.308	0.344	-0.120	92.0%
Muhu	Paenase	23.1536	58.6412	19	500	0.310	0.367	-0.189	90.9%
Muhu	Võiküla 1	23.385	58.5447	20	500	0.309	0.368	-0.196	90.6%
Muhu	Võiküla2	23.3087	58.551	19	20	0.289	0.344	-0.196	84.2%
Muhu	Üügu	23.2383	58.6711	19	3000	0.311	0.352	-0.137	91.2%
Saaremaa	Asva1	23.0612	58.4453	6	100	0.249	0.310	-0.280	60.6%
Saaremaa	Kahtla1	23.24	58.4656	20	100	0.269	0.326	-0.212	78.3%
Saaremaa	Kõruse	21.9393	58.4465	19	400	0.287	0.339	-0.188	85.0%
Saaremaa	Lõu	22.2014	58.1221	20	5000	0.284	0.327	-0.156	78.2%
Saaremaa	Neeme	21.9465	58.4839	20	2000	0.298	0.342	-0.154	87.9%
Saaremaa	Neeme	21.927	58.4986	18	2000	0.304	0.355	-0.170	86.9%
Saaremaa	Orinõmme	23.023776	58.584617	20	2000	0.306	0.357	-0.169	87.7%
Saaremaa	Vanamõisa	22.6743	58.2432	20	1000	0.244	0.298	-0.226	81.0%
Saaremaa	Vanamõisa	22.685017	58.225494	15	300	0.251	0.291	-0.165	80.4%
Saaremaa	Võrsna	22.7467	58.3891	11	100	0.274	0.302	-0.108	81.6%
Muhu average						0.305	0.353	-0.164	88.7%
Saaremaa average						0.277	0.325	-0.183	80.8%

Table 3. Landscape information about the study populations of *Primula veris* in Muhu- and Saaremaa. Areas are in hectars (ha).

Region	Site	Historic habitat area	Current habitat area	Area loss	Proportional area loss	Historic forest area (5 km buffer)	Current forest area (5 km buffer)	Historic human population density	Current human population density
Muhu	Koguva	571.95	257.32	316.17	55.28	40.17	1187.40	0.22	0.90
Muhu	Lõetsa1	254.89	23.26	231.76	90.93	159.67	1322.65	0.39	1.64
Muhu	Mäla	1145.54	14.01	1131.61	98.78	1108.30	3540.25	0.30	1.41
Muhu	Nõmmküla	686.70	122.00	564.70	82.23	0.00	1537.37	0.26	1.31
Muhu	Nõmmküla	686.70	122.52	564.91	82.26	0.00	1603.24	0.28	1.47
Muhu	Paenase	202.81	101.84	101.59	50.09	244.61	2529.88	0.23	1.02
Muhu	Võiküla 1	461.04	151.46	310.49	67.34	88.39	1057.44	0.17	0.65
Muhu	Võiküla2	1145.54	44.09	1101.71	96.17	668.16	2815.93	0.17	0.54
Muhu	Üügu	686.70	86.98	600.24	87.41	0.00	1385.16	0.28	1.37
Saaremaa	Asva1	1258.48	66.23	1192.65	94.77	97.29	2541.80	0.20	0.53
Saaremaa	Kahtla1	711.46	11.51	700.01	98.39	183.67	1103.26	0.13	0.15
Saaremaa	Kõruse	104.85	14.99	89.94	85.79	676.40	2766.77	0.14	0.11
Saaremaa	Lõu	175.35	53.02	122.64	69.94	56.63	1632.11	0.27	0.82
Saaremaa	Neeme	0.00	1.67	0	0.00	940.59	3065.95	0.15	0.13
Saaremaa	Neeme	289.92	55.62	234.63	80.93	552.13	2193.73	0.13	0.11
Saaremaa	Orinõmme	160.20	7.45	152.80	95.38	748.57	2286.58	0.19	0.72
Saaremaa	Vanamõisa	532.05	118.24	414.51	77.91	231.69	1584.26	0.19	1.52
Saaremaa	Vanamõisa	532.05	236.57	296.87	55.80	0.00	689.96	0.24	1.65
Saaremaa	Võrsna	351.58	139.66	212.75	60.51	696.40	4289.93	0.20	0.42
Muhu average		649.10	102.61	547.02	78.94	256.59	1886.59	0.26	1.15
Saaremaa average		411.59	70.50	341.68	71.94	418.34	2215.44	0.18	0.62

3.2 Landscape genetic analyses

In total, 18 models were obtained for different measures of genetic diversity (14 were chosen for further analysis where AIC value differences per genetic diversity variable were less than ten; Table 4), three for unbiased expected heterozygosity, four for observed heterozygosity, three for inbreeding coefficient, and four for the percentage of polymorphic loci. Models were ranked according to AIC values.

According to the model supported by the AIC, unbiased expected heterozygosity significantly ($p < 0.05$) decreased with increasing historic habitat area and current human population density (Table 4). Observed heterozygosity significantly decreased with increasing area loss, current habitat area and current human population density (Table 4). Inbreeding coefficient significantly decreased with increasing area loss (Table 4). Proportion of polymorphic loci significantly decreased with increasing area loss, current habitat area, historic human population density and current forest area, but increased with increasing proportional area loss (Table 4). Population size and historic forest area did not have an effect on any measures of genetic diversity used. The negative effect of habitat area loss on genetic diversity was the most consistent and significant result (Fig. 4).

Table 4. Ranking of models per variable of genetic diversity in the study populations of *Primula veris* in Muhu- and Saaremaa for model selection. For each model rank, independent variables, their effect on the dependent variable, summary statistics, AIC and weighted AIC (WiAIC) are shown.

Dependent variable	Model rank	Independent variable	Effect	Statistics	p-value	AIC	WiAIC
Unbiased expected heterozygosity, uH_e	1	Historic habitat area	-	$F_{1,15} = 20.43$	<0.001	-75.492	0.450
		Current human population density	-	$F_{1,15} = 6.00$	0.027		
	2	Area loss	-	$F_{1,15} = 20.42$	<0.001	-74.514	0.276
		Current habitat area	-	$F_{1,15} = 9.55$	0.007		
	3	Historic habitat area	-	$F_{1,15} = 20.41$	<0.001	-74.503	0.274
		Current habitat area	-	$F_{1,15} = 4.83$	0.044		
Observed heterozygosity, H_O	1	Area loss	-	$F_{1,14} = 14.35$	0.002	-64.714	0.339
		Current habitat area	-	$F_{1,14} = 5.80$	0.030		
		Current human population density	-	$F_{1,14} = 8.60$	0.011		
	2	Historic habitat area	-	$F_{1,14} = 14.34$	0.002	-64.707	0.337
		Current habitat area	-	$F_{1,14} = 2.91$	0.120		
		Current human population density	-	$F_{1,14} = 8.60$	0.011		
	3	Area loss	-	$F_{1,14} = 15.72$	0.001	-63.243	0.162
		Current habitat area	-	$F_{1,14} = 11.74$	0.004		
		Historic human population density	-	$F_{1,14} = 6.45$	0.023		
	4	Historic habitat area	-	$F_{1,14} = 15.71$	0.001	-63.237	0.162
		Current habitat area	-	$F_{1,14} = 7.05$	0.019		
		Historic human population density	-	$F_{1,14} = 6.45$	0.023		
Inbreeding coefficient, F_{IS}	1	Area loss	-	$F_{1,16} = 4.71$	0.045	-49.384	0.500
	2	Historic habitat area	-	$F_{1,16} = 4.70$	0.045	-49.304	0.480
	3	Historic habitat area	-	$F_{1,16} = 4.53$	0.049	-42.890	0.019
		Historic human population density	+	$F_{1,16} = 4.60$	0.048		
Polymorphic loci, %P	1	Area loss	-	$F_1 = 1553.17$	<0.001	578.599	0.778
		Current habitat area	-	$F_1 = 47.00$	<0.001		
		Proportional area loss	+	$F_1 = 6.22$	0.011		
		Historic human population density	-	$F_1 = 125.00$	<0.001		
		Current forest area (5 km buffer)	-	$F_1 = 8.43$	<0.001		
	2	Historic habitat area	-	$F_1 = 1591.76$	<0.001	581.745	0.161
		Historic human population density	-	$F_1 = 112.93$	<0.001		
		Current forest area (5 km buffer)	-	$F_1 = 20.45$	<0.001		
	3	Area loss	-	$F_1 = 1546.61$	<0.001	584.963	0.032
		Current habitat area	-	$F_1 = 46.95$	<0.001		
		Proportional area loss	+	$F_1 = 6.14$	<0.001		
		Historic human population density	-	$F_1 = 125.72$	<0.001		
	4	Historic habitat area	-	$F_1 = 1592.81$	<0.001	585.188	0.029
		Current habitat area	+	$F_1 = 0.51$	<0.001		
		Proportional area loss	+	$F_1 = 6.14$	<0.001		
		Historic human population density	-	$F_1 = 125.71$	<0.001		

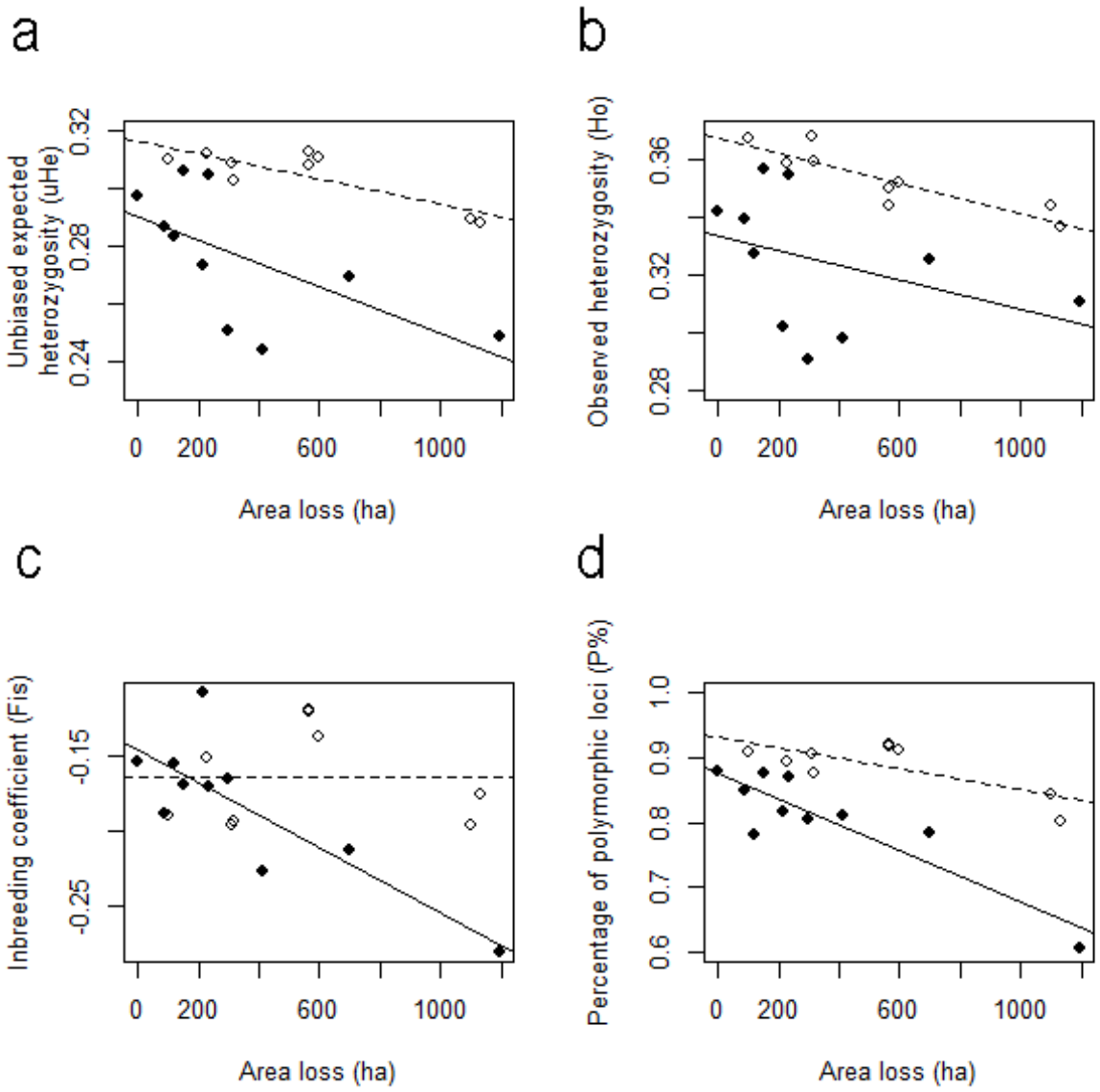

Figure 4. The influence of habitat area loss on different genetic diversity measures in the study populations of *Primula veris* – unbiased expected heterozygosity (a), observed heterozygosity (b), inbreeding coefficient (c) and percentage of polymorphic loci (d). Each dot represents a population. Empty dots represent populations from Muhumaa, filled dots from Saaremaa. Dashed and solid lines represent linear regressions between genetic diversity measures and habitat area loss in Muhumaa and Saaremaa, respectively.

Isolation by distance (IBD) analysis results suggest there is isolation by distance in Saaremaa (Fig. 5; Mantel statistic r: 0.4772, significance: 0.002) and Muhumaa (Fig. 5; Mantel statistic r: 0.5227, significance: 0.003). This suggests that populations in Saaremaa and Muhumaa which are geographically closer to each other are genetically more similar.

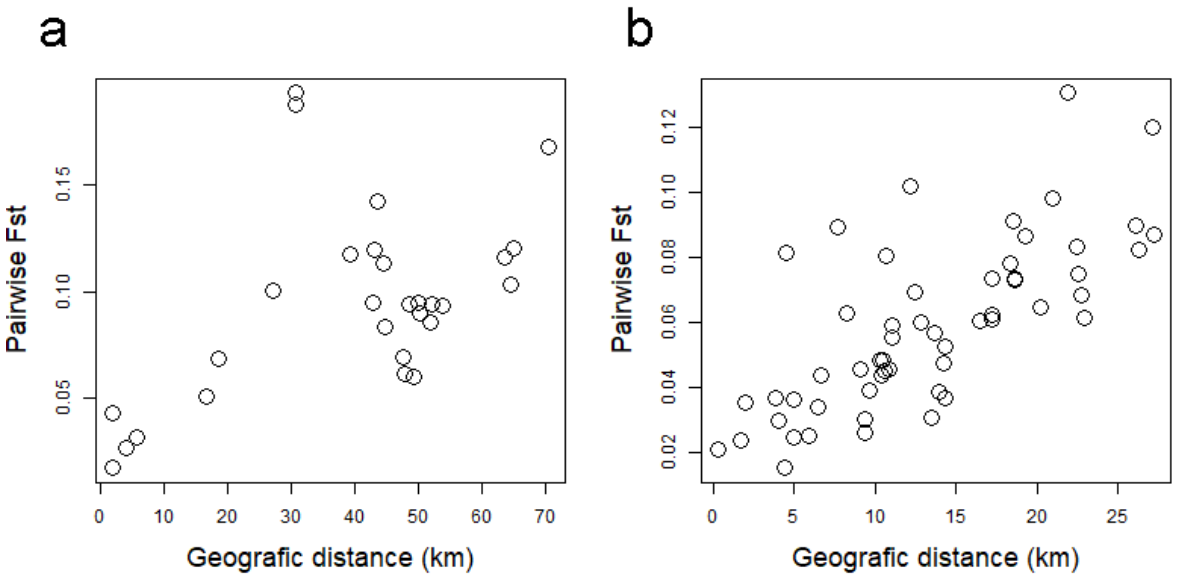

Figure 5. Isolation by distance plot in Saaremaa (a) and Muhumaa (b). Each dot represents a corresponding relationship for two populations. Pairwise F_{st} (genetic differentiation) represents genetic distance.

3.3 Analysis of genetic structure

A discriminant analysis of principal components (DAPC) resulted in the separation of study populations in seven clusters based on their genetic similarity (Fig. 6 and 7). 120 principal components and six discriminant functions were used. Populations in Muhumaa and East-Saaremaa were genetically more similar to each other than to populations in West- and central Saaremaa. Populations in Muhumaa had more individuals belonging to different clusters than in Saaremaa (Fig. 7).

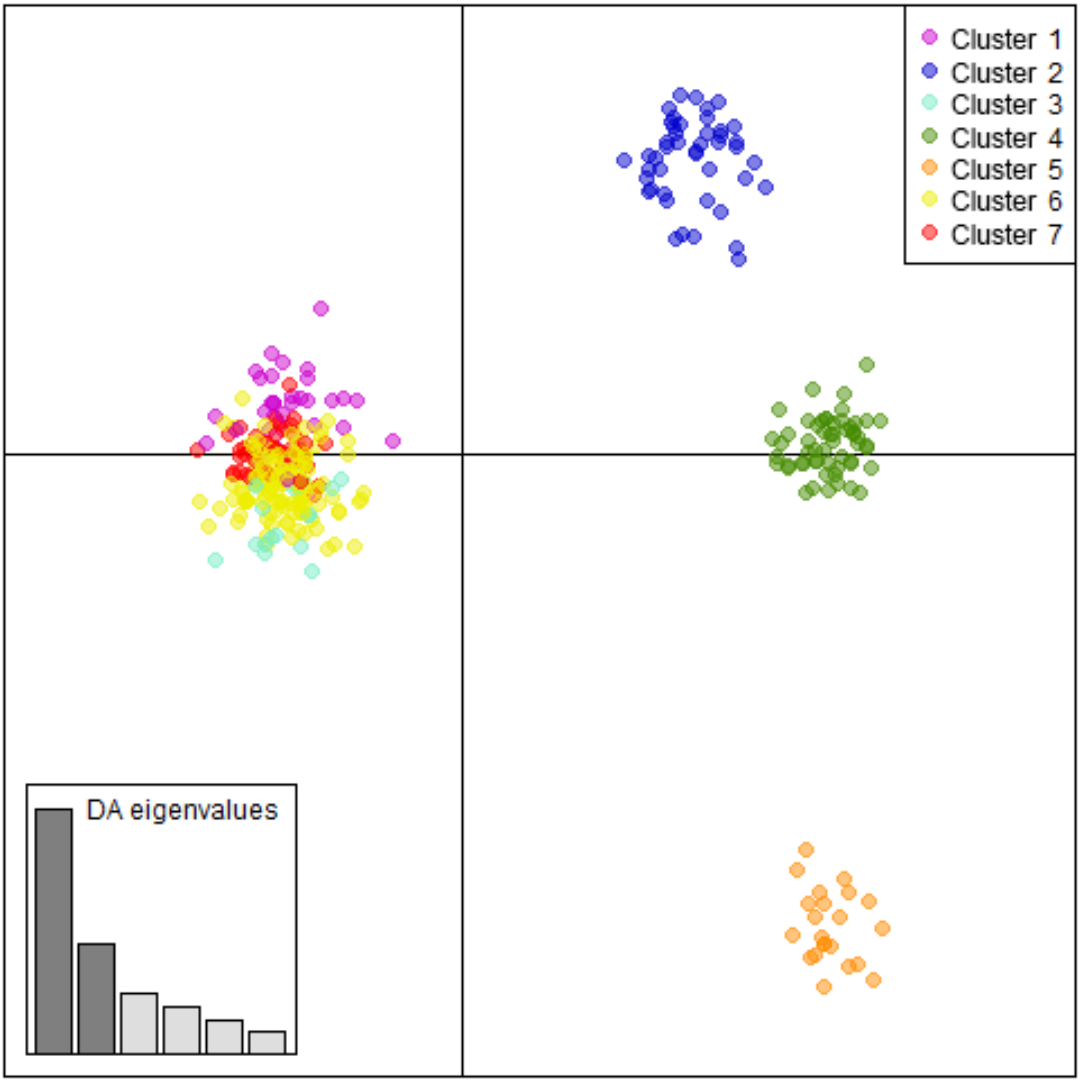

Figure 6. Discriminant analysis of principal components (DAPC) scatterplot representing the genetic structure of the study populations of *Primula veris* in Saare- and Muhumaa. Every dot represents an individual of *P. veris*. Individuals closer to each other are genetically more similar. DA eigenvalues stand for discriminant function eigenvalues indicating how well that function differentiates the clusters. Larger values indicate better differentiation ability.

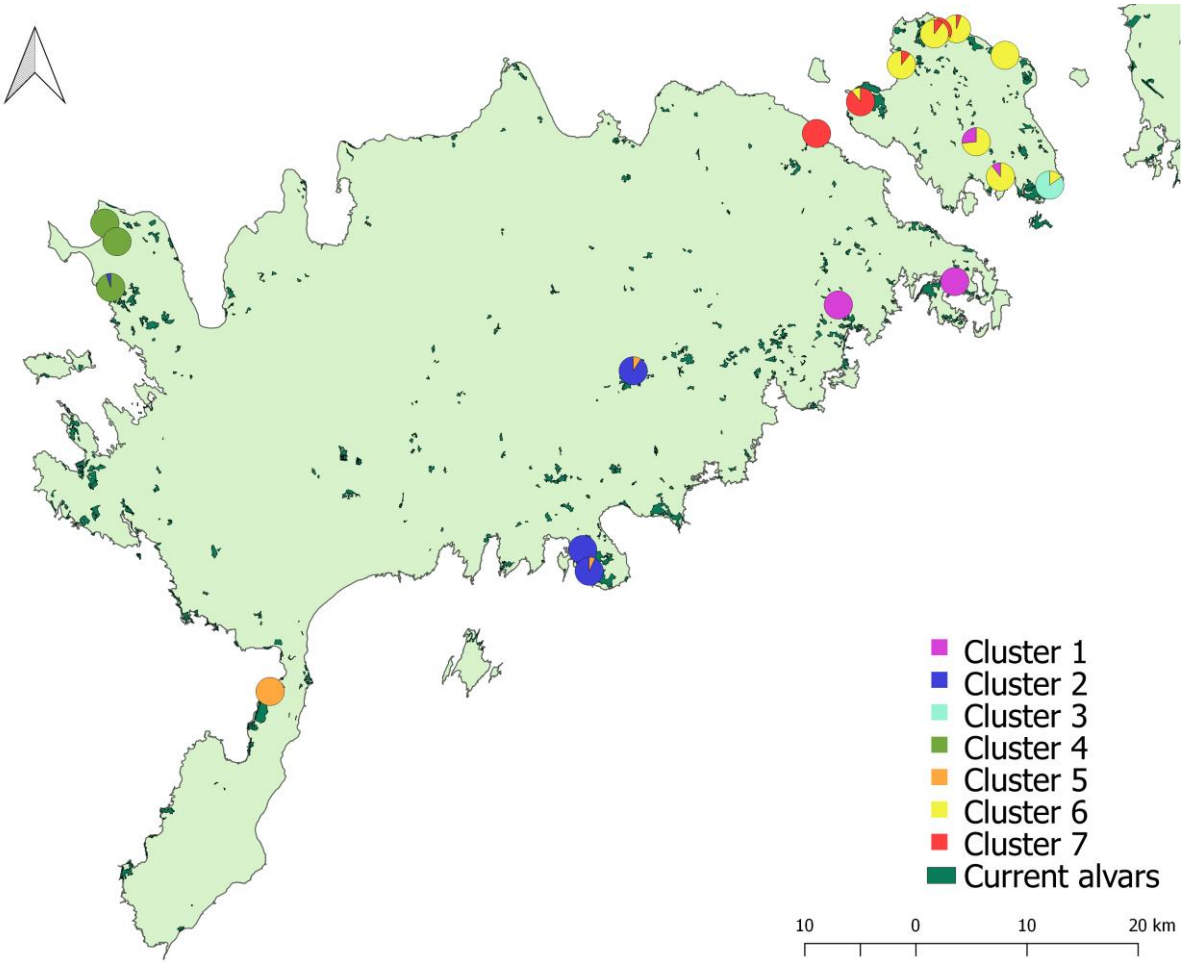

Figure 7. Map showing the distribution of individual memberships of *Primula veris* in Muhu- and Saaremaa to discriminant analysis of principal components (DAPC) genetic clusters. Each pie chart represents a population.

4. Discussion

Fragmentation and loss of habitat area are serious threats to genetic diversity of plants. Finding out which landscape elements influence genetic diversity and how fast genetic diversity reacts to landscape changes is thus very important and can help in guiding decision-making in environmental conservation policy. In my thesis, I asked if patterns of genetic diversity exhibit a lagged response to landscape change. I found that genetic diversity of *P. veris* is influenced by habitat area loss. However, the response of genetic diversity to changes in landscape configuration in Estonian alvars during the last century did not exhibit a time lag. I also asked if overgrowing of habitats (and thus increase in forest area) influence the genetic diversity of grassland plant populations. I found that for *P. veris*, current forest had a negative influence on genetic diversity, but historic forest area did not have an effect. Finally, I asked what influence does historic and current human population density have on genetic diversity. I found that both historic and current human population density have generally a negative influence on genetic diversity of *P. veris.*

4.1 The effect of habitat area

Current and historic habitat area had a negative effect on genetic diversity regardless of population size. This unexpected relationship could be driven by the correlation between historic habitat area and area change: historically bigger alvar grasslands have lost more area compared to historically smaller grasslands. Consequently, the negative effect of habitat area on genetic diversity might be rather an effect of the total amount of area lost, with area loss having a negative effect on genetic diversity. The negative effect of current habitat area on genetic diversity is also negligible, since the current habitat area obtained from the available map layer may not represent the quality of habitats adequately because it also includes patches, which have partially overgrown with junipers. This, in turn, may affect small-scale gene flow of *P. veris* via seed and pollen movement. Indeed, the negative effect of woody elements on one of the measures of genetic diversity (percentage of polymorphic loci) was also demonstrated in the current study (Chapter 4.2). Therefore, a more detailed landscape analysis is needed to determine the availability of suitable pathways for gene flow of *P. veris* in current alvar grasslands.

Genetic diversity of *P. veris* decreased with increasing habitat area loss. Habitat loss was comparable in Muhu- and Saaremaa. Loss of genetic diversity often indicates reduced gene flow between populations (Balkenhol *et al.* 2015). Changes in gene flow can be, among others,

related to pollen and seed vectors (Holderegger *et al.* 2010). On the one hand, pollinators as pollen vectors are known to be susceptible to landscape changes (Aguilar *et al.* 2006). Indeed, pollinator observations on LIFE to alvars project sites revealed that overgrowing of alvars with junipers and trees is accompanied by a significant loss in the diversity and abundance of important pollinator groups - butterflies and bumblebees (Prangel 2017). It is thus very likely, that changes in landscape structure, such as the overgrowth of alvars and consequent loss of suitable habitat, has led to deficiency of pollen vectors, which, in turn, has an overall negative effect on landscape-level gene flow as well as on within-population genetic diversity, as was observed in the current study. On the other hand, grazing of alvars by sheep as seed vectors was historically widespread, but decreased drastically in the last decades (Helm 2011). The latter is also accompanied by the lack of a rotational grazing system of sheep among alvar grasslands resulting in less gene flow via seed dispersal.

Knowing that habitat area loss is the main reason for genetic diversity loss should direct conservational efforts to maintain and restore suitable habitat conditions of *Primula veris* and other grassland species. Nevertheless, it should not be forgotten that restoring only structural connectivity, i.e. the recovery of patches of grasslands in the landscape, might not be enough and maintaining functional connectivity is also necessary (Jacquemyn *et al.* 2010). With semi-natural grasslands like alvars, achieving functional connectivity may be a little easier, since they need grazing (or mowing) to remain open. Thus, cattle or sheep are necessary to preserve semi-natural habitats such as alvars by preventing overgrowth of grasslands with shrubs and by spreading plant seeds within and among grassland patches, thus increasing functional connectivity among spatially distinct plant populations (Rico *et al.* 2014). Using sheep and cattle are the main management tools introduced for ensuring the maintenance of LIFE to alvars projects sites. It would therefore be highly interesting to monitor, whether the recovery of grazing in the landscape will promote the recovery of genetic diversity in grassland plant populations.

4.2. The effect of surrounding landscape characteristics

Genetic diversity of *P. veris* populations was higher in Muhumaa than in Saaremaa. Although habitat loss has been severe on both islands, historical and current area of grasslands in Muhumaa provides better conditions for maintaining functional connectivity and could thus be the reason for higher genetic diversity in Muhumaa. Analysis of the genetic structure of *P. veris* populations in Saaremaa and Muhumaa showed that populations in Muhumaa and populations

in Eastern-Saaremaa are more similar to each other than to populations in West- and central Saaremaa. One reason for this could be the geographic distance between populations - populations in Muhumaa are generally closer to each other. Isolation by distance analysis showed that populations that are geographically closer to each other are indeed genetically more similar to each other. Another reason could be related to a potential grazing system, i.e. that the same sheep herds grazed on alvar grasslands on Muhumaa (and Eastern-Saaremaa) facilitating gene flow by seed.

While area loss was an important factor influencing genetic diversity of *P. veris*, forest area in the proximity of studied alvar grasslands was not as important with current forest area having a significant negative effect only on the percentage of polymorphic loci of *P. veris*. This indicates that genetic diversity indices react differently (Leimu *et al.* 2006) and percentage of polymorphic loci might be more sensitive to changes in forest area around grasslands.

The effect of forest on genetic diversity of insect-pollinated grassland plants is usually assumed to be negative as it is believed to be a barrier for pollinators and thus to gene flow (Aavik *et al.* 2014; 2017). I found that for *P. veris*, current forest had a significant negative influence on genetic diversity, but historic forest area did not have an effect. The absence of current forest influence on most of genetic diversity parameters corresponds to previous findings by Hahn *et al.* (2013) where forest was found not to be a barrier for gene flow and thus not influencing genetic diversity. The general lack of a response of genetic diversity to forest area in the present study could mean forest is not a barrier for some pollinators such as bees and bumblebees (Kreyer *et al.* 2004; Zurbuchen *et al.* 2010) that are among most common pollinators for *P. veris* (Brys & Jacquemyn 2009). However, current forest area did not include shrub area, whilst it was shown that shrub area has a substantial negative influence on the diversity and abundance of pollinators in the same study landscapes and could thus potentially affect landscape-scale gene flow of insect-pollinated plants (Prangel 2017). Consequently, future studies should include both forest and shrub area in the analysis.

4.3 The role of human population density

Human population density was another important factor influencing genetic diversity of *P. veris*. Both current and historic human population density had a negative effect on genetic diversity with current human population density having a stronger effect than historic human population density. This can be expected since human settlements would decrease the habitat

area for grassland plants by demanding more land for agricultural purposes (Vitousek *et al.* 1997).

The negative effect of current human population density on genetic diversity can be expected as recent land use changes have resulted in decrease of semi-natural grasslands (Helm *et al.* 2009). However, the negative effect of historic human population density on genetic diversity was not expected, but this result could partly reflect the correlation between the two variables. Helm *et al.* (2009) have previously shown that current human population density had a negative effect whereas historic human population density had a positive effect on the genetic diversity of *Briza media* in alvar grasslands of Saare- and Muhumaa. The positive effect of historic human population density on genetic diversity would be expected as traditional land management supports the survival of semi-natural grasslands.

The difference in results of the current study and Helm *et al.* (2009) could be explained by contrasting plant species characteristics. *B. media* is a wind-pollinated plant whereas *P. veris* is an insect pollinated plant. It has been shown that modified landscape structure influences insect-pollinated species generally more than wind-pollinated species, because insect-pollinating species depend largely on their pollinators and pollinators can be sensitive to landscape changes (Aguilar *et al.* 2008). Consequently, it is harder for insect-pollinated species to maintain their genetic diversity in a fragmented landscape, because fragmentation decreases the amount of pollinating insects (Honnay & Jacquemyn 2007; Schmidt *et al.* 2009). This makes *P. veris* more vulnerable to landscape changes including the change in human population density.

4.4 Time lag

In the present thesis, I found no strong support for patterns of genetic diversity of *P. veris* exhibiting a lagged response to landscape structure. Nevertheless, I found that genetic diversity is influenced by habitat area loss. It could still be argued that there is in fact a lagged response in genetic diversity of *P. veris* reacting to landscape structure, because genetic diversity was related to historic habitat area, but since historic habitat area was strongly correlated with habitat area loss, it is reasonable to assume genetic diversity is actually influenced by habitat area loss and not historic habitat area. This indicates that genetic diversity of *P. veris* has most probably already reacted to landscape change. Helm *et al.* (2009) have previously come to the same conclusion for *Briza media* in alvar grasslands of the same region. Nevertheless, it is possible that there is a delayed response of the genetic diversity of *P. veris* to landscape changes, but it is shorter than what was analyzed here (about 80 years). It is possible that some specific

characteristics of *P. veris* influence the detection of a time lag response of genetic diversity to changes in landscape structure. One of these characteristics could be the age of plants used for analysis (Van Geert *et al.* 2008). It has been demonstrated that the analysis based on individuals from younger generations would show a differential response to landscape change compared to older individuals sampled from the same population (Van Geert *et al.* 2008). Lagged responses of plants to landscape change is an interesting avenue of research, which deserves further attention because a growing number of studies on grassland plants have demonstrated so-called genetic extinction debt (Münzbergová *et al.* 2013; Plue *et al.* 2017; Reisch *et al.* 2017).

Furthermore, having no time lag could be useful for conservational purposes. If there is no time lag in the reaction of genetic diversity to landscape structure, then genetic diversity of *P. veris* reacts fast to landscape change in the study system and no further decrease in genetic diversity can be expected. This suggests that the effect of restoration should be detectable in genetic diversity rather quickly as well thus helping to determine if restoration efforts have been successful.

4.5 Conclusions

In conclusion, the present study demonstrates that landscape change in Estonian alvars has had a major effect on patterns of genetic diversity of *P. veris*. Habitat area loss had the strongest influence for genetic diversity of *P. veris* with an overall negative effect. While historic habitat area had also a negative effect on genetic diversity, it is not possible to say for certain if there is a time lag for genetic diversity of *P. veris*, because historic habitat area was strongly correlated to habitat area loss. Uncertainty in detecting a time lag could mean that the actual time lag is shorter than what was studied here and different historic time points should be analyzed in future studies. Current forest had a slight negative influence on genetic diversity of *P. veris*, meaning forest may not be a strong barrier for gene flow. Nevertheless, future studies should also examine the impact of shrub on genetic diversity, because overgrowth of alvars with juniper shrubs is one of the main reasons for deteriorating habitat quality and resulting landscape-scale fragmentation. Both historic and current human population density had a negative effect on genetic diversity. This means human settlements decrease the habitat area for grassland plants by demanding more land for agricultural purposes. Findings of the present study will help to direct conservation efforts in Estonian alvars. For instance, maintaining landscape scale-connectivity is important for preserving genetic diversity. Furthermore, maintaining and restoring semi-natural grassland area is necessary to ensure genetic diversity

of grassland plants. Knowing how and what influences genetic diversity of grassland plants makes it also possible to protect and maintain genetic diversity that, in turn, provides adaptive potential of species to handle future environmental change.

Future steps for analyzing the genetic diversity of *P. veris* in alvars should be done. One focus should be adaptive genetic diversity since it could have a different response to landscape change than neutral genetic diversity used in this thesis. Furthermore, additional landscape analyses are needed for better understanding of the relationship between genetic diversity and landscape change. To advance knowledge on the dynamics of the change of genetic diversity, landscape data from different time periods could be included. Accounting for environmental variables could also provide new additional insights into the dynamics of genotype-environment relationships. Therefore, while the current study focused on relatively open grasslands, genetic patterns of plants on recently overgrown alvar grasslands should be examined as well.

Summary

Recent changes in environment and landscape have influenced natural ecosystems in many ways such as fragmentation of habitats and soil pollution. Not all organisms are able to survive habitat changes caused by these alterations. Habitat changes, in turn, could cause loss of genetic diversity making the species with lower genetic diversity more vulnerable to environmental changes. Genetic diversity could be influenced by many factors, such as the size of the population, habitat area, habitat connectivity, seed and pollen vectors and characteristics of plant species. These factors affect genetic diversity mainly through influencing genetic drift, gene flow and natural selection. All these factors could influence genetic diversity with a time lag. This means that the influence of loss of habitat area for example is not seen in genetic diversity immediately but after a certain time period. Understanding how landscape changes influence genetic diversity is thus necessary. Landscape genetics is a field studying such relationships through combining tools of population genetics and landscape ecology and forms the methodological basis for the current thesis.

In this thesis, I examined which landscape elements influence the genetic diversity within fragmented plant populations and if historic landscape characteristics have a stronger effect on current genetic diversity than current landscape features, i.e. do patterns of genetic diversity exhibit a lagged response to landscape change. I collected 338 *Primula veris* individuals from 19 alvar grasslands in Saare- and Muhumaa. I extracted DNA from the samples, prepared a library for sequencing and sequenced the library. Sequencing data was filtered using bioinformatics tools. I analyzed the data with discriminant analysis of principal components (DAPC) to examine the genetic structure of the populations. I used observed and expected heterozygosity, polymorphic loci and inbreeding coefficient as genetic diversity indices. I made linear mixed-effects models with landscape parameters (population size, current and historic habitat area, loss of habitat area, proportional habitat loss, current and historic forest area and current and historic human population density) as independent variables for statistical analysis.

The results from mixed-effects models showed that a loss of habitat area had a negative effect on genetic diversity as did historic and current habitat area. Current forest area had a slight negative effect whereas historic forest area had no effect on genetic diversity. Both historic and current human population density had a negative effect on genetic diversity of *P. veris*. DAPC analysis showed that populations in Muhumaa and Eastern-Saaremaa are genetically closer to each other than to the rest of populations.

These results show that that landscape change in Estonian alvars has had a major effect on patterns of genetic diversity of *P. veris*. It is most strongly influenced by habitat area loss. Although historic habitat area had a negative effect on the genetic diversity of *P. veris*, it can not be said for certain if genetic diversity of *P. veris* has a delayed reaction to landscape changes, because historic habitat area and habitat area loss were correlated with each other. The weak influence of current forest area could indicate that forest is not a strong barrier for gene flow. The negative effect of current human population density could result from intensified agriculture. This knowledge of how and what influences genetic diversity of *P. veris* in Estonian alvars will help to protect and maintain genetic diversity of grassland plant species which provides adaptive potential of species to handle future environmental changes.

Kokkuvõte

Hiljutised maastiku- ja keskkonnamuutused on looduslikke ökosüsteeme mõjutanud mitmel moel (nt kasvukohtade killustumine ning mullareostus). Kõik organismid ei suuda nende muutuste tõttu muutunud kasvukohas ellu jääda. See omakorda võib põhjustada geneetilise mitmekesisuse kao muutes madalama geneetilise mitmekesisusega liigid vastuvõtlikumaks keskkonnamuutustele. Seetõttu on vajalik mõista, kuidas keskkonnamuutused geneetilist mitmekesisust mõjutavad. Geneetilist mitmekesisust võivad mõjutada mitmed faktorid nagu populatsiooni suurus, kasvukoha pindala, kasvukoha sidusus, õietolmu ja leviste vektorid (nt tolmeldajad ja kariloomad) ja taimeliigi eripärad. Need faktorid mõjutavad geneetilist mitmekesisust peamiselt läbi geenitriivi, geenivoolu ja loodusliku valiku. Kõik need faktorid võivad geneetilist mitmekesisust mõjutada ajalise võlaga. See tähendab, et näiteks kasvukoha pindala vähenemise mõju geneetilisele mitmekesisusele ei ole nähtav koheselt, vaid teatud ajaperioodi möödudes. Seetõttu on vajalik mõista, kuidas maastikumuutused geneetilist mitmekesisust mõjutavad. Selliste mõjude uurimisega tegeleb maastikugeneetika, mis kombineerib populatsioonigeneetika ja maastikuökoloogia uurimismeetodeid ning moodustab käesoleva töö metoodilise aluse.

Käesolevas töös uurin, millised maastikuelemendid mõjutavad killustunud taimepopulatsioonide geneetilist mitmekesisust ja kas ajaloolistel maastiku struktuuril on geneetilisele mitmekesisusele tugevam mõju kui tänapäevasel, ehk kas maastikumuutustel on geneetilise mitmekesisuse mustritele hilinenud mõju. Üheksateistkümnest nurmenuku (*Primula veris*) populatsioonist Saare- ja Muhumaa loopealsetelt koguti geneetiliseks analüüsiks 338 nurmenuku proovi. Proovidest eraldati DNA, valmistati ette sekveneerimiseks vajalik raamatukogu ja sekveneeriti raamatukogu. Sekveneerimisel saadud andmed filtreeriti bioinformaatiliste tööriistade abil. Saadud geneetiliste andmetega viidi läbi peakomponentide diskriminantanalüüs (DAPC). Geneetilise mitmekesisuse indeksitena kasutati vaadeldud ja oodatavat heterosügootsust, polümorfsete lookuste protsenti ja inbriidingu koefitsienti. Tehti üldised lineaarsed segamudelid maastikuandmete mõju tuvastamiseks geneetilisele mitmekesisusele, kus sõltumatuteks muutujateks olid populatsiooni suurus, tänapäevane ja ajalooline loopealse pindala, loopealse pindala vähenemine, suhteline kasvukoha pindala vähenemine, tänapäevane ja ajalooline metsa pindala ning tänapäevane ja ajalooline inimasustuse tihedus.

DAPC analüüs näitas, et Muhu ja Ida-Saaremaa populatsioonid olid geneetiliselt üksteisega teiste populatsioonidega võrreldes sarnasemad. Üldiste segamudelite tulemused näitasid, et kasvukoha pindala kadu mõjus geneetilisele mitmekesisusele negatiivselt. Samasugune mõju oli ka ajaloolisel ja tänapäevasel kasvukoha pindalal. Tänapäevase metsa pindala omas kerget negatiivset mõju. Inimpopulatsiooni tihedus oli samuti negatiivse mõjuga.

Saadud tulemused näitavad, et maastikumuutustel on olnud suur mõju Eesti loopealsetel kasvavate nurmenukkude geneetilisele mitmekesisusele. Kõige tugevam mõju geneetilisele mitmekesisusele oli loopealse pindala kaotus. Kuigi ajaloolisel loopealse pindalal oli nurmenuku geneetilisele mitmekesisusele negatiivne mõju, ei saa ajaloolise loopealse pindala ja loopealse pindala kao vahelise korrelatsiooni tõttu kindlalt öelda, et nurmenuku geneetiline mitmekesisus reageerib maastikumuutustele ajalise võlaga. Metsa pindala nõrk mõju võib viidata, et mets ei ole geenivoolule väga tugevaks barjääriks. Tänapäevase inimasustuse tiheduse negatiivne mõju võib tuleneda põllumajanduse intensiivistumisest. Teades, kuidas ja mis mõjutavad nurmenuku geneetilist mitmekesisust Eesti loopealsetel, on võimalik ka paremini kaitsta geneetilist mitmekesisust, mis omakorda pakub paremat kohastumisvõimet tulevaste keskkonnamuutuste üleelamiseks.

Acknowledgements

I sincerely thank my supervisors Tsipe Aavik and Sabrina Träger, who have helped me with every step and provided valuable knowledge and guidance. I would also like to thank everyone who participated in the fieldwork for ’LIFE to alvars’ project and provided landscape data. I am thankful to Genetic Diversity Centre in Swiss Federal Institute of Technology in Zurich for the use of their laboratory and valuable support during laboratory work and bioinformatic analysis as well.

References

Aavik, T., Holderegger, R., Bolliger, J. 2014. „The Structural and Functional Connectivity of the Grassland Plant *Lychnis Flos-Cuculi*“. *Heredity* 112 (5): 471–78.

Aavik, T., Edwards, P. J., Holderegger, R., Graf, R., Billeter, R. 2012. „Genetic consequences of using seed mixtures in restoration: A case study of a wetland plant *Lychnis flos-cuculi*“. *Biological Conservation* 145 (1): 195–204.

Aavik, T., ja Helm, A. 2018. „Restoration of plant species and genetic diversity depends on landscape-scale dispersal“. *Restoration Ecology* 0 (0).

Aavik, T., Talve, T., Thetloff, M., Uuemaa, E., Oja, T. 2017. „Genetic consequences of landscape change for rare endemic plants – A case study of *Rhinanthus osiliensis*“. *Biological Conservation* 210 (Part A): 125–35.

Aguilar R., Ashworth, L., Galetto L., Aizen, M. A. 2006. „Plant reproductive susceptibility to habitat fragmentation: review and synthesis through a meta-analysis“. *Ecology Letters* 9 (8): 968–80.

Aguilar, R., Quesada, M., Ashworth, L., Herrerias-Diego, Y., Lobo, J. 2008. „Genetic Consequences of Habitat Fragmentation in Plant Populations: Susceptible Signals in Plant Traits and Methodological Approaches“. *Molecular Ecology* 17 (24): 5177–88.

Akaike, H. 1974. „A new look at the statistical model identification“. *IEEE Transactions on Automatic Control* 19 (6): 716–23.

Auffret, A. G., Rico, Y., Bullock, J. M., Hooftman, D. A. P., Pakeman, R. J., Soons, M. B., Suárez-Esteban, A., Traveset, A., Wagner, H.H., Cousins, S.A. 2017. „Plant functional connectivity – integrating landscape structure and effective dispersal“. *Journal of Ecology* 105 (6): 1648–56.

Balkenhol, N., Cushman, S., Storfer, A., Waits, L. 2015. *Landscape Genetics: Concepts, Methods, Applications*. John Wiley & Sons.

Bezdicek, D. F., Papendick, R. I., Lal, R. 1996. „Introduction: Importance of Soil Quality to Health and Sustainable Land Management“. *Methods for Assessing Soil Quality* sssaspecialpubl (methodsforasses): 1–8.

Bolger, A. M., Lohse, M., Usadel, B. 2014. „Trimmomatic: A Flexible Trimmer for Illumina Sequence Data“. *Bioinformatics* 30 (15): 2114–20.

Broeck, A. V., Ceulemans T., Kathagen, G., Hoffmann, M., Honnay, O., Mergeay, J. 2015. „Dispersal Constraints for the Conservation of the Grassland Herb *Thymus pulegioides*

L. in a Highly Fragmented Agricultural Landscape“. *Conservation Genetics* 16 (4): 765–76.

Brys, R., Jacquemyn, H. 2009. „Biological Flora of the British Isles: *Primula Veris* L.“ *Journal of Ecology* 97 (3): 581–600.

Campagne, P., Baumel, A., Affre, L., Juin, M., Duong, N., Roche, P., Tatoni, T. 2009. „Genetic signs of connectivity in *Primula vulgaris* (*Primulaceae*) in a hedgerow network landscape“. *Comptes Rendus Biologies* 332 (7): 652–61.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., Cresko, W. A. 2013. „Stacks: An Analysis Tool Set for Population Genomics“. *Molecular Ecology* 22 (11): 3124–40.

Catchen, J., Amores, A., Hohenlohe, P., Cresko, W. A., Postlethwait, J. H. 2011. „Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences“. *G3: Genes, Genomes, Genetics* 1 (3): 171–82.

Chung, M. Y., Nason, J. D., López-Pujol, J., Yamashiro, T., Yang, B., Luo, Y., Chung, M. G. 2014. „Genetic consequences of fragmentation on populations of the terrestrial orchid *Cymbidium goeringii*“. *Biological Conservation* 170 (veebruar): 222–31.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G.T., Sherry, S.T. and McVean, G. 2011. „The Variant Call Format and VCFtools“. *Bioinformatics* 27 (15): 2156–58.

DiLeo, M. F., Rico, Y., Boehmer, H. J., Wagner, H. H. 2017. „An ecological connectivity network maintains genetic diversity of a flagship wildflower, *Pulsatilla vulgaris*“. *Biological Conservation* 212 (august): 12–21.

DiLeo, M. F., Wagner, H. H. 2016. „A Landscape Ecologist's Agenda for Landscape Genetics“. *Current Landscape Ecology Reports* 1 (3): 115–26.

Epps, C.W., Keyghobadi, N. 2015. „Landscape Genetics in a Changing World: Disentangling Historical and Contemporary Influences and Inferring Change“. *Molecular Ecology* 24 (24): 6021–40.

ESRI 2016. ArcGIS Desktop: Release 10.4. Environmental Systems Research Institute, Redlands, CA.

Estonian Land Board https://geoportaal.maaamet.ee/est/Andmed-ja-kaardid/Arhiivimaterjalid/Kartograafilised-materjalid/Eesti-Baaskaart-150000-p31.html

Fahrig, L. 2013. „Rethinking Patch Size and Isolation Effects: The Habitat Amount Hypothesis“. *Journal of Biogeography* 40 (9): 1649–63.

Frankham, R, Ballou, J. D., Briscoe, D. A. 2004. *A Primer of Conservation Genetics*. Frankham, Richard; Macquarie University, James Cook University, Sydney, Australia, Australia.: Cambridge University Press.

Garrison, E, Marth, G. 2012. „Haplotype-based variant detection from short-read sequencing". *arXiv:1207.3907 [q-bio]*, july.

Hahn, T., Kettle, C. J., Ghazoul, J., Hennig, E. I., Pluess, A. R. 2013. „Landscape Composition Has Limited Impact on Local Genetic Structure in Mountain Clover, *Trifolium montanum* L." *Journal of Heredity* 104 (6): 842–52.

Heinaru, A. 2012. *Geneetika. Õpik kõrgkoolile*. Tartu: Tartu Ülikooli Kirjastus.

Helm A. 2003. Spatial and temporal dynamics of species richness of Saaremaa and Muhu alvars. Master Thesis, Institute of Botany and Ecology, Tartu University, Tartu.

Helm, A., 2011. Eesti loopealsed ja kadastikud. Juhend koosluste hooldamiseks ja taastamiseks.

Helm, A., Hanski, I., Pärtel, M.. 2006. „Slow Response of Plant Species Richness to Habitat Loss and Fragmentation". *Ecology Letters* 9 (1): 72–77.

Helm, A., Oja, T., Saar, L., Takkis, K., Talve, T., Pärtel, M. 2009. „Human Influence Lowers Plant Genetic Diversity in Communities with Extinction Debt". *Journal of Ecology* 97 (6): 1329–36.

Holderegger, R., Buehler, D., Gugerli. F., Manel, S. 2010. „Landscape genetics of plants". *Trends in Plant Science* 15 (12): 675–83.

Holsinger, K. E., Weir, B. S. 2009. „Genetics in Geographically Structured Populations: Defining, Estimating and Interpreting F_{ST}". *Nature Reviews Genetics* 10 (9): 639–50.

Honnay, O., Adriaens, D., Coart, E., Jacquemyn, H., Roldan-Ruiz, I. 2007. „Genetic Diversity within and between Remnant Populations of the Endangered Calcareous Grassland Plant *Globularia Bisnagarica* L." *Conservation Genetics* 8 (2): 293–303.

Honnay, O., Coart, E., Butaye, J., Adriaens, D., Van Glabeke, S., Roldán-Ruiz, I. 2006. „Low impact of present and historical landscape configuration on the genetics of fragmented *Anthyllis vulneraria* populations". *Biological Conservation* 127 (4): 411–19.

Honnay, O., Jacquemyn, H. 2007. „Susceptibility of Common and Rare Plant Species to the Genetic Consequences of Habitat Fragmentation". *Conservation Biology* 21 (3): 823–31.

Ilmateenistus. URL http://www.ilmateenistus.ee/kliima/kliimakaardid/ohutemperatuur/

Jacquemyn, H., Roldán-Ruiz, I., Honnay, O. 2010. „Evidence for Demographic Bottlenecks and Limited Gene Flow Leading to Low Genetic Diversity in a Rare Thistle". *Conservation Genetics* 11 (5): 1979–87.

Jombart, T., Devillard, S., Balloux, F. 2010. „Discriminant Analysis of Principal Components: A New Method for the Analysis of Genetically Structured Populations“. *BMC Genetics* 11 (october): 94.

Jones, A. G., Small, C. M., Paczolt, K. A., Ratterman, N. L. 2010. „A Practical Guide to Methods of Parentage Analysis“. *Molecular Ecology Resources* 10 (1): 6–30.

Kloss, L., Fischer, M., Durka, W. 2011. „Land-use effects on genetic structure of a common grassland herb: A matter of scale“. *Basic and Applied Ecology* 12 (5): 440–48.

Kramer, A. T., Fant, J. B., Ashley, M. V. 2011. „Influences of Landscape and Pollinators on Population Genetic Structure: Examples from Three Penstemon (*Plantaginaceae*) Species in the Great Basin“. *American Journal of Botany* 98 (1): 109–21.

Kreyer, D., Oed, A., Walther-Hellwig, K., Frankl, R. 2004. „Are forests potential landscape barriers for foraging bumblebees? Landscape scale experiments with *Bombus terrestris* agg. and *Bombus pascuorum* (*Hymenoptera, Apidae*)“. *Biological Conservation* 116 (1): 111–18.

Laasimer, L., 1965. *Eesti NSV taimkate.* Tallinn

Leimu, R., Mutikainen, P., Koricheva, J., Fischer, M. 2006. „How General Are Positive Relationships between Plant Population Size, Fitness and Genetic Variation?“ *Journal of Ecology* 94 (5): 942–52.

Leimu, R., Vergeer, P., Angeloni, F., Ouborg, N. J. 2010. „Habitat Fragmentation, Climate Change, and Inbreeding in Plants“. *Annals of the New York Academy of Sciences* 1195 (1): 84–98.

Li, H. 2013. „Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM“. *arXiv:1303.3997 [q-bio]*, march.

Li, H. 2014. „Toward better understanding of artifacts in variant calling from high-coverage samples“. *Bioinformatics* 30 (20): 2843–51.

Manel, S., Gaggiotti, O. E., Waples, R. S. 2005. „Assignment methods: matching biological questions with appropriate techniques“. *Trends in Ecology & Evolution* 20 (3): 136–42.

Manel, S., Schwartz, M. K., Luikart, G., Taberlet, P. 2003. „Landscape genetics: combining landscape ecology and population genetics“. *Trends in Ecology & Evolution* 18 (4): 189–97.

Menges, E. S., Dolan, R. W., Pickert, R., Yahr, R., Gordon, D. R. 2010. „Genetic Variation in Past and Current Landscapes: Conservation Implications Based on Six Endemic Florida Scrub Plants“. *International Journal of Ecology* 2010 (märts): e503759.

Münzbergová, Z., Cousins, S. A. O., Herben, T., Plačková, I., Mildén, M., Ehrlén, J. 2013. „Historical Habitat Connectivity Affects Current Genetic Structure in a Grassland Species“. *Plant Biology* 15 (1): 195–202.

Nowak, M. D., Russo, G., Schlapbach, R., Huu, C. N., Lenhard, M., Conti, E. 2015. „The draft genome of *Primula veris* yields insights into the molecular basis of heterostyly“. *Genome Biology* 16 (january): 12.

OpenWetWare URL https://openwetware.org/wiki/SPRI_bead_mix.

Peakall R., Smouse, P. E. 2005. „genalex 6: genetic analysis in Excel. Population genetic software for teaching and research“. *Molecular Ecology Notes* 6 (1): 288–95.

Peakall, R., Smouse, P. E. 2012. „GenAlEx 6.5: Genetic Analysis in Excel. Population Genetic Software for Teaching and Research—an Update“. *Bioinformatics* 28 (19): 2537–39.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., Hoekstra, H. E. 2012. „Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species“. Toimetanud Ludovic Orlando. *PLoS ONE* 7 (5): e37135.

Picó, F. X., Van Groenendael, J. 2007. „Large-Scale Plant Conservation in European Semi-Natural Grasslands: A Population Genetic Perspective“. *Diversity and Distributions* 13 (6): 920–26.

Plue, J., Cousins, S. A. O. 2013. „Temporal dispersal in fragmented landscapes“. *Biological Conservation* 160 (aprill): 250–62.

Plue, J., Vandepitte, K., Honnay, O., Cousins, S. A. O. 2017. „Does the seed bank contribute to the build-up of a genetic extinction debt in the grassland perennial *Campanula rotundifolia*?“ *Annals of Botany* 120 (3): 373–85.

Prangel E. 2017. Ökosüsteemi hüved avatud ja kinnikasvavatel loopealsetel. Master Thesis, Institute of Botany and Ecology, Tartu University, Tartu.

Prentice, H. C., Lönn, M., Rosquist, G., Ihse, M., Kindström, M. 2006. „Gene Diversity in a Fragmented Population of *Briza Media*: Grassland Continuity in a Landscape Context“. *Journal of Ecology* 94 (1): 87–97.

Project LIFE to alvars. URL http://life.envir.ee/english-project-life-alvars.

Puritz, J. B., Hollenbeck, C. M., Gold, J. R. 2014. „dDocent: A RADseq, Variant-Calling Pipeline Designed for Population Genomics of Non-Model Organisms“. *PeerJ* 2 (june): e431.

Puritz, J. B., Matz, M. V., Toonen, R. J., Weber, J. N., Bolnick, D. I., Bird, C. E. 2014. „Demystifying the RAD Fad“. *Molecular Ecology* 23 (24): 5937–42.

Pärtel, M., Zobel, M. 1999. „Small-Scale Plant Species Richness in Calcareous Grasslands Determined by the Species Pool, Community Age and Shoot Density“. *Ecography* 22 (2): 153–59.

Reisch, C., Schmidkonz, S., Meier, K., Schöpplein, Q., Meyer, C., Hums, C., Putz, C., Schmid, C. 2017. „Genetic Diversity of Calcareous Grassland Plant Species Depends on Historical Landscape Configuration“. *BMC Ecology* 17 (1): 19.

R Core Team 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rico, Y., Holderegger, R., Boehmer, H. J., Wagner, H. H. 2014. „Directed Dispersal by Rotational Shepherding Supports Landscape Genetic Connectivity in a Calcareous Grassland Plant“. *Molecular Ecology* 23 (4): 832–42.

Schmidt, T., Arens, P., Smulders, M. J. M., Billeter, R., Liira, J., Augenstein, I., Durka, W. 2009. „Effects of landscape structure on genetic diversity of *Geum urbanum* L. populations in agricultural landscapes“. *Flora - Morphology, Distribution, Functional Ecology of Plants* 204 (7): 549–59.

Schmitt, T., Varga, Z., Seitz, A. 2000. „Forests as dispersal barriers for *Erebia medusa* (*Nymphalidae, Lepidoptera*)“. *Basic and Applied Ecology* 1 (1): 53–59.

Selkoe, K. A., Toonen, R. J. 2006. „Microsatellites for Ecologists: A Practical Guide to Using and Evaluating Microsatellite Markers“. *Ecology Letters* 9 (5): 615–29.

Smulders, M. J. M., Cobben, M. M. P., Arens, P., Verboom, J. 2009. „Landscape Genetics of Fragmented Forests: Anticipating Climate Change by Facilitating Migration“. *iForest - Biogeosciences and Forestry* 2 (4): 128.

Statistikaamet. URL http://www.stat.ee

Zurbuchen, A., Bachofen, C., Müller, A., Hein, S., Dorn, S. 2010. „Are Landscape Structures Insurmountable Barriers for Foraging Bees? A Mark-Recapture Study with Two Solitary Pollen Specialist Species“. *Apidologie* 41 (4): 497–508.

Taberlet, P., Zimmermann, N. E., Englisch, T., Tribsch, A., Holderegger, R., Alvarez, N., Niklfeld, H. Coldea, G., Mirek, Z., Moilanen, A., Ahlmer, W. 2012. „Genetic diversity in widespread species is not congruent with species richness in alpine plant communities“. *Ecology Letters* 15 (12): 1439–48.

Takkis, K., Pärtel, M., Saar, L., Helm, A. 2013. „Extinction Debt in a Common Grassland Species: Immediate and Delayed Responses of Plant and Population Fitness“. *Plant Ecology* 214 (7): 953–63.

Tammekann, A. 1929. Outlines of the Distribution of Population in Estonia. With a Population Map 1:1 000 000. Tartu Ülikooli ja Loodusuurijate Seltsi Aruanded XXXV, Tartu.

Tewksbury, J. J., Levey, D. J., Haddad, N. M., Sargent, S., Orrock, J. L., Weldon, A., Danielson, B. J., Brinkerhoff, J., Damschen, E. I., Townsend, P. 2002. „Corridors Affect Plants, Animals, and Their Interactions in Fragmented Landscapes“. *Proceedings of the National Academy of Sciences* 99 (20): 12923–26.

Tin, M. M. Y., Rheindt, F. E., Cros, E., Mikheyev, A. S. 2015. „Degenerate Adaptor Sequences for Detecting PCR Duplicates in Reduced Representation Sequencing Data Improve Genotype Calling Accuracy“. *Molecular Ecology Resources* 15 (2): 329–36.

Vandepitte, K., Jacquemyn, H., Roldán-Ruiz, I., Honnay, O. 2007. „Landscape Genetics of the Self-compatible Forest Herb *Geum Urbanum*: Effects of Habitat Age, Fragmentation and Local Environment“. *Molecular Ecology* 16 (19): 4171–79.

Van Geert, A., Van Rossum, F., Triest, L. 2008. „Genetic Diversity in Adult and Seedling Populations of Primula Vulgaris in a Fragmented Agricultural Landscape“. *Conservation Genetics* 9 (4): 845.

Vellend, M., Lajoie, G., Bourret, A., Múrria, C., Kembel, S. W., Garant, D. 2014. „Drawing Ecological Inferences from Coincident Patterns of Population- and Community-Level Biodiversity“. *Molecular Ecology* 23 (12): 2890–2901.

Whitlock, R. 2014. „Relationships between Adaptive and Neutral Genetic Diversity and Ecological Structure and Functioning: A Meta-Analysis“. *Journal of Ecology* 102 (4):

Vitousek, P. M., Mooney, H. A., Lubchenco, J., Melillo, J. M. 1997. „Human Domination of Earth’s Ecosystems“. *Science* 277 (5325): 494–99.

Vranckx, G., Jacquemyn, H., Muys, B., Honnay, O. 2012. „Meta-Analysis of Susceptibility of Woody Plants to Loss of Genetic Diversity through Habitat Fragmentation“. *Conservation Biology* 26 (2): 228–37.

Völler, E., Auge, H., Bossdorf, O., Prati, D. 2013. „Land Use Causes Genetic Differentiation of Life-History Traits in *Bromus Hordeaceus*“. *Global Change Biology* 19 (3): 892–99.

Young, A., Boyle, T., Brown, T. 1996. „The population genetic consequences of habitat fragmentation for plants“. *Trends in Ecology & Evolution* 11 (10): 413–18.

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Iris Reinula

(sünnikuupäev 04.06.1994)

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose
„The impact of landscape change on the genetic diversity of the grassland plant Primula veris“,

mille juhendajad on Tsipe Aavik ja Sabrina Träger,

1.1. reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;

1.2. üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace´i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.

2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.

3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, 23.05.2018