Università degli Studi di Ferrara

DOCTORAL COURSE IN
EVOLUTIONARY BIOLOGY AND ECOLOGY

In cooperation with:
Università degli Studi di Parma
Università degli Studi di Firenze

CYCLE XXXIII

DIRECTOR Prof. BARBUJANI Guido

Patterns of Genomic Variation in Three Species of Alpine Grouse: Conservation and Management Using SNPs

Scientific/Disciplinary Sector (SDS) BIO/18

Candidate
Dott. FRASER Alice Kate

Supervisor
Prof. GHIROTTO Silvia

Years 2017/2021

COLLABORATORS AND ACKNOWLEDGEMENTS

CO-SUPERVISORS

Giorgio Bertorelle

Barbara Crestanello

Heidi C. Hauffe

COLLABORATORS

Andrea Brunelli

Francesco Maroso

Emiliano Trucchi

Stefano Casari

Margherita Collini

Roberta Covi

Matteo Girardi

Diego Micheletti

Stéphanie Manel

Laura Benestan

G.B., **B.C.** and **H.C.H**. obtained the funding for the doctoral scholarship and provided supervision during the thesis. **B.C.** obtained the funding for genetic and genomic analyses. The technologists and technicians of the Conservation Genomics Research Unit at Fondazione Edmund Mach, **B.C**., **S.C., M.C**., **R.C.** and **M.G.,** generated the 1995-2010 STR and mtDNA datasets; **A.K.F.** generated the 2015-2017 STR dataset, and selected and prepared samples to be sent for GBS analysis. **E.T.**, **A.B., F.M.** and **D.M.** provided training and tutoring for genomic bioinformatics. **A.B., F.M.** and **G.B.** provided advice regarding the population genetic analysis of the data. **A.K.F.** designed, wrote and tested bioinformatic pipelines, and selected pipelines for final datasets; analysed the mtDNA, STR and SNP data, and produced the thesis. **S.M.** and **L.B.** hosted **A.K.F**. during the period of long mobility and provided mentorship in landscape genetic analyses.

The following associations also provided samples and funding for the SNP and STR analyses:
ALCOTRA project (Piedmont and Aosta Valley)
Associazione Cacciatori Bellunesi
Associazione Cacciatori Trentini
Comprensori Alpini di Caccia di Tirano, Sondrio, Morbegno, Alta Valtellina, Val Chiavenna
Dolomiti Bellunesi National Park
Provincia Autonoma di Trento (PAT)
Servizio Caccia e Pesca della Provincia di Sondrio
Servizio Faunistico dela Provincia di Belluno
Unione Nazionale Cacciatori Zona Alpi (UNCZA)

FUNDING
The project has been funded with support from the Fondazione Edmund Mach, University of Ferrara, the Autonomous Province of Trento, and the Unione Nazionale Cacciatori Zona Alpi (UNCZA).
A. Fraser was funded by the Fondazione E. Mach (FIRST PhD Scholarship) and the University of Ferrara.

I would like to thank everyone that has been involved in this project for your expertise and contributions; including all of the members of the research groups both in the Department of Life Sciences and Biotechnology at the University of Ferrara and the Department of Biodiversity and Molecular Ecology at Fondazione Edmund Mach. I would particularly like to extend this thanks to my supervisors for their support and guidance throughout my studies and for giving me the opportunity to be a part of this project.
I would also like to express my gratitude to my proofreaders; Colin Fraser and Barbara Ilardo for taking the time to read sections of this thesis.
Additionally, I would like to thank my friends and family for their support throughout my PhD, especially Tom, Moki and Lily for never failing to cheer me up.
Finally, to my parents, Claire and Dougal, for their unfaltering encouragement, support and for always believing in me no matter what challenges I may face.

TABLE OF CONTENTS

ABSTRACT

In light of rapidly decreasing levels of biodiversity, conservation efforts towards the sustainable management of species and their ecosystems is becoming increasingly relevant. The black grouse (*Lyrurus tetrix*), rock ptarmigan (*Lagopus muta*) and western capercaillie (*Tetrao urogallus*) are listed as 'Least Concern' by the International Union for Conservation of Nature. However, at the southern edges of their ranges, these species are often more fragmented and some populations are classified as threatened. Importantly for conservation actions, the genetic 'health' of these elusive species in these regions is poorly understood. Here we use a multi-species, multi-marker approach to identify common factors affecting genetic patterns. Samples were collected predominantly from the Italian Alps over a 20 year period. Data from traditional markers (mitochondrial DNA and microsatellites) as well as genome-wide Single Nucleotide Polymorphisms (SNPs) were generated and compared. SNPs were typed using the Genotyping-by-Sequencing technique to investigate their suitability for this type of conservation application, and applied for the first time for these tetraonids. The black grouse was characterized by a strong pattern of isolation by distance across the Italian Alps, and the results also suggest that there may a barrier to their movement in a heavily urbanized Lombardy Region. The ptarmigan results suggest that this species forms a continuous panmictic population across the study area, confirming that dispersal distances for this species have been underestimated in the Alps. Unexpectedly given its low density and previously estimated range of movement, gene flow is high among capercaillie populations in Trentino-Alto Adige. No genetically isolated populations were identified for any of the species in the Italian Alps; in addition, genetic variability of the black grouse did not decrease over the study period, suggesting that current management practices (including the hunting of both sexes of ptarmigan and male black grouse) are suitable for maintaining current levels of genetic diversity. In general, the SNP results produced more detailed geographical patterns for the data (although they did not contrast with STR results). Therefore, given that the technique was fairly successful even for non-invasive samples, the periodic use of these genomic markers is recommended for the study of populations across all three species ranges, to aid future conservation efforts.

RIASSUNTO

La forte perdita di biodiversità in tutto il pianeta, frequentemente indicata come “sesta estinzione di massa” richiede nuovi e sempre più intensi sforzi di conservazione per la gestione delle specie e degli ecosistemi. Il fagiano di monte (*Lyrurus tetrix*), la pernice bianca (*Lagopus muta*) e il gallo cedrone (*Tetrao urogallus*) sono classificati come "Minor Preoccupazione" dall'Unione Internazionale per la Conservazione della Natura. Tuttavia, ai margini meridionali delle loro aree di distribuzione, le popolazioni sono piccole e frammentate, e quindi a rischio di estinzione locale. Per gestire nel modo migliore le azioni di conservazione in queste regioni, la "salute" genetica è importante ma ancora poco analizzata e compresa. Qui usiamo un approccio multi-specie e multi-marcatore per identificare i fattori comuni che influenzano i pattern genetici. I campioni sono stati raccolti prevalentemente nelle Alpi italiane in un periodo di circa vent’anni. I dati relativi ai marcatori tradizionali (DNA mitocondriale e microsatelliti) sono stati confrontati con i polimorfismi a singolo nucleotide (SNPs). Per la prima volta in queste specie, gli SNPs sono stati tipizzati usando la tecnica del Genotyping-by-Sequencing, e la loro idoneità per questo tipo di applicazioni in conservazione è stata indagata. Per il fagiano di monte, un forte pattern di isolamento per distanza è presente nelle Alpi, e i risultati suggeriscono anche che potrebbe esserci una barriera al loro movimento in una regione lombarda fortemente antropizzata. Inaspettatamente, i risultati per la pernice bianca suggeriscono che la specie è costituita da una unica popolazione panmittica in tutta l'area di studio, indicando che le distanze di dispersione sono state sottostimate in passato. Ancora più sorprendenti sono stati i risultati del gallo cedrone, che suggeriscono anche in questo caso che il flusso genico sia elevato tra le popolazioni del Trentino-Alto Adige, nonostante il loro range di movimento precedentemente stimato fosse ridotto. Nessuno dei risultati ottenuti indica la presenza di popolazioni geneticamente isolate nelle Alpi italiane in nessuna specie. Il fagiano di monte, per il quale è stato possibile fare un confronto temporale, non sembra aver perso variabilità genetica negli ultimi 20 anni, e questo suggerisce che le attuali pratiche di gestione della specie e delle attività venatorie non stanno compromettendo la diversità genetica. In generale, i risultati ottenuti con i marcatori SNPs hanno prodotto inferenze compatibili, ma più dettagliate, di quelle possibili con i marcatori classici. Pertanto, dato che la tecnica ha avuto successo anche per i campioni non invasivi, si raccomanda l'uso periodico di questi marcatori per il

monitoraggio delle popolazioni in tutte e aree di distribuzione delle tre specie, allo scopo di identificare eventuali perdite di diversità o frammentazione in futuro e guidare eventuali modifiche ai piani gestionali.

1. INTRODUCTION

1.1 CONSERVATION OF ANIMAL BIODIVERSITY

Current extinction rates of animal species greatly exceed estimates of natural background values, indicating that the planet is currently experiencing a sixth mass extinction (Ceballos *et al.*, 2015). Over 142,500 species are listed on the IUCN (International Union for Conservation of Nature) Red List with over 40,000 threatened with extinction. This represents, in part, 41% of amphibians, 34% of conifers, 26% of mammals and 14% of birds studied (IUCN, 2022). Conservation efforts are therefore essential to protect biodiversity and, as a result, the bioresources that humans depend on for food and medicines, as well as ecosystems services such as pollination and oxygen production (Pimentel *et al*., 1997; Frankham *et al.,* 2010).

The need for conservation actions to be prioritized has led to the concept of Evolutionary Significant Units (ESU). ESUs were first described by Ryder (1986) to indicate a discrete population or group within the range of a species that is prioritised in conservation efforts. ESUs are typically identified based on the adaptive and genetic differentiation of the population, although the exact definition has been debated (Robertson *et al.,* 2014). Management Units, or MUs, have a similar purpose, however, these represent smaller populations, with multiple MUs being represented by a single ESU (Moritz, 1994). Moritz (1994) describes MUs as populations with a significant divergence of allele frequencies. These units of conservation have been used to classify populations for a variety of different species of both flora and fauna (Mäkinen and Merilä, 2008; Muñoz-Fuentes *et al*., 2009; Abbasi *et al*., 2016). ESUs aid the development of management strategies and are used to determine distinct population segments under the U.S. Endangered Species Act, exampled by Waples (1991) in populations of Pacific salmon (*Oncorhynchus* spp.; Robertson *et al.,* 2014)

The definition of ESUs and MUs depends on knowledge of genetic diversity. A high level of genetic diversity enables a species to respond to environmental pressures and adapt to new conditions, and is therefore crucial to species survival (Dirzo and Raven, 2003; Frankham *et al*., 2010). With rapidly decreasing levels of diversity due to both the indirect and direct effects of human activities, it is becoming increasingly

important to manage not only species and ecosystems, but also their genetic legacy (Frankham *et al.*, 2010; Hohenlohe *et al.,* 2021).

1.2. CONSERVATION GENETICS

Conservation genetics is the application of molecular techniques to aid conservation biology and reduce extinction rates (Frankham, 2010). These tools can be used for taxonomic purposes, identifying and monitoring species diversity (McNeely et al., 1990; Frankham, 1995; Frankham *et al.*, 2010), as well as for estimating genetic indices of population genetic 'health' such as genetic drift, inbreeding and genetic diversity. Conservation genetics also investigates the past and present effects of habitat fragmentation and climate change on these genetic aspects, relying on the availability of different markers with various mutation rates to answer these questions (Allendorf et al., 2010). For example, with advances in massively parallel sequencing technologies, which are becoming increasingly available to wildlife biology thanks to rapidly decreasing costs and complexity, conservation genetics is transitioning into 'conservation genomics' allowing variation across the entire genome to be estimated, using markers such as Single Nucleotide Polymorphisms (SNPs) or Whole Genomic Sequencing (WGS) (Allendorf *et al*., 2010; Angeloni *et al.*, 2011; Allendorf, 2017; Supple and Shapiro, 2018). These genomic tools can be applied in the management of small populations, the reintroduction of a species, as well as for biotechnological applications of conservation such as assisting the adaption of a species to a changing environment (Segelbacher *et al.*, 2021). Genomic tools can also be used in the identification of conservation units, such as ESUs and MUs; genomic markers typically provide more detailed results compared to traditional genetic markers, such as mitochondrial DNA (mtDNA) and nuclear microsatellites (single tandem repeats; STR) which are currently used to identify of these conservation units (Hohenlohe *et al.,* 2021).

1.3. GENETIC MARKERS

Mitochondrial DNA and STRs started to replace the use of allozymes in 1979, as mtDNA allowed for maternal inheritance patterns to be monitored and STRs provided the ability to examine a larger portion of the genome than possible with allozymes or mtDNA (Allendorf, 2017). These markers have been the most common

genetic markers for investigating genetic diversity up to a few years ago (often in combination to observe both past (mtDNA) and recent (STR) changes), but very few of these markers are used in each study and are only representative of a tiny, and typically neutral, portion of the genome (Angeloni *et al.*, 2011; Allendorf, 2017). These traditional markers are being replaced by tens or hundreds of thousands of SNPs obtained from genome sequencing methods (Restriction site Associated DNA Sequencing – RAD-Seq; Whole Genome Sequencing; Genotyping by Sequencing – GBS; Ouborg et al., 2010; Elshire *et al*., 2011; Guigo and de Hoon, 2018). Because SNPs represent genetic diversity across the entire genome, they are considered to represent a much larger proportion of individual diversity, both neutral and adaptive (Kirk and Freeland, 2011). SNPs are often used to investigate the recent evolution of the genome that may have occurred due to both natural and anthropic processes, such as climate change and urbanisation, which can lead to changes in distribution and/or adaptation to a new environment. Therefore the ability to predict changes in genomic diversity affecting species' survival can be measured with more accuracy (Morin, *et al.,* 2004; Kirk and Freeland, 2011).

Conservation genomics techniques have, to date, been developed for numerous wild animal species. However, aligning fragments produced by RAD-Seq and GBS protocols can often be challenging and time-consuming without a reference genome; due to sequencing errors, missing data and repetitive regions in the fragments produced in the sequencing protocols (Lischer and Shimizu, 2017). Although most animal genomes sequenced thus far are mammals, the availability of reference genomes for additional taxa is increasing (NCBI, 2022). In some cases, reference genomes of domestic species phylogenetically related to the target wild species can be used to assist with the alignment in a reference guided assembly (Galla *et al.*, 2018). In the absence of a reference genome, fragments can be aligned de novo, but to combat the challenges arising from this alignment method, sample types producing high quality and quantities of DNA such as blood or tissue are preferred. However, these sample types are not often available for endangered and protected or for elusive species. Given the effort required to obtain DNA of sufficient quality for genomic analyses compared to the well-established STRs, it has been suggested that the efficiency and added value of SNP techniques should be tested before they can be used for non-invasive samples for threatened species (Carroll *et al.,* 2018).

1.4. STUDY TAXA

Grouse are a distinct group in the Order Galliformes with eighteen recognised species in 52 countries worldwide, predominantly distributed in the northern hemisphere (Storch, 2007). Fourteen of these species are listed in the National Red Lists for at least one country in which they are present (Storch, 2007). A study by Lucchini *et al*. (2001) suggests that the grouse phylogeny originated in North America with the *Bonasa* genus. Grouse can be distinguished from other galliforms due to their feathered feet and nostrils as well as a lack of spurs. All grouse species are non-migratory, ground-dwelling and adapted to cold climates, and although they are present in a wide variety of habitats, from Arctic tundra to coniferous forests, each species of grouse has relatively limited habitat preferences and are vulnerable to changes within these habitats (Johnsgard, 1983 as cited by Hovick *et al.,* 2014; Song *et al.,* 2021). Due to this environmental sensitivity, grouse are considered 'indicator species', i.e. indicative of intact habitats suitable for a variety of species and therefore, efforts are made to protect these areas (Suter, *et al.,* 2002; Rowland *et al.*, 2006; Storch, 2007). The ranges of some grouse species are known to overlap, which can lead to hybridization in some cases, although the extent of hybridization, and its impact on local adaptation, is unknown for most sympatric species (Quintela *et al*., 2010; Chunco, 2014; Ottenburghs, 2019).

1.5. BLACK GROUSE (*Lyrurus tetrix*)

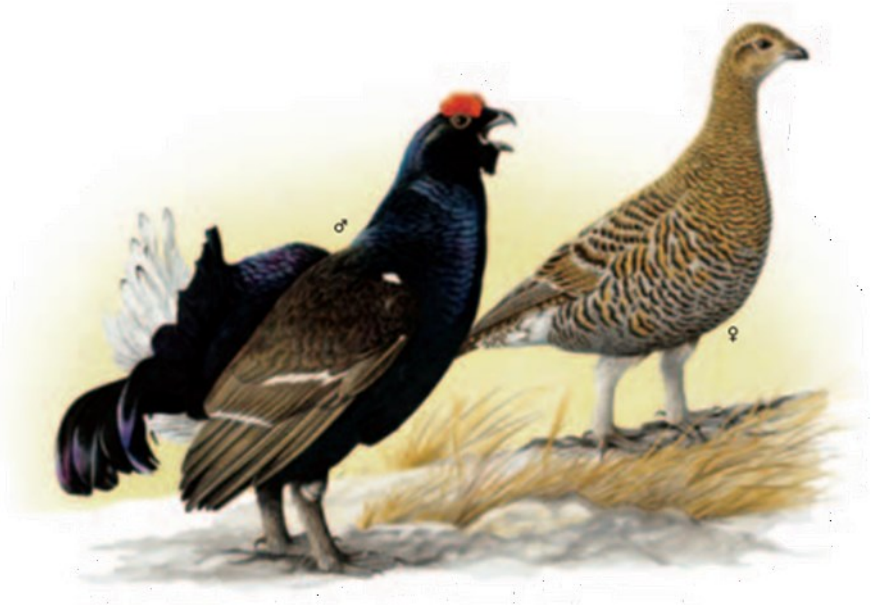


Figure 1.5.1 – Black grouse (*Lyrurus tetrix*). Both the adult male (left - 1000-1750g) and adult female (right - 750-1100g) of the species are shown. Young males have a similar appearance to females until four months of age (Demartin and Flaim, 2012). From Bocca (2004a).

1.5.1. DISTRIBUTION AND HABITAT

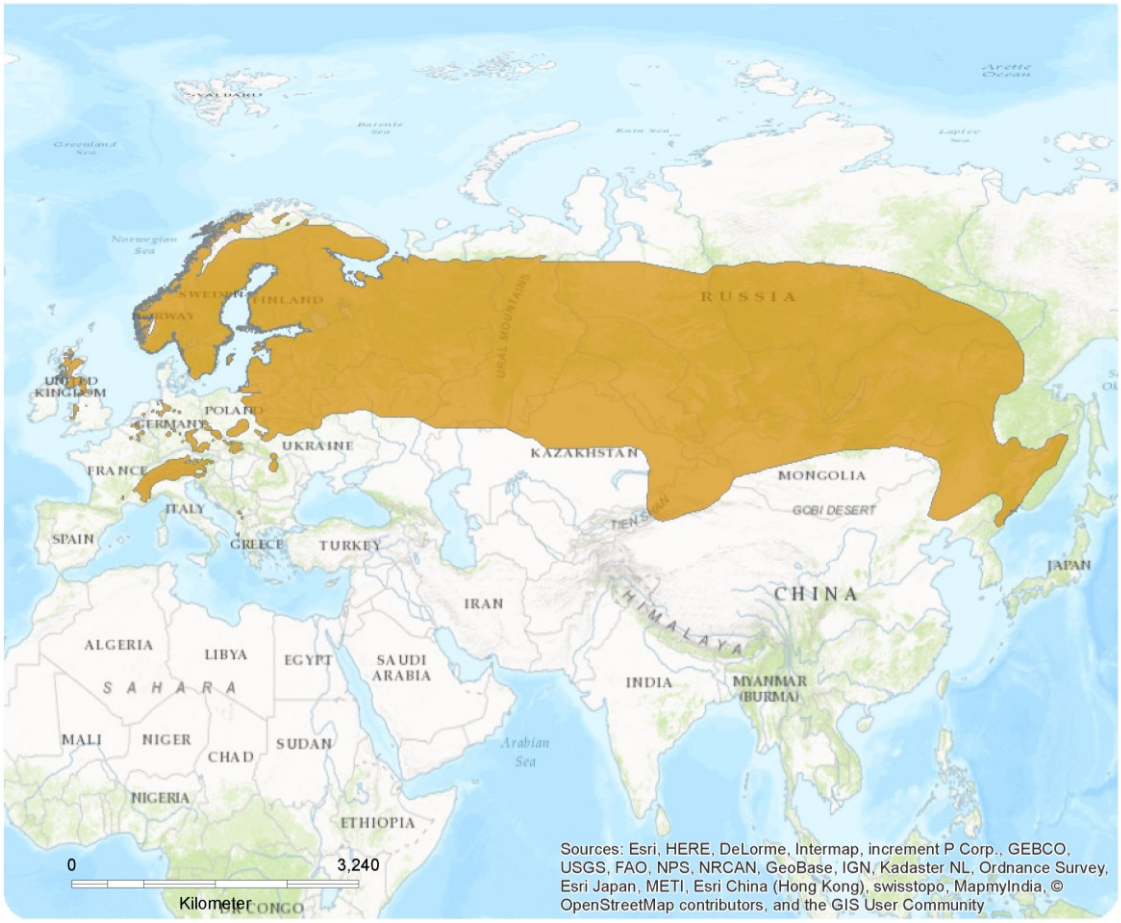


Figure 1.5.2 – The distribution of the black grouse. The species range is shown across northern Eurasia in orange. From BirdLife International (2016a).

The highest densities of black grouse are found in northern Eurasia (Figure 1.5.2). The species has a continuous distribution across the northern areas of its ranges from Scandinavia to southeastern Siberia, however, in the southern and western areas of its range this distribution is more fragmented, with the largest population in these fragmented areas in the Alps (Storch, 2007; BirdLife International, 2016a). The preferred altitude of this species varies widely depending on latitude and climate but is mainly within the range of 900-2,300 m above sea level (a.s.l.) (Caizergues *et al.,* 2003a; Bocca 2004a). The species has broad habitat requirements: in boreal regions, the black grouse is present at the edge of coniferous forests where vegetation is in the early stages of forest succession. Outside of these areas, the species can be found in habitats that are structurally similar such as heaths,

treelines, alpine pastures and meadows. In general, the black grouse is found in open areas, avoiding habitats with dense tree cover (Storch and Segelbacher, 2000; Angelstam, 2004; Storch, 2007; Wegge and Kastdalen, 2008; BirdLife International, 2016a).

1.5.2. BEHAVIOUR

The black grouse is a polygamous species. In April the males join leks and display themselves to visiting females. Typically the largest and dominant male is chosen by multiple females (Rintamäki *et al.,* 1995). After mating, females move away and lay a clutch of 6-10 eggs at the end of May, which hatch by the end of June, producing chicks that are immediately active. Chicks usually become independent after three months, however, during this time there is a high mortality rate linked with adverse conditions such as low temperature and heavy rainfall (Storch, 2007; Demartin and Flaim, 2012; Viterbi *et al.,* 2015). The males of the species tend to be philopatric or disperse only short distances (typically up to 1km), with a maximum distance recorded by Caizergues and Ellison (2002) of 8.5km. However, females make short migrations of up to 14km between summer and winter habitats in the Alps (Caizergues and Ellison, 2002; Warren and Baines, 2002; Lebigre *et al.*, 2008). Marjakangas and Kiviniemi (2005) recorded a maximum female yearling dispersal of 33.2km in Finland. The black grouse feed opportunistically on buds, leaves and flowers of shrubs and herbs, such as rhododendron and marigold as well as fruits such as blueberries and cranberries. They also feed on the needles and twigs from trees such as birch, alder, spruce and pine. The birds change their diet seasonally, with hens requiring a protein-rich diet during the spring; in the summer, areas with a higher abundance of invertebrates are sought as these are eaten by the chicks. In the winter the black grouse tunnels under the show to search for food as well as to shelter from low temperatures, and to maintain their precious energy reserves (Beeston *et al.,* 2005; Storch, 2007; Wegge and Kastdalen, 2008; Demartin and Flaim, 2012; BirdLife International, 2016a).

1.5.3. CONSERVATION STATUS

The black grouse has a global population estimation of 8,000,000-14,000,000 mature individuals (BirdLife International, 2016a), although population sizes are

subject to periodic fluctuations of 15-20 year cycles (Demartin and Flaim, 2012); generation time is about 6.4 years globally and 4.1 years in its European populations (BirdLife International, 2016a; BirdLife International, 2021a). Given its extensive northern populations, the black grouse is listed globally as 'Least Concern' (IUCN 2022; BirdLife International, 2016a); however, the EU28 regional assessment which represents the fragmented populations at the southern and western edges of its range, lists the black grouse as 'Vulnerable' and these populations are a target for conservation efforts (Storch, 2000; BirdLife International, 2021a). Globally, the black grouse has a decreasing population trend, however, it is not decreasing at a rate that would allow it to fall into the 'Vulnerable' category; in the EU28, the population trend has estimated to have decreased approximately 44% in the last 11 years (BirdLife International, 2016a; BirdLife International, 2021a). While the overall population size of the black grouse is large, in the fragmented areas of its range, populations may be isolated from each other, and many habitat patches are small which puts them at risk of low genetic variability and inbreeding depression (Keyghobadi, 2007; BirdLife International, 2016a).

1.5.4. THREATS

The main threats this species faces in western and central Europe are habitat fragmentation and destruction due to human activity, including land-use changes, like afforestation which disturbs the species' natural habitat, as well as the intensification of agriculture and the grazing of livestock (Bowker *et al.,* 2007; Storch 2007). Human disturbance from infrastructure associated with tourist and leisure activities, such as skiing, also pose a threat to the species, since many birds are killed due to collisions with power lines, fences and ski lift cables and pylons. The increase in stress from disturbance, especially in winter, has been shown to reduce the species' resistance to disease (Storch, 2007; Formenti *et al*., 2015). In addition, the black grouse is a popular game bird, and although black grouse are protected in western and central Europe, and hunting is generally highly regulated (and mainly restricted to male birds), in some countries, harvesting pressure is high through both legal and illegal hunting of the species; in some areas, campaigns attempt to limit the disturbance of these species (Storch, 2007). Changes in climate (increased rainfall in June and increased December temperature) are also thought to affect the population trends of the black grouse, including breeding success (Summers *et al.,*

2004; Viterbi *et al.,* 2015). Efforts have been made to reintroduce captive-bred birds to wild populations, however, these have had little success to date (Ludwig *et al.,* 2008; Walker, 2010). Proposed conservation actions include further research investigating habitat fragmentation, the impacts of hunting, dispersal rates, and population dynamics. It was also been suggested that there is a need for the restoration of the species habitat and spatial connectivity between habitat patches (Höglund *et al*., 2007; White *et al.,* 2013; Warren *et al.,* 2020).

1.5.5. PREVIOUS GENETIC RESEARCH

Most previous genetic studies of black grouse have used neutral markers, such as mtDNA and STRs, to characterize population genetic diversity across the species range (Larsson *et al*., 2008; Höglund *et al*., 2011; Rutkowski *et al*., 2018). Larsson *et al.* (2008) and Höglund *et al*. (2011) used these markers to investigate populations in Britain, both studies showed that these populations were isolated from the other populations in the range and Höglund *et al.* (2011) concluded that the British populations represent a distinct Management Unit. Höglund *et al*. (2007) and Sittenhaler *et al*. (2018) used STRs on both invasive and non-invasive samples to investigate genetic variability and fine-scale genetic structure, as well as the effect of fragmentation in Alpine populations, concluding that Alpine populations exhibited a similar level of genetic diversity to populations in the north of the species range. However, Caizergues *et al*. (2003a) found that the Alpine populations had a higher level of differentiation and suggested these Alpine populations are subject to limited gene flow. Strand *et al*. (2012) compared genetic diversity calculated with SNP, STR and major histocompatibility complex (MHC) markers to study the black grouse, they found that genetic diversity of the MHC markers was lower compared to the diversity found with STR and SNP markers. The SNPs were obtained by sequencing 24 protein-coding regions and not WGS techniques.

Other genomic resources have been generated by Wang *et al*. (2014), who produced a black grouse genome contig assembly and Li *et al*. (2016) who sequenced the complete mitochondrial genome allowing for these resources to be used in future studies.

1.6. ROCK PTARMIGAN (*Lagopus muta*).

Figure 1.6.1 – Rock ptarmigan (*Lagopus muta*). The female bird (summer plumage) is shown to the left, with the males on the right (summer plumage upper and winter plumage lower), respectively. Sexual dimorphism in size is minimal: males weigh between 380-520g and females between 350-470g (Demartin and Flaim, 2012). From Bocca (2004b).

1.6.1. DISTRIBUTION AND HABITAT

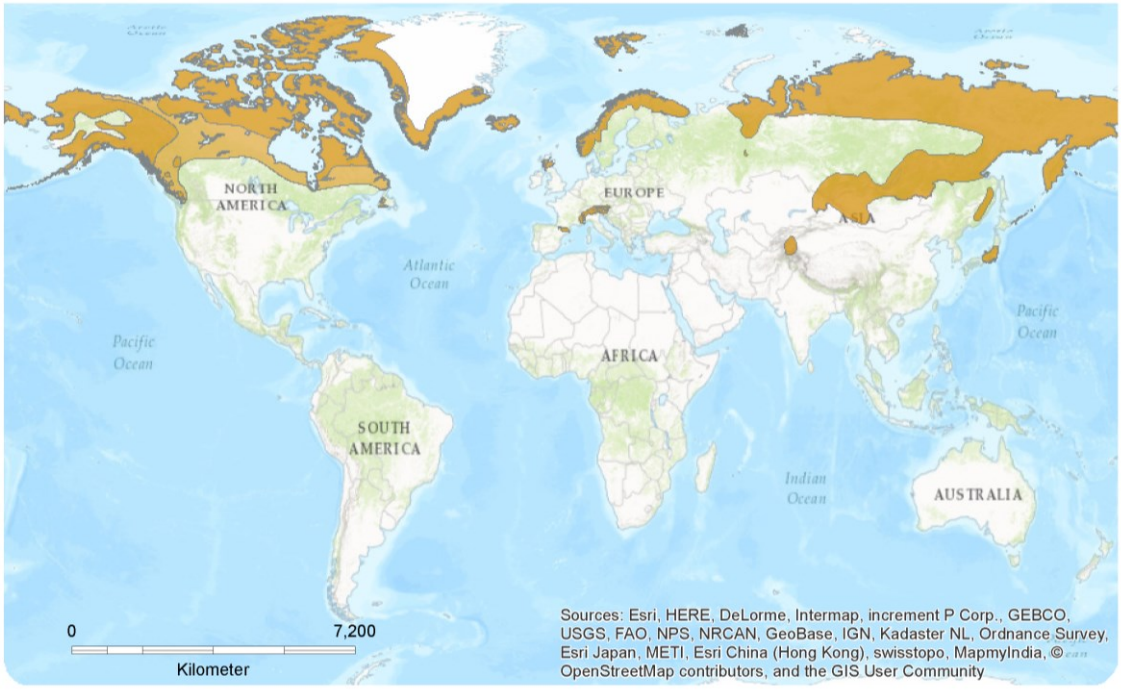


Figure 1.6.2 – The distribution of the rock ptarmigan. The species resident range across the northern hemisphere is shown in orange with its non-breeding range shown in beige for North America. For Eurasian populations the species summer and winter ranges are the same. From BirdLife International (2016b).

The rock ptarmigan has the widest latitudinal distribution of all the grouse species, spanning across northern North America and Eurasia between 42°N in the Pyrenees and 83°N in Greenland (Figure 1.6.2; Storch, 2007). The species inhabits areas of dry, rocky tundra, or alpine summits, and in the winter the species makes seasonal migrations to higher latitudes or altitudes. It prefers areas of sparse vegetation, predominantly grasses and mosses. In winter, the species stays in areas where ground vegetation remains accessible in snow-swept areas (Storch, 2007; BirdLife International, 2016b). In the Alps, the ptarmigan is typically found at altitudes over 1,800m and very rarely traverses below the upper limit of arboreal vegetation (Cattadori and Hudson, 1999; Bocca, 2004b; Pernollet *et al.,* 2015).

1.6.2. BEHAVIOUR

The ptarmigan was believed to migrate relatively short distances. Nilsen *et al*. (2020) found a mean migration distance of 20.3 km in populations in central Norway; however, Storch (2007) reported a seasonal migration of more than 1,000km

between Greenland and North America. Ptarmigan are monogamous and mating pairs typically occupy their territory in April, which the male defends with its calls and intimidating performances from higher altitudes. In June, females lay between 5-8 eggs, often in reasonably open areas. Chicks are immediately active when they hatch around three weeks later, becoming independent after approximately three months; chick survival is linked to adverse weather conditions during incubation and to predation (Scherini *et al*., 2003; Demartin and Flaim, 2012). The species diet varies seasonally but consists of buds, twigs, shoots and leaves of trees and shrubs as well as berries in the winter (García-González *et al.,* 2016)

1.6.3. CONSERVATION STATUS

The IUCN list the rock ptarmigan globally as 'Least Concern' with a decreasing population size estimated at over 8,000,000 individuals (BirdLife International, 2016b). Population sizes are reported to fluctuate in 10-year cycles (Storch, 2007), and the generation time for ptarmigans is estimated as 4.2 years globally and 3.4 years in Europe (BirdLife International, 2016b; BirdLife International, 2021b). In 2015 European populations were assessed as 'Near Threatened' by the IUCN, this classification has been reduced to 'Least Concern' for the latest assessment and the European population trend is listed as stable. However, populations in the southern edges of the ptarmigan's European range (e.g. Italy, Slovenia and Lichenstein) show a decreasing population trend (BirdLife International, 2021b).

1.6.4. THREATS

As this species is usually present in areas with low human population density, it is relatively well-protected. However, the species does face local threats including habitat degradation, over-hunting, overgrazing and the effects of tourism (Watson and Moss, 2004; BirdLife International, 2016b). Overhead power lines also pose a threat to the species as collisions often result in fatalities (Bevanger and Brøseth, 2001). Climate change is also a threat to the species, since it is adapted to cold climates, and climate change has led to an increase in temperature and shift of the treeline (Revermann *et al.,* 2012). As a result, the ptarmigan has shifted its altitudinal range, with birds being found at increasingly higher altitudes in a study in the Swiss Alps (Pernollet *et al.* 2015). The IUCN recommends species monitoring in areas of

high hunting pressure and implementing programmes similar to conservation efforts in Germany to minimise the disturbance of the ptarmigan in heavily used tourist areas (BirdLife International, 2016b). More research on climate change and habitat changes has also been suggested (Revermann *et al.,* 2012).

1.6.5. PREVIOUS GENETIC RESEARCH

Genetic research in this species has primarily consisted of investigating the patterns of variation using mtDNA and STRs in populations from across the species range, mainly from North America and Europe (Quintela *et al.,* 2010; Bech *et al.,* 2013; Lagerholm *et al.*, 2017). Holder *et al.,* (2004) used mtDNA to study populations of Nearctic rock ptarmigan, their results identified at least six ESUs, four of which corresponded to recognised subspecies. Lagerholm *et al*. (2017) used ancient mtDNA collected from palaeontological sites for the rock ptarmigan and willow ptarmigan (*Lagopus* lagopus) to show a genetic continuity across the European populations examined; the authors suggested that this is due to continuous habitat availability as well as the birds' ability to fly. Quintela *et al.* (2010) studied the possibility of hybridisation between the rock and willow ptarmigans using both mtDNA and STRs, they found that hybridisation does occur between these species and that plumage identification is not a reliable indicator of a hybrid individual. Bech *et al.* (2013) suggested that a major bottleneck is responsible for the lack of genetic variation in Europe, while Caizergues *et al.* (2003b) found a significant pattern of isolation by distance in males from the French Alps.

To date, no genomic research has been published on the ptarmigan using SNPs obtained using WGS approaches. However, Costanzi et al. (2018) used NGS techniques to develop new microsatellite markers specific to the rock ptarmigan and willow grouse (*Lagopus lagopus*), and Kozma *et al*. (2019) used outlier loci to investigate adaption in three species of *Lagopus*, concluding that the genomic regions of these species likely vary due to speciation and adaption. However, the complete mitochondrial DNA has been sequenced for the rock ptarmigan as well as for the Japanese subspecies *L. m. japonica* (Sveinsdóttir and Magnísson, 2017; Yonezawa and Nishibori, 2020).

1.7. CAPERCAILLIE (*Tetrao urogallus*)

Figure 1.7.1 – Western capercaillie (*Tetrao urogallus*). The male (left) and female (right) show the marked sexual dimorphism of the species (females: 1400-2300g, males: up to 5000g). Young males are more similar to females (Demartin and Flaim, 2012). From De Franceschi (2004).

1.7.1. DISTRIBUTION AND HABITAT

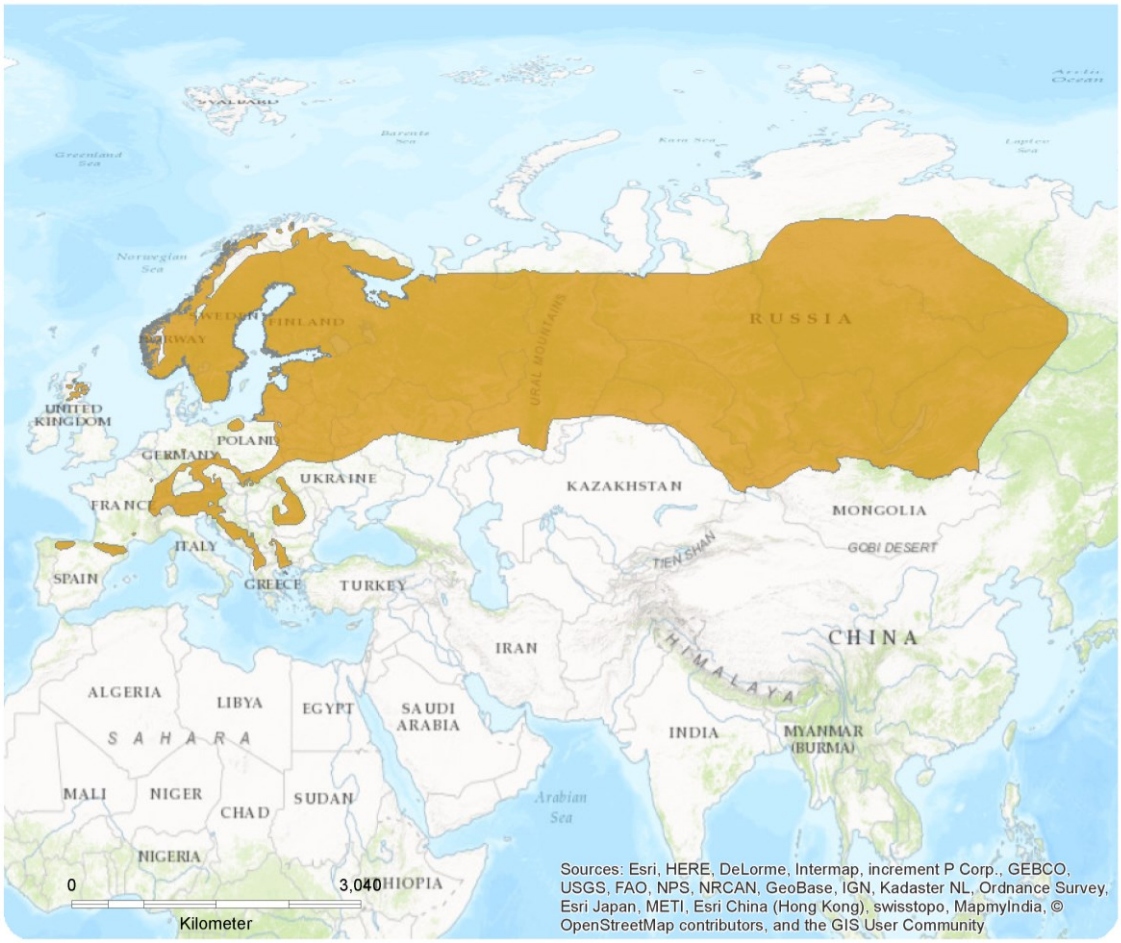


Figure 1.7.2 – Distribution of the western capercaillie. The resident range is shown in orange, representing both the summer and winter distribution. From BirdLife International (2016c).

The capercaillie is present across Eurasia in a mostly continuous distribution from Scandinavia to Siberia (Figure 1.7.2). Populations are more fragmented in western and central Europe. The species' preferred habitat consists of boreal forests: coniferous or mixed forests with open structures, but some canopy cover (BirdLife International, 2016c; Storch, 2007). The capercaillie appears to prefer wetter areas, such as bogs, or areas interspersed with young trees in the successional forest at altitudes between 1,000 and 2,000m a.s.l. (Cattadori and Hudson, 1999; Demartin and Flaim, 2012; BirdLife International, 2016c).

1.7.2. BEHAVIOUR

The capercaillie is polyamorous, and males begin to form poorly defined leks in April, with the dominant male typically mating with more than one female. Mated females move away from the leks to lay 7-9 eggs in a small depression on the ground, often at the base of a tree, lining the nest with feathers and plant material. After an incubation period of just under four weeks, chicks hatch and remain close to the female, depending on her entirely for the first few days. Capercaillie chicks become independent after 5 or 6 months, with mortality depending on weather conditions, predation and disturbance. Males reach sexual maturity at three years however the females typically reach reproductive state sooner (Demartin and Flaim, 2012). The capercaillie and black grouse are also known to create potentially fertile hybrid individuals (Kleven *et al.,* 2020): This hybridisation most commonly occurs between a male black grouse and female capercaillie to produce fertile male offspring (female hybrids are typically sterile); hybrid males have also been observed participating in black grouse leks (Porkert *et al.*, 1997).

The capercaillie is an elusive species and males are generally philopatric; if disturbed the species usually retreats on foot, but can also fly. In fact, Moss *et al*., (2006) observed a median dispersal distance for young hens of 11km. In the winter months, the capercaillie inhabits more open coniferous forests where its diet mostly consists of conifer needles. In the summer months, this species prefers areas with a higher availability of berries, but they also feed on the leaves, buds and flowers of these herbs and shrubs (Storch, 2007; Wegge and Kastdalen, 2008; Demartin and Flaim, 2012).

1.7.3. CONSERVATION STATUS

The IUCN classify the capercaillie as ‘Least Concern’ in both its global and European distributions given the species range and population size. The global population trend for the species is decreasing; however, European populations are increasing. The global population size for mature individuals is estimated at 3,000,000 to 5,500,000 and the species has a generation length of 6.4 years globally, this is reduced to 5.6 years in its European distribution (BirdLife International, 2016c; BirdLife International, 2021c). Some national Red Lists, such as that of Italy, classify the capercaillie as ‘Vulnerable’ due to its declining

populations and highly fragmented habitat (Peronace *et al*., 2012). In addition, the capercaillie is considered an indicator species for mature boreal forest, and therefore, for both these reasons is a focus of conservation efforts (Suter *et al.*, 2002).

1.7.4. THREATS

The primary threats for this species include habitat destruction and degradation as well as reduced population sizes in the areas where species habitat is fragmented. Overhunting has also led to species decline. Despite apparently strict hunting regulations, it appears many more birds are shot illegally (BirdLife International, 2016c). As with the black grouse and ptarmigan, collisions with fences and power lines lead to numerous fatalities annually. Disturbance from skiing facilities and tourism as well as predation and climate change also threaten the species (Summers *et al.,* 2004; Čas, 2010; Coppes *et al.*, 2017). Habitat management is key for this species (e.g. in Scotland; Broome *et al.,* 2014), some areas of the species range already fall under protected areas such as national parks, however, Mikoláš *et al.* (2017) found that suitable habitat was still lost in these protected areas by human activities such as logging. Reintroductions of captive-bred individuals have been attempted but, for the most part, these have been unsuccessful, except in Scotland, where the capercaillie went extinct in 1785 and a reintroduction in 1836-1837 succeeded in re-establishing the species. However, these populations are again in decline and increased protective measures are being proposed to protect them, such as reducing the number of fences in the species habitat (Moss *et al.,* 2000; Moss, 2001; BirdLife International, 2016c). As many areas of the species range indicate the need for distinct management units, more research is needed to understand population dynamics, as well as the conservation significance of 'subspecies' in these areas (Segelbacher and Piertney 2007; Klinga *et al*., 2015; Fameli *et al*., 2017). The IUCN also propose further restoration and protection of capercaillie habitat, particularly in areas where the species is already threatened (BirdLife International, 2016c).

1.7.5. PREVIOUS GENETIC RESEARCH

Many genetic studies have focussed on the endangered Cantabrian subspecies,

using mtDNA and STRs to investigate genetic variability and structure within this taxon. These studies concluded that the Cantabrian population is an isolated, but panmictic, population with little genetic variability and that the area should be defined as an ESU (Duriez *et al.,* 2007; Rodrigues-Muñoz *et al.,* 2007; 2015). Although Duriez *et al.* (2007) found, using mtDNA, that the Cantabrian and Pyrenees populations were closely related to each other, Segelbacher and Piertney (2007) argued that mtDNA phylogeographic analyses suggest the Pyrenees populations should be classified as a separate ESU. The phylogenetic results of Segelbacher and Piertney (2007), also suggest that the 12 subspecies are unfounded, since they only found two distinct clades across the species' European range. Rutowski *et al*. (2017) also proposed that the Polish Carpathian and lowland populations should be classified as MUs. Two studies by Segelbacher *et al.* (2003; 2007) used STRs to estimate dispersal distances, concluding that there was no significant population genetic structure between closely spaced (4km) leks in Russia, but that populations in the Bavarian Alps over 10km apart showed significant genetic differentiation.

The complete mitochondrial DNA has been sequenced for this species; however, this is not the case for the nuclear genome (Aleix-Mata *et al.,* 2019). To date, no genomic studies have been published on the species.

1.8. AIMS AND MOTIVATION FOR THIS THESIS

The black grouse, ptarmigan and capercaillie represent three of the four grouse species present in Italy (the fourth being the hazel grouse, *Bonasa bonasia*), and the Alpine populations of all three in Italy represent the southern edge of these species' ranges, which as mentioned above, are typically more fragmented than those in their northern ranges. In fact, both the ptarmigan and capercaillie are listed as 'Vulnerable' in the Italian Red List (Peronace *et al.* 2012). Although the capercaillie is a protected species in Italy, limited numbers of black grouse males and both sexes of rock ptarmigan are legally hunted in most Italian regions of the Alps from 1st October to 30th November based on spring censuses (Federazione Italiana Della Caccia, n.d.). Populations of these species are regularly monitored to ensure that hunting does not impact overall species numbers; however, hunting pressure is high, and its impact on population status is unknown. No genomic work

has been published on these species from Italy to date, despite these techniques potentially enabling a more in-depth analysis of the genetic health of the species, as well as gene flow between populations, and other population parameters useful to deciding management interventions.

For these elusive grouse species, non-invasive samples will be the preferred sampling type to avoid disturbing the birds; however, genomic techniques, while providing more detailed estimates, often require high quality and quantities of DNA which may not be achievable with non-invasive samples. Blood and tissue samples can be collected from black grouse and the ptarmigan hunting bags, and feathers can be collected from leks and preferred habitats; however, for the capercaillie, samples are predominantly non-invasive (feathers and faecal pellets from lekking arenas). The combination of high availability of dual samples types, as well as their vulnerable and fragmented status and range of dispersal capabilities, make this taxon suitable for testing the benefits of applying next-generation techniques to conservation issues, before applying them to poorer quality or more valuable samples of endangered species.

Using over 1,200 samples from all three species, collected primarily across the Italian Alps over a period of 20 years, this thesis aimed to type SNP markers for a large set of black grouse, ptarmigan and capercaillie samples using GBS techniques. The genetic structure identified with these markers was then compared to patterns available from previously generated mtDNA and STR results to determine the benefits of a genomic approach. For the black grouse, the environmental or anthropic factors potentially affecting genomic variation were also tested, by comparing genomic variation over time or by comparing environmental patterns across the sampling area to the genomic patterns. The potential impact of these results on effective management and conservation of these endangered species is discussed.

To achieve the above aims, the following hypotheses were tested:

H1: SNP datasets are capable of producing finer-scale patterns compared to the traditionally used STR markers;

H2: Black grouse populations will be isolated by distance, with evident impacts of heavily urbanized valleys, and loss of genetic variation through time from

overhunting.

H3: Ptarmigan populations are genetically isolated due to the species preference for habitats at high altitudes (over 1,800m) and relatively low dispersal distances;

H4: Capercaillie populations are genetically isolated even within the Trentino Alto-Adige region, with distinct patterns present on either side of the Adige Valley as found in mammals with lower dispersal distances by Vernesi *et al.* (2016).

2. METHODS

2.1. STUDY AREA AND SAMPLES

Biological samples of all three species; black grouse, capercaillie and ptarmigan, were collected from 1995 to 2017 across the Italian Alps (Figure 2.1), from seven regions: Liguria, Piedmont, Aosta Valley, Lombardy, Trentino-Alto Adige, Veneto and Friuli-Venezia Giulia. Additional samples for the ptarmigan and capercaillie were also collected from areas of the species wider ranges i.e. Austria, Switzerland, Finland, Sweden and Iceland. Within northern Italy, for the black grouse and ptarmigan, samples were assigned to 'populations' corresponding to the International Standardized Mountain Subdivision of the Alps (SOIUSA; Marazzi, 2005) as shown by the different coloured sections in Figure 2.1. Capercaillie samples were collected from protected areas within Trentino-Alto Adige and divided into 'populations' based on the national parks and localities where they were collected.

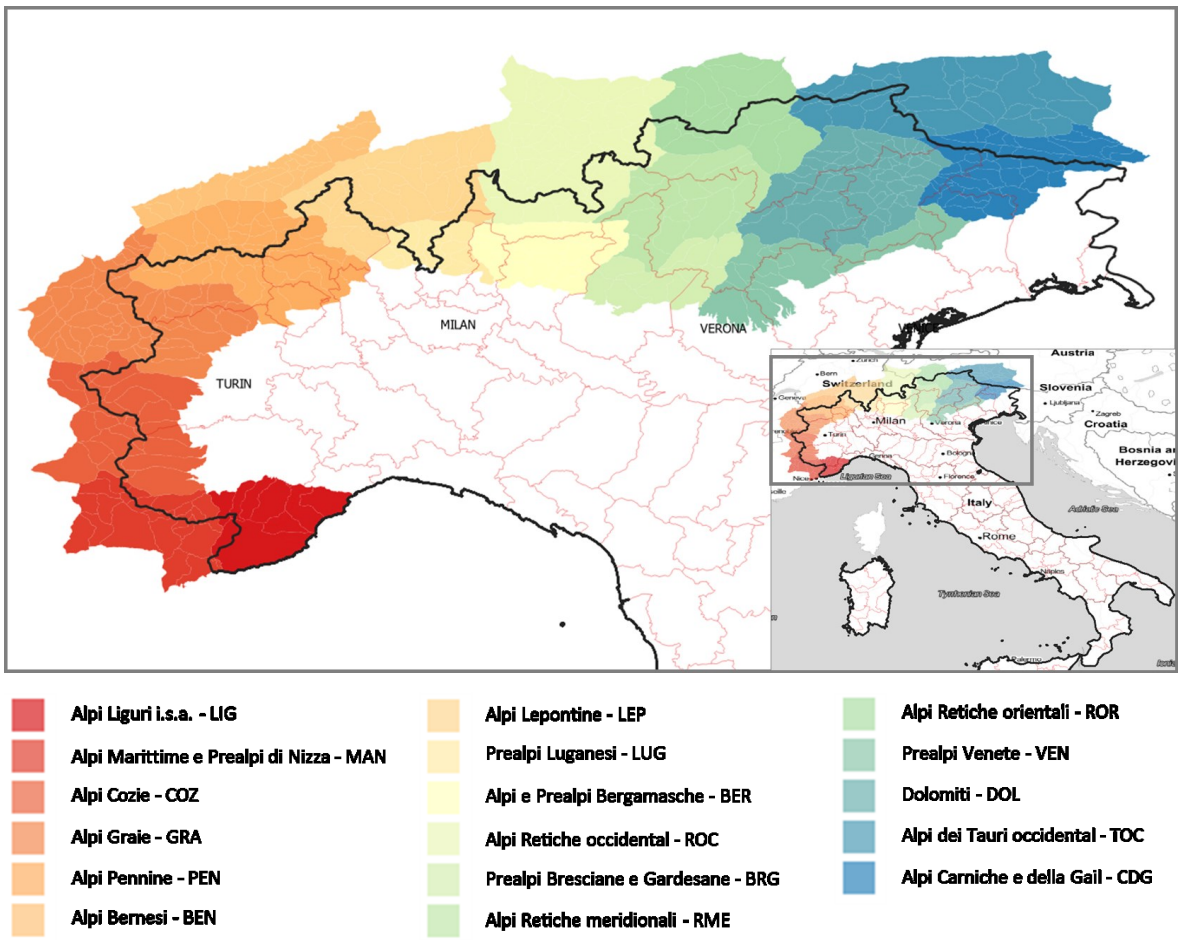


Figure 2.1 – Map of northern Italy showing the SOIUSA (International Standardized Mountain Subdivision of the Alps) mountain groups (shaded areas). Red lines delineate the Provinces; the black line indicates the national border. Created in QGIS (QGIS Development Team, 2021) with data from Accorsi (n.d.).

2.2. SAMPLE COLLECTION

Tissue, feather and faecal samples were available in the biobanks of the Conservation Genomics Unit, Fondazione E. Mach, having been collected across the Italian Alps in collaboration with the National Union of Alpine Hunters (*Unione Nazionale Cacciatori Zona Alpi*), Trentino Hunter's Association (*Associazione Cacciatori Trentini*), Wildlife Service of the Province of Sondrio (*Servizio Caccia e Pesca della Provincia di Sondrio*), Valtellina Hunter's Association (*Associazione Cacciatori Valtellinesi*), Wildlife Service of the Province of Belluno, Dolomiti Bellunesi National Park, Belluno Hunter's Association (*Associazione Cacciatori Bellunesi*), ALCOTRA project (Regions Piedmont and Aosta Valley). Tissue samples and feathers originated from bagged black grouse and ptarmigan with permission from hunters, during voluntary or obligatory inspection of hunting bags

by provincial wildlife services. Small (up to 200mg) samples were taken using gloves and sterile scissors from thoracic muscles during inspections on the same day as the animal was bagged and immediately frozen in sterile 1.5ml Eppendorf tubes at -20°C. Shed feathers and faeces were collected for all species by researchers, wildlife technicians, park rangers and hunters during censuses conducted in the post-reproductive season, in collaboration with the Servizio Foreste e Fauna (Wildlife Services) of the Autonomous Province of Trento (PAT) and Trentino Hunter's Association (Associazione Cacciatori Trentini), as well as opportunistically by PAT wildlife technicians. **Black grouse** samples collected in the Autonomous Province of Trento were divided into three timeframes; 1995-1999, 2009-2010 and 2015-2017, to investigate any changes in genetic variation over time. No temporal analyses were performed for the ptarmigan or capercaillie as, typically, samples were only collected once per population throughout the sampling period. As a result of strict hunting laws, for black grouse, samples were predominantly from male birds. Sex is easily determined for feather samples due to the sexual dimorphism in the plumage of the black grouse.

2.3. DNA EXTRACTION AND GENOTYPING

Whole DNA was extracted from all samples and sample types using Qiagen Extraction kits (DNeasy Blood & Tissue, QIAamp DNA Investigator and QIAamp® DNA Stool Mini Kit; Qiagen Inc., Hilden, Germany) following the manufacturer's instructions. Extractions were performed in a separate pre-PCR laboratory in a dedicated area with specific pipettes and filter tips to reduce the probability of contamination. The quality and quantity of extracted DNA was measured using agarose gel electrophoresis and a Qubit fluorometer (Invitrogen, Carlsbad, United States), respectively.

In order to compare the results of genetic (mtDNA, STRs) and genomic (SNP) markers, the entire D-loop region (1235bp) was amplified for the **ptarmigan** with the primers PHDL (5'-AGGACTACGGCTTGAAAAGC-3') and PHDH (5'-CATCTTGGCATCTTCAGTGCC-3') as described in Randi and Lucchini (1998). Samples that could not be sequenced with this primer pair (probably as a result of DNA degradation) were amplified with overlapping d-loop internal primers PH-L400 (5'-ATTTATTGATCGTCCACCTCACG-3') and PH-H521 (5'-TTATGTGCTTGACCGAGGAACCAG-3'; Randi and Lucchini, 1998). For the

capercaillie primers, PHDL and PH-H521 were used, resulting in a trimmed fragment length of 388bp. Each individual was assigned to a haplotype for both the long and short fragments, and these haplotype sequences were used for further analyses. Since previous research on **black grouse** has already focused on patterns of mitochondrial DNA haplotypes, this marker was not reanalysed here.

Black grouse and **ptarmigan** samples from across the Alps were genotyped with ten previously developed tetranucleotide microsatellite (STR) markers (Table 2.1), seven of which were designed for the black grouse (BG10, BG12, BG15, BG16, BG18, BG19, BG20; Piertney and Höglund, 2001) and three for the capercaillie (TUT1, TUT2, TUT3; Segelbacher *et al.,* 2000). **Capercaillie** samples were also genotyped with the same ten microsatellites, but BG19 was replaced by TUT4 (Segelbacher *et al*., 2000). Samples were first tested with BG10 and BG12 to confirm that they were of a sufficient quality for genotyping. To improve the potentially high error rate often shown by non-invasive samples, a multi-tube approach was used, where two independent amplifications were conducted at each locus to allow for a confirmation of allele calls in heterozygous individuals and four in homozygous individuals. Additional replicates for missing loci were carried out for individuals for which six loci were successfully typed. Loci that were repeatedly unsuccessful after eight replicates were recorded as missing data. The majority of the mtDNA and STR genotyping occurred prior to the start of this project following the protocol described in Collini (2011), except the black grouse from Trentino-Alto Adige sampled in the 2015-2017 time frame, which were genotyped here to complete the black grouse STR dataset for all three timeframes. These loci were amplified following the methods previously used with two multiplexes of five microsatellite loci each with fluorescent-labelled forward primers (G5 matrix; Applied Biosystems). The PCRs were performed in a total volume of 10µl, with Taq Hot Master (5-Prime), in a Veriti® Thermal Cycler (Applied Biosystems). The thermocycler conditions and amount of primer are shown in Table 2.2. In the case of poor amplification, a single primer PCR reaction was used. An ABI3130 Genetic Analyzer (Applied Biosystems) was used to separate the PCR product. Allele lengths were measured using GeneMapper software v 3.7 (Applied Biosystems).

Table 2.1 – The microsatellite markers used in this study, showing the forward and reverse primer sequence, the repeated motif and the species for which each marker was used.*

Locus	Primer Sequence 5'-3'	Motif	Ref.	Species
BG19	CAAGGCGCAACATTAAGATTC TGTATTTTGGAAACTCTGTGTGC	$(GATA)_{13}$	1	B P
BG20	AAGCACTTACAATGGTGAGGAC TATGTTTTCCTTTTCAGTGGTATG	$(GATA)_{17}$	1	B P C
TUT1	GGTCTACATTTGGCTCTGACC ATATGGCATCCCAGCTATGG	$(CTAT)_{12}$	2	B P C
TUT2	CCGTGTCAAGTTCTCCAAAC TTCAAAGCTGTGTTTCATTAGTTG	$(GATA)_{12}$	2	B P C
TUT3	CAGGAGGCCTCAACTAATCACC CGATGCTGGACAGAAGTGAC	$(TATC)_{11}$	2	B P C
TUT4	GAGCATCTCCCAGAGTCAGC TGTGAACCAGCAATCTGAGC	$(TATC)_{8}$	2	C
BG10	ATGTTTCATGTCTTCTGGAATAG ATTTGGTTAGTAACGCATAAGC	$(GATA)_{17}$	1	B P C
BG12	TCTCCTTCTAAACCAGTCATTC TAGTTTCCACAGAGCACATTG	$(GATA)_{29}$	1	B P C
BG15	AAATATGTTTGCTAGGGCTTAC TACATTTTTCATTGTGGACTTC	$(CTAT)_{16}$	1	B P C
BG16	GTCATTAGTGCTGTCTGTCTATCT TGCTAGGTAGGGTAAAAATGG	$(CTAT)_{15}$	1	B P C
BG18	CCATAACTTAACTTGCACTTTC CTGATACAAAGATGCCTACAA	$(CTAT)_{17}$	1	B P C

* References (Ref.): 1 – Piertney and Höglund 2001, 2 – Segelbacher *et al.* 2000;

Species: B – black grouse, P – ptarmigan, C – capercaillie.

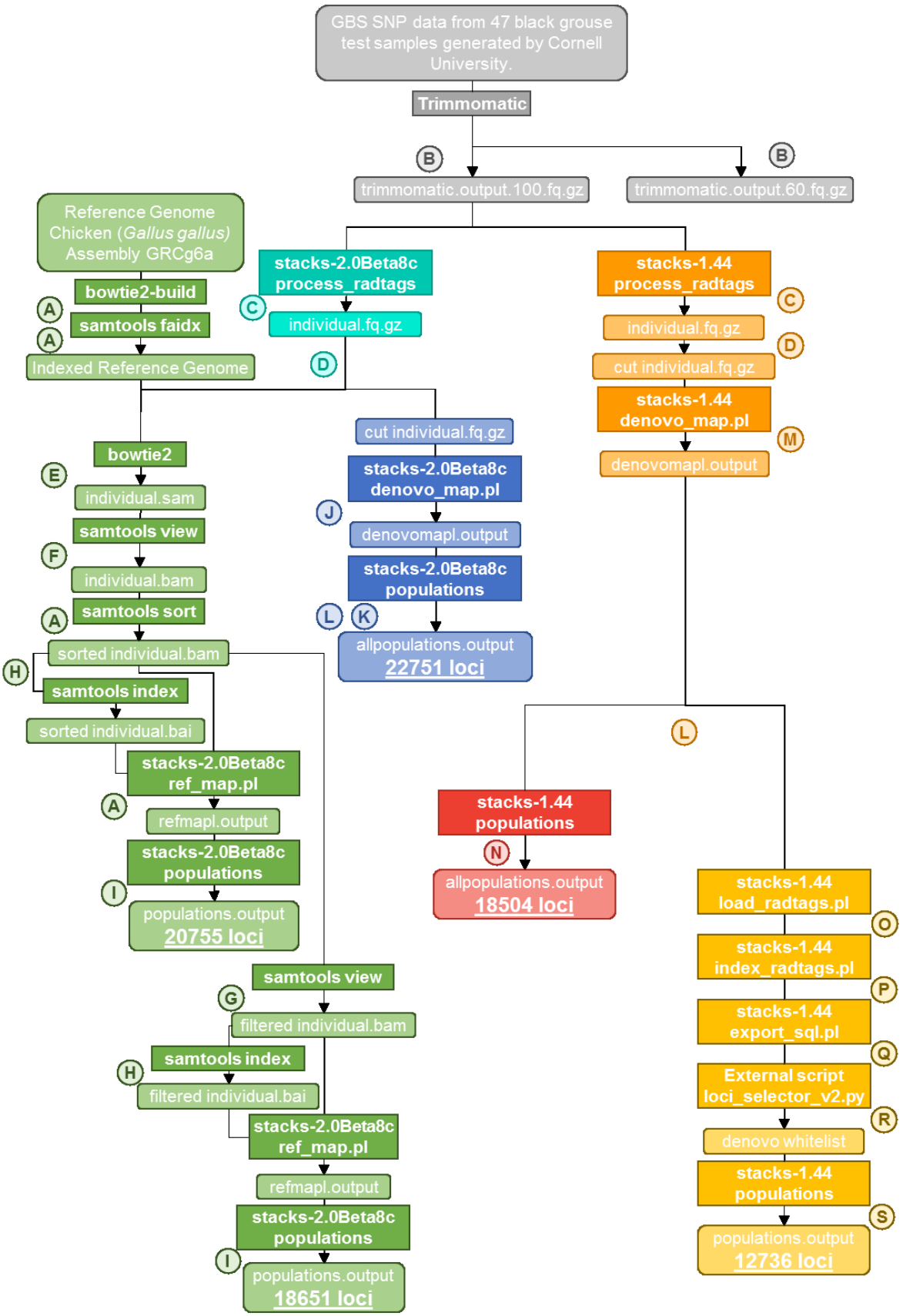
Table 2.2 – Multiplex reactions used microsatellite loci, showing the quantity of each primer, and thermocycler conditions.

Locus	Primer (μM)	Thermocycler Conditions
BG19	0.08	
BG20	0.1	**Multiplex 1** 95°C(120") [94°C(60")-60°C(30")-70°C(30") x 30] 60°C(120")
TUT1	0.08	
TUT2	0.13	
TUT3	0.2	
BG10	0.08	
BG12	0.14	**Multiplex 2** 95°C(120") [94°C(60")-56°C(30")-70°C(30") x 30] 60°C(120")
BG15	0.1	
BG16	0.08	
BG18	0.1	

Genotyping by Sequencing (GBS) was chosen to investigate patterns of genomic variation as this approach typically requires less DNA than other genomic methods (Elshire *et al.,* 2011). However, a greater DNA quality and quantity is still required for GBS analysis compared to other genetic markers. Therefore, samples with non-degraded DNA (determined using agarose gel electrophoresis) and a DNA quantity greater than 10ng/μl (measured using a Qubit fluorometer; Invitrogen, Carlsbad, United States) were identified. From this subset of samples, for all species, samples were selected for SNPs genotyping if they were already genotyped with STRs (to allow for a direct comparison of results from two different marker types), and an equal number of samples was chosen from each sampling area. For the black grouse and ptarmigan, the majority of samples used for the GBS analysis were tissue or feathers (from which blood could be extracted) from hunted animals. However, for the capercaillie samples used were feathers collected by non-invasive methods; therefore, only a limited number of samples produced an adequate quality and quantity of DNA for GBS.

A pilot study was conducted at the outset of the project to test the suitability of non-invasive samples for GBS with regards to the DNA quality and quantity: DNA from a sample subset (referred to as **'trial samples'**: 47 black grouse, 37 ptarmigan and 10 capercaillie) was sent for GBS analysis with the HiSeq Illumina Platform at Cornell University (BRC Genomic Diversity Facility; Ithaca, New York, USA) and sequenced using the method described by Elshire *et al*. (2011) using the restriction enzyme EcoT22I. This enzyme was chosen after fragment size distributions of GBS libraries, for three enzymes (ApeKI, EcoT22I and PstI), were compared; EcoT22I was recommended as it produced the smallest fragment pool and therefore would be the best for reducing genome complexity.

SNP data from the **black grouse trial samples** were used to test a processing pipeline in order to obtain the best quality SNPs, this pipeline became the baseline for the final datasets (details in Figure 2.2). Preliminary analyses were conducted using the yellow pipeline in Figure 2.2. For the final dataset, however, this pipeline was adjusted to incorporate a more recent version of the Stacks (Catchen *et al.,* 2013; Figure 2.3) and filtering was conducted with VCFtools (Danecek *et al*., 2011) as opposed to using the loci_selector_V2.py script (created by Emiliano Trucchi, available at: http://www.emilianotrucchi.it/done.html).


GBS SNP data from 47 black grouse test samples generated by Cornell University.
Trimmomatic
B
B
trimmomatic.output.100.fq.gz
trimmomatic.output.60.fq.gz
Reference Genome Chicken (Gallus gallus) Assembly GRCg6a
bowtie2-build
A
samtools faidx
A
Indexed Reference Genome
stacks-2.0Beta8c process_radtags
C
individual.fq.gz
D
stacks-1.44 process_radtags
C
individual.fq.gz
D
cut individual.fq.gz
stacks-1.44 denovo_map.pl
M
denovomapl.output
bowtie2
E
individual.sam
samtools view
F
individual.bam
samtools sort
A
sorted individual.bam
H
samtools index
sorted individual.bai
cut individual.fq.gz
stacks-2.0Beta8c denovo_map.pl
J
denovomapl.output
stacks-2.0Beta8c populations
L
K
allpopulations.output
22751 loci
stacks-2.0Beta8c ref_map.pl
A
refmapl.output
stacks-2.0Beta8c populations
I
populations.output
20755 loci
L
stacks-1.44 populations
N
allpopulations.output
18504 loci
stacks-1.44 load_radtags.pl
O
stacks-1.44 index_radtags.pl
P
stacks-1.44 export_sql.pl
Q
External script loci_selector_v2.py
R
denovo whitelist
stacks-1.44 populations
S
populations.output
12736 loci
samtools view
G
filtered individual.bam
samtools index
H
filtered individual.bai
stacks-2.0Beta8c ref_map.pl
refmapl.output
stacks-2.0Beta8c populations
I
populations.output
18651 loci

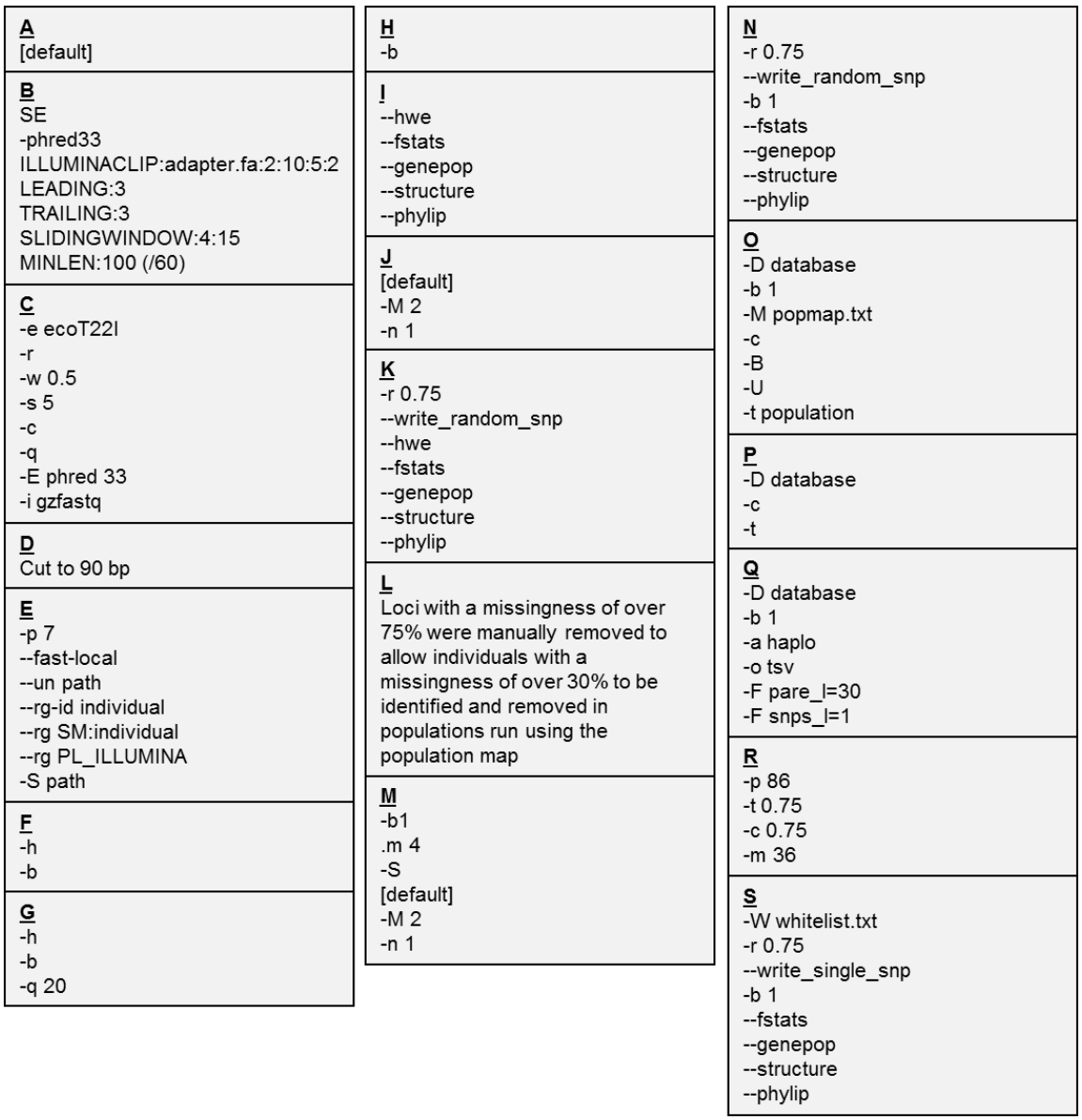


Figure 2.2 – Flowchart of the main pipelines trialled using the 47 black grouse test samples. The darker rectangular boxes with **bold** text list the software package or step in the pipeline used; adjacent circles indicate the parameters, with letters corresponding to the legend. The lighter boxes with rounded corners list the files produced and used for each following stage. The final number of loci produced by the pipelines are also shown.

After confirming that the DNA quality and quantity of the non-invasive samples was suitable for GBS techniques:

1) **Black grouse samples** (including trial samples) were sent to NOVOGENE (Beijing, China) for GBS analysis using the restriction enzymes MseI and HaeIII (following Elshire *et al.,* 2011) on a NextSeq Illumina Platform. *(NB: This change in laboratory was necessary because Cornell University could no longer provide GBS services due to the reissue of a key patent.)* New

restriction enzymes were selected based on the *in silico* enzyme digestion evaluation conducted by NOVOGENE. To improve overall coverage, GBS of these samples had to be repeated, to reach a final coverage of 20X;

2) **Ptarmigan and capercaillie samples** were sent to the University of Minnesota Genomics Centre (Minneapolis, Minnesota, USA) for GBS sequencing with EcoT22I and the NextSeq Illumina Platform. *(NB. This second change of laboratory was due to NOVOGENE no longer providing GBS services due to a company service transformation; however, the use of EcoT22I improved the coverage and quality of the GBS analysis compared to the black grouse)*. The data sequenced by the University of Minnesota were merged with the data from the trial samples sequenced at Cornell University. Six ptarmigan and three capercaillie were repeated in both the Cornell University and University of Minnesota sequencing efforts. Due to the different enzymes, and subsequent pipelines, used for each species, the data could not be analysed together, but results could still be compared between species.

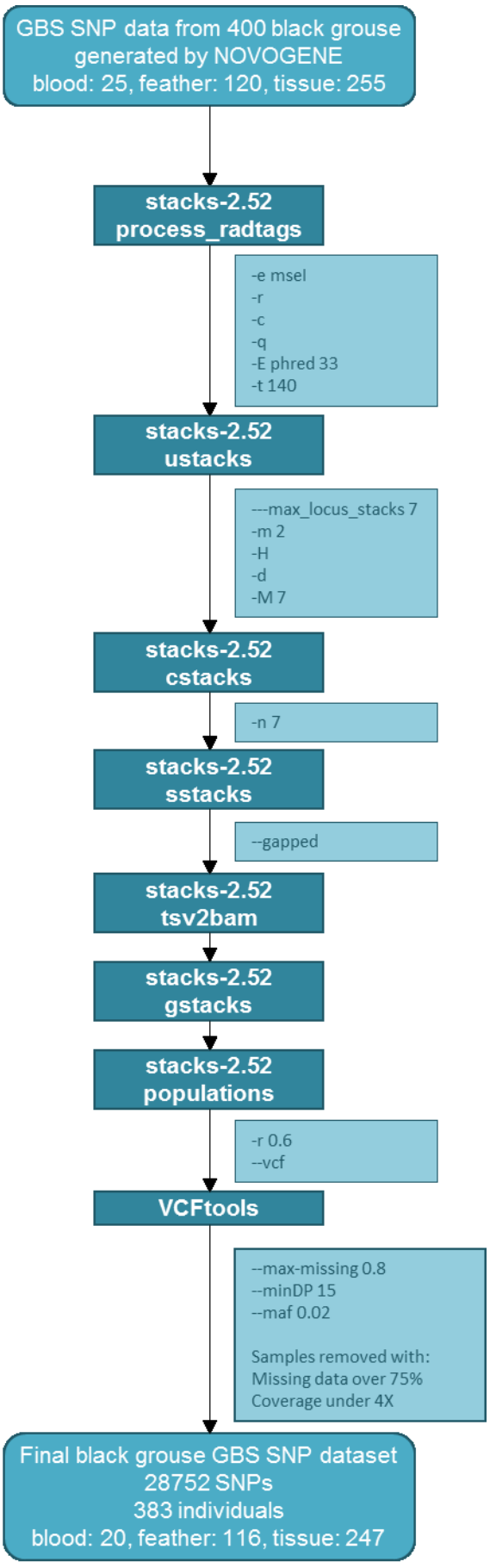
For the **black grouse** NOVOGENE data, raw sequences were processed with Trimmomatic (Bolger *et al*., 2014) to remove adapter sequences, and Stacks version 2.52 (Catchen *et al.,* 2013) was used to filter and align the sequences (Figure 2.3). This pipeline was run manually due to the large amount of data obtained in the sequencing. The final Stacks pipeline assembled the sequences with the following specifications: gapped alignment with paired sequences, deleveraging algorithm enabled, and up to seven mismatches allowed between stacks within and between individuals. A minimum number of two identical, raw reads were required to create a stack, and a maximum number of seven stacks were allowed at a single locus.

For the **ptarmigan and capercaillie**, Trimmomatic (Bolger *et al*., 2014) was used to remove the adapter sequences from the samples sequenced at Cornell University, while gbstrim (Garbe, 2019) was used to trim the adapter sequences from the dataset processed by the University of Minnesota. Different programs were used because Trimmomatic did not recognise the adapters for the University of Minnesota dataset as additional adapters are added to the standard adapters for better read quality. The outcome for both programmes was, therefore, the same.

The *denovo_map.pl* wrapper pipeline was used for both of these species in Stacks version 2.0Beta8c (Catchen *et al*., 2013) with the deleveraging algorithm and gapped alignment disabled, a max distance of two allowed between stacks, a maximum of three stacks allowed per locus and one mismatch allowed between stacks (Figure 2.3).

For all species, the *populations* pipeline in Stacks was used to generate the input file for VCFtools version 0.1.15 (Danecek *et al*., 2011). VCFtools was used to further filter the data to obtain SNPs which shared between at least 80% of within-species individuals for the black grouse and ptarmigan, and 70% of individuals for the capercaillie, and had a minimum depth greater than or equal to 15X. For the black grouse, a Minor Allele Frequency greater or equal to 0.02 was used, while for the ptarmigan and capercaillie Minor Allele Counts of 3 and 2 were used for filtering, respectively. Individuals with over 75% missing data and average coverage below 4X were also removed with VCFtools for the black grouse, while 70% missing data and 15X coverage were used as the thresholds for the ptarmigan and capercaillie. Different filtering options were tested in order to optimise the number and quality of the SNPs produced for each species (Figure 2.3).

A


GBS SNP data from 400 black grouse
generated by NOVOGENE
blood: 25, feather: 120, tissue: 255
stacks-2.52
process_radtags
-e msel
-r
-c
-q
-E phred 33
-t 140
stacks-2.52
ustacks
---max_locus_stacks 7
-m 2
-H
-d
-M 7
stacks-2.52
cstacks
-n 7
stacks-2.52
sstacks
--gapped
stacks-2.52
tsv2bam
stacks-2.52
gstacks
stacks-2.52
populations
-r 0.6
--vcf
VCFtools
--max-missing 0.8
--minDP 15
--maf 0.02
Samples removed with:
Missing data over 75%
Coverage under 4X
Final black grouse GBS SNP dataset
28752 SNPs
383 individuals
blood: 20, feather: 116, tissue: 247

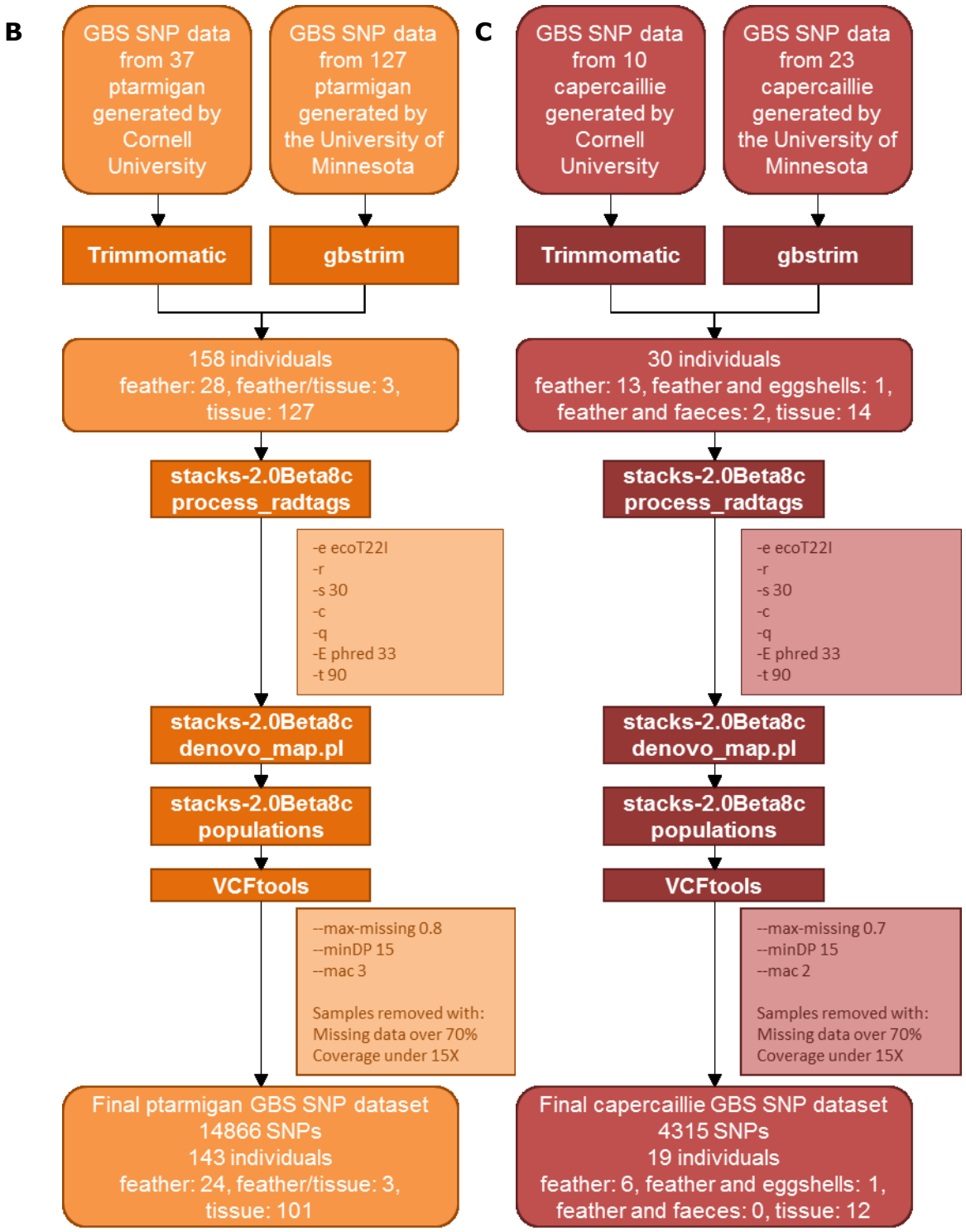


Figure 2.3 – Final GBS Pipelines used for the black grouse (**A**), ptarmigan (**B**) and capercaillie (**C**) showing the software package and step in the pipeline shown in **bold** in the dark coloured rectangle. The parameters used in Stacks (Catchen *et al.,* 2013) and VCFtools (Danecek *et al.,* 2011) are shown in pale rectangles adjacent to the corresponding stage. When no parameters are shown, default parameters were used. The rounded rectangles detail the dataset used, the number of individuals, the breakdown of sample types and the number of SNPs produced.

2.4. POPULATION GENETIC ANALYSIS

The following analyses were performed on **all three species** for both STR and SNPs data unless otherwise stated. A summary of the software, conditions, parameters and source for each analysis is listed in Table 2.3.

Duplicate individuals and samples with no known locality were removed from some analyses where this information was required. Summary statistics (expected heterozygosity (H_E), gene diversity (calculated at the haplotype level), fixation index (F_{ST}), mean number of alleles and allelic size range, were calculated using Arlequin (Excoffier and Lischer, 2010).

For **ptarmigan and capercaillie**, the distribution of mtDNA haplotypes between populations was calculated using a Median Joining Network (Bandelt *et al.*, 1999) and visualized using PopART (Leigh and Bryant, 2015).

A Principal Component Analysis (PCA) was performed to examine the genetic structure of the SNP and STR data; however, as a PCA requires clusters to be defined *a priori,* a Discriminant Analysis of Principal Components (DAPC) was also performed to investigate other possible group definitions within the data. Both the PCA and DAPC were performed with the R package *adegenet* (Jombart, 2008; R Core Team, 2018), developed for multivariate analysis of genetic markers. The summary statistics and PCA were also repeated for each temporal dataset for the black grouse samples collected from PAT.

STRUCTURE (Pritchard et al., 2000) was used to further investigate the population genetic structure of each species using a 10,000 step burn-in period and 20,000 MCMC iterations after the burn-in. The number of assumed populations (K) varied by species, however for all analyses ten replicate runs were performed for each value of K. For the **black grouse** SNP data, K was set from 1 to 30, for the black grouse STR data, Ks 1 to 15 were tested, this was reduced due to the reduction in populations present for this dataset. Both the SNP and STR values of K were set to 1-20 for the **ptarmigan** and 1-5 for the **capercaillie**. The results for each STRUCTURE run were merged using CLUMPP (Jakobsson and Rosenberg, 2007) and the most supported K for the data was estimated using the Evanno method (Evanno et al, 2005) and STRUCTURE HARVESTER (Earl and vonHoldt, 2012).

An Analysis of MOlecular VAriance (AMOVA), run with 1,000 permutations, was used to examine hierarchal population structure and confirm the results of the previous analyses. The **black grouse** and **ptarmigan** populations were split into

two groups at the border between SOIUSA mountain groups LEP-LUG and BER-ROC (Figure 2.1), effectively dividing the samples between the 'eastern' and 'western' mountain groups, based on the results shown for the black grouse DAPC and STRUCTURE. For the **capercaillie**, samples were divided into two groups east and west of the Adige Valley between SOIUSA mountain groups RME-BRG and DOL-VEN (Figure 2.1) since previous population genetic patterns of several mammalian herbivores in the same area have indicated that this valley is an effective barrier to gene flow in species with low dispersal capability (see Vernesi *et al*., 2016).

Effective population size (N_e) for each population, as well as an overall estimate, was calculated using the molecular co-ancestry method as described by Nomura (2008) using NeEstimator (Do *et al.,* 2014) with the linkage disequilibrium model (Waples and Do, 2008) under a random mating system and a 0.05 critical value.

Mantel tests, conducted using the R package *vegan* (Oksanen *et al*., 2019), were used to determine if either STR or SNP data followed a pattern of isolation by distance (IBD). The linearized F_{ST} calculated with Arlequin between pairs of populations was compared to the logarithm of the Euclidian distance calculated between coordinates for these locations with the *raster* package (Hijmans et al., 2019) in R.

The Mantel test was also used to assess the effect of altitude and urban areas as a barrier to movement for the **black grouse**. These isolation by resistance (IBR) tests were conducted using the F_{ST}s and coordinates used in the IBD test with the geographical distances calculated using the R *gdistance* package (van Etten, 2017) and the random walk method. Since this tetraonid species originated from areas where dispersal entails navigating mountainous terrain, this method was chosen to measure geographic distance taking into account all possible routes between two points, rather than a straight line. Both altitude and urban areas were tested as possible barriers to movement, with higher values of resistance assigned to altitudes outside of the black grouse's typical range (over 2,500m a.s.l.) and more densely populated areas. Values of resistance were adjusted to examine their effect on the value of Mantel's R produced. The actual geographic distances between populations were then calculated taking these barriers into account, to give a more realistic measure of dispersal distance between two sites for this species.

The software Estimated Effective Migration Surfaces (EEMS; Petkova *et al*., 2016), which allows the visualisation of changes in effective migration rates across an area,

was used to identify possible barriers to gene flow between populations in the Italian Alps. The pairwise dissimilarity matrix used for the SNP data was calculated in Arlequin. Three independent analyses were run for both the SNP and STR data with 3,000 and 1,500 demes respectively, and both analyses used a 1,000,000 burn-in and 2,000,000 iterations. The results were combined for each marker and visualised using the R package rEEMSplots (Petkova, 2016).

A redundancy analysis (RDA) was used to examine the variation in response to various environmental factors using the R package vegan (Oksanen *et al*., 2019). This was performed on the **black grouse SNP** data with individuals collected with coordinates, examining the effect of longitude, latitude, year, altitude, human population density, land cover type, mean rainfall in June and mean temperature in June. The **black grouse** SNP dataset was also used to examine if outlier loci, which may be involved in local adaption, could be detected. The detection of outliers was performed with PCAdapt (Luu *et al.,* 2019). Three methods were used to test for the detection of outliers; q-values, the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995) and the more conservative Bonferroni correction (Dunn, 1961).

Table 2.3 – Software and, where applicable, the conditions used for each analysis. Unless stated otherwise, each analysis was performed for all three species for both STR and SNP datasets. For the motivation behind the use of each analysis, please see the above text.

Analysis	Software and Conditions	Reference
Summary Statistics *Including mtDNA*	Arlequin	Excoffier and Lischer, 2010
Haplotype Network *mtDNA only*	PopART Median-Joining Network	Leigh and Bryant, 2015 Bandelt *et al*., 1999
Principal Component Analysis (PCA)	R package: adegenet	Jombart, 2008
Discriminant Analysis of Principal Components (DAPC)	R package: adegenet	Jombart, 2008

Population Structure	STRUCTURE: 10,000 burn-in, 20,000 MCMC, 10 replicates CLUMPP Evanno STRUCTURE HARVESTER	Pritchard *et al.*, 2000 Jakobsson and Rosenberg, 2007 Evanno *et al*, 2005 Earl and vonHoldt, 2012
AMOVA	Arlequin	Excoffier and Lischer, 2010
Effective Population Size	NeEstimator Model: linkage disequilibrium Critical value: 0.05	Do et al., 2014 Waples and Do, 2008
Genetic Distances	Arlequin	Excoffier and Lischer, 2010
Euclidian Distances	R package: raster	Hijmans *et al.,* 2019
Corrected Geographical Distances *Black grouse only*	R package: gdistance Random walks method	van Etten, 2017
Mantel test	R package: vegan	Oksanen *et al*., 2019
Estimated Effective Migration Surfaces	EEMS: 3,000/1,500 demes, 1,000,000 burn-in, 2,000,000 iterations rEEMSplots	Petkova *et al*., 2016
RDA *Black grouse SNP only*	R package: vegan	Oksanen *et al*., 2019
Outlier Detection *Black grouse SNP only*	PCAdapt	Luu *et al.,* (2019)

3. RESULTS

3.1. BLACK GROUSE

3.1.1. SAMPLE COLLECTION

A total of 628 black grouse samples were collected over the 20-year sampling period (Figure 3.1.1, Table 3.1.1). Each sample was assigned to a SOIUSA mountain group ('population') using the coordinates provided by the collector. In cases where coordinates were not available (N=18), samples were assigned to populations based on the hunting area where the animal was bagged. Only 11 individuals could not be assigned to a mountain group. Samples collected from across the study area were used in the SNP analysis; however, for the STR dataset, no samples were used from populations in the centre of the study area (LUG-ROC). Trentino-Alto Adige samples for both datasets were also divided into three time-frames (Figure 3.1.2, Table 3.1.2).

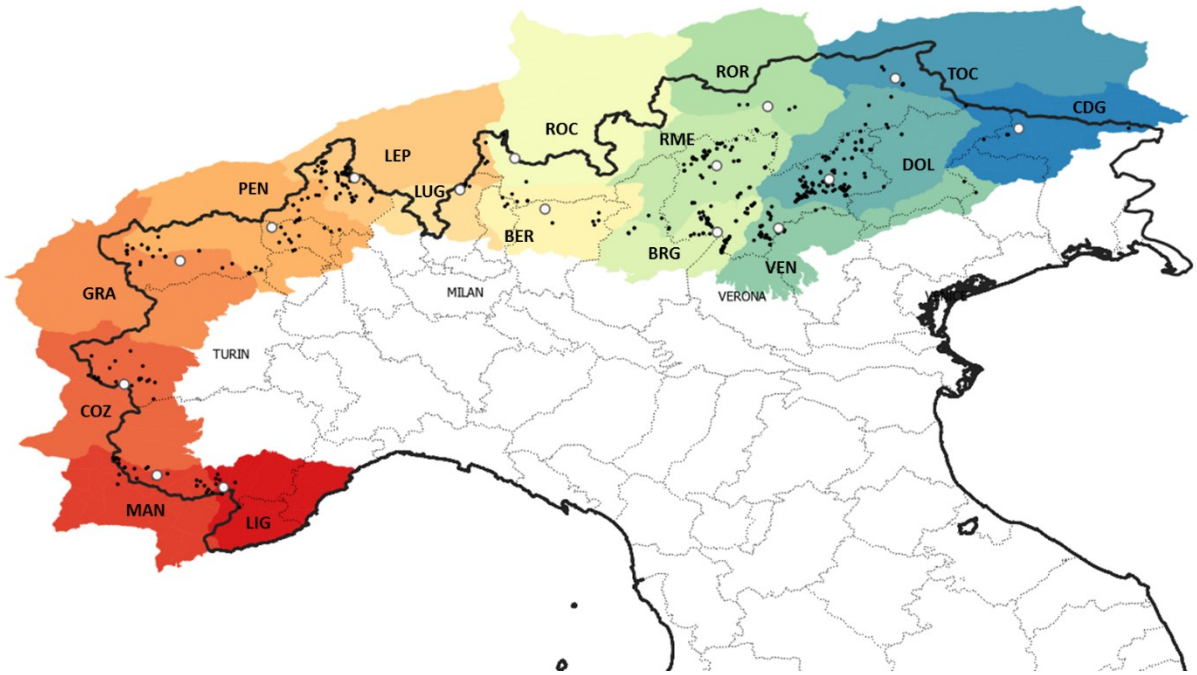


Figure 3.1.1 – Map of northern Italy showing the black grouse sample locations. Each point represents a single sample; white dots represent the mean centre of each population. The thick lines indicate the Italian borders, while the thin lines indicate the Italian provinces. Each coloured area represents a SOIUSA mountain group (see Table 3.1.1 and Figure 2.1). Map created in QGIS (QGIS Development Team, 2021).

Table 3.1.1 – Black grouse samples collected in the Italian Alps, showing the number collected for each population and those analysed to generate Single Tandem Repeats (or Microsatellites: STR) and Single Nucleotide Polymorphism (SNP) genetic data.

SOIUSA Group (Population)	Abbreviation	N. Samples	SNP	STR
Alpi Liguri i.s.a.	**LIG**	9	6	8
Alpi Marittime e Prealpi di Nizza	**MAN**	24	16	21
Alpi Cozie	**COZ**	26	21	25
Alpi Graie	**GRA**	19	15	14
Alpi Pennine	**PEN**	34	25	15
Alpi Lepontine	**LEP**	75	31	55
Prealpi Luganesi	**LUG**	7	7	0
Alpi e Prealpi Bergamasche	**BER**	13	13	0
Alpi Retiche occidentali	**ROC**	2	2	0
Prealpi Bresciane e Gardesane	**BRG**	93	71	88
Alpi Retiche meridionali	**RME**	77	48	74
Alpi Retiche orientali	**ROR**	4	4	0
Prealpi Venete	**VEN**	51	47	44
Dolomiti	**DOL**	172	76	155
Alpi dei Tauri occidentali	**TOC**	6	6	0
Alpi Carniche e della Gail	**CDG**	5	5	0
Undefined Mountain Group	**UNK**	11	7	8
Totals		**628**	**400**	**507**

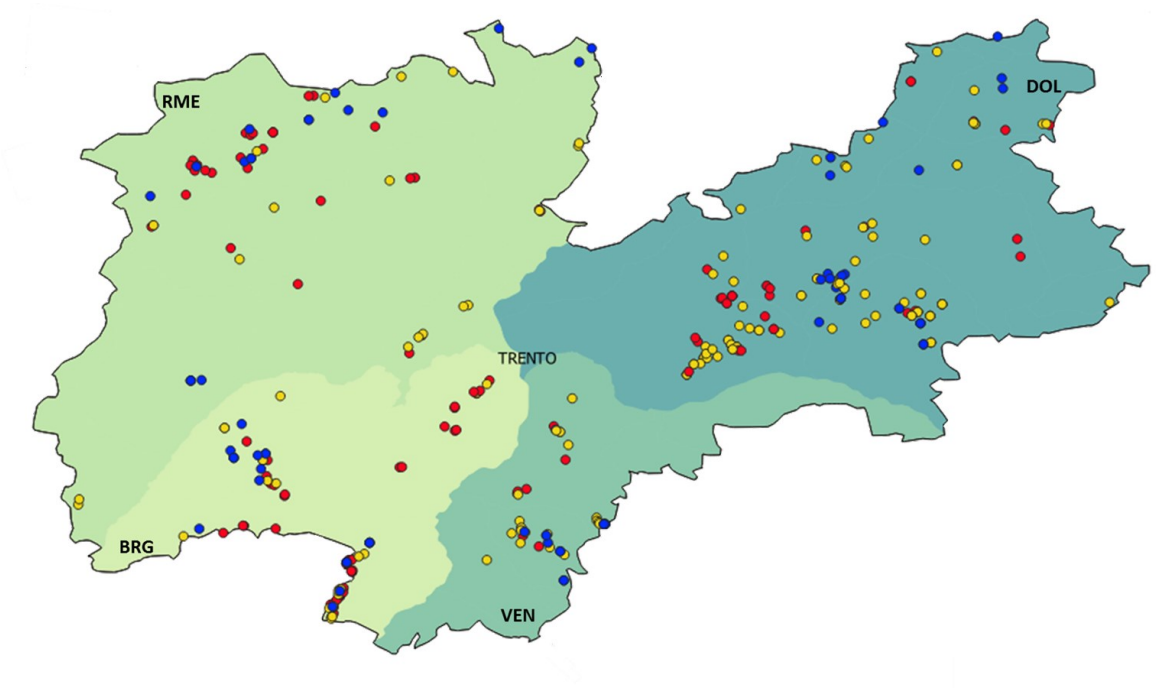


Figure 3.1.2 – The Autonomous Province of Trento showing black grouse samples available for three different time frames (1995-1999: red; 2009-2010: yellow and 2015-2017: blue) and in each area for both SNPs and STRs. Map created in QGIS (QGIS Development Team, 2021).

Table 3.1.2 – The number of black grouse samples represented in each period for each genetic marker in the Autonomous Province of Trento.

SOIUSA Group	Total n. samples		1995-1999		2009-2010		2015-2017	
	SNP	STR	SNP	STR	SNP	STR	SNP	STR
Prealpi Bresciane e Gardesane (BRG)	62	88	22	44	20	24	20	20
Alpi Retiche meridionali (RME)	45	74	16	35	12	22	17	17
Prealpi Venete (VEN)	40	44	10	11	18	21	12	12
Dolomiti (DOL)	64	155	18	40	25	102	21	13
Totals	**211**	**361**	**66**	**130**	**75**	**169**	**70**	**62**

3.1.2. INDIVIDUAL GENOTYPING

Ten STRs were characterised for a total of 507 black grouse. One individual was removed after preliminary analysis suggested it was an outlier for each of the analyses, probably due to a misidentification of the species. The overall level of missingness across all loci was 0.1% and no locus had a missingness level greater than 1% (range 0-0.99%). The mean total H_E was 0.74 (SD ±0.16), ranging from 0.6 in LIG to 0.75 in PEN. The mean number of alleles for each population ranged from 4.1 for LIG to 8.7 for DOL (Table 3.1.3). Significant F_{ST} results ($P \leq 0.05$) ranged from 0.01 between BRG and DOL to 0.16 between GRA and VEN (Table 3.1.4a).

For the SNP analysis, after removing two individuals for which no sequencing data was produced, the Stacks pipeline produced over 500,000 SNPs for 398 individuals, with a mean coverage of 9.8X. A further 15 individuals were removed from the SNP dataset because of high levels of missing data (> 75%) and low coverages (< 4X coverage). After filtering in VCFtools, the final SNP dataset consisted of 383 individuals with 28,752 SNPs, with a mean missingness level of 16.83% and a mean coverage of 27.1X. Mean population H_E ranged from 0.21 in LIG to 0.25 for ROR, with an overall mean H_E of 0.26 (SD ±0.16) for the pooled dataset. F_{ST} ranged from 0 between GRA and PEN to 0.12 between LIG and BRG (Table 3.1.4b).

Table 3.1.3 – Summary statistics calculated for both the STR and SNP datasets, showing expected heterozygosity (H_E), and the mean number of alleles and size range for the STR dataset.

	H_E		Number of alleles (STR only)	Allele Size Range (STR only)
	STR	SNP		
LIG	0.60	0.21	4.1	6.222
MAN	0.65	0.22	5.5	8.556
COZ	0.71	0.23	6.7	8.7
GRA	0.71	0.24	5.6	6.4
PEN	0.75	0.25	6.6	9
LEP	0.72	0.25	7.7	9.4
LUG	-	0.24	-	-
BER	-	0.25	-	-
ROC	-	0.25	-	-
BRG	0.71	0.25	7.8	9.6
RME	0.73	0.25	7.4	9.2
ROR	-	0.25	-	-
VEN	0.71	0.25	6.4	8.4
DOL	0.73	0.25	8.7	10.1
TOC	-	0.24	-	-
CDG	-	0.24	-	-
UNK	0.75	0.26	5.4	8
Totals	**0.74**	**0.26**	**11**	**12**

Table 3.1.4 – Values of F_{ST} shown below the diagonal with p-values shown above for the STR (**A**) and SNP (**B**) datasets. Significant values of F_{ST} are shown in bold.

A	LIG	MAN	COZ	GRA	PEN	LEP	BRG	RME	VEN	DOL	UNK
LIG	-	0.468	0.108	0.000	0.216	0.595	0.027	0.009	0.018	0.072	0.180
MAN	0.006	-	0.351	0.063	0.189	0.252	0.027	0.000	0.009	0.009	0.090
COZ	0.033	0.004	-	0.144	0.405	0.234	0.000	0.000	0.000	0.000	0.009
GRA	**0.119**	0.046	0.015	-	0.063	0.009	0.000	0.000	0.000	0.000	0.000
PEN	0.019	0.020	0.002	0.029	-	0.432	0.072	0.000	0.054	0.036	0.324
LEP	-0.014	0.010	0.006	**0.041**	0.000	-	0.000	0.000	0.000	0.000	0.099
BRG	**0.061**	**0.033**	**0.060**	**0.121**	0.032	**0.050**	-	0.018	0.018	0.000	0.270
RME	**0.119**	**0.052**	**0.094**	**0.125**	**0.072**	**0.091**	**0.017**	-	0.000	0.000	0.036
VEN	**0.076**	**0.077**	**0.111**	**0.157**	0.049	**0.083**	**0.019**	**0.037**	-	0.108	0.622
DOL	0.025	**0.045**	**0.071**	**0.120**	**0.033**	**0.048**	**0.012**	**0.044**	0.005	-	0.784
UNK	0.038	0.089	**0.089**	**0.184**	0.007	0.045	0.018	0.084	-0.013	-0.019	-

B	LIG	MAN	COZ	GRA	PEN	LEP	LUG	BER	ROC	BRG	RME	ROR	VEN	DOL	TOC	CDG	UNK
LIG	-	0.865	0.000	0.000	0.000	0.000	0.000	0.000	0.018	0.000	0.000	0.009	0.000	0.000	0.009	0.000	0.000
MAN	0.000	-	0.000	0.000	0.000	0.000	0.000	0.000	0.036	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
COZ	**0.019**	**0.014**	-	0.000	0.000	0.000	0.000	0.000	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
GRA	**0.063**	**0.033**	**0.018**	-	0.036	0.432	0.000	0.000	0.018	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.018
PEN	**0.066**	**0.057**	**0.039**	**0.000**	-	0.000	0.000	0.000	0.027	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.135
LEP	**0.056**	**0.066**	**0.052**	0.000	**0.002**	-	0.000	0.000	0.550	0.000	0.000	0.027	0.000	0.000	0.000	0.000	0.838
LUG	**0.081**	**0.088**	**0.074**	**0.027**	**0.031**	**0.012**	-	0.000	0.324	0.000	0.000	0.018	0.000	0.000	0.000	0.009	0.045
BER	**0.088**	**0.090**	**0.077**	**0.035**	**0.041**	**0.027**	**0.025**	-	0.072	0.000	0.000	0.009	0.000	0.000	0.000	0.000	0.090
ROC	**0.106**	**0.085**	**0.071**	**0.051**	**0.036**	0.002	0.001	0.012	-	0.036	0.027	0.081	0.009	0.009	0.045	0.036	0.144
BRG	**0.118**	**0.109**	**0.095**	**0.057**	**0.058**	**0.040**	**0.052**	**0.033**	**0.052**	-	0.000	0.000	0.000	0.000	0.000	0.000	0.000
RME	**0.108**	**0.100**	**0.086**	**0.050**	**0.050**	**0.032**	**0.043**	**0.024**	**0.043**	**0.015**	-	0.009	0.000	0.000	0.000	0.000	0.018
ROR	**0.103**	**0.081**	**0.065**	**0.048**	**0.033**	**0.011**	**0.024**	**0.015**	0.003	**0.016**	**0.009**	-	0.000	0.018	0.072	0.045	0.072
VEN	**0.109**	**0.107**	**0.092**	**0.052**	**0.057**	**0.045**	**0.054**	**0.040**	**0.048**	**0.010**	**0.019**	**0.012**	-	0.000	0.000	0.000	0.000
DOL	**0.107**	**0.104**	**0.089**	**0.049**	**0.053**	**0.041**	**0.050**	**0.037**	**0.047**	**0.018**	**0.018**	**0.008**	**0.012**	-	0.000	0.000	0.027
TOC	**0.108**	**0.109**	**0.094**	**0.058**	**0.062**	**0.050**	**0.051**	**0.046**	**0.028**	**0.036**	**0.033**	0.000	**0.033**	**0.025**	-	0.009	0.063
CDG	**0.113**	**0.106**	**0.094**	**0.068**	**0.065**	**0.044**	**0.049**	**0.043**	**0.038**	**0.038**	**0.036**	**0.012**	**0.030**	**0.018**	**0.023**	-	0.009
UNK	**0.063**	**0.044**	**0.026**	**0.003**	0.000	0.000	**0.004**	0.002	0.000	**0.010**	**0.007**	0.000	**0.007**	**0.005**	0.006	**0.012**	-

3.1.3. POPULATION GENETIC ANALYSIS

The PCA scatterplot (Figure 3.1.3A) revealed a weak pattern of genetic structure for the STR data, with overlapping clusters of individuals from each population and no outlying groups. However, individuals from the ‘western’ populations (LIG-LEP; Figure 3.1.1) were closer to the righthand side of the scatterplot, while individuals from the ‘eastern’ populations (BRG-DOL) were closer to the left. The DAPC analysis of STRs (Figure 3.1.4A) gave a similar result, where cluster 5 was primarily represented in ‘eastern’ populations BRG to DOL, while cluster 4 was mainly present in ‘western’ populations LIG to LEP.

The PCA of the SNP data (Figure 3.1.3B) showed a more definite population structure so that individuals within populations cluster closer together, and there were several cluster groups: LIG to COZ, GRA to ROC, and BRG to CDG. In the DAPC, individuals were split between two clusters, ‘western’ LIG to LUG (primarily represented by cluster 2; Figure 3.1.4B) and ‘eastern’ BER to CDG (cluster 1). The ROC population for this analysis contained two individuals, each representing a different cluster. However, despite these patterns, the low values of variance for the first two PCA axes (PC1 and PC2) for both datasets imply low levels of diversity between the clusters.

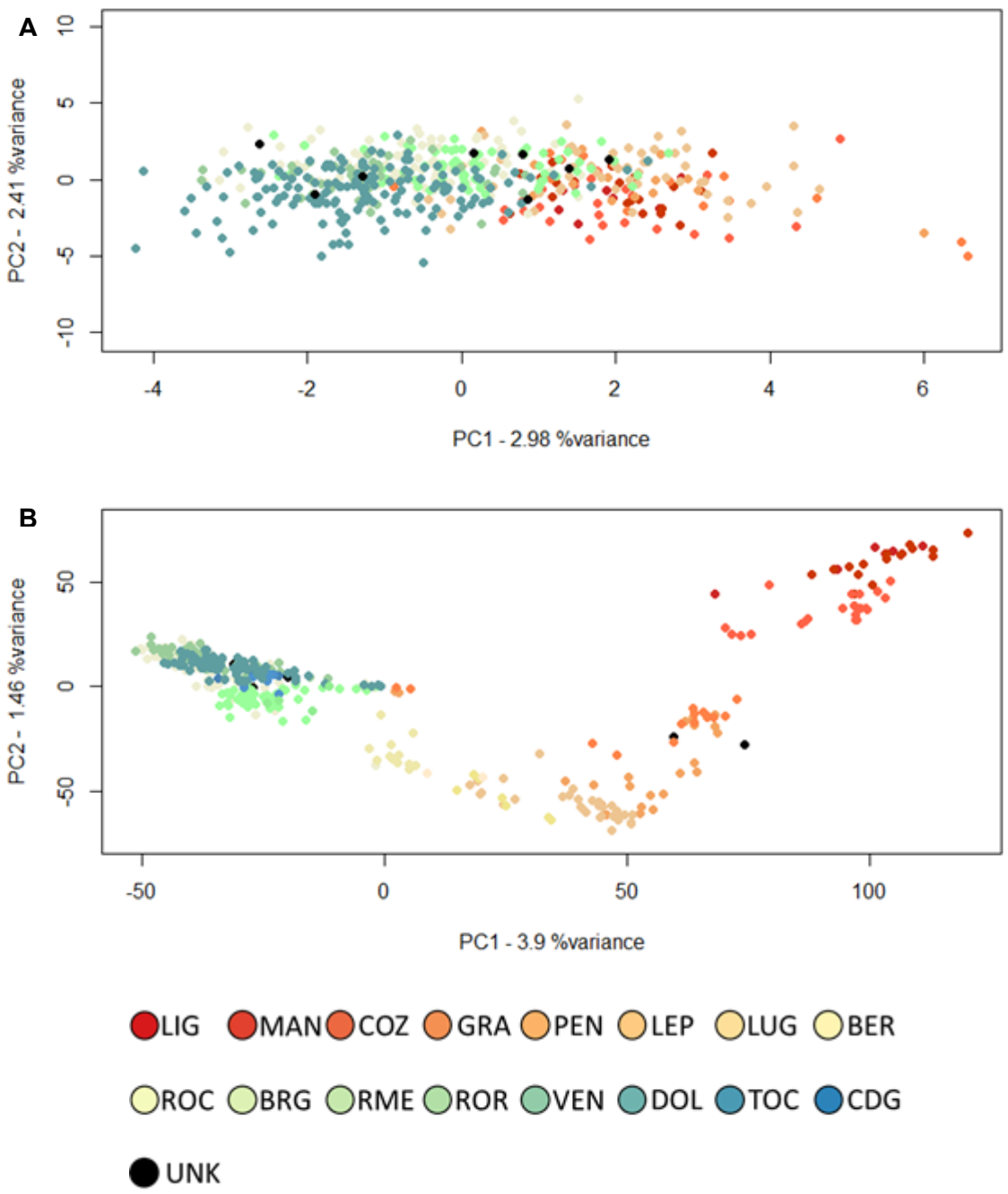


Figure 3.1.3 – Principal Component Analysis (PCA) for the black grouse. **A** STR data (n=506); **B** SNP data (n=383). Each colour represents a different mountain group (population) as defined in Figure 3.1.1.

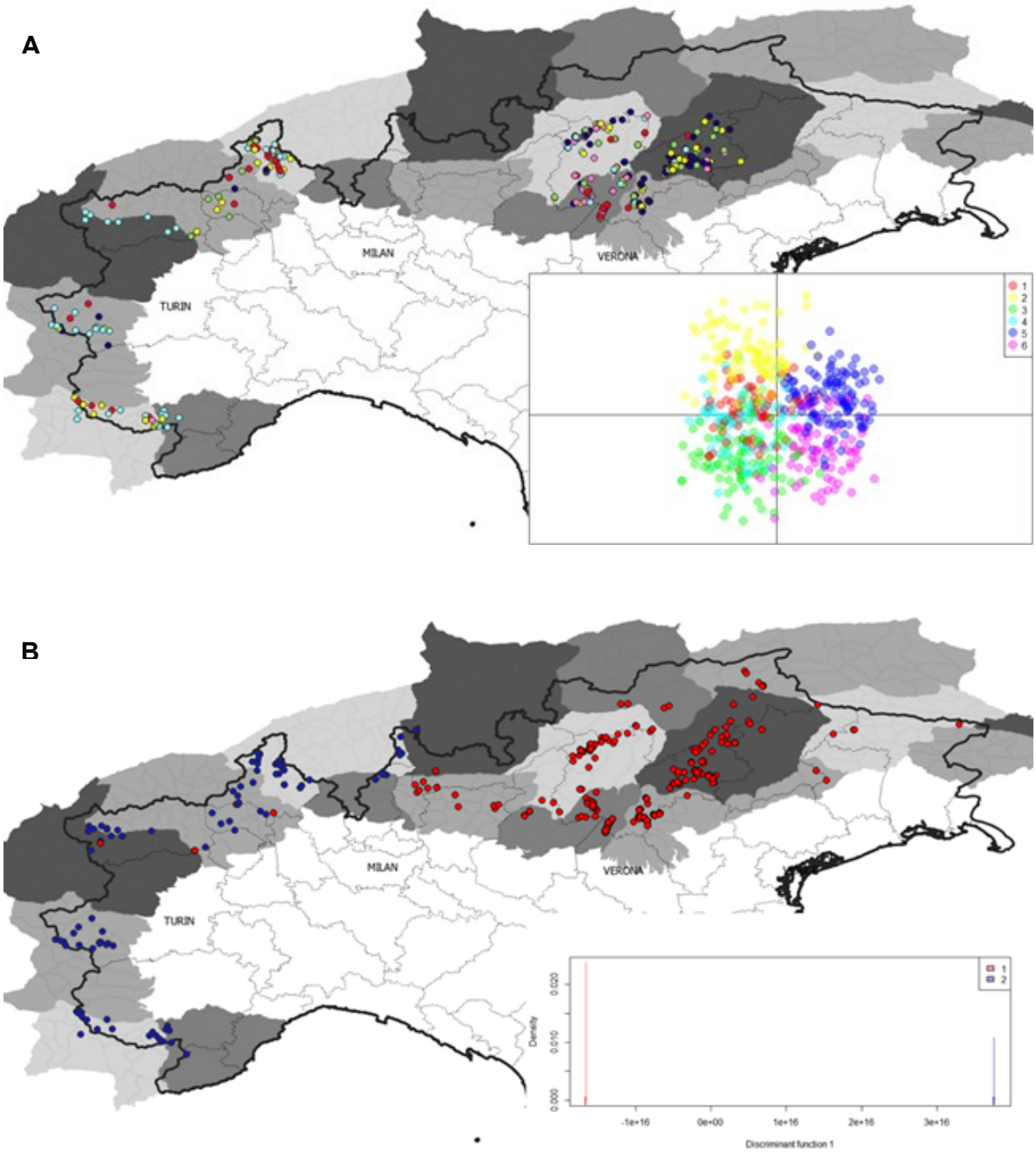


Figure 3.1.4 – Discriminant Analysis of Principal Components (DAPC) for the Black Grouse: **A** STR data (6 clusters, n=506); **B** SNP data (2 clusters, n=383). The DAPC plot is shown in the lower right corner of each map and the maps show the geographical representation of the clusters. Maps created in QGIS (QGIS Development Team, 2021).

The results of STRUCTURE and Structure Harvester suggested that the most supported number of genetic clusters for both the STR and SNP datasets was K=2 (delta K = 31.41 and 2928.81 respectively). To investigate any underlying patterns within the STRUCTURE results, the plot for K=3 was also examined. While all individuals were admixed between both clusters, the STR STRUCTURE plot (Figure

3.1.5A) shows the first cluster (green) primarily represented populations BRG to DOL (‘eastern populations’) and the second cluster (yellow) represents populations LIG to LEP (‘western’ populations); however, the RME population (in the central Alps) was admixed, with individuals representing both clusters. For K=3, BRG is predominantly represented by the third cluster (purple) and other ‘eastern’ samples are admixed with the first and third clusters; RME has the greatest proportion of cluster 2 (yellow) in these populations.

It should be noted that one of the clusters for the SNP K=2 STRUCTURE results (shown in green in Figure 3.1.5B) does not represent a single population; instead, the individuals represented by this cluster are those with the highest proportion of missing data. The yellow cluster for this plot represents the remaining individuals as a single continuous population. In the K=3 plot the samples with high levels of missing data are again represented by the first cluster (green), the second (yellow) cluster now represented more by populations LIG to ROC, while the third (purple) represented populations BRG to CDG. However, as with the STR data, all individuals have some level of admixture with other clusters. Even when samples with greater than 20, 30, 40, 50 and 60% missing data were removed from the dataset each analysis found that individuals with high levels of missing data continued to cluster separately at K=2 and other K values (data not shown).

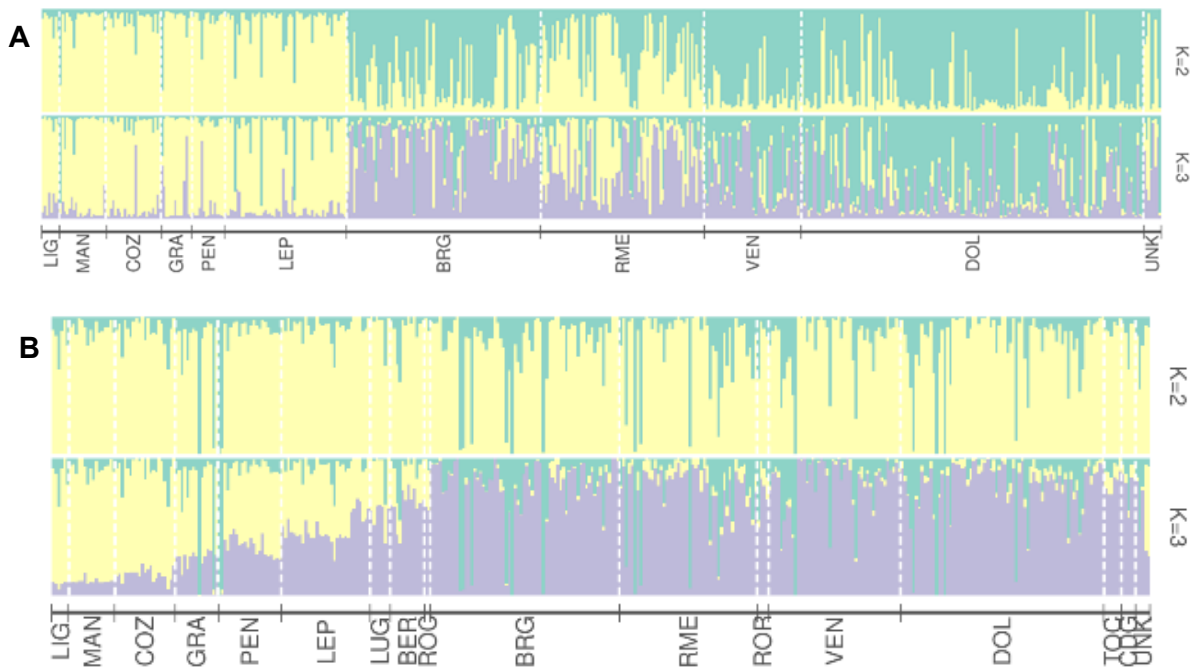


Figure 3.1.5 – STRUCTURE plots showing plots for K=2 and K=3 for the STR data (**A**, n=506) and SNP dataset (**B**, n=383). White dashed lines separate populations. Plots were visualised in Pophelper version 1.0.10 (Francis, 2016).

The AMOVA analysis (Table 3.1.5) was used to investigate the potential 'western' (LIG to LUG) and 'eastern' (ROC to CDG) division of populations (referred to as groups for the AMOVA) as shown in the results of the STRUCTURE and DAPC analyses (see also Figure 3.1.4B). The AMOVA showed that the highest proportion of variation for both datasets was found within individuals (93.15% and 94.50% for the STR and SNP datasets, respectively). However, this result for the STR dataset was not significant at a threshold of p=0.05. The second-largest proportion of variation for both datasets was among groups (STR: 4.97% and SNP: 4.00%) followed by among populations within groups (STR: 1.79% and SNP: 2.67%). Among individuals within populations had the lowest amount of variation with 0.092% in the STR dataset, with a non-significant result for the SNP dataset.

Effective population size (N_e) varied greatly across the populations for both markers (Table 3.1.6). For both markers, N_e was highest in PEN (STR: 9465.6 and SNP: 763.1), while the lowest estimation of N_e was 12.9 (GRA) and 49 (LUG) for STRs and SNPs, respectively. Estimations of N_e for the species as a whole was 234.7 (STRs) and 121.8 (SNPs). In general, the confidence intervals (CI) for the SNP dataset were narrower than those for STRs.

Table 3.1.5 – AMOVA results for the STR and SNP data. Genetic structure among groups (Va/FCT), among groups within populations (Vb/FSC), among individuals within populations (Vc/FIS), and within individuals (Vd/FIT) estimated using Arlequin (Version 3.5).

Marker	Variation Source	Sum of Squares	Variance Components	Percentage of Variation	F Index	P-value
STR	Among groups	1819.26	3.99	4.97	0.050	0.0020
	Among populations within groups	1620.81	1.44	1.79	0.019	0
	Among individuals within populations	36352.69	0.074	0.092	0.00099	0.49
	Within individuals	37019.50	74.78	93.15	0.069	0.087
	Total	76812.25	80.28			
	Among groups	50611.05	150.90	4.00	0.040	0
	Among populations within groups	99718.55	100.89	2.67	0.028	0
	Among individuals within populations	1039602.09	-44.14	-1.17	-0.013	1
	Within individuals	1123936.50	3567.06	94.50	0.055	0
	Total	2313868.19	3774.72			

Table 3.1.6 – Effective population size (N_e) estimates for the black grouse from STR and SNP datasets with 95% confidence intervals (CI). Inf: infinite.

Population	STR		SNP	
	N_e	95% CI	N_e	95% CI
LIG	Inf	12.6-Inf	Inf	Inf
MAN	Inf	54.6-Inf	Inf	Inf
COZ	45.6	25.2-136.4	570.8	555.6-586.9
GRA	12.9	7.3-27.3	73.1	72.6-73.7
PEN	9465.6	38.6-Inf	763.1	738-789.9
LEP	136.2	75.6-441	365.7	362.3-369.3
LUG	-	-	49	48.5-49.4
BER	-	-	Inf	Inf
ROC	-	-	Inf	Inf
BRG	131	84.2-251	169.2	168.8-169.7
RME	152.8	89.2-397.3	95.3	95.1-95.5
ROR	-	-	Inf	Inf
VEN	142.8	65-Inf	179.1	178.4-179.7
DOL	79.5	64.8-99.3	315.8	314.6-317
TOC	-	-	108.1	105.3-111
CDG	-	-	Inf	Inf
UNK	480.8	13.9-Inf	Inf	Inf
Overall	**234.7**	**193-289.7**	**121.8**	**121.7-121.8**

The Mantel tests for both the STR and SNP datasets showed strong and significant patterns of isolation by distance (IBD) when the logarithm of the great-circle distance and the linearized F_{ST}s were compared (STR dataset: R=0.54, p<0.01, n=498; SNP dataset: R=0.82, p<0.01, n=378). The isolation by resistance (IBR) results were significant for both altitude (STR dataset: R=0.37, p<0.02; SNP dataset: R=0.84, p<0.01) and urban areas STR dataset: R=0.64, p<0.01; SNP dataset: R=0.8, p<0.01) using the resistance values outlined in Table 3.1.7 and Figure 3.1.6. For these analyses, negative and non-significant values of F_{ST} were replaced with values of 0. When the resistance values were adjusted, the R values continued to show a significant pattern of IBR.

Table 3.1.7 – Environmental variables tested in the isolation by resistance (IBR) analysis. The resistance values assigned to each variable and the ecological justification for this assignment are listed. The criteria for altitude ranges were defined based on the findings of Caizergues *et al.* (2003a) and Bocca (2004a).

Environmental Factor	Criteria	Resistance Values	Ecological Justification
Altitude	<900m	2	Atypical, occasional range
	≥900m <1800m	1	Typical range
	≥1800m ≤2500m	5	Rarely reported at this range
	>2500m	10	Unsuitable habitat
Urbanisation	Urban Areas	10	Unsuitable habitat; avoided
	Rural Areas	1	Typical habitat

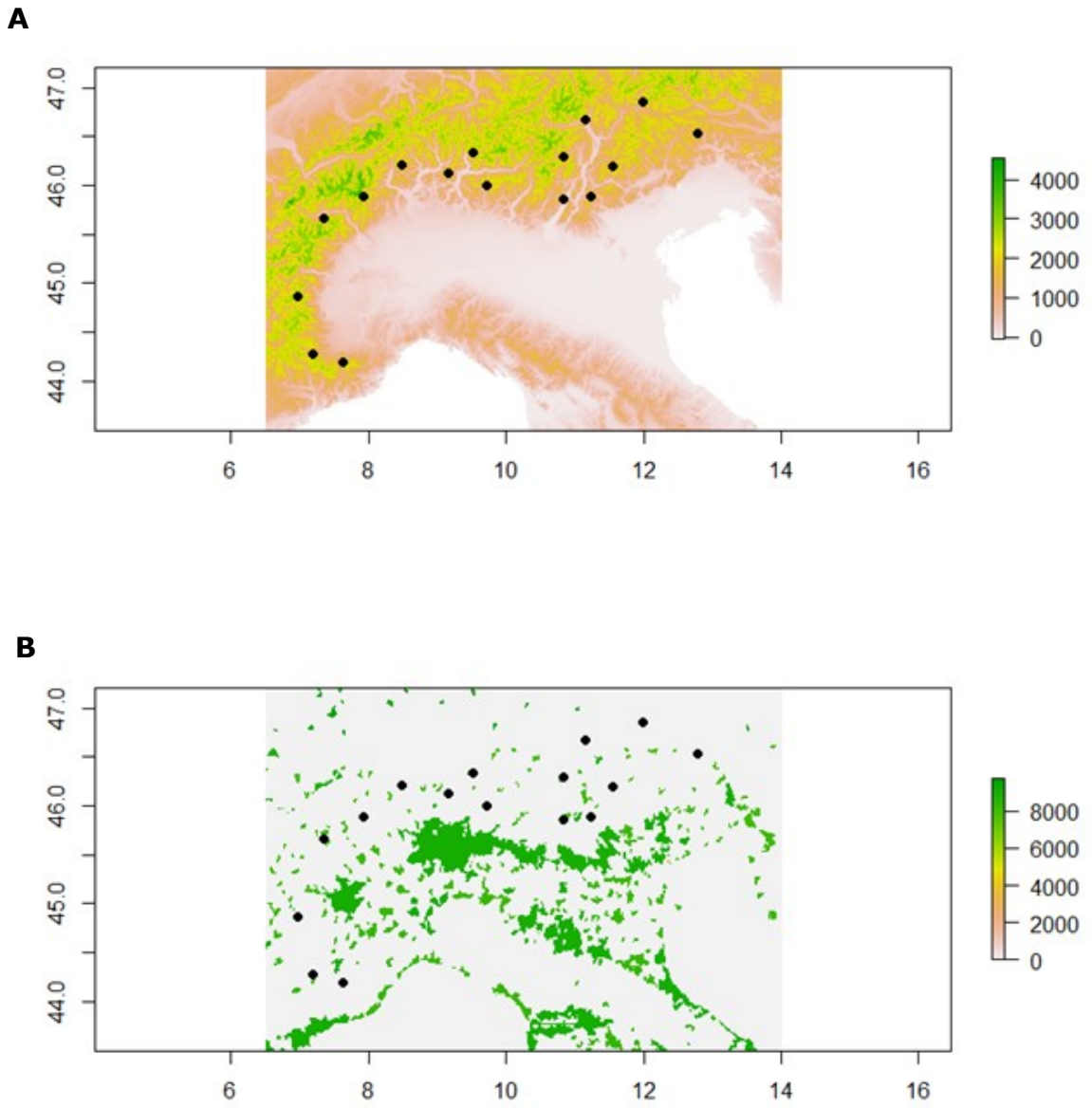
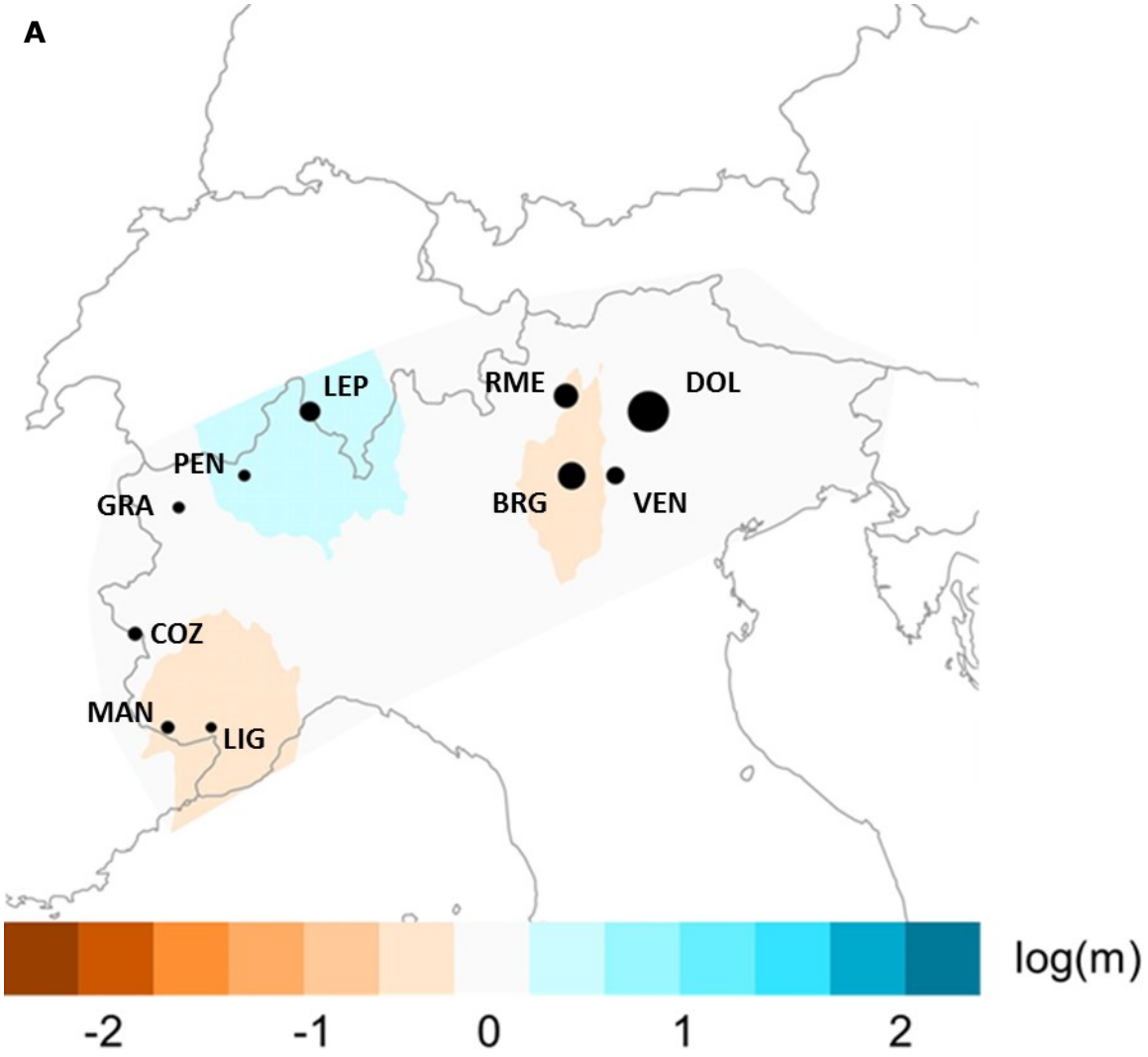


Figure 3.1.6 – Isolation by resistance maps for altitude (A) and urban areas (B). Points represent the mean centre of each mountain group population.

For the STR dataset, the EEMS (Figure 3.1.7A) showed that there were low levels of migration in both the LIG and MAN populations, as well as the Trentino-Alto Adige populations west of the Adige River (BRG and RME), while PEN and LEP show greater rates of migration in comparison to the others. The SNP dataset (Figure 3.1.7B) confirmed the low levels of migration rates for the southwestern populations in the study area (LIG, MAN and COZ). In addition, the SNP results show lower migration rates between LUG and ROC (coinciding with the 'eastern' and 'western' Alpine divide indicated in the DAPC). However higher levels of migration were

indicated between the remaining populations, with a suggested migration pathway connecting GRA and PEN to the ‘eastern’ populations, again consistent with the patterns shown in the DAPC in Figure 3.1.4B.



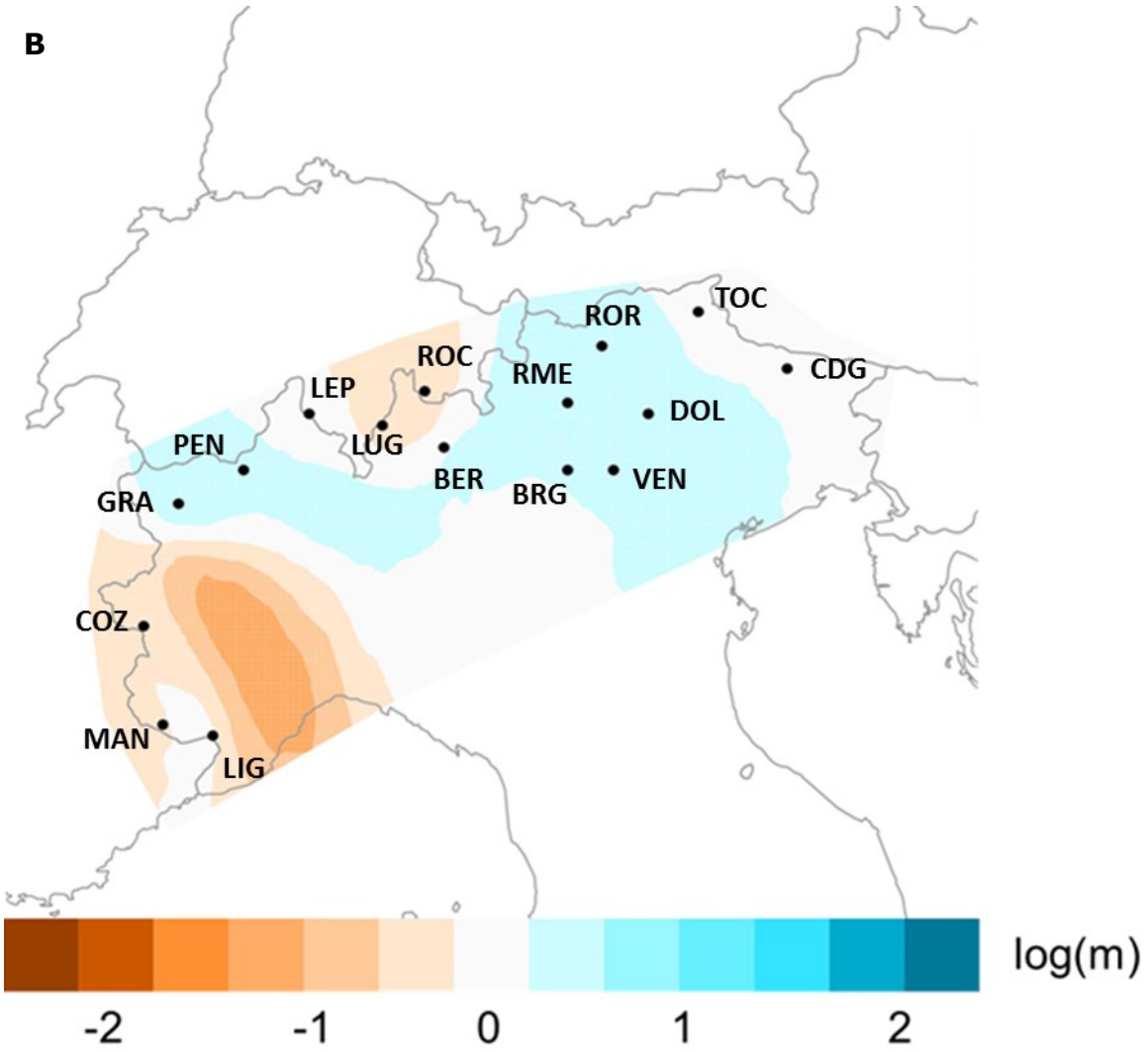


Figure 3.1.7 – Posterior mean migration rates (m) for the black grouse on the log10 scale for the STR (**A**) and SNP (**B**) datasets. The shading represents effective rates of migration. Calculated using Estimated Effective Migration Surfaces (EEMS).

For the RDA, latitude and altitude were removed as parameters as they were strongly correlated with other environmental factors (longitude and mean June temperature, respectively). The RDA plot (Figure 3.1.8) showed samples separated along the longitude vector, with samples from the 'eastern' Alps on the left of the plot and samples from the west on the right. The y-axis at 0 on this plot separates individuals in a similar area to that indicated by the DAPC and STRUCTURE analyses. Samples from the 'western' populations were also shown to be in regions with higher human population density compared to those in the east. Genetic diversity in the populations in the lower-left area of the plot, such as BRG and RME, are positively associated with high June temperature and low June rainfall, the reverse can be said for populations from the upper right side (i.e. LUG and LEP).

Individuals negatively correlated with the landcover vector were primarily collected in areas classified as “tree cover”, while many individuals positively correlated with this vector were collected from areas classified as “cultivated and managed areas”. PCAdapt identified outliers with each of the methods tested: 703 outliers were detected using both the q-value and Benjamini-Hochberg Procedure, while the more conservative Bonferroni correction found a total of 448 outliers.

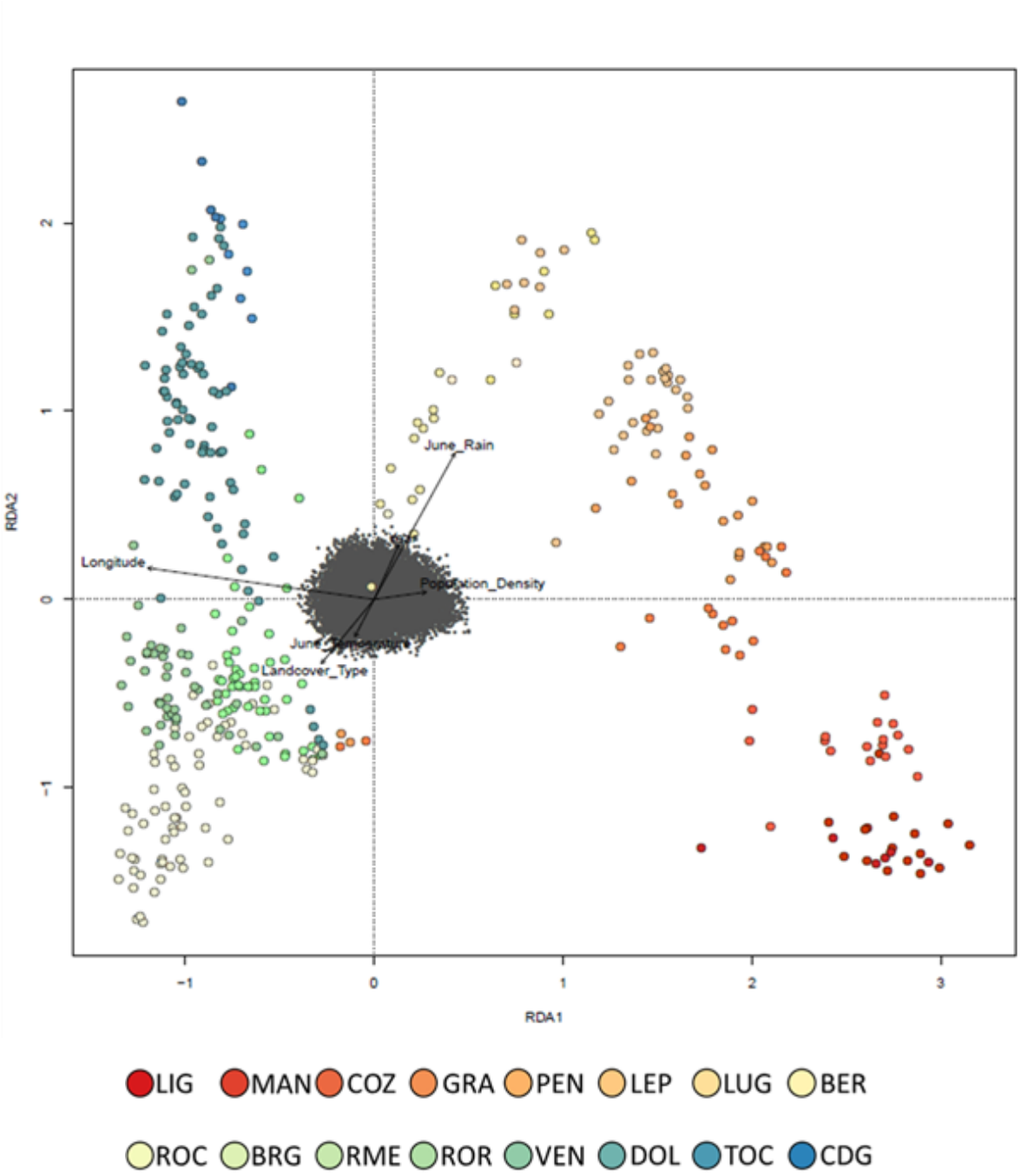


Figure 3.1.8 – Redundancy Analysis (RDA) for the black grouse SNP data (n=368). Grey points at the centre of the plot represent the SNPs; individuals are colour coded based on the mountain group population (see Figure 3.1.1). The black vectors represent the environmental predictors tested: year sampled, mean June rainfall and human population density in the upper right quadrant, mean June temperature and landcover category in the lower left quadrant, and Longitude in the upper right quadrant.

3.1.4. TEMPORAL ANALYSIS

For the STR data, the mean total H_E was 0.73 (SD ±0.15), while H_E for each time period were very similar: 0.74, 0.72 and 0.73 for 1995-1999, 2009-2010 and 2015-2017, respectively. The mean total number of alleles was 9.6 across all loci and individuals (1995-1999: 8.8; 2009-2010: 8.6; 2015-2017: 8.1). In contrast, for the SNP data, the mean total H_E and for each time frame was 0.25 (SD $\pm$0.16). The lack of variation in H_E and a similar number of alleles in all time frames was also reflected in the PCA. Figure 3.1.9 indicated for both the STR and SNP data there is no clustering of individuals based on the year that the samples were collected, and the variance included in the PCA plot is low. Estimations of N_e for both datasets were the lowest for the 1995-1999 time frame (STR: 87.6, 95%CI: 68.8-115.2; STR: 102.5, 95%CI: 102.3-102.7). The calculated value for N_e for the STR dataset was highest for the 2009-2010 samples (154.3, 95%CI: 114-222.9) compared to the 2015-2017 samples (147.4, 95%CI: 80.7-509). For the SNP dataset the estimated N_e calculated increased over time (2009-2010: 214.6, 95%CI: 214.1-215.1; 2015-2017: 322.1, 95%CI: 320.8-323.4). Compared to the STR dataset, the 95% confidence intervals for the SNP data were narrower and did not overlap.

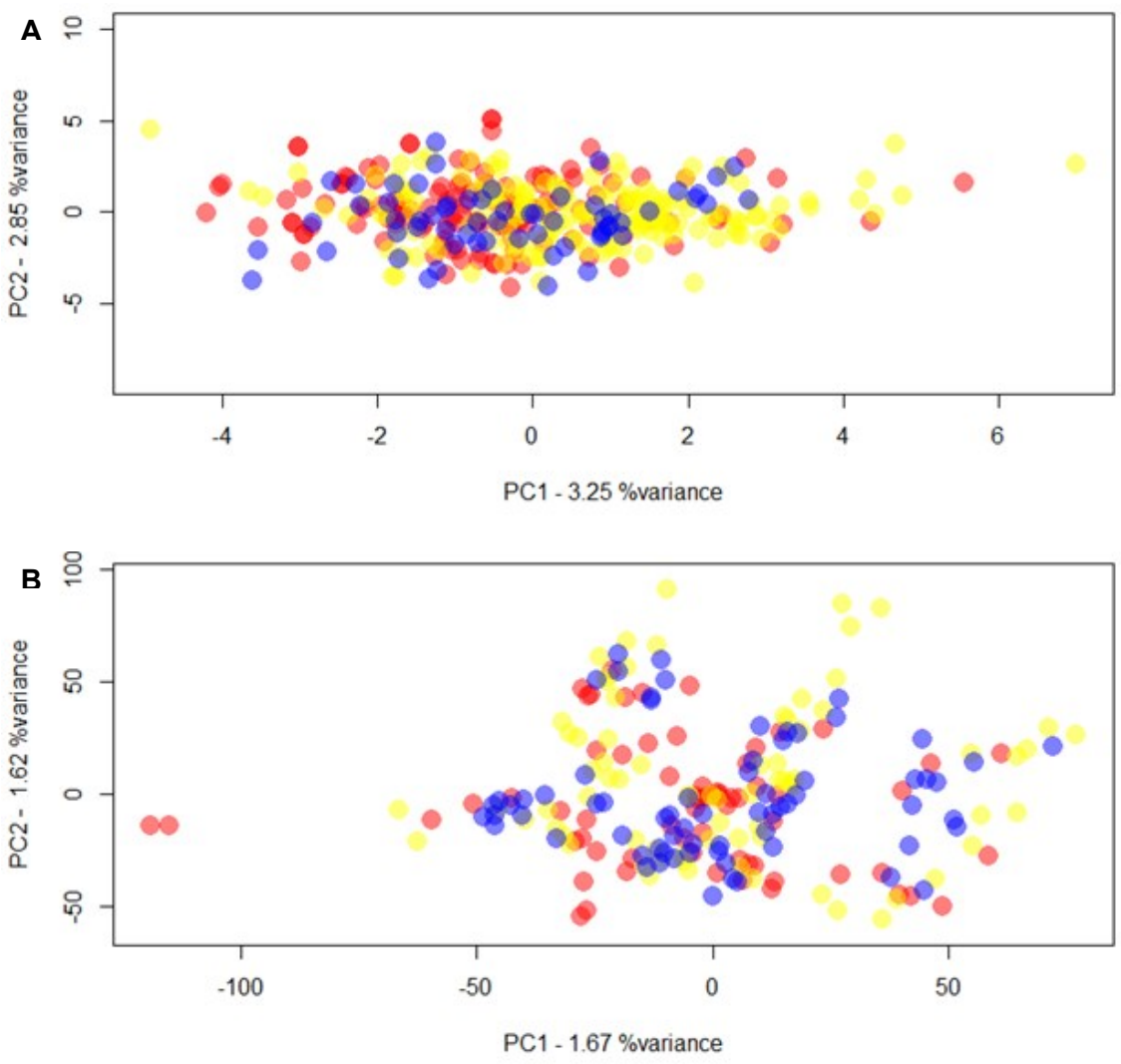


Figure 3.1.9 – PCA for the temporal analysis showing the results for the STR (**A**, n=360) and SNP (**B**, n=211) analyses. Red points represent samples from the 1995-1999 time frame, yellow points from the 2009-2010 time frame and blue points from the 2015-2017 time frame.

3.2. PTARMIGAN

3.2.1. SAMPLE COLLECTION

In total, 407 ptarmigan samples were collected between 1996 and 2017, from mountain groups across the Alps as well as from three populations in the northern range of the species (Figure 3.2.1 and Table 3.2.1). To ensure a comparison of results between species was possible, the same population definitions were used as for the black grouse.

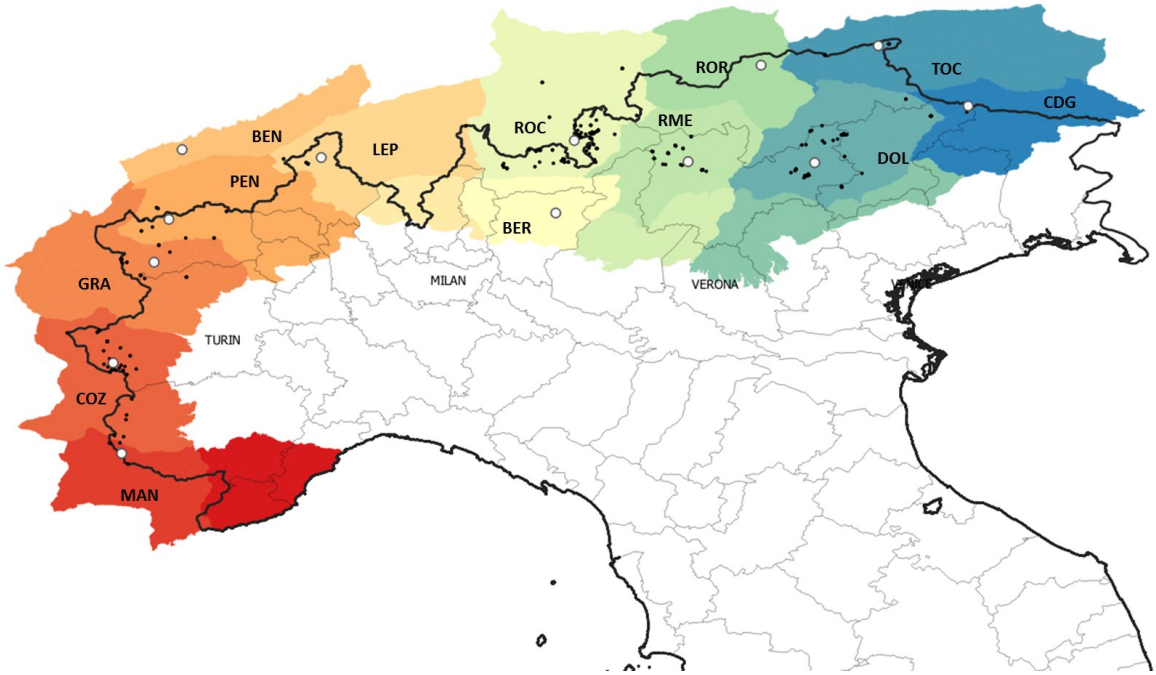


Figure 3.2.1 – Map of northern Italy showing the locations of the ptarmigan samples. Points represent single samples and white dots represent the mean population centre. Thick black lines represent the Italian border; thin grey lines Italian provinces. The shaded areas represent SOIUSA mountain groups (see Table 3.2.1 and Figure 2.1). Map created in QGIS (QGIS Development Team, 2021).

Table 3.2.1 – Origin of ptarmigan samples used to generate SNP, STR and mtDNA (long and short fragment) datasets.

Mountain Group (Population)	Abbreviation	Total	SNP	STR	mtDNA long	mtDNA short
Alpi Marittime e Prealpi di Nizza	**MAN**	4	1	4	4	4
Alpi Cozie	**COZ**	43	20	36	34	36
Alpi Graie	**GRA**	21	15	15	12	13
Alpi Pennine	**PEN**	10	4	6	2	5
Alpi Bernesi	**BEN**	8	5	8	8	8
Alpi Lepontine	**LEP**	8	8	4	4	4
Alpi e Prealpi Bergamasche	**BER**	1	0	1	1	1
Alpi Retiche occidentali	**ROC**	114	53	97	94	94
Alpi Retiche meridionali	**RME**	41	12	32	28	34
Alpi Retiche orientali	**ROR**	1	0	1	1	1
Dolomiti	**DOL**	106	17	72	50	82
Alpi dei Tauri occidentali	**TOC**	12	11	12	7	8
Alpi Carniche e della Gail	**CDG**	27	3	20	8	12
Iceland	**ISL**	3	3	0	0	0
Lapland	**LAP**	1	1	0	0	0
Sweden	**SWE**	5	5	0	0	0
Undefined Mountain Group	**UNK**	2	0	2	0	0
Total		**407**	**158**	**310**	**253**	**302**

3.2.2. INDIVIDUAL GENOTYPING

Mitochondrial DNA sequences were trimmed to two lengths: 410bp ('short fragment') and 1035bp ('long fragment'). Twenty-three haplotypes were found for 302 individuals for the short fragment, and 26 haplotypes were identified for 253 individuals for the long fragment (Table 3.2.2). The majority of values calculated for gene diversity for the long fragment ranged between 0.67 (SD±0.20; MAN) and 0.86 (SD±0.06; GRA) excluding values calculated as 1 (SD±0; BER and ROR) and 0 (SD±0; PEN). For the short fragment, PEN had the lowest levels of gene diversity (0.6; SD±0.18), again BER and ROR had a gene diversity of 1 (SD±0) and GRA had a gene diversity of 0.87 (SD±0.05) (Table 3.2.3). Significant F_{ST} values for the short fragment ranged from 0.03 (between COZ and ROC) and 0.48 (between PEN and RME); for the long fragment, significant results ranged between 0.06 and 0.71 for the same pairs of populations (Table 3.2.4).

Table 3.2.2 – Haplotypes for the ptarmigan mitochondrial DNA datasets. The long fragment is shown in **A** and the short in **B**. The frequency of each haplotype present in each population is also shown.

A	Population													Total
Haplotype	MAN	COZ	GRA	PEN	BEN	LEP	BER	ROC	RME	ROR	DOL	TOC	CDG	
HLm01	2	18	1	0	2	1	0	35	11	0	24	2	4	**100**
HLm02	0	1	3	2	1	0	0	8	0	0	1	0	0	**16**
HLm03	0	0	0	0	0	0	0	2	0	0	0	0	0	**2**
HLm04	0	0	0	0	0	0	0	2	0	0	0	0	0	**2**
HLm05	0	0	0	0	1	0	0	5	4	0	6	0	0	**16**
HLm06	0	0	0	0	0	0	0	5	0	0	0	2	1	**8**
HLm07	0	0	0	0	0	0	0	7	0	1	1	2	1	**12**
HLm08	0	0	0	0	0	0	0	3	0	0	1	0	0	**4**
HLm09	0	3	3	0	4	0	1	7	7	0	7	0	0	**32**
HLm10	0	0	0	0	0	0	0	0	2	0	3	0	0	**5**
HLm11	0	1	0	0	0	0	0	2	0	0	1	0	0	**4**
HLm12	0	1	0	0	0	1	0	10	0	0	0	0	0	**12**
HLm13	0	0	0	0	0	0	0	1	0	0	0	0	0	**1**
HLm14	0	0	0	0	0	0	0	3	0	0	0	0	0	**3**
HLm15	0	7	1	0	0	2	0	2	0	0	0	0	0	**12**
HLm16	0	0	0	0	0	0	0	0	3	0	2	0	1	**6**
HLm17	0	0	0	0	0	0	0	1	0	0	0	0	0	**1**
HLm18	0	0	0	0	0	0	0	0	0	0	2	0	0	**2**
HLm20	0	0	0	0	0	0	0	0	0	0	0	0	1	**1**
HLm21	0	0	0	0	0	0	0	1	0	0	1	0	0	**2**
HLm22	2	2	3	0	0	0	0	0	0	0	0	0	0	**7**
HLm23	0	0	1	0	0	0	0	0	0	0	0	0	0	**1**
HLm24	0	1	0	0	0	0	0	0	0	0	0	0	0	**1**
HLm25	0	0	0	0	0	0	0	0	0	0	0	1	0	**1**
HLm26	0	0	0	0	0	0	0	0	1	0	0	0	0	**1**
HLm27	0	0	0	0	0	0	0	0	0	0	1	0	0	**1**

B	Population													Total
Haplotype	MAN	COZ	GRA	PEN	BEN	LEP	BER	ROC	RME	ROR	DOL	TOC	CDG	
HLm01c	2	18	2	0	2	1	0	36	16	0	36	3	6	**122**
HLm02c	0	1	3	3	1	0	0	8	0	0	1	0	0	**17**
HLm05c	0	0	0	0	1	0	0	5	4	0	10	0	0	**20**
HLm06c	0	0	0	0	0	0	0	5	0	0	0	2	1	**8**
HLm07c	0	0	0	2	0	0	0	7	0	1	1	2	1	**14**
HLm08c	0	0	0	0	0	0	0	3	0	0	1	0	0	**4**
HLm09c	0	5	3	0	4	0	1	9	11	0	14	0	0	**47**
HLm10c	0	0	0	0	0	0	0	0	2	0	5	0	2	**9**
HLm11c	0	1	0	0	0	0	0	2	0	0	1	0	0	**4**
HLm12c	0	1	0	0	0	1	0	10	0	0	0	0	0	**12**
HLm13c	0	0	0	0	0	0	0	1	0	0	0	0	0	**1**
HLm14c	0	0	0	0	0	0	0	3	0	0	0	0	0	**3**
HLm15c	0	7	1	0	0	2	0	4	0	0	0	0	0	**14**
HLm18c	0	0	0	0	0	0	0	0	0	0	4	0	0	**4**
HLm20c	0	0	0	0	0	0	0	0	0	0	0	0	1	**1**
HLm21c	0	0	0	0	0	0	0	1	0	0	3	0	0	**4**
HLm22c	2	2	3	0	0	0	0	0	1	0	0	0	0	**8**
HLm23c	0	0	1	0	0	0	0	0	0	0	0	0	0	**1**
HLm24c	0	1	0	0	0	0	0	0	0	0	0	0	0	**1**
HLm25c	0	0	0	0	0	0	0	0	0	0	0	1	0	**1**
HLm27c	0	0	0	0	0	0	0	0	0	0	5	0	0	**5**
HLm28c	0	0	0	0	0	0	0	0	0	0	1	0	0	**1**
HLm29c	0	0	0	0	0	0	0	0	0	0	0	0	1	**1**

A total of 10 STR loci were characterised for 310 individuals collected from across the Italian Alps. Two individuals were removed from the analysis from the UNK population as these were found to be outliers in the preliminary analysis and may have been misidentified. The mean total missingness per locus for the STR data was 0.3% ranging from 0-1.29%. The H_E ranged between 0.6 in BER and ROR and 0.78 in ROC with a mean total H_E of 0.76 (SD ±0.11). The mean total number of alleles was 11.1 ranging from 1.6 (BER and ROR) to 8.4 (ROC) and the total allelic size range was 14 (1.67: ROR to 8.8: RME; Figure 3.2.3). Significant F_{ST} for the STR data ranged from 0.03 between DOL and ROC and 0.28 for BER and TOC (Figure 3.2.4).

SNP data was generated for a total of 158 individuals as some samples were repeated by both companies, and over 530,000 SNPs were produced with a mean coverage of 15.1X. A total of 15 individuals were removed due to low levels of coverage and high levels of missing data, and upon filtering with vcftools, the number of SNPs was reduced to 14,866, with a mean missingness per individual of 10.3% and a mean coverage of 29.6X. From some of the analyses, the 9 samples collected from outside of the Alps were removed, as well as an additional 6 samples that were identified as outliers. Across the Alps, the mean H_E ranged from 0.14 (MAN) to 0.15 (GRA), samples from the northern extent of the species range had a mean H_E of 0.07 in ISL, 0.11 in LAP and 0.15 in SWE. The mean total H_E was 0.15 (SD ±0.14). Significant levels of F_{ST} across the Alps ranged from 0 between GRA and PEN and 0.02 between DOL and BEN, however, the levels of F_{ST} were greater between Alpine and northern populations of the species, with values up to 0.4 between SWE and ISL (Figure 3.2.4).

Table 3.2.3 – Summary statistics showing gene diversity (*h*) calculated from mtDNA haplotype data; expected heterozygosity (H_E) calculated for the STR and SNP datasets; the mean number of alleles and allelic size range is also shown for the STR dataset.

Population	*h*		H_E		Number of alleles (STR only)	Allele size range (STR only)
	mtDNA long	mtDNA short	STR	SNP		
MAN	0.67	0.67	0.68	0.14	3.5	3.4
COZ	0.68	0.71	0.75	0.15	6.8	6.9
GRA	0.86	0.87	0.76	0.15	6.5	6.8
PEN	0	0.60	0.74	0.15	4.9	5.9
BEN	0.75	0.75	0.71	0.15	4.8	5
LEP	0.83	0.83	0.77	0.15	4	3.7
BER	1	1	0.60	-	1.6	3.5
ROC	0.83	0.82	0.78	0.15	8.4	8.2
RME	0.77	0.68	0.75	0.15	7.8	8.8
ROR	1	1	0.60	-	1.6	1.7
DOL	0.74	0.76	0.70	0.15	7.3	8.4
TOC	0.86	0.82	0.72	0.15	5.3	5.2
CDG	0.79	0.76	0.70	0.14	5.5	5.1
ISL	-	-	-	0.07	-	-
LAP	-	-	-	0.11	-	-
SWE	-	-	-	0.15	-	-
UNK	-	-	0.88	-	3.3	7.7
Total	**-**	**-**	**0.76**	**0.15**	**11.1**	**14**

Table 3.2.4 – Pairwise F_{ST} for the long (**A**) and short (**B**) mitochondrial DNA fragments, the microsatellite dataset (**C**) and SNP dataset (**D**) shown below the diagonal and calculated significant above the diagonal.

A	MAN	COZ	GRA	PEN	BEN	LEP	BER	ROC	RME	ROR	DOL	TOC	CDG
MAN	-	0.369	0.937	0.108	0.712	0.153	0.991	0.559	0.117	0.991	0.216	0.468	0.225
COZ	0.010	-	0.027	0.000	0.550	0.162	0.432	0.036	0.063	0.991	0.000	0.036	0.000
GRA	-0.131	**0.087**	-	0.054	0.252	0.036	0.991	0.027	0.000	0.991	0.000	0.036	0.009
PEN	0.207	**0.355**	0.209	-	0.000	0.072	0.180	0.054	0.000	0.991	0.000	0.081	0.045
BEN	-0.062	-0.018	0.003	**0.298**	-	0.018	0.991	0.324	0.505	0.991	0.523	0.135	0.090
LEP	0.234	0.182	**0.248**	0.421	**0.247**	-	0.514	0.063	0.000	0.991	0.009	0.045	0.009
BER	-0.336	-0.328	-0.417	0.307	-0.510	-0.430	-	0.676	0.595	0.991	0.550	0.712	0.432
ROC	-0.026	**0.031**	**0.065**	0.123	0.007	0.125	-0.283	-	0.000	0.991	0.000	0.315	0.234
RME	0.131	0.043	**0.204**	**0.478**	-0.025	**0.443**	-0.005	**0.070**	-	0.991	0.414	0.000	0.009
ROR	0.334	0.259	0.292	0.000	0.295	0.091	1.000	-0.183	0.450	-	0.991	0.991	0.991
DOL	0.033	**0.046**	**0.147**	**0.296**	-0.016	**0.332**	-0.147	**0.048**	-0.001	0.119	-	0.045	0.171
TOC	0.020	**0.122**	**0.127**	0.166	0.079	**0.230**	-0.192	0.002	**0.199**	-0.430	**0.087**	-	0.306
CDG	0.088	**0.094**	**0.179**	**0.155**	0.088	**0.315**	0.090	0.010	**0.117**	-0.138	0.028	0.026	-

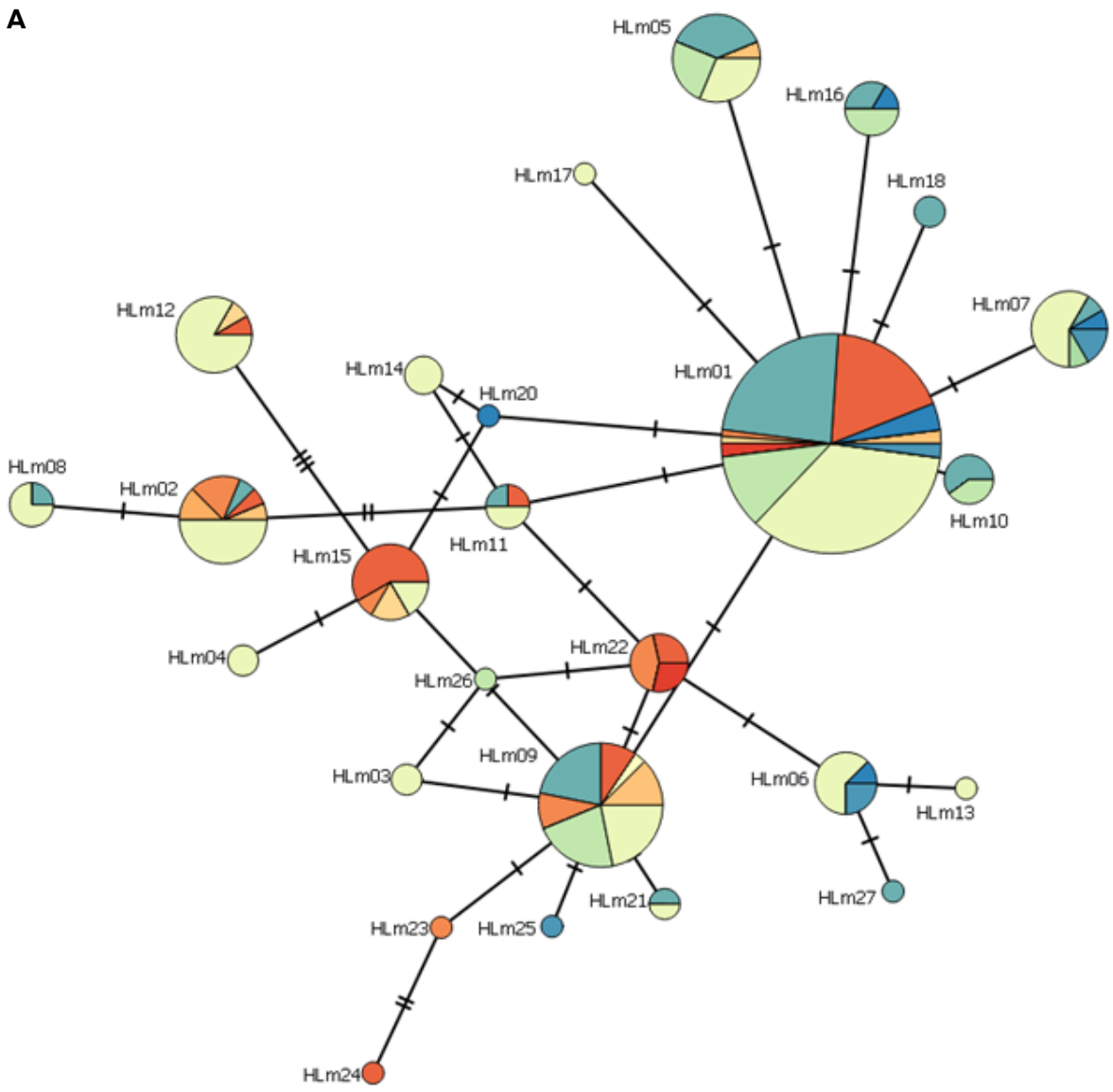
B	MAN	COZ	GRA	PEN	BEN	LEP	BER	ROC	RME	ROR	DOL	TOC	CDG
MAN	-	0.378	0.730	0.081	0.766	0.261	0.991	0.523	0.144	0.991	0.027	0.378	0.306
COZ	0.002	-	0.009	0.009	0.541	0.027	0.505	0.081	0.063	0.991	0.000	0.027	0.261
GRA	-0.091	**0.118**	-	0.180	0.288	0.036	0.991	0.027	0.000	0.991	0.000	0.045	0.027
PEN	0.680	**0.618**	0.385	-	0.108	0.045	0.351	0.027	0.000	0.991	0.027	0.027	0.045
BEN	-0.067	-0.015	0.013	0.608	-	0.045	0.991	0.369	0.369	0.991	0.171	0.117	0.135
LEP	0.234	**0.163**	**0.224**	**0.659**	**0.223**	-	0.613	0.099	0.000	0.991	0.000	0.045	0.027
BER	-0.334	-0.319	-0.455	1.000	-0.572	-0.429	-	0.649	0.658	0.991	0.468	0.991	0.541
ROC	-0.033	0.022	**0.075**	**0.401**	-0.001	0.104	-0.331	-	0.000	0.991	0.000	0.315	0.622
RME	0.101	0.052	**0.237**	**0.712**	0.003	**0.389**	-0.044	**0.065**	-	0.991	0.541	0.009	0.297
ROR	0.334	0.201	0.230	1.000	0.214	0.091	1.000	-0.236	**0.311**	-	0.991	0.991	0.991
DOL	**0.092**	**0.065**	**0.240**	**0.656**	0.024	**0.399**	0.034	**0.058**	-0.006	0.201	-	0.000	0.387
TOC	0.034	**0.143**	**0.145**	**0.607**	0.101	**0.217**	-0.213	0.013	**0.215**	-0.482	**0.191**	-	0.288
CDG	0.074	0.031	**0.192**	**0.666**	0.059	**0.259**	0.047	-0.019	0.033	-0.143	0.001	0.031	-

C	MAN	COZ	GRA	PEN	BEN	LEP	BER	ROC	RME	ROR	DOL	TOC	CDG	UNK
MAN	-	0.018	0.450	0.550	0.099	0.126	0.189	0.081	0.144	0.991	0.072	0.009	0.000	0.126
COZ	**0.065**	-	0.090	0.108	0.468	0.243	0.027	0.144	0.171	0.991	0.000	0.036	0.000	0.009
GRA	0.045	0.019	-	0.703	0.441	0.315	0.586	0.216	0.667	0.991	0.144	0.306	0.135	0.009
PEN	0.004	0.040	-0.017	-	0.252	0.667	0.351	0.036	0.144	0.991	0.090	0.261	0.009	0.135
BEN	0.108	-0.001	0.011	0.023	-	0.514	0.991	0.180	0.541	0.991	0.117	0.496	0.018	0.000
LEP	0.115	0.028	**0.043**	-0.005	0.014	-	0.207	0.090	0.162	0.991	0.000	0.225	0.000	0.018
BER	0.342	**0.169**	-0.007	0.106	0.261	0.368	-	0.153	0.342	0.991	0.117	0.045	0.018	0.604
ROC	0.055	0.003	0.009	**0.045**	0.015	0.043	0.106	-	0.243	0.991	0.000	0.000	0.000	0.000
RME	0.047	0.007	-0.006	0.030	-0.006	0.043	0.056	0.005	-	0.991	0.243	0.135	0.144	0.018
ROR	0.304	-0.006	-0.086	-0.147	0.083	0.216	0.121	0.006	-0.075	-	0.991	0.505	0.991	0.622
DOL	0.053	**0.035**	0.016	0.061	0.025	**0.123**	0.161	**0.028**	0.006	0.015	-	0.000	0.009	0.000
TOC	**0.143**	**0.029**	0.022	0.029	-0.004	0.042	**0.277**	**0.033**	0.020	0.059	**0.071**	-	0.009	0.009
CDG	**0.194**	**0.037**	0.031	**0.100**	**0.059**	**0.187**	**0.264**	**0.053**	0.010	0.004	**0.037**	**0.093**	-	0.000
UNK	0.666	**0.755**	**0.650**	0.587	**0.708**	**0.623**	0.185	**0.746**	**0.712**	0.423	**0.804**	**0.715**	**0.799**	-

D	MAN	COZ	GRA	PEN	BEN	LEP	ROC	RME	DOL	TOC	CDG	SWE	LAP	ISL
MAN	-	0.306	0.081	0.180	0.225	0.090	0.991	0.117	0.991	0.090	0.324	0.171	0.991	0.351
COZ	0.011	-	0.000	0.108	0.577	0.081	0.000	0.586	0.000	0.000	0.171	0.000	0.991	0.000
GRA	0.023	**0.002**	-	0.009	0.324	0.018	0.000	0.009	0.000	0.000	0.108	0.009	0.991	0.009
PEN	0.004	0.000	**0.001**	-	0.279	0.378	0.027	0.000	0.000	0.027	0.063	0.018	0.171	0.009
BEN	0.029	-0.007	-0.006	-0.010	-	0.775	0.207	0.000	0.000	0.072	0.117	0.018	0.991	0.000
LEP	0.017	0.003	**0.003**	-0.004	-0.008	-	0.000	0.117	0.000	0.000	0.279	0.000	0.991	0.009
ROC	0.033	**0.010**	**0.011**	**0.008**	-0.001	**0.008**	-	0.991	0.000	0.000	0.342	0.000	0.991	0.000
RME	0.081	-0.004	**-0.001**	**0.004**	**0.004**	-0.004	-0.012	-	0.000	0.459	0.000	0.000	0.991	0.000
DOL	0.081	**0.013**	**0.016**	**0.018**	**0.023**	**0.014**	**0.007**	**0.007**	-	0.000	0.117	0.000	0.991	0.009
TOC	0.026	**0.015**	**0.016**	**0.009**	0.006	**0.013**	**0.011**	-0.007	**0.005**	-	0.423	0.000	0.991	0.000
CDG	0.030	0.005	0.009	0.005	0.004	0.001	0.002	**0.020**	0.008	-0.014	-	0.045	0.432	0.054
SWE	0.308	**0.205**	**0.214**	**0.231**	**0.233**	**0.215**	**0.203**	**0.230**	**0.233**	**0.215**	**0.259**	-	0.108	0.009
LAP	0.087	0.066	0.078	0.062	0.072	0.058	0.075	0.124	0.115	0.055	0.094	0.299	-	0.991
ISL	0.415	**0.186**	**0.203**	**0.235**	**0.249**	**0.199**	**0.183**	**0.247**	**0.240**	**0.197**	0.311	**0.395**	0.437	-

3.2.3. POPULATION GENETIC ANALYSIS

Neither of the haplotype networks for the short and long mtDNA fragments showed a strong pattern of population clustering (Figure 3.2.2). However, some haplotypes such as HLm05(c), 07(c), and 16 primarily represent ‘eastern’ Alpine populations (BER to CDG) and haplotypes HLm15(c) and 22(c) represent ‘western’ populations (MAN to LEP). In addition, there are very few base pair differences between the haplotype sequences (a maximum of three for HLm12 and HLm15 for the long fragment; Figure 3.2.2A). A similar lack of association between haplotype and population of origin is shown in Table 3.2.3 where haplotypes are present across the Italian Alps regardless of sampling location.



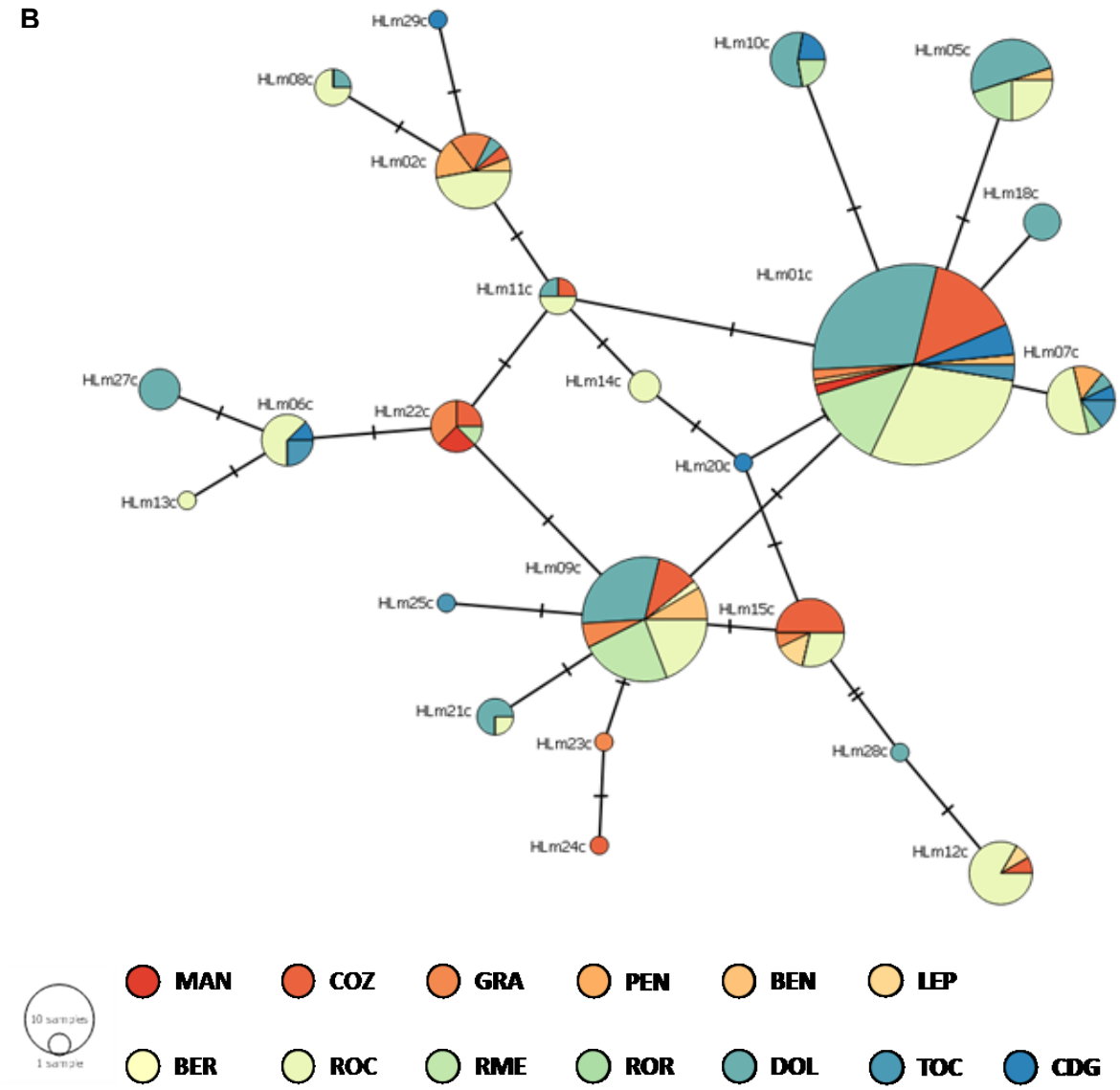


Figure 3.2.2 – Haplotype Network for the Ptarmigan mtDNA. Neighbour-joining networks are shown for both the long (**A**, n=253) and short (**B**, n=302) fragments. Each circle represents a unique haplotype and is coloured based on the number of individuals represented in each. The vectors represent the connections between each haplotype; perpendicular marks on each vector represent the proportional number of base pair differences between each haplotype.

The PCA for the STR data does not show any population structure; that is, all samples appear to cluster together regardless of the population in which they were collected (Figure 3.2.3). The same pattern is shown in the DAPC, where seven clusters were found to best explain the dataset (Figure 3.2.4). However, none of these clusters were associated with any geographical area. Instead, for the SNP data, there is a clear separation between the 'eastern' and 'western' samples within the plot. The clustering occurs on either side of the LEP mountain group. The

‘eastern’ cluster also showed some internal geographical structure, where samples from DOL, TOC and CDG cluster at the bottom of the plot, while samples from ROC cluster at the top. One individual collected from LEP also clustered with ROC. For both datasets, the first two axes of the PCA represent low proportions of variance, suggesting that the patterns shown are not strongly supported. clustering suggested by the SNP data is also repeated in the DAPC results. Two groups were found to best represent the data, from the same geographical area as in the PCA with the blue group representing the ‘western’ populations and the red group representing the ‘eastern’ populations.

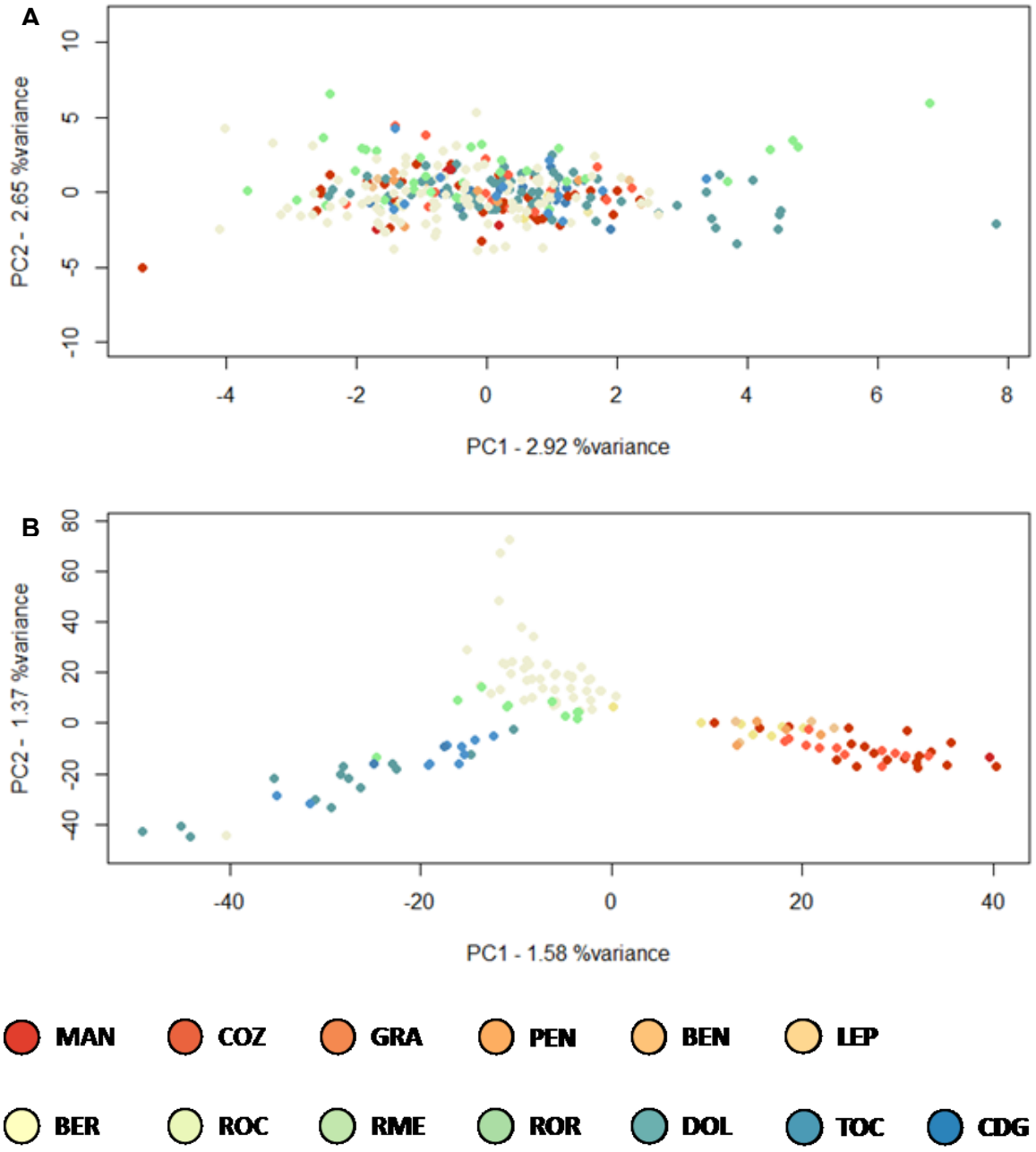


Figure 3.2.3 – Principal Component Analysis for the ptarmigan STR (**A**, n=308) and SNP (**B**, n=128) data. Samples are coloured according to the mountain group from which they were collected.

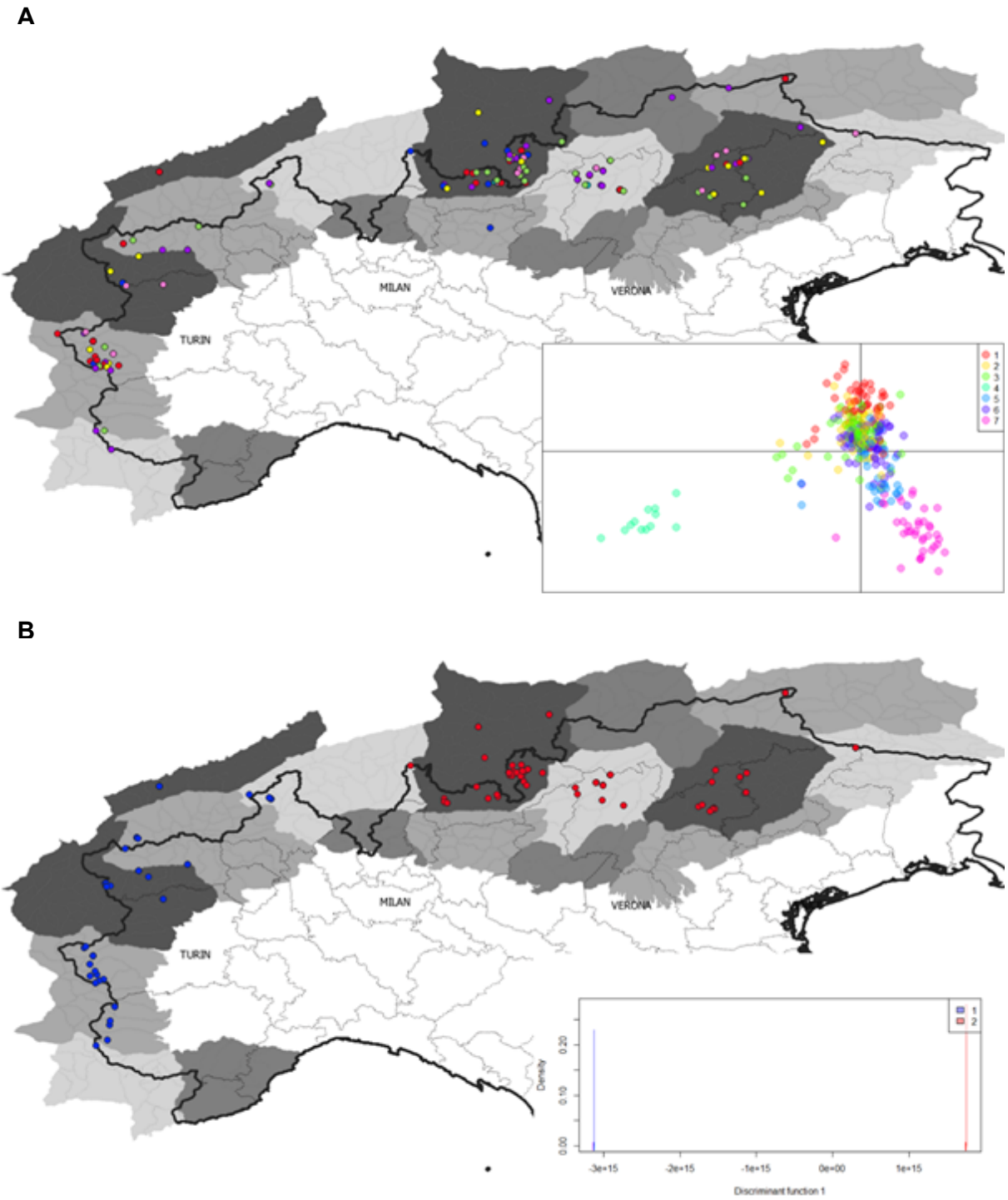


Figure 3.2.4 – Discriminant Analysis of Principal Components (DAPC) for the ptarmigan STR (**A**, n=308) and SNP (**B**, n=128) datasets. The DAPC plot is shown in the lower right corner of each figure; geographical sampling points are coloured based on the DAPC cluster.

The Evanno method found that K=2 best explained both the STR and SNP STRUCTURE results (Figure 3.2.5). However, both K=2 and K=3 are shown here to examine any underlying structure within the datasets. In both datasets, all individuals are admixed with each cluster and there does not appear to be any populations that are primarily represented by a single cluster.

Figure 3.2.5 – STRUCTURE plot showing K=2 and 3 for both the STR (**A**, n=310) and SNP (**B**, n=143) datasets. White lines represent the division of populations. Plots were visualised in Pophelper version 1.0.10 (Francis, 2016).

An AMOVA was used to investigate the possibility of an east/west division in the populations (Chapter 3.1). Samples were divided into two groups, MAN to LEP and BER to CDG; an additional group was used for the SNP data that included populations from the northern extent of the species range (ISL, LAP and SWE). The AMOVAs for both datasets show the greatest percentage variation is found within

individuals (89.4%: STR and 92.9%: SNP; Table 3.2.5). For the STR dataset, the second greatest source of variation was among the individuals within populations with a value of 8.3% with the lowest significant variation at 2.4% among populations within groups. For the SNP dataset, the source of variation among populations within groups was the second greatest source of variation (3%) and the lowest source was among individuals within populations (1.4%). For both datasets, all AMOVA tests were found to be significant with a P-value of 0 (Table 3.2.5).

Values of N_e for Alpine populations (Table 3.2.6) ranged from 3.1 (MAN) and 589.3 (COZ) for the STR dataset, and from 207.6 (ROC) to 826.5 (COZ) for the SNPs. Overall N_e for the STR dataset was estimated 321.1, while for the SNP data this was 490. When samples from outside of the Alpine range were included, N_e for the SNP data decreased to 264.5.

Table 3.2.5 – AMOVA Analysis for the ptarmigan STR and SNP datasets. The 'eastern' and 'western' Alpine populations were used as groups for the purpose of this analysis. The genetic structure was estimated among groups (Va/FCT), among groups within populations (Vb/FSC), among individuals within populations (Vc/FIS), and within individuals (Vd/FIT) using Arlequin (Version 3.5).

Marker	Variation Source	Sum of squares	Variance components	Percentage of variation	F Index	P-value
STR	Among groups	9.58	-0.0023	-0.06	-0.00059	0.41
	Among populations within groups	83.00	0.090	2.37	0.024	0
	Among individuals within populations	1183.38	0.31	8.27	0.085	0
	Within individuals	1043.00	3.40	89.42	0.11	0
	Total	2318.96	3.80			
SNP	Among groups	7426.27	29.52	2.83	0.028	0
	Among populations within groups	16090.60	30.75	2.95	0.030	0
	Among individuals within populations	114054.026	14.27	1.37	0.015	0
	Within individuals	124400.50	967.46	92.85	0.072	0
	Total	261971.40	1041.99			

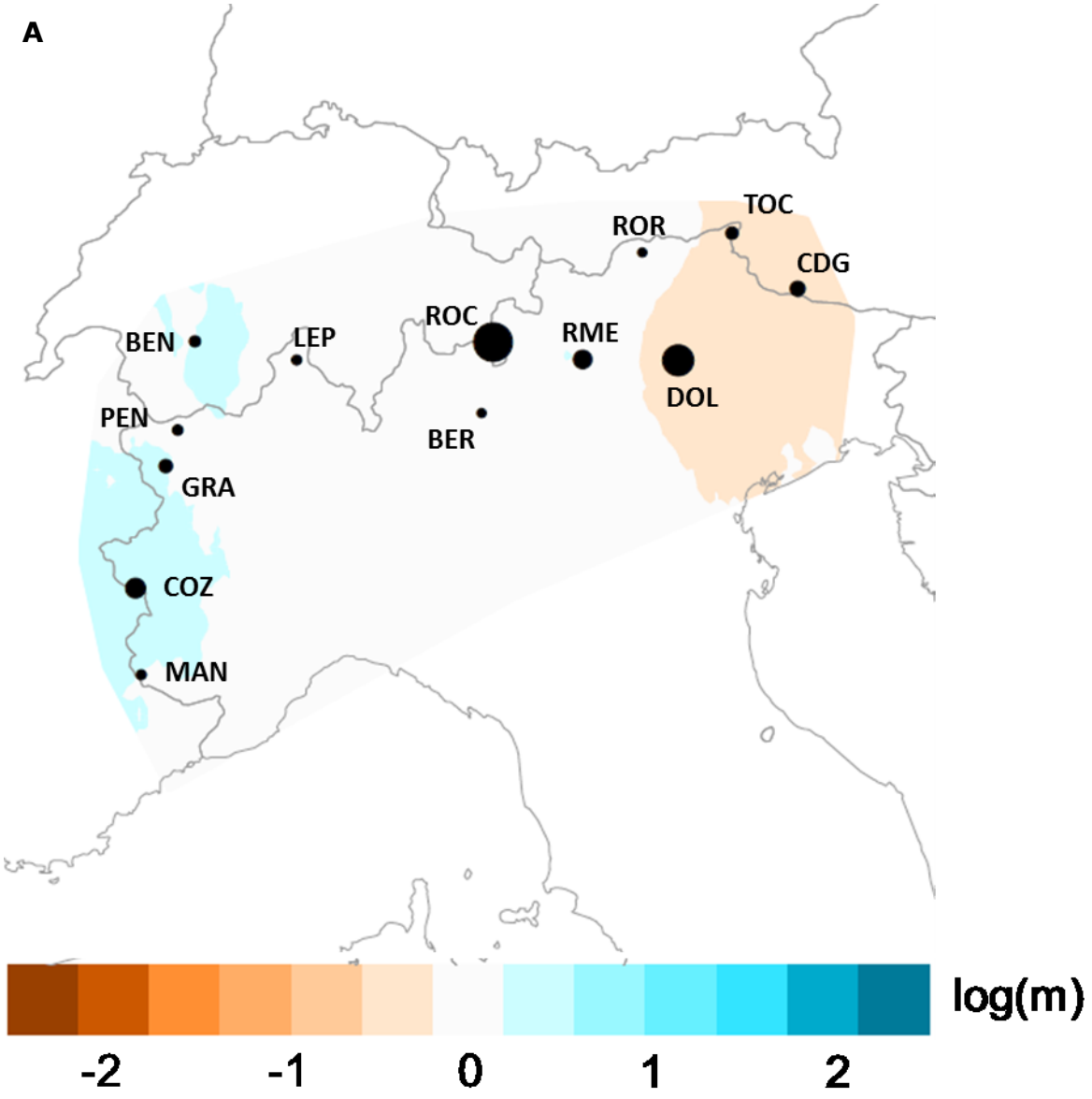
Table 3.2.6 – Estimates of effective population size (N_e) for the ptarmigan STR and SNP datasets with 95% confidence intervals (CI). Inf: infinite.

	STR		SNP	
	N_e	95% CI	N_e	95% CI
MAN	3.1	1-Inf	Inf	Inf
COZ	589.3	101.2-Inf	826.5	756.8-910.3
GRA	35.1	16.7-287.9	Inf	Inf
PEN	Inf	Inf	Inf	Inf
BEN	59	9.1-Inf	Inf	Inf
LEP	Inf	3-Inf	Inf	Inf
BER	Inf	Inf	-	-
ROC	284.8	147.6-1482.3	207.6	205.7-209.6
RME	13.4	10.2-17.8	Inf	Inf
ROR	Inf	Inf	-	-
DOL	66.7	46.5-105.3	384.6	350.6-425.9
TOC	40.6	13-Inf	536.2	475.6-614.3
CDG	39.8	18.5-361	Inf	Inf
ISL	-	-	Inf	Inf
LAP	-	-	Inf	Inf
SWE	-	-	13.1	12.7-13.5
UNK	Inf	Inf	-	-
Overall (Alps only)	**321.1**	**235.3-472.9**	**490**	**486.3-493.8**
Overall (All populations	**-**	**-**	**264.5**	**263.4-265.6**

In the IBD tests, non-significant and negative values of F_{ST} were adjusted to 0, to ensure that only significant values of F_{ST} were included. Samples from LAP, ISL, SWE and UNK were removed from these analyses. IBD tests for the mtDNA fragments and the STR datasets were not significant (Long DNA fragment: R=0.1, p=0.18; Short DNA fragment: R=0.04, p=0.35; STR: 0.17, p=0.06). However, the SNP data did show a significant pattern of isolation by distance R=0.14 (p<0.03).

For the EEMS, the STR results suggest low levels of migration in the 'eastern' populations and high levels of migration in the 'western' populations. Samples from the central mountain groups (LEP, ROC, BER, RME and ROR) showed no change in effective migration rate (Figure 3.2.6A). The results for the SNP dataset suggest an area of low migration around the MAN and PEN mountain groups, and south of

TOC, while high levels of migration appear to be present between ROC and RME (Figure 3.2.6B).



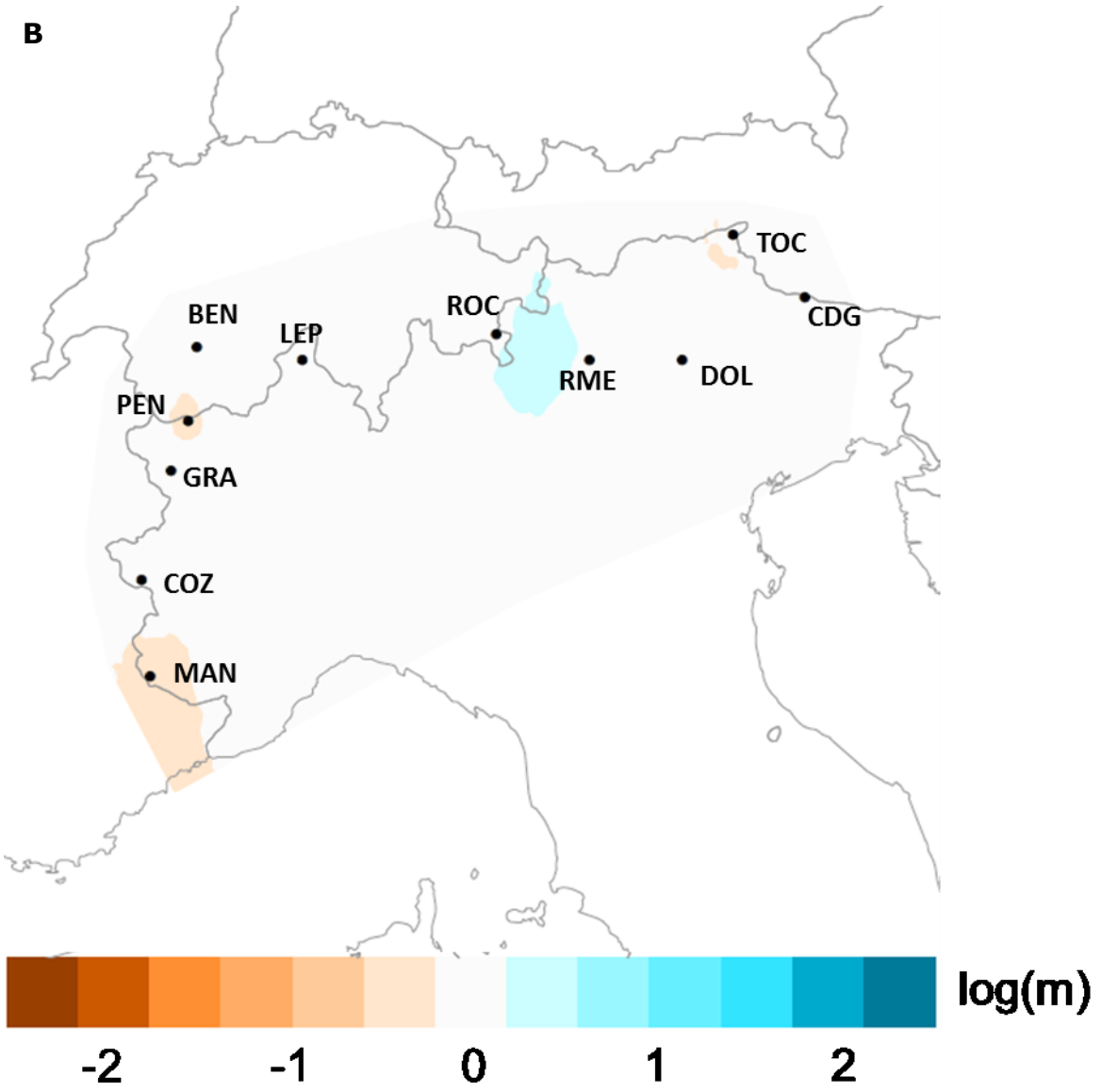


Figure 3.2.6 – Changes in effective migration rates shown for the STR (**A**) data and the SNP data (**B**). The logarithm of posterior mean migration rates (m) is shown by the shaded areas on the maps, with each point indicating the mean centre of the population. Calculated with Estimated Effective Migration Surfaces (EEMS).

3.3. CAPERCAILLIE

3.3.1. SAMPLE COLLECTION

A total of 261 capercaillie samples were collected from the Alps, with the majority representing Italian populations from the Trentino-Alto Adige Region (Figure 3.3.1). Three samples were obtained from the Austrian Alps. The majority of samples were collected from three parks in Trentino-Alto Adige: Parco Naturale Monte Corno, Parco Nazionale Dello Stelvio and Parco Naturale Paneveggio – Pale Di San Martino, which were considered 'populations' for the following analyses. Three individuals not collected from parks were considered separate populations. Despite the low numbers of individuals present in some populations, the SOIUSA mountain groups were not used to define populations, in order to examine regional genetic structure.

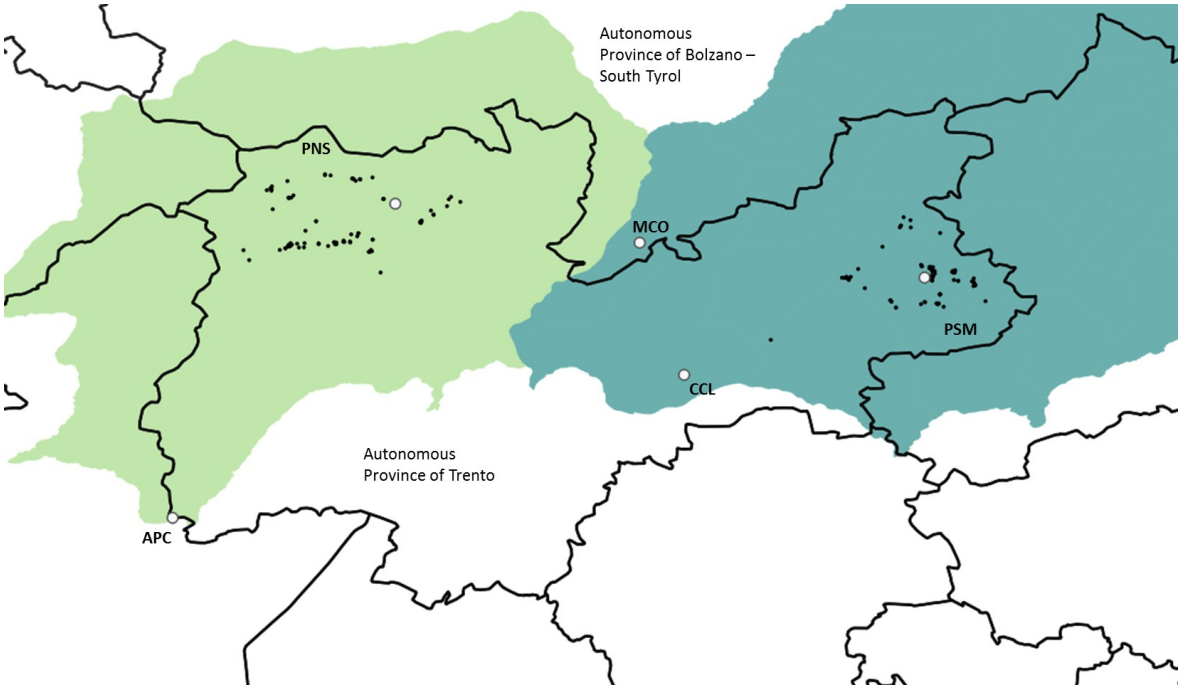


Figure 3.3.1 – Capercaillie samples collected from the Trentino Alto Adige. Points represent individuals and white points mark the mean centre of the population. As no coordinates were collected for the MCO population the mean centre of Parco Naturale Monte Corno was used. The coloured areas represent the SOIUSA Mountain groups, DOL and RME, used for the black grouse and ptarmigan analyses; the black lines show the borders of the Italian provinces.

Table 3.3.1 – The distribution of capercaillie sampling across populations and genetic markers analysed.

Park or Locality	Population	Total	SNP	STR	mtDNA long	mtDNA short
Parco Naturale Monte Corno	**MCO**	14	0	14	12	12
Parco Nazionale Dello Stelvio	**PNS**	140	19	127	129	130
Parco Naturale Paneveggio – Pale Di San Martino	**PSM**	97	2	94	94	1
Adamello, Presanella, Destra Chiese	**APC**	1	1	0	0	0
Cima d'Asta, Croce, Lagorai	**CCL**	1	1	0	0	0
Val Campelle	**VCA**	1	0	1	1	1
Austrian Alps	**ATA**	3	3	0	0	0
Unknown Sampling Location	**UNK**	4	4	0	0	0
Total		**261**	**30**	**236**	**236**	**144**

3.3.2. INDIVIDUAL GENOTYPING

The analyses used two sequence lengths of mitochondrial DNA: 254bp and 388bp. 13 unique haplotypes represented 144 individuals with the shorter mtDNA fragment, while 17 haplotypes were found for 236 individuals for the long fragment (Table 3.3.2). Gene diversity (Table 3.3.3) for the long fragment ranged from 0.70 (SD±0.04; PSM) and 0.92 (SD±0.01; PNS). For the short fragment, MCO had a gene diversity of 0.85 (SD±0.07) and for PNS this was 0.91 (SD±0.01). The VCA population had a value of 1 (SD±0) in both fragments, while the gene diversity of PSM was also 1 (SD±0) for the short fragment. For each of the fragment lengths,

only one population pair had a significant level of F_{ST} of 0.09 (MCO and PNS) for the short fragment, 0.02 (PNS and PPSM) for the long fragment (Table 3.3.4).

Table 3.3.2 – Per population haplotype frequencies for both the long (**A**) and short (**B**) mitochondrial DNA fragments.

A	Population				Total
Haplotype	MCO	PNS	PSM	VCA	
nsH7	4	7	7	0	18
nsH15	0	3	2	0	5
nsH17	0	0	1	0	1
TS1014	0	5	7	0	12
Tu9	0	14	11	0	25
Tu10	0	3	3	0	6
Tu11	2	22	40	0	64
Tu12	1	14	8	1	24
Tu13	1	2	0	0	3
Tu15	0	4	0	0	4
Tu16	0	7	1	0	8
Tu22	3	9	3	0	15
TuS01	0	11	3	0	14
TuS05	0	11	0	0	11
TuS14	1	7	0	0	8
TuS15	0	0	7	0	7
TuS42	0	10	1	0	11

B	Population				Total
Haplotype	MCO	PNS	PSM	VCA	
TS1014	0	5	0	0	5
Tu5	3	9	0	0	12
Tu9	0	14	0	0	14
Tu10	0	10	0	0	10
Tu11	2	22	0	0	24
Tu12	1	18	0	1	20
Tu13	1	2	0	0	3
Tu15	0	4	0	0	4
Tu24	0	10	0	0	10
TuB03	4	7	1	0	12
TuS01	0	11	0	0	11
TuS05	0	11	0	0	11
TuS14	1	7	0	0	8

Ten STR loci were characterised for 236 individuals collected from four populations. Expected heterozygosity ranged from 0.3 and 0.66 in VCA and PSM, respectively. The mean total H_E was 0.66 (SD±0.09). The total mean number of alleles was 6.8 with an allelic size range of 7.8. The mean number of alleles ranged from 1.2 (VCA) and 6 (PNS) and allelic size ranged from 1.7 to 6 in the same populations (Table 3.3.3). F_{ST} for the STR data ranged from 0.03 (PSM and MCO) and 0.08 (PNS and MCO; (Table 3.3.4).

Thirty individuals were analysed using GBS. After running the Stacks pipeline, over 280,000 loci were identified with a mean coverage of 15.9X. After filtering with vcftools, one of each of the three duplicate individuals was removed based on coverage and missingness and 11 individuals were removed due to low coverage and high levels of missingness. The final number of SNPs was 4,315 with 19 individuals with a mean level of missingness over the remaining individuals of 3.2% and a mean coverage of 34.1X. In some analyses, one individual was also removed as it was identified as an outlier. Mean H_E ranged from 0.15 in APC and 0.22 in PNS (Table 3.3.3), with a mean total H_E of 0.18 (SD±0.13). No values of F_{ST} for the SNP data were significant (Table 3.3.4).

Table 3.3.3 – Summary statistics showing the gene diversity (*h*) calculated for mtDNA haplotypes, expected heterozygosity (H_E) for the STR and SNP datasets and the mean number of alleles and allelic size range calculated for the STRs.

	h		**H_E**		**Number of alleles (STR only)**	**Allele Size Range (STR only)**
	mtDNA long	**mtDNA short**	**STR**	**SNP**		
MCO	0.85	0.85	0.64	-	4	4.4
PNS	0.92	0.91	0.64	0.22	6	6
PSM	0.79	1	0.66	0.12	5.6	5.4
APC	-	-	-	0.11	-	-
VCA	1	1	0.30	-	1.2	1.667
ATA	-	-	-	0.11	-	-
UNK	-	-	-	0.12	-	-
Total	**-**	**-**	**0.66**	**0.18**	**6.8**	**7.8**

Table 3.3.4 – Pairwise F_{ST} for the long (**A**) and short (**B**) mitochondrial fragments as well as the STR (**C**) and SNP (**D**) datasets shown below the diagonal with the corresponding significance value above.

A	PNS	MCO	VCA	PSM
PNS	-	0.009	0.991	0.991
MCO	**0.088**	-	0.991	0.991
VCA	0.185	0.467	-	0.991
PSM	0.175	-0.136	1	-

B	PNS	MCO	VCA	PSM
PNS	-	0.054	0.991	0.009
MCO	0.077	-	0.991	0.144
VCA	0.124	0.466	-	0.991
PSM	**0.021**	0.038	0.252	-

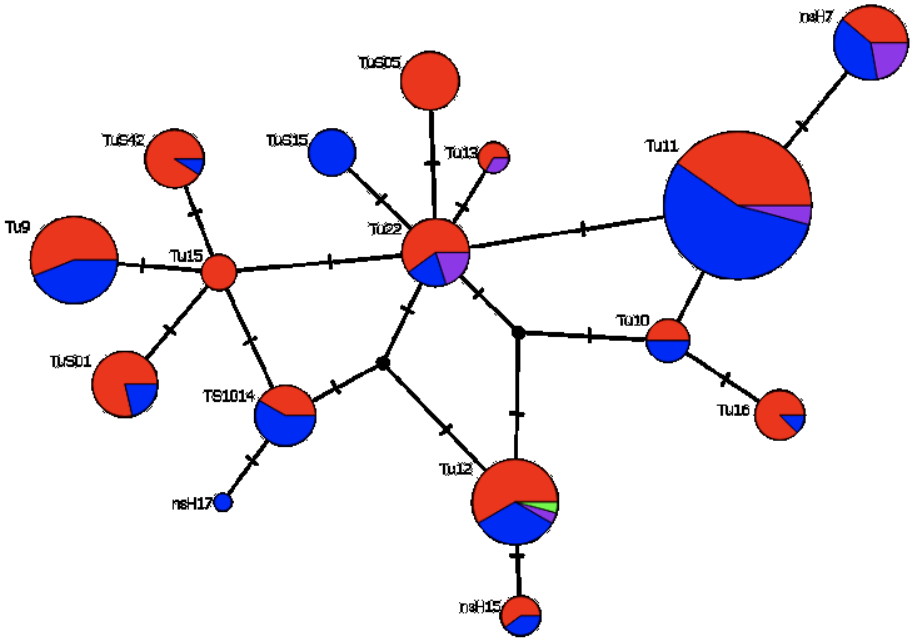
C	PNS	MCO	VCA	PSM
PNS	-	0	0.991	0
MCO	**0.084**	-	0.991	0.045
VCA	-0.015	0.205	-	0.991
PSM	**0.058**	**0.032**	0.161	-

D	PNS	APC	PSM	ATA	UNK
PNS	-	0.991	0.901	0.324	0.523
APC	-0.099	-	0.991	0.991	0.991
PSM	-0.064	0.015	-	0.207	0.991
ATA	0.016	0.058	0.024	-	0.144
UNK	-0.035	-0.009	-0.029	0.028	-

3.3.3. POPULATION GENETIC ANALYSIS

Mitochondrial DNA haplotypes had very similar sequences, and for both fragment lengths, network analysis found no relationships between geographic locality and haplotype. However, the majority of the samples for the short haplotype were collected from PNS, therefore the lack of geographical patterns was not surprising. Table 3.3.2 also confirmed these results for the mtDNA, since haplotypes were present across the sampling area. However, in the long fragment, the haplotype Tu11 appeared to be primarily present in the PSM population.

A


nsH7
TuS05
TuS42
TuS15
Tu13
Tu11
Tu9
Tu15
Tu22
Tu10
TuS01
TS1014
Tu16
Tu12
nsH17
nsH15

B

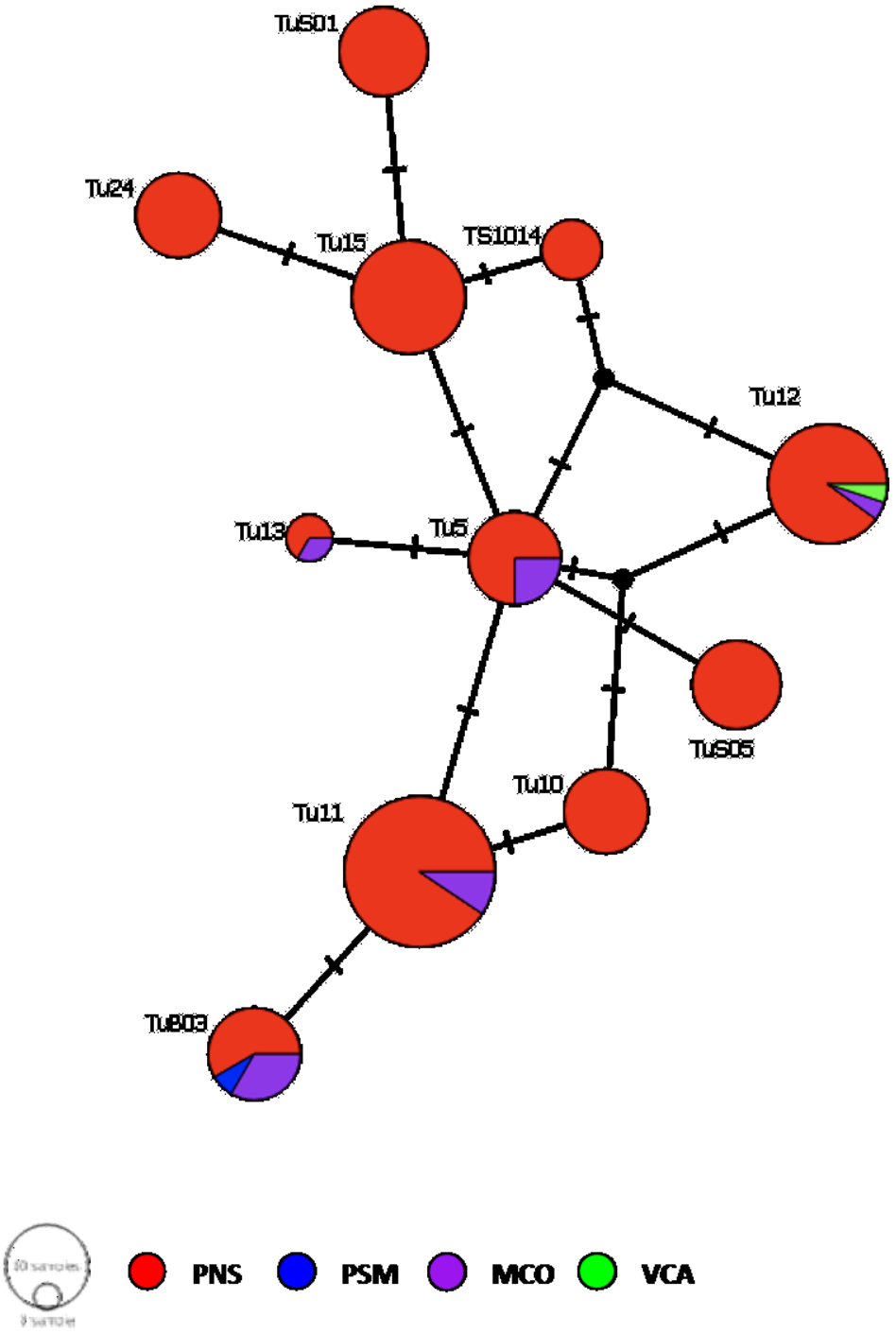


Figure 3.3.2 – Haplotype network for the capercaillie. Both the long (**A**, n=236) and short (**B,** n=144) mtDNA fragments are shown. Each node represents a unique haplotype; node colours represent populations (PNS: Parco Nazionale dello Stelvio; PSM: Parco Naturale Paneveggio – Pale Di San Martino; MCO: Parco Naturale Monte Corno; VCA: Val Campelle). The marks on the vectors connecting haplotypes show the proportional number of base pair differences between each haplotype.

Similarly, the PCA results indicated very little geographical structure in the STR and SNP datasets (Figure 3.3.3). Neither PCA plot showed distinct clusters of populations with the exception of samples from the Austrian Alps in the SNP dataset. However, samples collected to the east of the Adige Valley from PSM (blue) tend to cluster towards the righthand side of the plot while samples from PNS to the west of the Adige valley cluster towards the left. The same geographical structure is found in the DAPC results (Figure 3.3.4): for the STR data, five clusters best represented the data, however, while these clusters are present on both sides of the Adige valley, cluster 5 was primarily represented in the 'western' populations. For the SNP data, the samples were divided into two clusters: the red cluster is not shown on the map as it characterises the individuals from the Austrian Alps (no coordinates available), however, all of the Italian samples were characterised by the blue cluster.

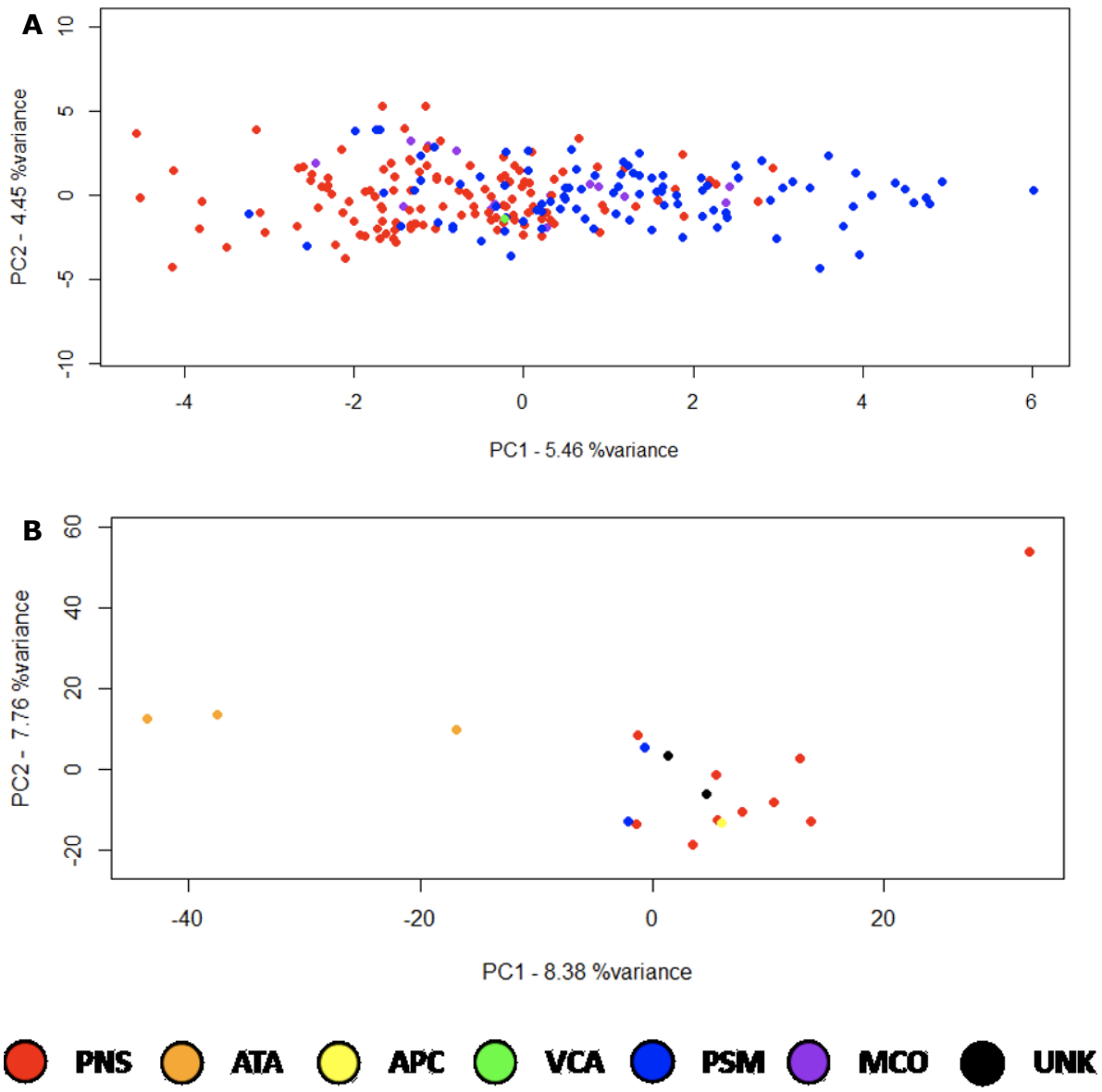


Figure 3.3.3 – Principal Component Analysis (PCA) for the STR (**A,** n=236) and SNP (**B**, n=18) datasets of the capercaillie. Each coloured point represents an individual and their respective population.

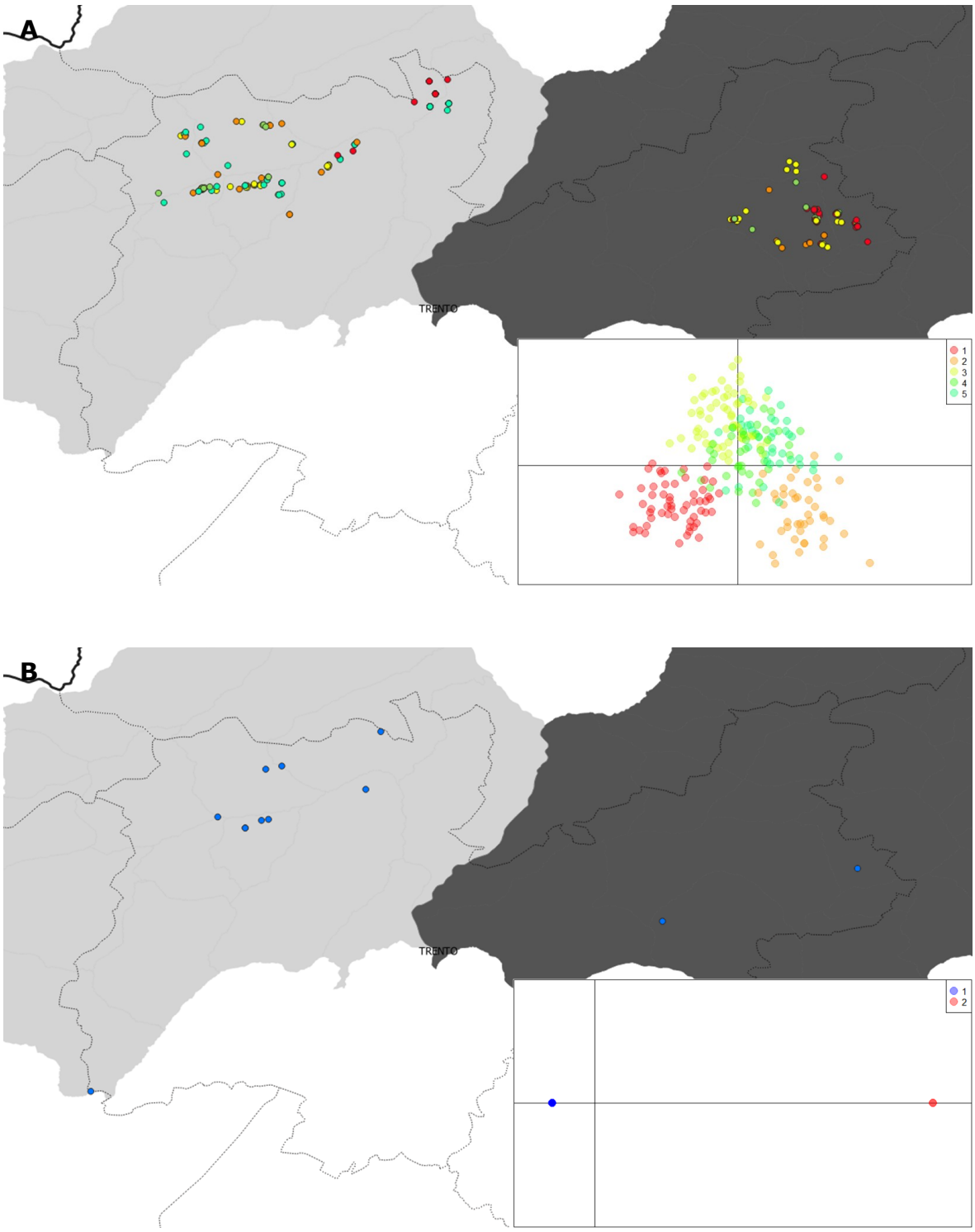


Figure 3.3.4 – Discriminant Analysis of Principal Components (DAPC) for the capercaillie data. Showing the clusters that best represent that data for both the STR (**A**, n=236) and SNP (**B**, n=18) datasets, DAPC plots are shown in the bottom right of each figure. The map shows the geographical origin of each individual from the plot. Due to a lack of coordinates, cluster 2 for the SNP dataset is not shown on the map, these individuals are from the Austrian Alp population (ATA – no coordinates given).

Although K=1 best represented the data, a value for Delta K is not given at this number of clusters, so the STRUCTURE plots for both K=2 and K=3 are shown here to evaluate any underlying patterns in the study area (Figure 3.3.5). The STR results show a high level of admixture between populations and no single cluster represented a population. Samples from PSM are, however, primarily represented by one cluster (yellow) different from those of the PNS population. For the SNP analyses, a single sample from PNS separated out from the rest at K=2 (Green cluster); this sample was found to be an outlier in other analyses and was removed from these, thus this individual was likely misidentified upon collection. All other samples primarily correspond to cluster 2. At K=3 the outlier individual represents its own cluster (green), however, the other samples all primarily represent cluster 2 (yellow) with five individuals showing admixture with cluster 3 (purple).

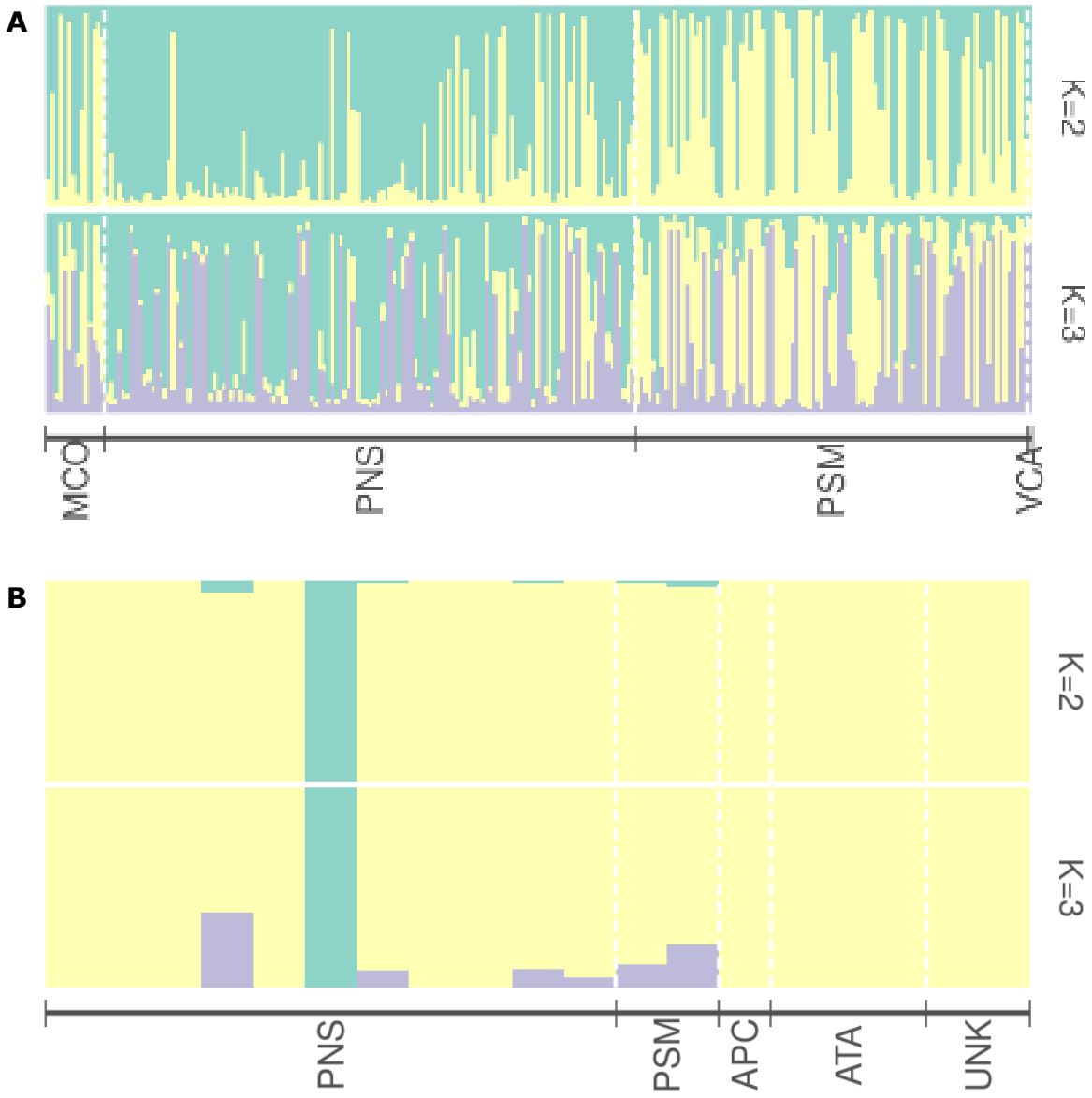


Figure 3.3.5 – STRUCTURE plot for the capercaillie STR (**A**, n=236) and SNP (**B**, n=19) datasets. K=2 and K=3 are shown for both datasets; dashed vertical lines separate the populations. Plots were visualised in Pophelper version 1.0.10 (Francis, 2016).

For both the STR (86.6%) and SNP (60.4%) AMOVA results the primary source of variation was within individuals; these results were significant for both datasets (Table 3.3.5). Both STR and SNP datasets also showed that the second greatest cause of variation was among individuals within populations (9.5% and 58.1% respectively). The lowest source of variation came from among groups for the STR and among populations within groups for the SNPs. However, this result was not significant for either dataset.

Effective population size (Table 3.3.6) ranged from 4.5 (MCO) to 82.4 (PNS) for the STR dataset with an overall estimated N_e of 52.6. For the SNP dataset, the estimation was only possible for PNS (0.4).

Table 3.3.5 – AMOVA for the capercaillie STR and SNP datasets. 'Groups' for this analysis were defined as the populations east and west of the Adige valley. Genetic structure was estimated among groups (Va/FCT), among groups within populations (Vb/FSC), among individuals within populations (Vc/FIS), and within individuals (Vd/FIT) using Arlequin (Version 3.5).

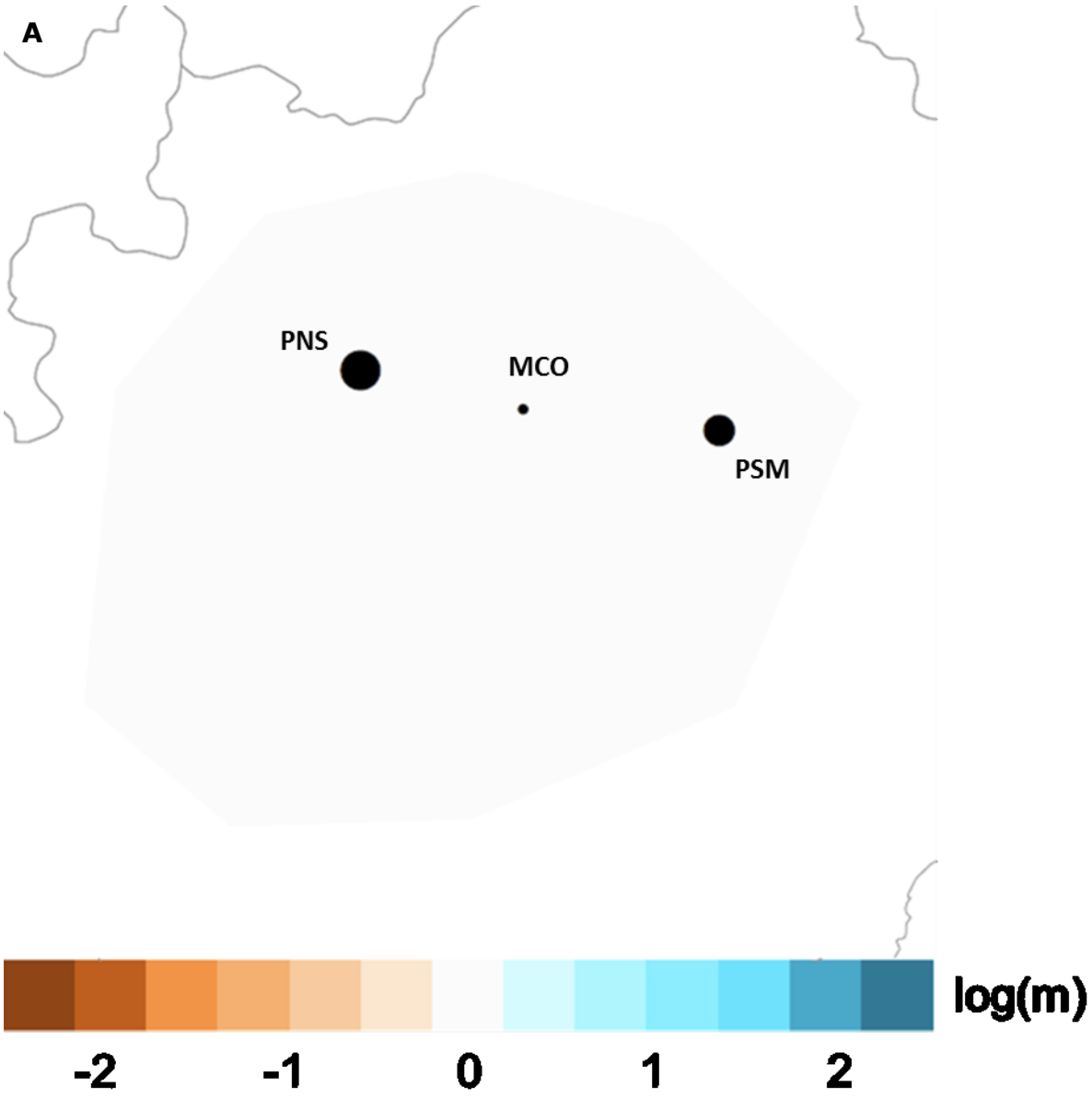
Marker	Variation Source	Sum of squares	Variance components	Percentage of variation	F Index	P-value
STR	Among groups	26.79	-0.034	-1.08	-0.011	0.73
	Among populations within groups	14.23	0.15	4.95	0.049	0
	Among individuals within populations	747.38	0.29	9.49	0.10	0
	Within individuals	625.00	2.69	86.64	0.13	0
	Total	1413.40	3.10			
SNP	Among groups	310.77	35.65	9.07	0.40	0
	Among populations within groups	297.95	-108.43	-27.59	0.49	0
	Among individuals within populations	7223.73	228.45	58.13	-0.30	1
	Within individuals	3162.50	237.35	60.39	0.09	0.001
	Total	10994.94	393.019			

Table 3.3.6 – Effective population size (N_e) estimates calculated for the capercaillie STR and SNP datasets with 95% confidence intervals (CI). Inf: Infinite.

Population	STR		SNP	
	N_e	95% CI	N_e	95% CI
MCO	4.5	2.4-10.3	-	-
PNS	82.4	60.3-118.7	0.4	0.4-0.4
PSM	23.3	18.3-29.8	Inf	Inf
APC	-	-	Inf	Inf
VCA	Inf	Inf	-	-
ATA	-	-	Inf	Inf
UNK	-	-	Inf	Inf
Overall	**52.6**	**43.7-63.5**	**0.4**	**0.4-0.4**

For the capercaillie, the IBD tests were conducted both with the original data, and edited negative or non-significant values of F_{ST} which were set at 0. Samples from UNK, as well as from ATA and VCA were removed from the analyses. None of the IBD tests found a significant pattern, therefore only the results for the original F_{ST} values will be discussed. For the edited IBD tests, all F_{ST} values for the SNP dataset were changed to 0, therefore, the tests could not be conducted. For the STR dataset, all F_{ST} values were significant and positive, therefore there was no difference in the results between the original and edited datasets. For the original F_{ST} IBD tests, the SNP data had an IBD value of R=0.99 (p=1.7), for the STR R=-0.19 (p=0.67). The long fragment of mtDNA also gave a negative R-value of -0.84 (p=1) and for the short fragment R=0.64 (p=0.33).

The STR EEMs results do not suggest any variable rates of migration in the study areas; however, for the SNP dataset, the results show low levels of migration around the PNS population and high levels of migration around the PSM population (Figure 3.3.6).


A
PNS
MCO
PSM
log(m)
-2
-1
0
1
2

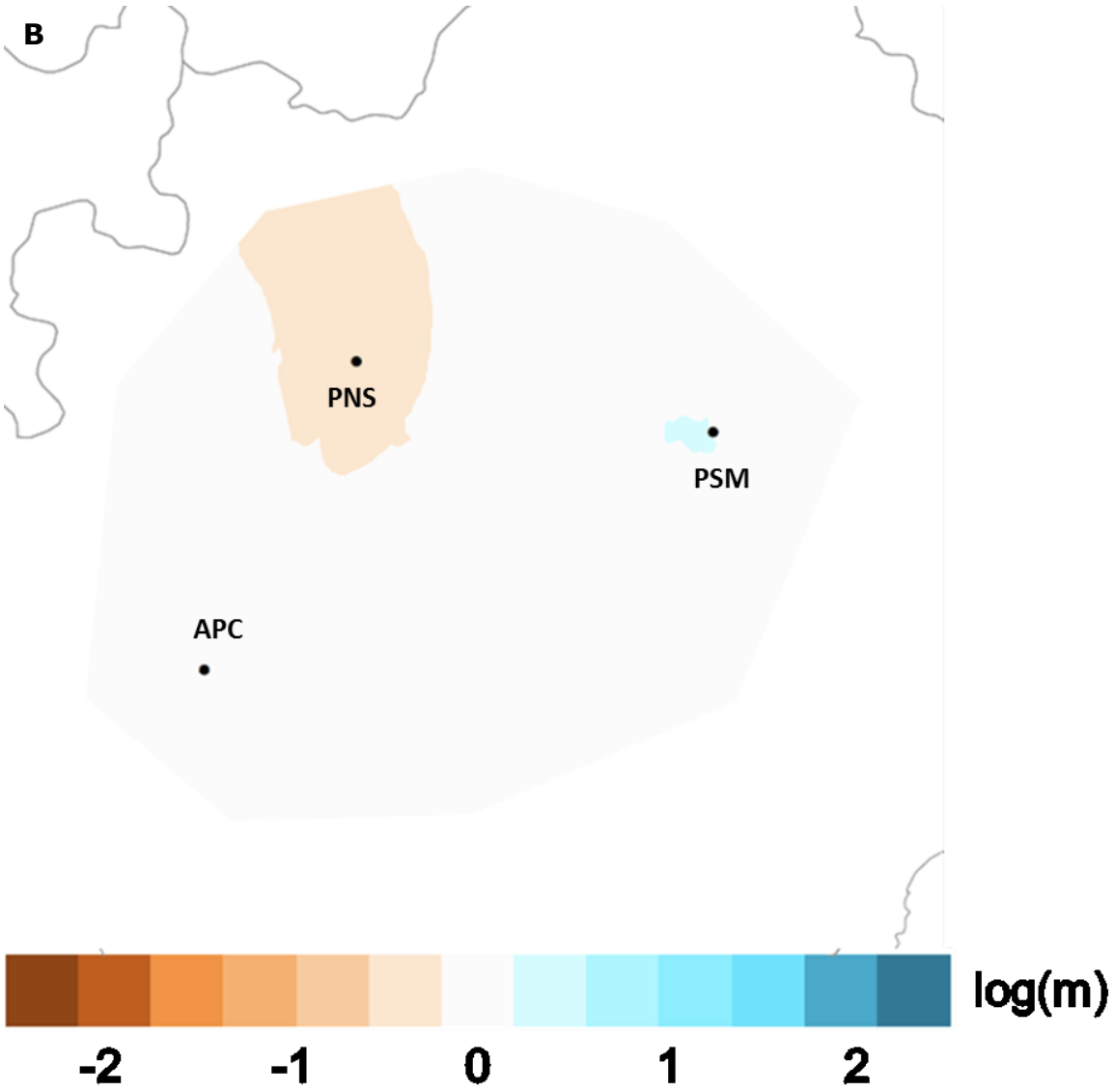


Figure 3.3.6 – Posterior mean migration rates (m) on the log10 scale for the STR (**A**) and SNP (**B**) capercaillie datasets. The shading represents effective rates of migration and each point indicates the mean centre of the population. Calculated using Estimated Effective Migration Surfaces (EEMS).

4. DISCUSSION

For this thesis, a total of 1,298 invasive and non-invasive samples from three galliform species were collected across the Italian Alps over a 20-year period. These samples were analysed with previously published mtDNA primers and microsatellite markers in conjunction with modern whole-genome SNP data, to investigate their past and current genetic structures, and the climatic, environmental and anthropogenic factors that may impact them. Results from the different marker types were compared to confirm these genetic patterns, and also to assess their suitability for conservation genetic applications. Unexpectedly for these game birds (subject to both legal and illegal hunting), results indicate that although disturbance affects patterns of genetic variability, Alpine populations of all three species are well-connected, suggesting that dispersal abilities were previously underestimated and current management strategies are adequate to preserve both species and genetic diversity. These results are also of immediate practical use for improving management and conservation strategies for the species across their Italian distribution.

4.1. TECHNICAL CONSIDERATIONS

This study has successfully produced SNPs using GBS methods on tissue and feather samples: almost 30,000 SNPs were produced for the black grouse with a final coverage of more than 25X (out of 400 invasive samples, only 17 had to be discarded due to low DNA quantity/quality); 15,000 SNPs with a coverage of almost 30X for the ptarmigan (158 individuals with 15 invasive samples removed) and just over 4,000 SNPs with more than 34X coverage were produced for the capercaillie (although 11 out of 30 noninvasive feather and faecal samples could not be used). For the most part, these numbers of SNPs are consistent or greater than those used in other studies. Zimmerman *et al.* (2020) conducted their research on the Gunnison sage-grouse (*Centrocercus minimus*) with approximately 15,000 loci; Leyhausen *et al.* (2022) used over 24,000 SNPs in their study of the hazel dormouse (*Muscardinus avellanarius*) obtained using GBS, and Peters *et al*. (2016) used a double digest-RAD sequencing approach to produce over 3,000 SNPs for mottled ducks (*Anas fulvigula*).

A serendipitous result of this project has been to confirm the importance of choosing the correct enzyme for GBS projects, as demonstrated by the differences in coverage produced by two different companies using the black grouse test samples. After filtering the data, a total of 28,752 SNPs were produced for the black grouse dataset typed by the NOVOGENE company (enzymes: MseI and HaeIII) with a coverage of 27.1X. However, prior to the resequencing efforts by the company (on our insistence) to improve this dataset, only 2,442 SNPs were produced during the first attempt with a coverage of 3.9X. In contrast, the black grouse pilot samples produced 12,736 SNPs with 25.6X coverage, typed using the GBS services Cornell University (enzyme: EcoT22I). These results illustrate the importance of the restriction enzymes in the initial GBS steps. Therefore, it is recommended that GBS studies trial several possible enzymes on a small subset of samples, then use a larger trial dataset to ensure that an appropriate enzyme has been chosen, before investing in GBS of the entire sample set. As population genomics and whole-genome sequencing of wild species becomes more common, it is predicted that enzyme choice will be less uncertain, as more information on the frequency of restriction enzyme cut-sites on these genomes becomes available.

One of the main aims of this project was to show that DNA quality and quantity for non-invasive samples met the threshold for GBS analysis. While only 4.2% (17/400) black grouse and 9.4% (15/158) ptarmigan samples did not meet the threshold (even though many of these samples had been collected and archived at -80°C up to 22 years previously), 36.7% (11/30) noninvasive feather and faecal samples of capercaillie were unusable, despite repeated DNA extractions. For future GBS studies, DNA quantity and quality for non-invasive samples should be improved. Dai *et al.* (2015) emphasise the need to collect non-invasive samples which are as fresh as possible to improve DNA yield, and they did not recommend the use of faecal samples if a high yield is needed; however, for this study on elusive species, sample collection relied on the collaboration of volunteers that were trained to recognize and collect the freshest samples, but in any case, very fresh samples are rare. This same author recommends swabbing freshly laid eggs, which gave the best yield and quality of DNA for females (although there is a risk of cross-contamination with the male). In fact, since this thesis began, swabs are being used much more extensively for non-invasive genotyping from faecal pellets (see also Vallant *et al.,* 2018), but swabbing eggs of these ground-dwelling birds, although relatively simple, would not

be recommended as they are easily disturbed (Moss *et al.,* 2014). Many other studies have focused on optimising the DNA yield and data quality from non-invasive samples such as faeces, feathers and eggshells by adjusting DNA extraction protocols or by using novel amplification techniques, such as a primerless PCR developed by Peters *et al.* (2020). The modifications to existing extraction protocols typically consist of increasing the volume of sample used (e.g. for feathers: Vallant *et al.* 2018), increasing the digestion time, adjusting temperatures of reagents to optimise their function, and adjusting the volume of reagents to maximise the yield of DNA (De Volo *et al.,* 2008; Costa *et al.,* 2016). However, very recently, Cumer *et al.* (2021), found that small quantities of DNA are more suitable for double-digest RAD sequencing techniques, but also that when the quantity of DNA was increased to 250ng the probability of experimental failure also increased. Therefore while most DNA extraction of samples used in this study occurred prior to the start of this thesis, the above suggestions should be taken into consideration for future research. However, the techniques purported to improve DNA yield need to be tested further for their feasibility in SNP studies.

4.2. H1: SNP DATASETS ARE CAPABLE OF PRODUCING FINER-SCALE PATTERNS COMPARED TO THE TRADITIONALLY STR MARKERS

In general, the SNP datasets produced here gave clearer and more detailed geographical patterns than STR data, although results from the two markers did not contrast with one another. For example, the PCA results for both the **black grouse** and **ptarmigan** SNPs show clear geographical patterns across the Alps with samples from neighbouring mountain groups clustering together (Figure 3.1.3 and Figure 3.2.3). PCA results from both species also clustered the data into two groups, and samples from each mountain group clustered together. The same is the case for the **capercaillie** data, however, clustering of individuals was less defined; this may be in part due to the lower number of individuals used and the smaller geographical area covered in these analyses. The DAPC results, which do not require any prior location information, also confirm the results of the PCAs, with two groups present in each SNP analysis. In comparison, the DAPC for STR datasets showed a very weak geographical structuring across the three species and the mtDNA results for this project did not show any geographical structuring at all. The more detailed patterns observed with the SNP data indicate that genomic

techniques such as GBS provide more appropriate baseline data for observing changes in genomic patterns due to environmental changes or human harvesting pressure, as well as management intervention, and should be considered for monitoring more species of conservation concern, including endangered species.

For many species of conservation interest, STR and mtDNA markers were established as relatively cheap and rapid genetic monitoring for at least two decades; they are easily optimised for non-invasive DNA, and can be used to answer questions regarding phylogeny, recent changes in population size and individual identification (Allendorf, 2017). However, a large set of SNPs from across a species' range would potentially allow for additional analyses to be conducted for conservation purposes. For example, in order to study how and if populations are adapted to certain environments, mapping outlier SNPs to an annotated reference genome would enable investigation into adaptive traits, if the outliers can be linked to a specific gene with known function. These techniques could be applied to many endangered species to aid in their conservation, although this would require reference genomes to be made available for a wider variety of species, as is the focus of several global initiatives such as the European Reference Genome Atlas (ERGA, n.d.). Although a genome assembly has been produced for the black grouse by Wang *et al.* (2014), and this fragmented draft genome is still incomplete, it has already been used by Kozma *et al.* (2019) together with the chicken genome to investigate adaption in three species of grouse (including the rock ptarmigan) using outlier loci. This study found that outliers corresponded to seven genes relating to stress response as well as the development of the limbs, and olfactory, gut and neural systems. These techniques would be beneficial for the populations studied here, in comparison to other individuals in the species range, to investigate the differences in adaption in different areas of the distribution. A complete reference genome would also be useful in future studies to align contigs against a reference rather than *denovo* as this can increase the number of SNPs identified (Fuentes-Pardo and Ruzzante, 2017). Reference genomes would also allow for a variety of additional genetic applications that can contribute to the field of conservation genetics, such as the development of SNP arrays. This technology enables rapid individual genotyping and the identification of hybrids, as well as measures of population genetic diversity. The development of SNP arrays (and relevant bioinformatic pipelines) could improve the accessibility of conservation managers to

SNP analyses and enable the method to become widely used across a species' distribution (Brandies *et al.*, 2019).

4.3. H2: BLACK GROUSE POPULATIONS WILL BE ISOLATED BY DISTANCE, WITH EVIDENT IMPACTS OF HEAVILY URBANIZED VALLEYS, AND LOSS OF GENETIC VARIATION THROUGH TIME FROM OVERHUNTING.

The STR patterns for the black grouse found here are consistent with those reported in previous research on the species; for example, a significant pattern of IBD was also shown by Caizergues *et al.* (2003a) in both the Alps and Finland, Lebigre *et al.* (2008) in Finland and Sittenthaler *et al.* (2018) in the Austrian Alps, suggesting this pattern is present in both the continuous northern and fragmented southern limits of the black grouse's distribution. Both the ranges of population pairwise F_{ST} (0-0.12) and the mean number of alleles (0.16-8.4) were also similar to those found in other studies of Alpine black grouse (Caizergues et al., 2003a; Sittenhaler et al., 2018). The studies by Lebigre *et al.* (2008) in Finland and Corrales and Höglund (2012) in Sweden found mean allele numbers to be 10.3 and 9.2, respectively, which again are similar results to the mean total number found here (11.1). The study by Lebigre *et al*. (2008) also showed low levels of genetic variation with F_{ST} ranging from 0-0.016. The high level of H_E (0.74) found in this study is also within the range found in other studies of black grouse, both in the Alps and the species wider range (0.66-0.76; Caizergues *et al*., 2003a; Corrales and Höglund, 2012; Höglund *et al*., 2007; Lebigre *et al.*, 2008; Rutkowski *et al.*, 2018; Sittenthaler *et al.*, 2018). Overall this suggests that, despite the purportedly fragmented nature of the Alpine populations, there is no evidence of reduced genetic diversity or genetic isolation in these populations of the black grouse in comparison to more continuous populations in northern Europe.

The SNP PCA, DAPC and STRUCTURE analyses for black grouse all indicated that this species is divided into two genetically distinct clusters. The geographical location of the division between these two clusters runs north-south between the mountain groups LUG-LEP and ROC-BER, along the northern arm of Lake Como and Val Chiavenna (Province of Sondrio, Lombardy). While Lake Como itself and two state highways running either side could contribute to a barrier between the dispersal of individuals, this distance (c. 10km) is less than the maximum known dispersal distance of the species in the Alps (29km; Caizergues and Ellison, 2002),

and Val Chiavenna is not particularly wide or urbanized. Therefore, the capacity of the black grouse to disperse between the two clusters cannot explain the reason for them.

Although the AMOVA analysis found that only a small proportion of variation was represented by the 'east' and 'west' groups analysed (STR: 4.97% and SNP: 4%; and a similar result (3.1%) was found between subpopulations in Sittenthaler et al (2018) in a study of black grouse STR in Austria, a clue to the reason for these two clusters may lie in the fact that the same north-south boundary between them also denotes the 'western' and 'eastern' Alps, two tectonic units which differ in the age of their formation (Figure 4.1; Schmidt *et al.,* 2004; Marazzi, 2005). One reason for these two genetic groups may be that various environmental factors have led to local adaptation to these two Alpine areas. Alternatively, there may be a lack of preferred habitat in this contact area. For example, Figure 4.2B shows that where the black grouse divides into two clusters, there is higher precipitation than the rest of the Alps. Summers *et al.* (2004) and Viterbi *et al.* (2015) found a decrease in black grouse productivity with an increase in rainfall. This hypothesis is supported by the results of the RDA, where June rainfall was found to be a key environmental factor affecting the genetic variation based on the variables tested. Therefore, this area of increased rainfall may lead to a decrease in breeding success, lowering the density of the species in this area. Low densities, or 'density troughs', may cause a barrier to gene flow, genetically segregating these populations from one another (Barton and Hewitt, 1985). Finally, the largest area outside the species' preferred altitudinal range also occurs at Lake Como (Figure 4.2A; Figure 4.2C); since altitude and urban areas show a significant pattern of IBR for both of these barriers; the lake and unfavourable altitude combined might inhibit the crossing of individuals. However, the IBD results are also significant. Therefore, the results of the IBR may be an artefact of isolation by distance patterns rather than an indication of a true barrier, as when resistance values were adjusted, the significant IBR result remained. More intensive sampling and genomic analysis focussed on this geographical area of interest, could help to resolve the conundrum and determine whether these two clusters should be considered separate MUs.

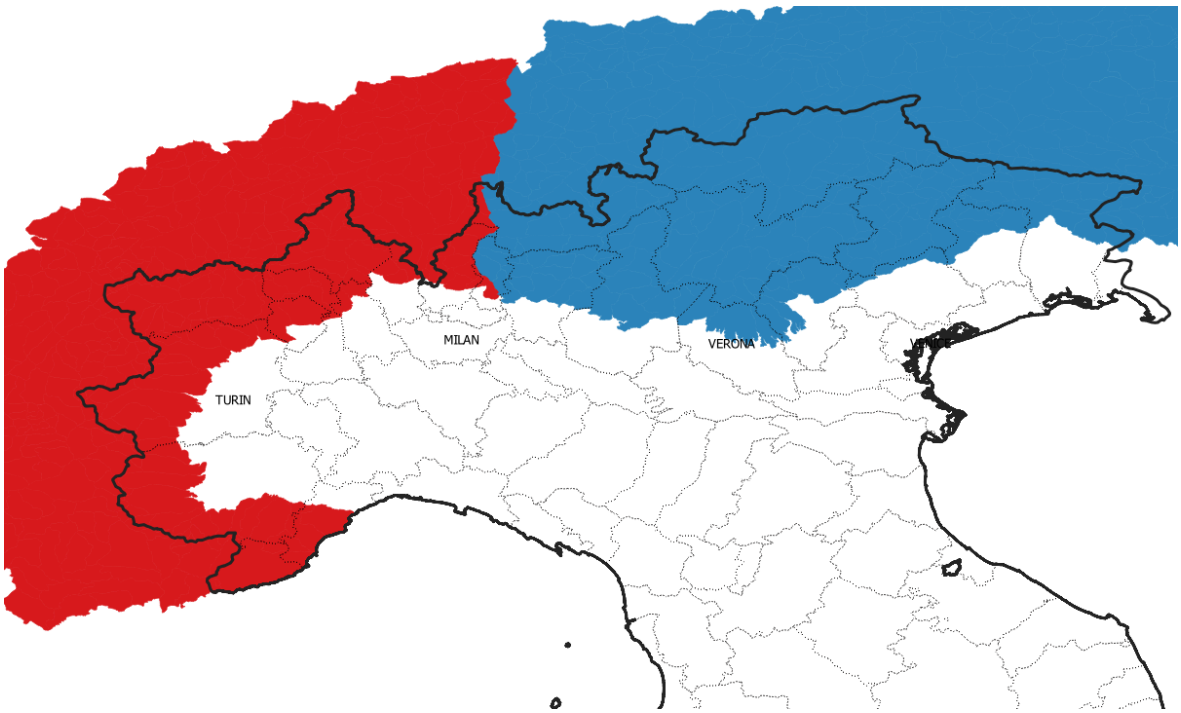
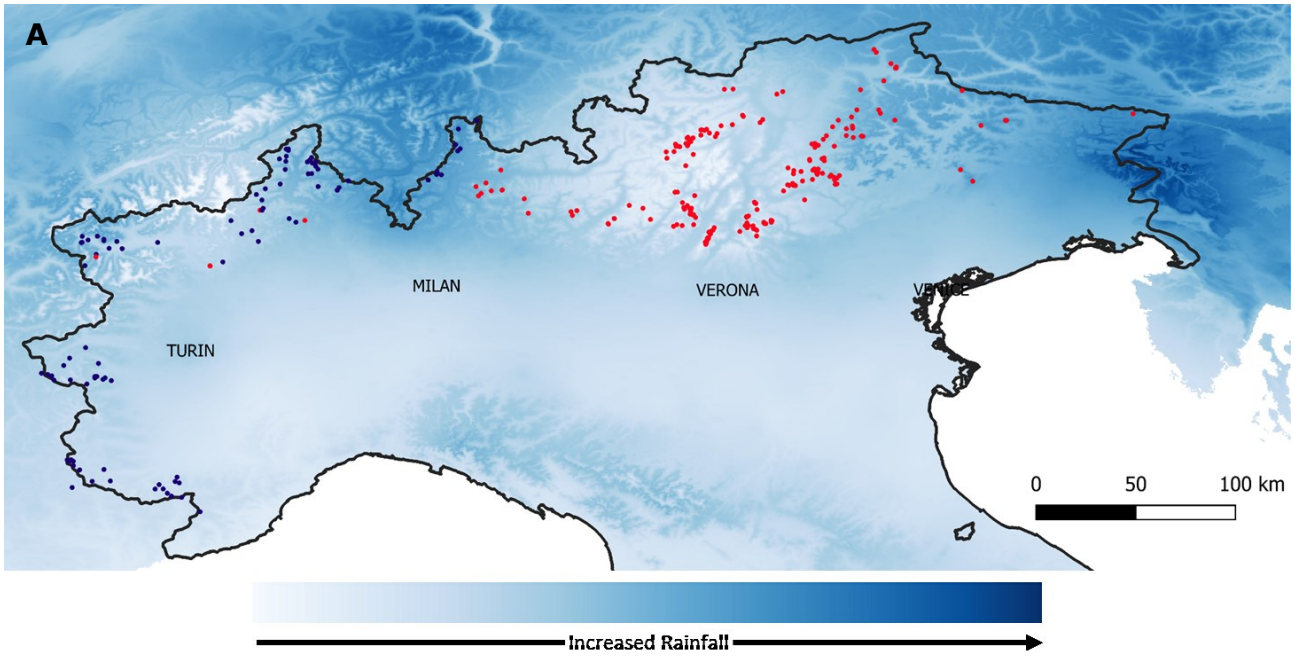


Figure 4.1 – The primary partition of the Alps following the International Standardized Mountain Subdivision of the Alps (SOIUSA). The 'western' Alps (Alpi Occidentali) are shown in red, while the 'eastern' Alps (Alpi Orientali) are shown in blue as defined in Marazzi (2005). Created in QGIS (QGIS Development Team, 2021) with data from Accorsi (n.d.).



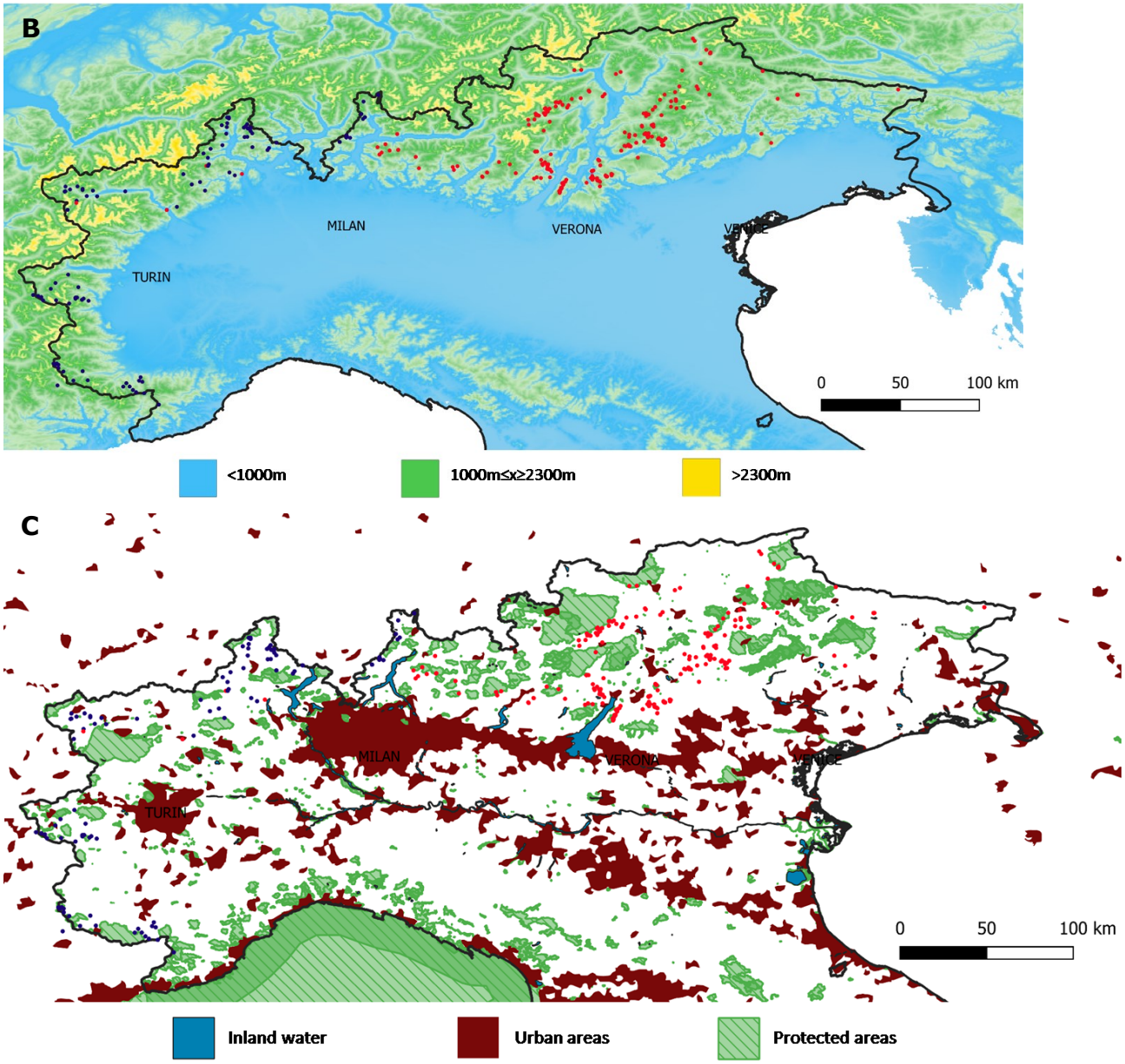


Figure 4.2 – Potential environmental factors affecting the genetic variation of Alpine grouse. The 'east-west' division proposed occurs along the northern arm of Lake Como and Val Chiavenna, north of Milan. The red and blue dots in each map represent the sampling points of the black grouse SNP and their corresponding DAPC group, as shown in Figure 3.1.4B. These points are included so that the locality of the 'east-west' divide is more visible. **A** shows the average precipitation levels in June in the Alps, with the intensity of colour increasing with greater rainfall. **B** shows the altitude across the Italian Alps, the green areas show the preferred range of the black grouse, blue represents below and yellow represents the altitudes above the preferred range. **C** shows protected areas represented in green, inland bodies of water in blue and urban areas represented in brown. Maps created in QGIS (QGIS Development Team, 2021), with data from DIVAGIS, WorldClim (Fick and Hijmans, 2017), Natural Earth, and WDPA and WD-OECM.

In fact, other analyses such as the PCA (STR) and IBD (STR and SNP) suggested a pattern of isolation by distance, or a continuous population across the Alps. Using radiotracking methods, Caizergues and Ellison (2002) estimated that the natal dispersal of female black grouse was up to 29km (mean 8km), and males were found to be more philopatric in nature (mean 1.5 km). However, since many of the populations in the present study were greater than 30km apart, and the genetic and genomic data suggest gene flow between most of them is high, the maximum dispersal distances recorded by Caizergues and Ellison may provide a more accurate representation of black grouse dispersal in the Alps. This also implies that Alpine populations that are classified as fragmented, may actually have a relatively high connectivity. This is confirmed by the temporal analysis of black grouse data, which suggested that genetic variation has remained stable over the last 20 years, representing multiple generations; a loss of genetic diversity would have indicated isolated populations, which was not the case here. It is recommended that dispersal distances and gene flow are monitored to prevent future isolation of populations.

4.4. H3: PTARMIGAN POPULATIONS ARE GENETICALLY ISOLATED DUE TO THE SPECIES PREFERENCE FOR HABITATS AT HIGH ALTITUDES (OVER 1,800M) AND RELATIVELY LOW DISPERSAL DISTANCES.

For the ptarmigan, STR markers suggested that this species is present in a continuous panmictic population across the Italian Alps; similarly, mtDNA results suggested little correlation between sampling location and genetic variation. No pattern of IBD is shown with this dataset, consistent with the studies by Holder *et al*. (1999) and Pruett *et al.* (2010) for populations in Canada and Alaska. Values of H_E (0.6-0.78) were within the range of those found by Pruett *et al.* (2010), who used STRs to evaluate populations of rock ptarmigan from the Aleutian-Commander archipelago (H_E=0.45-0.89). The results of H_E and mitochondrial gene diversity found in this study (*h*=0.6-0.87) were also in the range of the results found by Quintela *et al.* (2010; 0.653 and 0.655, respectively). The levels of haplotype diversity found by Holder *et al*. (2004) for Nearctic rock ptarmigan (0.83) were also were within the range found by this study. This suggests that the Alpine populations studied here have a similar level of genetic diversity compared to other populations in the more continuous parts of the species range, contrary to the results from Holder *et al.* (1999) where mean mitochondrial gene diversity was found to be 0.235 in the

individual Aleutian Islands populations studied in Alaska.

In contrast to STRs, the SNP results for the ptarmigan show a significant pattern of IBD, confirmed by the PCA where most populations cluster together with their neighbours. However, the first two axes of the PCA account for a low percentage variation, supported by the low H_E for this marker, signifying low levels of genetic variability which are indicative of a single mixing population. Therefore, the genetic variability of the ptarmigan in the Alps should be monitored periodically to prevent further decrease in variability, which could lead to the eradication of this species in this area. In addition, F_{ST} values between the Alps and the northern populations (Iceland and Sweden) were greater than 0.18 and the F_{ST} calculated between Sweden and Iceland was 0.4, suggesting that these populations are more genetically distinct from the Alpine populations and part of different MUs.

Similarly to the black grouse, the PCA and DAPC plots for the ptarmigan showed a genetic differentiation between the 'eastern' (ROC to CDG) and 'western' (MAN to LEP) parts of the species' Alpine range. The ptarmigan's greater dispersal distance (mean 20.3km; Nilsen *et al*., 2020) means that geographical separations such as these pose less of a barrier, however, the distance measured between two sampling locailities was approximately 87km indicating that the dispersal distances may have been previously underestimated in this area. Unfavourable conditions to dispersal such as the A2 motorway in Switzerland, nearby cities (Como and Lugano), large bodies of water (Lake Maggiore, Lake Lugano and Lake Como), as well as high numbers of ski resorts and tourist destinations, may mean the actual distance that an individual needs to fly around these disturbances is even greater than the direct distance measured. However, in the case of the ptarmigan, the area separating populations occurred where there is also a sampling gap. As the SNP results suggest that samples are in isolation by distance, the most likely cause for this division is a lack of samples across the entire LEP mountain group. Further confirmation of this conclusion is that none of the three markers suggested that ptarmigan in the Alps are present in small isolated groups.

4.5. H4: CAPERCAILLIE POPULATIONS ARE GENETICALLY ISOLATED EVEN WITHIN THE TRENTINO ALTO-ADIGE REGION, WITH DISTINCT PATTERNS PRESENT ON EITHER SIDE OF THE ADIGE VALLEY.

In this study, the capercaillie was primarily studied in the Italian Region of Trentino-Alto Adige, as a result of previous research funded by associations and parks interested in knowing more about this elusive species, and thanks to the willingness of various volunteers to collect many non-invasive samples from across the territory. All three marker types showed low levels of variation than ptarmigan and black grouse: H_E for the STR dataset was 0.66, consistent with other studies reporting a range of 0.52 to 0.66 (Segelbacher *et al.*, 2003; 2007; 2008; Regnaut *et al.,* 2006). Duriez *et al.* (2007) studied Eurasian populations of the capercaillie using mtDNA and found similar levels of haplotype diversity (0.5-0.93) to those found in this study (0.79-0.92). The Alpine population tested by Duriez *et al.* (2007) had a haplotype diversity of 0.93, while in Finnish populations, *h* was 0.77 (Liukkonen-Anttila *et al.* 2004). However, F_{ST} values were also low for all markers, with no significant values exceeding 0.1; in addition, the majority of the pairwise genetic distances were not significant and no marker found the data to be in isolation by distance. These results suggest that capercaillie populations are not isolated in Trentino-Alto Adige and geneflow is continuous. This result is contrary to those of Vernesi *et al.* (2016) for five mammal species using mtDNA and STRs. The authors found that the Adige Valley represented a strong barrier both past and present, for four of the five Alpine species investigated, hypothesizing that glaciers occupying the valley during the Last Glacial Maximum, and now dense human infrastructure has separated populations. These patterns are not present in the results for the capercaillie studied here: populations east and west of the Adige Valley were grouped together for the AMOVA analysis which found that variation between individuals best represented the data and the east-west groups proposed representing less than 10% of the variation for each dataset. Similarly, the PCA and STRUCTURE for the STR dataset show only a very weak east and west geographical clustering, and DAPC, SNP STRUCTURE and PCA indicate no geographical structure is present in the study area. The distribution of sampling may be a factor in the patterns shown in this data, as the majority of samples were collected from Parco Nazionale dello Stelvio and in the cases of the short mtDNA and SNP datasets only one and two individuals respectively represented Parco Naturale Paneveggio – Pale Di San Martino. This sampling bias prevented any geographical structure from being examined fully.

4.6. ADDITIONAL CONSERVATION AND MANAGEMENT IMPLICATIONS

Multiple studies have demonstrated the effects of climate change on grouse populations. Brommer (2004) investigated the locations of range margins in Finland for a 12-year period and hypothesised that the changes observed were due to climate change. Climate change may pose a problem for the ptarmigan in the future, as the species is primarily found at high altitudes where increasing mean temperatures are predicted to shift treelines to higher altitudes, decreasing the habitat available for the species, and leading to the extinction of many populations, and the isolation of others. In a study by Revermann *et al.* (2012) the authors concluded that a temperature rise of over 4°C could result in a decrease in the ptarmigan habitat of up to two-thirds by 2070. In addition, Kozma et al. (2016) noted that, when comparing historic climate fluctuations with effective population size, they observed that as temperatures increase, the N_e of the ptarmigan decreased. Therefore, it is important to maintain the corridors that, based on the results of this study, appear to still be present between populations of this species, as well as to protect the remaining habitat from disturbance. Black grouse may also be affected: Ludwig *et al.* (2006) noted that black grouse populations studied did not always adjust their annual reproductive activity to fit the changing environment, which under climate change scenarios, could potentially lead to a mismatch between breeding and optimal rearing conditions. While Viterbi *et al.* (2015) suggest that climate change is unlikely to be the main limiting factor for the black grouse populations studied, they do recommend that human disturbance activities should be limited in years with unfavourable weather conditions. As with the black grouse and ptarmigan, rainfall was negatively correlated to capercaillie productivity (Summers *et al.*, 2004), however, a study by Wegge and Rolstad (2017) found that breeding success increased with warmer temperatures, the authors speculated that this was due to the relationship between climate and food abundance. Viterbi *et al*. (2015) recommend that hunting quotas should be reduced when populations are expected to decline to prevent local extinctions due to unfavourable conditions.

Outside of protected areas, hunting regulations play a key role in the conservation of a species and its ecosystem. In Italy, hunting protocols are currently managed at regional and provincial levels. For the Autonomous Province of Trento, species are

divided into two categories: *contingentate* and *non contingentate* (quota and non-quota). Quota species are hunted with specific hunting programs based on the number of animals that can be killed in a hunting season; non-quota species are subject to daily bag limits. The black grouse and ptarmigan fall into the quota category for the Autonomous Province of Trento. Hunting limits are determined on the basis of numbers and dynamics of the species populations. Spring and autumn data is collected through censuses in collaboration with hunters (Provincia Autonoma di Trento, n.d.). The temporal analysis conducted in this study implies that the hunting of black grouse in the Autonomous Province of Trento is not reducing genetic diversity and the measures that are in place to protect diversity are effective. Spanò and Salvidio (2012) also showed, using censuses, that the abundance of black grouse remained constant over the 31-year period in the French Maritime Alps, despite changes in hunting pressures. However, Gregersen and Gregersen (2009) found declines in hunting bags since the 1970s which they proposed was due to climate changes and forestry management. Tracking effective population sizes in relation to climate, as shown by Kozma *et al.,* (2016), and using this data in conjunction with the censuses produced each year could enable more accurate hunting bag counts and quotas to be determined. This would be more beneficial towards the conservation and maintenance of grouse populations. N_e was calculated for this study and, for a large proportion of the populations assessed across all three species (overall estimates for STR and SNP respectively for the black grouse: 234.7 and 121.8, ptarmigan: 321.1 and 490, and capercaillie: 52.6 and 0.4), estimates were similar to or greater than those calculated for other avian species and galliforms. In Athrey *et al*. (2018) for the red junglefowl (*Gallus gallus murgha*) N_e was calculated as 7.3 for STR and 0.2 for SNP, for the lesser prairie-chicken (*Tympanuchus pallidicinctus*) Pruett *et al.* (2011) estimated N_e of 57.6 and 69.4 for STR, and Vázquez *et al.* (2013) estimated an N_e of 17 using STRs for the Cantabrian populations of capercaillie. This indicates that the Alpine populations examined here require a larger population size in order to maintain the current genetic diversity. This should be taken into account in the management of the Alpine populations. However, Athrey *et al.* (2018) discuss the biases that can incur with these estimates due to small population sizes, as well as performing the analysis using data from a small section of the genome. In addition, the studies exampled here use multiple methods to calculate N_e which produced varying results. Consistency in methodology, as well as sampling, is important to ensure that

estimates are accurate.

It should be noted that levels of heterozygosity differed at STR and SNP markers for all three species, so that H_E was 3-5 times higher in the STR dataset compared to SNPs. This difference in H_E between STR and SNP data has already been noted in studies of the sage grouse, where a lower number of multi-allelic markers (microsatellites) estimated a higher level of H_E (Davis *et al*., 2015; Oyler-McCane *et al*., 2015; Row *et al*., 2015; Cross *et al*., 2018; Zimmerman *et al*., 2020). For example, Zimmerman *et al*. (2020) estimated an H_E of ~0.5 with 22 microsatellites for 254 individuals for the Gunnison sage-grouse, while for the SNP loci, the H_E was ~0.2. Similar differences in H_E was also shown in other avian SNP/STR studies (variegated fairy-wren, *Malurus lamberti*: Thrasher *et al.*, 2018; black-throated blue warbler, *Setophaga caerulescens*; Kaiser *et al*., 2017). This difference may be due to the bi-allelic nature of SNPs so that, despite the low numbers of markers used in STR studies, there is the possibility of a proportionally larger number of alleles indicating greater genetic diversity. As more species are genotyped for large SNP datasets, the level of H_E considered 'natural' or 'healthy' will become more evident. For now, this difference needs to be taken into account when interpreting genetic data for conservational purposes, as results may differ depending on the markers used.

Applying these next-generation sequencing markers to populations outside of the Alps will allow for a comparison of SNP N_e and H_E, across the species' entire range. This information will enable units of conservation to be determined for future management strategies. However, in the meantime, the results from this study indicate that there does not appear to be any areas of conservation concern across the Italian Alps. It has yet to be determined whether the Alps as a whole could represent a single MU for any of these grouse species in comparison to their wider ranges. These populations should be regularly monitored in a standardised way to prevent loss of biodiversity and to monitor the effects of hunting, in the case of the black grouse and ptarmigan, as changes in genetic diversity can take multiple generations to be visible in the genome. If genomic results from outside of the Alps indicate that the Alps are a single MU, then the management of this area will need to be transboundary, in contrast to the current system where each country and region manages populations individually, including a single set of hunting

regulations across the Alps as a whole.

4.7. CONCLUSIONS

Overall, this project has successfully typed SNPs using the GBS method for three species of Alpine grouse (black grouse, rock ptarmigan and western capercaillie). These SNP results showed a finer scale geographical pattern, compared to traditionally used microsatellites and mitochondrial DNA. The results for all three species tested showed no isolated populations in the Italian Alps and appeared to either represent a continuous population or one in isolation by distance. The SNP results for the black grouse and ptarmigan showed a genetic clustering into two groups in the ‘eastern’ and ‘western’ Alps. An increased sampling effort will be necessary for the ptarmigan to confirm if the separation indicated is due to environmental factors or a sampling gap. All species should be continuously monitored as low genetic diversity was present across the Alps for some markers, this is also necessary to ensure that populations do not become isolated.

LITERATURE CITED

Abbasi, S., Afsharzadeh, S., Saeidi, H. and Triest, L. (2016). Strong Genetic Differentiation of Submerged Plant Populations across Mountain Ranges: Evidence from *Potamogeton pectinatus* in Iran. *Plos One* **11:**e0161889.

Accorsi, M. (n.d). *topografia_dati_di_base/SOIUSA (MapServer)*. [online] Webgis.arpa.piemonte.it. Available at: <http://webgis.arpa.piemonte.it/ags101free/rest/services/topografia_dati_di_base/SOIUSA/MapServer> [Accessed 25 November 2020].

Aleix-Mata, G., Ruiz-Ruano, F., Pérez, J., Sarasa, M. and Sánchez, A. (2019). Complete mitochondrial genome of the Western Capercaillie Tetrao urogallus (Phasianidae, Tetraoninae). *Zootaxa*, **4550**:585.

Allendorf, F. W. (2017). Genetics and the conservation of natural populations: Allozymes to genomes. *Molecular Ecology*, **26**:420-430.

Allendorf, F. W., Hohenlohe, P. A. and Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature Reviews Genetics* **11**:697-709.

Angeloni, F., Wagemaker, N., Vergeer, P. and Ouborg, J. (2012). Genomic toolboxes for conservation biologists. *Evolutionary Applications* **5**:130-143.

Angelstam, P. (2004). Habitat thresholds and effects of forest landscape change on the distribution and abundance of black grouse and capercaillie. *Ecologicasl Bulletins,* **51**:173-187.

Athrey, G., Faust, N., Hieke, A. C. and Brisbin, I. L. (2018). Effective population sizes and adaptive genetic variation in a captive bird population. *PeerJ* **6**:e5803.

Bandelt H., Forster P., Röhl A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**:37–48.

Barton, N. H. and Hewitt, G. M. (1985). Analysis of Hybrid Zones. *Annual Review of Ecology and Systematics* **16**:113-148.

Bech, N., Barbu, C., Quéméré, E., Novoa, C., Allienne, J. and Boissier, J. (2013). Pyrenean ptarmigans decline under climatic and human influences through the Holocene. *Heredity*, **111**:402-409.

Beeston, R., Baines, D. and Richardson, M. (2005). Seasonal and between-sex differences in the diet of Black Grouse *Tetrao tetrix*. *Bird Study,* **52**:276-281.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B*, **57:**289-300.

Bevanger, K. and Brøseth, H. (2001). Bird collisions with power lines — an experiment with ptarmigan (Lagopus spp.). *Biological Conservation*, **99**:341-346.

BirdLife International (2016a). *Lyrurus tetrix*. *The IUCN Red List of Threatened Species 2016*:e.T22679480A85944601. http://dx.doi.org/10.2305/IUCN.UK.2016-3.RLTS.T22679480A85944601.en

BirdLife International (2016b). *Lagopus muta*. *The IUCN Red List of Threatened Species 2016*:e.T22679464A113623562. http://dx.doi.org/10.2305/IUCN.UK.2016-3.RLTS.T22679464A89358137.en

BirdLife International (2016c). *Tetrao urogallus. The IUCN Red List of Threatened Species2016*: e.T22679487A85942729. http://dx.doi.org/10.2305/IUCN.UK.2016-3.RLTS.T22679487A85942729.en

BirdLife International (2021a). *Lyrurus tetrix*. *The IUCN Red List of Threatened Species 2021:* e.T22679480A166187432. https://dx.doi.org/10.2305/IUCN.UK.2021-3.RLTS.T22679480A166187432.en

BirdLife International (2021b). *Lagopus muta. The IUCN Red List of Threatened Species 2021:* e.T22679464A166187006. https://dx.doi.org/10.2305/IUCN.UK.2021-3.RLTS.T22679464A166187006.en

BirdLife International (2021c). *Tetrao urogallus. The IUCN Red List of Threatened Species 2021:* e.T22679487A166188330. https://dx.doi.org/10.2305/IUCN.UK.2021-3.RLTS.T22679487A166188330.en

Bocca M. (2004a). 'Fagiano di monte *Tetrao tetrix* Linnaeus, 1758', in Spagnesi M., Serra L., (ed). *Uccelli d'Italia.* Quaderni per la Conservazione della Natura 21. Ministero dell'Ambiente – Istituto Nazionale per la Fauna Selvatica. Pp. 82-83.

Bocca M. (2004b). 'Pernice bianca *Lagopus mutus* Montin, 1776', in Spagnesi M., Serra L., (ed). *Uccelli d'Italia.* Quaderni per la Conservazione della Natura 21. Ministero dell'Ambiente – Istituto Nazionale per la Fauna Selvatica. Pp. 80-81.

Bolger, A. M., Lohse, M. and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics* **30:**2114-2120.

Bowker, G., Bowker, C. and Baines, D. (2007). Survival Rates and Causes of Mortality in Black Grouse *Tetrao tetrix* at Lake Vyrnwy, North Wales, UK. *Wildlife Biology,* **13**:231-237.

Brandies, P., Peel, E., Hogg, C. and Belov, K. (2019). The Value of Reference Genomes in the Conservation of Threatened Species. *Genes*, **10**:846.

Brommer, J. E. (2004). The range margins of northern birds shift polewards. *Annales Zoologici Fennici,* **41**:391-397.

Broome, A., Connolly, T. and Quine, C. P. (2014). An evaluation of thinning to improve habitat for capercaillie (*Tetrao urogallus*). *Forest Ecology and Management*, **314**:94-103.

Caizergues, A. and Ellison. L. (2002). Natal dispersal and its consequences in Black Grouse *Tetrao tetrix. Ibis* **144**:478-487.

Caizergues, A., Rätti, O., Helle, P., Rotelli, L., Ellison. L. and Rasplus, J. (2003a). Population genetic structure of male black grouse (*Tetrao tetrix L.*) in fragmented vs. continuous landscapes. *Molecular Ecology* **12**: 2297-2305.

Caizergues, A., Bernard-Laurent, A., Brenot, J., Ellison, L. and Rasplus, J. (2003). Population genetic structure of rock ptarmigan *Lagopus mutus* in Northern and Western Europe. *Molecular ecology,* **12**:2267-74.

Carroll, E., Bruford, M., DeWoody, J., Leroy, G., Strand, A., Waits, L. and Wang, J. (2018). Genetic and genomic monitoring with minimally invasive sampling methods. *Evolutionary Applications*, **11**:1094-1119.

Čas, M. (2010). Disturbances and predation on capercaillie at leks in Alps and Dinaric Mountains. *Sumarski List*, **134**:487-494.

Catchen, J., Hohenlohe, P., Bassham, S., Amores, A. and Cresko, W. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology* **22:**3124-3140.

Cattadori, I. and Hudson, P. (1999). Temporal dynamics of grouse populations at the southern edge of their distribution. *Ecography*, **22**:374-383.

Ceballos, G., Ehrlich, P., Barnosky, A., García, A., Pringle, R. and Palmer, T. (2015). Accelerated modern human–induced species losses: Entering the sixth mass extinction. *Science Advances*, **1**.

Chunco, A. J. (2014). Hybridization in a warmer world. *Ecology and Evolution***4**:2019-2031.

Collini, M. (2011). *La pernice bianca alpina (Lagopus muta helvetica): una sottospecie endemica in declino. Filogeografia, variabilità genetica e aspetti biologici.* Tesi di Laurea, Università degli studi di Padova

Coppes, J., Ehrlacher, J., Thiel, D., Suchant, R. and Braunisch, V. (2017). Outdoor recreation causes effective habitat reduction in capercaillie *Tetrao urogallus*: a major threat for geographically restricted populations. *Journal of Avian Biology*, **48**:1583-1594.

Corrales, C. and Höglund, J. (2012). Maintenance of gene flow by female-biased dispersal of Black Grouse *Tetrao tetrix* in northern Sweden. *Journal of Ornithology* **153**:1127-1139.

Costa, V., Rosenbom, S., Monteiro, R., O'Rourke, S. and Beja-Pereira, A. (2016). Improving DNA quality extracted from fecal samples—a method to improve DNA yield. *European Journal of Wildlife Research*, **63**.

Costanzi, J., Bergan, F., Sæbø, M., Jenkins, A. and Steifetten, Ø. (2018). Development and evaluation of 16 new microsatellite loci for the rock ptarmigan (Lagopus muta) and cross-species amplification for the willow grouse (L. lagopus). *BMC Research Notes*, **11**.

Cross, T. B., Schwartz, M. K., Naugle, D. E., Fedy, B. C., Row, J. R. and Oyler-McCance, S. J. (2018). The genetic network of greater sage-grouse: Range-wide identification of keystone hubs of connectivity. *Ecology and Evolution* **8**:5394-5412.

Cumer, T., Pouchon, C., Boyer, F., Yannic, G., Rioux, D., Bonin, A. and Capblancq, T. (2021). Double-digest RAD-sequencing: do pre- and post-sequencing protocol parameters impact biological results? *Molecular Genetics and Genomics*, **296**:457-471.

Dai, Y., Lin, Q., Fang, W., Zhou, X. and Chen, X. (2015). Noninvasive and nondestructive sampling for avian microsatellite genotyping: a case study on the vulnerable Chinese Egret (Egretta eulophotes). *Avian Research*, **6**.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., DePristo, M. A., Handsaker, R. E., Lunter, G. *et al.* (2011). The variant call format and VCFtools. *Bioinformatics* **27:**2156-2158.

Davis, D. M., Reese, K. P., Gardner, S. C. and Bird, K. L. (2015). Genetic structure of Greater Sage-Grouse (*Centrocercus urophasianus*) in a declining, peripheral population. *The Condor* **117**:530-544.

De Franceschi P. F. (2004). 'Gallo cedrone *Tetrao urogallus* Linnaeus, 1758', in Spagnesi M., Serra L., (ed). *Uccelli d'Italia.* Quaderni per la Conservazione della Natura 21. Ministero dell'Ambiente – Istituto Nazionale per la Fauna Selvatica. Pp. 84-85.

De Volo, S., Reynolds, R., Douglas, M. and Antolin, M. (2008). An Improved Extraction Method to Increase DNA Yield from Molted Feathers. *The Condor*, **110**:762-766.

Demartin, P. and Flaim S. (2012). Galliformi alpini e Lepre bianca. *UNCZA.*

Dirzo, R. and Raven, P. (2003). Global State of Biodiversity and Loss. *Annual Review of Environment and Resources*, **28**:137-167.

Diva-gis.org. n.d. Download data by country | DIVA-GIS. [Online] Available at: https://www.diva-gis.org/gdata [Accessed 1 August 2021].

Do, C., Waples, R. S., Peel, D., Macbeth, G. M., Tillett, B. J. and Ovenden, J. R. (2014)

NeEstimator V2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources* **14**: 209-214.

Dunn OJ. (1961). Multiple comparison among means. *Journal of the American Statistical Association*, **56**:52–64.

Duriez, O., Sachet, J., Ménoni, E., Pidancier, N., Miquel, C. and Taberlet, P. (2007). Phylogeography of the capercaillie in Eurasia: what is the conservation status in the Pyrenees and Cantabrian Mounts? *Conservation Genetics*, **8**:513-526.

Earl, D. A. and vonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4**:359-361.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S. and Mitchell, S. E. (2011). A robust, simple, Genotyping-by-Sequencing (GBS) approach for high diversity species. *PLoS ONE* **6**:e19379.

ERGA (n.d.). A genome atlas of european biodiversity. [online] Available at: https://www.erga-biodiversity.eu/ [Accessed March 20, 2022].

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 14:2611–2620.

Excoffier, L. and Lischer, H.E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*. **10**: 564-567.

Fameli, A., Morán-Luis, M., Rodríguez-Muñoz, R., Bañuelos, M.J., Quevedo, M., Mirol, P. (2017). Conservation in the southern edge of Tetrao urogallus distribution: Gene flow despite fragmentation in the stronghold of the Cantabrian capercaillie. *European Journal of Wildlife Research,* **63**.

Federazione Italiana Della Caccia (n.d.). Specie cacciabili. [online] Available at: https://www.federcaccia.org/specie-cacciabili/ [Accessed March 22, 2022].

Fick, S. E. and Hijmans, R. J. (2017). WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* **37**: 4302-4315.

Formenti, N., Viganó, R., Bionda, R., Ferrari, N., Trogu, T., Lanfrachi, P. and Palme, R. (2015). Increased hormonal stress reactions induced in an Alpine Black Grouse (*Tetrao tetrix*) population by winter sports. *Journal of Ornithology,* **156**:317-321.

Francis, R. M. (2016). POPHELPER: An R package and web app to analyse and visualise population structure. *Molecular Ecology Resources*, **17**:27-32.

Frankham, R. (1995). Conservation Genetics *Annual Reviews Genetics,* **29**:305-327.

Frankham, R. (2010). Where are we in conservation genetics and where do we need to go? *Conservation Genetics*, **11**:661-663.

Frankham, R., Ballou, J. and Briscoe, D. (2010). *Introduction to conservation genetics*. 2nd ed. Cambridge, UK: Cambridge University Press.

Fuentes-Pardo, A. and Ruzzante, D. (2017). Whole-genome sequencing approaches for conservation biology: Advantages, limitations and practical recommendations. *Molecular Ecology*, **26**:5369-5406.

Galla, S., Forsdick, N., Brown, L., Hoeppner, M., Knapp, M., Maloney, R., Moraga, R. *et al.* (2018). Reference Genomes from Distantly Related Species Can Be Used for Discovery of Single Nucleotide Polymorphisms to Inform Conservation Management. *Genes*, **10**.

Garbe, J. (2019). Gbstrim. Perl script https://bitbucket.org/jgarbe/gbstrim/src/master/

García-González, R., Aldezabal, A., Laskurain, N., Margalida, A. and Novoa, C. (2016). Factors Affecting Diet Variation in the Pyrenean Rock Ptarmigan (*Lagopus muta pyrenaica*): Conservation Implications. *PLoS ONE,* **11:**e0148614.

Gregersen, F. and Gregersen, H. (2009). Ongoing population decline and range contraction in Norwegian forest grouse. *Ornis Norvegic,* **32**:179-189

Guigo, R. and de Hoon, M. (2018). Recent advances in functional genome analysis. *F1000Research* **7:**1968.

Hijmans, R. J. (2019). raster: Geographic Data Analysis and Modelling. R package version 2.8-19. https://CRAN.R-project.org/package=raster.

Höglund, J., Larsson, J. K., Jansman, H. A. H., Segelbacher, G. (2007). Genetic variability in European black grouse (*Tetrao tetrix*). *Conservation Genetics* **8**:239-243.

Höglund, J., Larsson, J. K., Corrales, C., Santafé, G., Baines, D. and Segelbacher, G. (2011). Genetic structure among black grouse in Britain: implications for designing conservation units. *Animal Conservation* **14**:400-408.

Hohenlohe, P. A., Funk, W. C. and Rajora, O. P. (2021). Population genomics for wildlife conservation and management. *Molecular Ecology,* **30**:62-82.

Holder, K., Montgomerie, R. and Friesen, V. (1999). A Test of the Glacial Refugium Hypothesis using Patterns of Mitochondrial and Nuclear DNA Sequence Variation in Rock Ptarmigan (Lagopus mutus). *Evolution*, **53**:1936-1950.

Holder, K., Montgomerie, R. and Friesen, V.L. (2004). Genetic diversity and management of Nearctic Rock Ptarmigan (*Lagopus mutus*). *Canadian Journal of Zoology*

82:564–575.

Hovick, T. J., Elmore, R. D., Dahlgren, D. K., Fuhlendorf, S. D. and Engle, D. M. (2014). REVIEW: Evidence of negative effects of anthropogenic structures on wildlife: a review of grouse survival and behaviour. *Journal of Applied Ecology*, **51**: 1680-1689.

IUCN (2022). *The IUCN Red List of Threatened Species. Version 2021-3*. https://www.iucnredlist.org. Downloaded on [20 November 2021].

Jakobsson, M. and Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23:** 1801-1806.

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**: 1403-1405. doi: 10.1093/bioinformatics/btn129.

Kaiser, S. A., Taylor, S. A., Chen, N., Sillett, T. S., Bondra, E. R., Webster, M. S. (2017). A comparative assessment of SNP and microsatellite markers for assigning parentage in a socially monogamous bird. *Molecular Ecology Resources* **17**:183-193.

Keyghobadi, N. (2007). The genetic implications of habitat fragmentation for animals. *Canadian Journal of Zoology,* **85**:1049-1064.

Kirk, H. and Freeland, J. (2011). Applications and Implications of Neutral versus Non-neutral Markers in Molecular Ecology. *International Journal of Molecular Sciences*, **12**:3966-3988.

Kleven, O., Brøseth, H., Jonassen, K. and Pedersen, H. (2020). Backcrossing of a capercaillie × black grouse hybrid male in the wild revealed with molecular markers. *European Journal of Wildlife Research*, **66**.

Klinga, P., Mikoláš, M., Zhelev, P., Höglund, J. and Paule, L. (2015). Genetic differentiation of western capercaillie in the Carpathian Mountains: the importance of post glacial expansions and habitat connectivity. *Biological Journal of the Linnean Society*, **116**: 873–889.

Kozma, R., Melsted, P., Magnússon, K. P. and Höglund, J. (2016). Looking into the past - the reaction of three grouse species to climate change over the last million years using whole genome sequences. *Molecular Ecology* **25**:570-580.

Kozma, R., Rödin-Mörch, P. and Höglund, J. (2019). Genomic regions of speciation and adaptation among three species of grouse. *Scientific Reports*, **9**.

Lagerholm, V., Sandoval-Castellanos, E., Vaniscotte, A., Potapova, O., Tomek, T., Bochenski, Z., Shepherd, P. *et al.* (2017). Range shifts or extinction? Ancient DNA

and distribution modelling reveal past and future responses to climate warming in cold-adapted birds. *Global Change Biology*, **23**:1425-1435.

Larsson, J. K., Jansman, H. A. H., Segelbacher, G., Höglund, J., Koelewijn, H. P. (2008). Genetic impoverishment of the last black grouse (*Tetrao tetrix*) population in the Netherlands: detectable only with a reference from the past. *Molecular Ecology* **17**:1897-904.

Lebigre, C., Alatalo, R. V., Forss, H. E. And Siitari, H. (2008). Low levels of relatedness on black grouse leks despite male philopatry. *Molecular Ecology* **17**:4512-4521.

Leigh, J. W., Bryant D. (2015). PopART: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution* **6**:1110–1116.

Leyhausen, J., Cocchiararo, B., Nowak, C., Ansorge, H., Bertolino, S., Büchner, S., Fietz, J. *et al.* (2022). Genotyping-by-sequencing based SNP discovery in a non-model rodent, the endangered hazel dormouse. *Conservation Genetics Resources, **3**.*

Li, B., Zhu, C., Ding, P., Bai, S. and Cui, J. (2016). Complete mitochondrial genome of black grouse (*Lyrurus tetrix*). *Mitochondrial DNA Part A* **27**:134-135.

Lischer, H. E. L. and Shimizu, K. K. (2017). Reference-guided *de novo* assembly approach improves genome reconstruction for related species. *BMC Bioinformatics* **18**.

Liukkonen-Anttila, T., Rätti, O., Kvist, L., Helle, P. and Orell, M. (2004). Lack of genetic structuring and subspecies differentiation in the capercaillie (*Tetrao urogallus*) in Finland. *Annales Zoologici Fennici.* **41**:619-633.

Lucchini, V., Höglund, J., Klaus, S., Swenson, J. and Randi, E. (2001). Historical Biogeography and a Mitochondrial DNA Phylogeny of Grouse and Ptarmigan. *Molecular Phylogenetics and Evolution*, **20**:149-162.

Ludwig, G. X., Alatalo, R. V., Helle, P., Lindén, H., Lindström and Siitari, H. (2006). Short- and long-term population dynamical consequences of asymmetric climate change in black grouse. *Proceedings of The Royal Society of Biological Sciences,* **273**:2009-2016.

Ludwig, T., Storch, I. and Wübbenhorst, J. (2008). How the Black Grouse was lost: historic reconstruction of its status and ditribution in Lower Saxony (Germany). *Journal of Ornithology,* **149**:587-596.

Luu, K., Blum, M., Privé, F., Bazin, E. Duforet-Frebourg, N. (2019). Pcadapt: Fast Principal Component Analaysis for Outlier Detection. R package version 4.1.0 https://github.com/bcm-uga/pcadapt

Mäkinen, H. S. and Merilä, J. (2008). Mitopchrondrial DNA phylogeography of the three-

spined stickleback (*Gasterosteus aculeatus*) in Europe – Evidence for multiple glacial refugia. *Molecular Phylogenetics and Evolution* **46**:167-182.

Marazzi, S. (2005). *Atlante orografico delle Alpi : SOIUSA : suddivisione orografica internazionale unificata del sistema alpino.* Turin: Priuli & Verlucca.

Marjakangas, A. and Kiviniemi, S. (2005). Dispersal and migration of female Black Grouse Tetrao tetrix in eastern central Finland. *Ornis Fennica,* **82**:107-116.

McNeely, J. A., Miller, K. R. , Reid, W. V., Mittermeier, R. A., Werner, T. B. (1990). *Conserving the World's Biological Diversity*. IUCN, World Resources Institute, Conservation International, WWF-US and the World Bank: Washington, DC.

Mikoláš, M., Tejkal, M., Kuemmerle, T., Griffiths, P., Svoboda, M., Hlásny, T., Leitão, P. *et al*. (2017). Forest management impacts on capercaillie (*Tetrao urogallus*) habitat distribution and connectivity in the Carpathians. *Landscape Ecology*, **32:**163-179.

Morin, P. A., Luikart, G., Wayne, R. K. and the SNP workshop group (2004). SNPs in ecology, evolution and conservation. *TRENDS in Ecology and Evolution*, **19**:208-216.

Moritz, C. (1994). Defining 'Evolutionarily Significant Units' for conservation. *Trends in Ecology & Evolution* **9:** 373-375.

Moss, R. (2001). Second extinction of capercaillie (*Tertao urogallus*) in Scotland? *Biological Conservation,* **101**:255-257.

Moss, R., Picozzi, N., Summers, R. W., and Baines, D. (2000). Capercaillie *Tetrao urogallus* in Scotland - demography of a declining population. *Ibis*, **142**:259-267.

Moss, R., Picozzi, N., Catt, D. C. (2006). Natal dispersal of capercaillie *Tetrao urogallus* in northeast Scotland. *Wildlife Biology* **12**:227–232.

Moss, R., Leckie, F., Biggins, A., Poole, T., Baines, D. and Kortland, K. (2014). Impacts of Human Disturbance on Capercaillie Tetrao urogallus Distribution and Demography in Scottish Woodland. *Wildlife Biology,* **20**:1-18.

Muñoz-Fuentes, V., Darimont, C. T., Wayne, R. K., Paquet, P. C. and Leonard, J. A. (2009). Ecological factors drive differentiation in wolves from British Columbia. *Journal of Biogeography* **36**:1516-1531.

National Center for Biotechnology Information (NCBI) (2022). Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] Available at: https://www.ncbi.nlm.nih.gov/ [Accessed 19 March 2022].

Naturalearthdata.com. n.d. Natural Earth - Free vector and raster map data at 1:10m,

1:50m, and 1:110m scales. [online] Available at: https://www.naturalearthdata.com/ [Accessed 1 August 2021].

Nilsen, E. B., Moa, P. F., Brøseth, H., Pedersen, H. C. and Hagen, B. R. (2020). Survival and Migration of Rock Ptarmigan in Central Scandinavia. *Frontiers in Ecology and Evolution*, **8.**

Nomura, T. (2008) Estimation of effective number of breeders from molecular coancestry of single cohort sample. *Evolutionary Applications* **1**: 462-474.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre,P., McGlinn, D., Minchin, P.R. *et al.* (2019). vegan: Community Ecology Package. R package version 2.5-4. https://CRAN.R-project.org/package=vegan.

Ottenburghs, J. (2019). Mutlispecies hybridization in birds. *Avian Research*, **10.**

Ouborg, N. J., Pertoldi, C., Loeschcke, V., Bijlsma, R. K. And Hedrick, P. W. (2010). Conservation genetics in transition to conservation genomics. *Trends in Genetics* **26**:177-187.

Oyler-McCance, S. J., Cornman, R. S., Jones, K. L. and Fike, J. A. (2015). Z chromosome divergence, polymorphism and relative effective population size in a genus of lekking birds. *Heredity* **115**:452-459.

Pernollet, C., Korner-Nievergelt, F. and Jenni, L. (2015). Regional changes in the elevational distribution of the Alpine Rock PtarmiganLagopus muta helveticain Switzerland. *Ibis*, **157**:823-836.

Peronace, V., Cecere, J. G., Gustin, M. and Rondinini, C. (2012). Lista Rossa 2011 degli Uccelli Nidificanti in Italia. *Avocetta* **36**:11-58.

Peters, C., Nelson, H., Rusk, B. and Muir, A. (2020). A novel method to optimise the utility of underused moulted plumulaceous feather samples for genetic analysis in bird conservation. *Conservation Genetics Resources*, **12**:457-467.

Peters, J., Lavretsky, P., DaCosta, J. M., Bielefeld, R. R., Feddersen, J. C. and Sorenson, M. D. (2016). Population genomic data delineate conservation units in mottled ducks (Anas fulvigula). *Biological Conservation,* **203**:272-281.

Petkova, D., Novembre, J. and Stephens, M. (2016). Visualizing spatial population structure with estimated effective migration surfaces. *Nature Genetics* **48**:94-100.

Piertney, S. B. and Höglund, J. (2001). Polymorphic microsatellite DNA markers in black grouse (*Tetrao tetrix*). *Molecular Ecology Notes* **1**:303-304.

Pimentel, D., Wilson, C., McCullum, C., Huang, R., Dwen, P., Flack, J., Tran, Q. *et al.* (1997). Economic and Environmental Benefits of Biodiversity. BioScience, ***47:***747–757.

Porkert, J., Solheim, R. and Flor, A. (1997). Behaviour of hybrid maleTetrao tetrix♂ ×T. urogallus♀ on black grouse leks. *Wildlife Biology*, **3**:169-176.

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics 155:945–959.

Provincia Autonoma di Trento (n.d.) Organizzazione della caccia. [online] Available at: <https://forestefauna.provincia.tn.it/Fauna/Caccia/Organizzazione-della-caccia> [Accessed 16 March 2022].

Pruett, C., Turner, T., Topp, C., Zagrebelny, S. and Winker, K. (2010). Divergence in an archipelago and its conservation consequences in Aleutian Island rock ptarmigan. *Conservation Genetics*, **11**:241-248.

Pruett, C. L. Johnson, J. A., Larsson, L. C. Wolfe, D. H. and Patten, M. A. (2011). Low effective population size and survivorship in a grassland grouse. *Conservation Genetics* **12**:1205–1214.

QGIS Development Team (2021). QGIS Geographic Information System. Open Source Geospatial Foundation Project. Available at: http://qgis.osgeo.org.

Quintela, M., Thulin, C. and Höglund, J. (2010). Detecting hybridization between Willow Grouse (L*agopus lagopus*) and Rock Ptarmigan (*L. muta*) in Central Sweden through bayesian admixture analyses and mtdna screening. Conservation Genetics **11**:557–569.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing. [Online]. Vienna, Austria. Available at: http://www.R-project.org/.

Randi, E., Lucchini, V. (1998). Organization and Evolution of the Mitochondrial DNA Control Region in the Avian Genus *Alectoris. Journal of Molecular Evolution* **47**: 449–462.

Regnaut, S., Christe, P., Chapuisat, M. and Fumagalli, L. (2006). Genotyping faeces reveals facultative kin association on capercaillie's leks. *Conservation Genetics*, **7**:665-674.

Revermann, R., Schmid, H., Zbinden, N., Spaar, R. and Schröder, B. (2012). Habitat at the mountain tops: how long can Rock Ptarmigan (*Lagopus muta helvetica*) survive rapid climate change in the Swiss Alps? A multi-scle approach. *Journal of Ornithology,* **153**:891-905

Rintamäki, P. T., Alatalo, R. V., Höglund, J. and Lundberg, A. (1995). Mate sampling behaviour of black grouse females (*Tetrao tetrix*). *Behavioural Ecology and Sociobiology,* **37**:209-215.

Robertson, J. M., Langin, K. M., Sillett, T. S., Morrison, S. A., Ghalambor, C. K. and Funk W. C. (2014). Identifying Evolutionarily Significant Units and Prioritizing Populations for Management on Islands. *Monographs of the Western North American Naturalist*, **7**:397-411.

Rodríguez-Muñoz, R., Mirol, P., Segelbacher, G., Fernández, A. and Tregenza, T. (2007). Genetic differentiation of an endangered capercaillie (Tetrao urogallus) population at the Southern edge of the species range. *Conservation Genetics*, **8**:659-670.

Rodríguez-Muñoz, R., del Valle, C., Bañuelos, M. and Mirol, P. (2015). Revealing the consequences of male-biased trophy hunting on the maintenance of genetic variation. *Conservation Genetics*, **16**:1375-1394.

Row, J. R., Oyler-McCance, S. J., Fike, J. A., O'Donnell, M. S., Doherty, K. E., Aldridge, C. L., Bowen, Z. H. *et al.* (2015). Landscape characteristics influencing the genetic structure of greater sage-grouse within the stronghold of their range: a holistic modeling approach. *Ecology and Evolution* **5**:1955-1969.

Rowland, M. M., Wisdom, M. J., Suring, L. H. and Meinke, C. W. (2006). Greater sage-grouse as an umbrella species for sagebrush-associated vertebrates. *Biological Conservation,* **129**:323-335.

Rutkowski, R., Zawadzka, D., Suchecka, E. and Merta, D. (2017). Conservation genetics of the capercaillie in Poland - Delineation of conservation units. *PLOS ONE*, **12**:e0174901.

Rutkowski, R., Palucki, A., Dulisz, B., Ciach, M., Nowak-Życzyńska, Z., and Kowalewska, K. (2018). Conservation genetics of the Black Grouse *Tetrao tetrix* in Poland — distribution of genetic diversity among the last populations. *Acta Ornithologica* **53**: 181-204.

Ryder, A. (1986). Species conservation and systematics: the dilemma of subspecies. *Trends in Ecology & Evolution* **1**: 9-10.

Scherini, G.C., Tosi, G. and Wauters, L. (2003). Social Behaviour, Reproductive Biology and Breeding Success of Alpine Rock Ptarmigan Lagopus Mutus Helveticus in Northern Italy. *Ardea,* **91**:11-23.

Schmid, S.M., Fügenschuh, B., Kissling, E. and Schuster, R. (2004). Tectonic map and overall architecture of the Alpine orogen. *Eclogae Geologicae Helvetiae,* **97**:93e117.

Segelbacher, G. and Piertney, S. (2007). Phylogeography of the European capercaillie (Tetrao urogallus) and its implications for conservation. *Journal of Ornithology*,

148:269-274.

Segelbacher, G., Paxton, R. J., Steinbrück, G., Tronteli, P. and Storch, I. (2000). Characterization of microsatellites in capercaillie *Tetrao urogallus* (AVES). *Molecular Ecology* **9**:1919-1952.

Segelbacher, G., Storch, I. and Tomiuk, J. (2003). Genetic evidence of capercaillie Tetrao urogallus dispersal sources and sinks in the Alps. *Wildlife Biology*, **9**:267-273.

Segelbacher, G., Wegge, P., Sivkov, A. and Höglund, J. (2007). Kin groups in closely spaced capercaillie leks. *Journal of Ornithology*, **148**:79-84.

Segelbacher, G., Manel, S. and Tomiuk, J. (2008). Temporal and spatial analyses disclose consequences of habitat fragmentation on the genetic diversity in capercaillie (Tetrao urogallus). *Molecular Ecology*, **17**:2356-2367.

Segelbacher, G., Bosse, M., Burger, P., Galbusera, P., Godoy, J. A., Helsen, P., Hvilsom, C. *et al.* (2021). New developments in the field of genomic technologies and their relevance to conservation management. *Conservation Genetics.*

Sittenthaler, M., Kunz, F., Szymusik, A., Grünschachner-Berger, V., Krumböck, S., Stauffer, C., Nopp-Mayr, U. (2018). Fine-scale genetic structure in an eastern Alpine black grouse *Tetrao tetrix* metapopulations. *Journal of Avian Biology* **e01681**.

Song, K., Gao, B., Halvarsson, P., Fang, Y., Klaus, S., Jiang, Y., Swenson, J. *et al.* (2021). Demographic history and divergence of sibling grouse species inferred from whole genome sequencing reveal past effects of climate change. *BMC Ecology and Evolution*, **21**.

Spanò, S. and Salvidio, S. (2012). Dynamics of a black grouse (*Tetrao tetrix*) population in the French Maritime Alps. *Bollettino dei Musei e degli Istituti Biologici dell'Università di Genova* **74**:55-66.

Storch, I. (2000). An Overview to Population Status and Conservation of Black Grouse Worldwide. *Cahiers d'Ethologie*, **20**:153-164.

Storch, I. (2007). Grouse: Status Survey and Conservation Action Plan 2006–2010. Gland, Switzerland: IUCN and Fordingbridge, UK: World Pheasant Association. 114p.

Storch, I. and Segelbacher, G. (2000). Genentic correlates of spatial population structure in central European capercaillie *Terao urogallus* and black grouse *T. tetrix*: a project in progress. *Wildlife Biology*, **6**:305-310.

Strand, T. M., Segelbacher, G., Quintela, M., Xiao, L. Axelsson, T. And Höglund, J.

(2012). Can balancing selection on MHC loci counteract genetic drift in small fragmented populations of black grouse? *Ecology and Evolution* **2**:341-353.

Summers, R. W., Green, R. E., Proctor, R., Dugan, D., Lambie, D., Moncrieff, R., Moss, R. *et al*. (2004). An experimental study of the effects of predation on the breeding productivity of capercaillie and black grouse. *Journal of Applied Ecology,* **41**:523-525.

Supple, M. A. and Shapiro, B. (2018). Conservation of biodiversity in the genomics era. *Genome Biology* **19**.

Suter, W., Graf, R. F., Hess, R. (2002). Capercaillie (T*etrao urogallus*) and Avian Biodiversity: Testing the Umbrella-Species Concept. *Conservation Biology* **16**:778-788.

Sveinsdóttir, M. and Magnússon, K. (2017). Complete mitochondrial genome and phylogenetic analysis of willow ptarmigan (Lagopus lagopus) and rock ptarmigan (Lagopus muta) (Galliformes: Phasianidae: Tetraoninae). *Mitochondrial DNA Part B*, **2**:400-402.

The World Database on Protected Areas (WDPA) and World Database on Other Effective Area-based Conservation Measures (WD-OECM) Cambridge, UK: UNEP-WCMC and IUCN. [Online] Available at: www.protectedplanet.net [Accessed: 14 July 2021].

Thrasher, D. J., Butcher, B. G., Campagna, L., Webster, M. S. and Lovette, I. J. (2018). Double-digest RAD sequencing outperforms microsatellite loci at assigning paternity and estimating relatedness: A proof of concept in a highly promiscuous bird. *Molecular Ecology Resources* **18**:953-965.

Vallant, S., Niederstätter, H., Berger, B., Lentner, R. and Parson, W. (2018). Increased Dna typing success for feces and feathers of capercaillie (*Tetrao urogallus*) and black grouse (*Tetrao tetrix*). *Ecology and Evolution* **8**:3941-3951.

van Etten, J. (2017). R Package gdistance: Distances and Routes on Geographical Grids. *Journal of Statistical Software* **76**: 1-21.

Vázquez, J. F., Perez, T., Albornoz, J. and Domínguez, A. (2013). Census and effective population size of the endangered Cantabrian capercaillie (Tetrao urogallus) estimated from non-invasive samples. *Grouse News,* 46**:**12-26.

Vernesi, C., Hoban, S. M., Pecchioli, E., Crestanello, B., Bertorelle, G., Rosà, R., Hauffe, H. C. (2016). Ecology, environment and evolutionary history influence genetic structure in five mammal species from the Italian Alps. *Biological Journal of the Linnean Society* **117**:428–446.

Viterbi, R. Imperio, S., Alpe, D., Bosser-Peverelli, V. and Provenzale, A. (2015). Climatic control and population dynamics of black grouse (*Tetrao Tetrix*) in the western Italian Alps. The *Journal of Wildlife Management,* **79**:156–166.

Walker, A. (2010). The reintroduction of black grouse to the Isle of Arran, Scotland. *Grouse News,* **40**:13-16.

Wang, B., Ekblom, R., Bunikis, I., Siitari, H. and Höglund, J. (2014). Whole genome sequencing of the black grouse (Tetrao tetrix): reference guided assembly suggests faster-Z and MHC evolution. *BMC Genomics*, **15**:180.

Waples, R.S. (1991). Pacific salmon, *Oncorhynchus* spp., and the definition of "Species" under the Endangered Species Act. *Marine Fisheries Review,* **53**:11-22.

Waples, R.S. and Do C. (2008). ldne: a program for estimating effective population size from data on linkage disequilibrium. *Molecular Ecology Resources* **8**:753–756.

Warren, P. K. and Baines, D. (2002). Dispersal, survival and causes of mortality in black grouse *Tertrao tetrix* in norther England. *Wildlife Biology*, **8**:91-97.

Warren, P., Land, C., Hesford, N. and Baines, D. (2020). Conserving Black Grouse Lyrurus tetrix in southern Scotland: evidence for the need to retain large contiguous moorland habitat within a forest-moorland landscape. *Bird Study*, **66**:1-9.

Watson, A. and Moss, R. (2004). Impacts of ski-development on ptarmigan (*Lagopus mutus*) at Cairn Gorm, Scotland. *Biological Conservation*. **116**:267-275.

Wegge, P. and Kastdalen, L. (2008). Habitat and diet of young grouse broods: resource partitioning between Capercaillie (*Tetrao urogallus*) and Black Grouse (*Tetrao tetrix*) in boreal forests. *Journal of Ornithology,* **149**:237-244.

Wegge, P. and Rolstad, J. (2017). Climate change and bird reproduction: Warmer springs benefit breeding success in boreal forest grouse. *Proceedings of the Royal Society B: Biological Sciences*, **284**:20171528

White, P. J. C., Warren, P. and Baines, D. (2013). Forest expansion in Scotland and its potential effects on black grouse *Tetrao tetrix* conservation. *Forest Ecology and Management,* **308**:145-152**.**

Yonezawa, T. and Nishibori, M. (2020). The complete mitochondrial genome of the Japanese rock ptarmigan (Lagopus muta japonica Clark, 1907). *Mitochondrial DNA Part B*, **5**:1648-1649.

Zimmerman, S. J., Aldridge, C. L. and Oyler-McCance, S. J. (2020). An empirical comparison of population genetic analyses using microsatellite and SNP data for species of conservation concern. *BMC Genomics* **21**.