2022

Genetic Variation in Mountain Bumblebees

FINAL REPORT TO SWEDISH ENVIRONMENTAL PROTECTION AGENCY (NATURVÅRDSVERKET)

MATTHEW WEBSTER & YUANZHEN LIU

UPPSALA UNIVERSITET

Investigation of genetic variation in mountain bumblebees in Sweden
Undersökning av genetisk variation hos fjällhumlor

Matthew Webster, Yuanzhen Liu
Dept. Medical Biochemistry and Microbiology, Uppsala University

Final report for Swedish Environmental Protection Agency (Naturvårdsverket)

Agreement number (Överenskommelse/avtal nr): **225-21-003**
Case number (Naturvårdsverkets ärendenummer): **NV-04697-21**

Project leader (Projektledare, utförare): **Matthew Webster**
Responsible for quality (Kvalitetsansvarig): **Matthew Webster**

Program area: **Knowledge of genetic diversity**
Programområde: **Kunskap om genetisk mångfald**

Section: **Pollinators - population monitoring**
Delprogram: **Pollinatörer - populationsövervakning**

Responsible administrator at SEPA (Ansvarig handläggare på Naturvårdsverket):
Kristin Löwenborg

Contact:
Prof. Matthew Webster
IMBIM, BMC D11:3, Box 582, Uppsala University, 75 123 Uppsala
email: matthew.webster@imbim.uu.se
tel: 018 471 4391

Cover picture: *Bombus alpinus* © Göran Holmström

REPORT SUMMARY

- We used whole-genome sequencing to assess genetic variation in seven species of bumblebees in the Sweden: six that are only found in mountain habitats and one that is widespread.
- Estimates of effective population size (N_E) indicate that none of the species are immediately threatened with extinction, but genetic diversity is substantially lower in species with restricted distributions at higher elevations and latitudes.
- We do not find evidence of population structure in the mountain bumblebee species, indicating the absence of long-term barriers to gene flow.
- We identified an inbred sterile diploid male of the rare specialist species *Bombus hyperboreus* collected near the southern limits of its range, indicating that inbreeding is negatively impacting this population.
- Reconstruction of historical fluctuations of N_E reveals that two Arctic-restricted species (*Bombus hyperboreus* and *Bombus pyrrhopygus*) have declined since the last ice age.
- Most or all of the suitable habitat for these mountain species in Scandinavia is predicted to be lost due to climate change within a century.
- We propose continued genetic monitoring of these species and analysis of genetic variation in historical populations represented in museum collections. This would enable detection of ongoing or future population declines and inform conservation actions.

ABSTRACT

The abundance of many species of bumblebees is declining and climate change is a major contributing factor. The distributions of bumblebees are shifting northwards to track suitable climates across continents. Some of the species that are most vulnerable inhabit the Arctic and have no possibility to move further north. An important indicator of the viability of animal populations is genetic diversity, which can be tracked over time by genetic monitoring to identify populations that are declining or at risk. Species of arctic bumblebees found in the Swedish mountains are considered high priority for inclusion in genetic monitoring due to the threats they face from climate change. In this report we use whole-genome sequencing to assess genetic variation in seven species of bumblebee that inhabit the Swedish mountains. Six of the species are restricted to arctic and mountain environments, whereas one is widespread throughout Europe. In total, we sequenced 333 samples of these species from across their ranges in Sweden. Estimates of effective population size (N_E) predicted from levels of genetic variation vary between ~55,000 for the species with the most restricted high alpine distributions, to 200,000 for the more common mountain species, and 220,000 for the widespread species. Population fragmentation is generally very low or undetectable in Swedish mountains, suggesting an absence of barriers to gene flow between sub-populations. None of the species appear to be at immediate risk of negative genetic effects due to high levels of genetic drift. Reconstruction of historical fluctuations in N_E indicate that the arctic specialist species *Bombus hyperboreus* has already experienced substantial population declines since the last ice age and we detected one highly inbred sample of this species, suggesting it may be particularly vulnerable. The ranges of all of the mountain bumblebee species are predicted to shrink drastically due to the effects of climate change. Future genetic monitoring would be a useful way to detect whether populations have been adversely affected by declines or fragmentation that could threaten their viability. The data presented here serves as an important first step towards future genetic monitoring programs for these species.

BACKGROUND SUMMARY

- Intraspecific genetic variation is recognised as a crucial component of biodiversity but it is only monitored in a tiny fraction of the species on the planet.
- Genetic variation is crucial for populations to adapt to changing environment and an important indicator of population viability.
- Due to technological advances, it is now feasible to assay genetic variation using whole-genome sequencing.
- Genetic monitoring of pollinators is considered a high priority due to their ecosystem services and benefits to agriculture.
- Bumblebees are vital pollinators in temperate regions but are threatened by loss of suitable habitats due to climate change. Bumblebees in arctic and alpine environments have a particularly poor prognosis.
- In this study we aimed to assess genetic variation in seven bumblebee species that inhabit the arctic and alpine environments in Sweden using whole-genome sequencing. Six of these species are restricted to these environments whereas one is found widely all over Europe.
- The study had the following aims:
    - Detection of putative cryptic species
    - Assessment of measures of intraspecific genetic variation (H_{OBS}, θ_W)
    - Estimation of effective population size (N_E)
    - Assessment of population structure (PCA, Admixture)
    - Reconstruction of historical changes in N_E (SMC++)
    - Analysis of inbreeding using the distribution of runs of homozygosity (ROH) in each sample
- The results of this assessment can be used to assess vulnerability of each species and to design future genetic monitoring programmes.

BACKGROUND

Intraspecific genetic variation is a crucial but neglected component of biodiversity

We are currently experiencing a global crisis in biodiversity loss that has been described as a sixth mass extinction (Ceballos et al. 2017; IPBES 2019). It is widely recognised that levels of genetic variation within species are a crucial component of biodiversity, supporting essential ecological functions and nature's contribution to people (Des Roches et al. 2021). Genetic variation within and among populations is crucial for species to adapt, evolve and survive over time. Loss of genetic variation associated with population declines and fragmentation is a central threat to many wild populations, which can reduce the resilience of populations to challenges such as habitat loss and degradation, environmental pollution and climate change (Willi et al. 2022). However despite the importance of genetic variation, global conservation policy lacks specific efforts aimed at monitoring and protecting it (Hoban et al. 2021; Laikre et al. 2020). For instance, only 1.1% of species evaluated by the International Union for Conservation of Nature (IUCN) are assessed below the species level (Des Roches et al. 2021).

A key role of genetic variation is that it encodes phenotypic variation, which is important for environmental adaptation. This variation is necessary for species to adapt to a changing environment or to survive across a range in which the environment is spatially variable. The

presence of a diverse set of genotypes can also be advantageous for resisting pathogens or invasive species, for stable habitat construction, for ensuring population stability, and for maintaining ecosystem services (Des Roches et al. 2021). Preserving variation is therefore important to maximise resilience of a species to challenges such as future environmental change. In addition, knowledge of how variation is partitioned in species distributions is important for making management decisions. For example, species can be split into locally adapted subpopulations, which are important to maintain to preserve resilience or to make decisions about genetic rescue or supplementing populations.

Genetic variation is also an important indicator of population viability (Willi et al. 2022). Low levels of genetic variation are indicative of small effective population size (N_E), which can indicate that populations are at risk of decreased fitness. There are three main reasons for this: 1) Inbreeding is more likely to occur within small populations, which leads to decreased fitness due to deleterious alleles appearing in the homozygous state. 2) Natural selection is less effective in small populations, which means that deleterious mutations accumulate, increasing genetic load. 3) Smaller populations are less likely to adapt to environmental change due to loss of functional genetic variation and a lower efficacy of natural selection.

Monitoring genetic diversity is a useful tool for species conservation

Assessments of genetic variation at single time points enable estimation of key population genetic parameters that reflect population viability. These include measures of N_E, inbreeding, and population fragmentation. These assessments can identify populations that are vulnerable or at risk for extinction. A standard rule in conservation genetics is that populations with N_E less than 500 run a risk of decreased fitness and are considered to be of conservation concern (Allendorf et al. 2022).

There is greater power to infer whether populations are threatened or declining if genetic variation is assayed at multiple time points (Díez-del-Molino et al. 2018). This is important because a species may have low N_E due to intrinsic life history characteristics rather than recent population declines (Ellegren & Galtier 2016; Webster et al. 2023). In addition, some populations of insects have been shown to maintain high N_E even though they are small and fragmented (Hoffmann et al. 2021). Genetic monitoring quantifies temporal changes in a time series of genetic measurements (Schwartz et al. 2007). This can specifically be used to determine whether N_E is changing over time, whether populations are becoming more fragmented, and the level of migration and gene flow between populations. Monitoring levels of genetic diversity within and between wildlife populations contributes to tracking their long-term viability.

The importance of monitoring genetic diversity has been recognised by SEPA who have published a proposal for genetic monitoring of species in Sweden (Posledovich et al. 2021a). This proposal uses several criteria of importance and feasibility to generate a list of 167 species, of which 60 are marked as high priority, including vertebrates, invertebrates and plants, for inclusion in a Swedish program for genetic monitoring, with a separate report specifically focussing on pollinators (Posledovich et al. 2021b).

A key first step in any genetic monitoring program is a genetic assessment of the populations of interest. This provides an estimate of several key population genetic parameters and identifies populations under immediate threat. In addition, this initial estimation of population genetic parameters enables evaluation of priority and feasibility of including the species in

genetic monitoring program and provides a first set of measurements to form part of such a program.

Whole-genome sequencing can be used to assay genetic variation

The past two decades has seen a revolution in DNA sequencing technology, which has made it possible to assay genetic variation on a whole genome scale cheaply and effectively. Conservation genetic analyses have traditionally relied on assaying variation at a limited number of genetic loci, such as by typing microsatellite markers or sequencing single loci mitochondrial or nuclear loci. Genetic variation in populations can now be assessed at the whole-genome scale using next-generation sequencing (NGS). In contrast to traditional genetic markers, NGS datasets can more accurately estimate population genetic parameters such as N_E and fixation index (F_{ST}) and perform additional analyses such as uncovering loci involved in adaptation, identifying deleterious mutations, or reconstructing demographic history (Webster et al. 2023; Hohenlohe et al. 2021).

The most effective way to use NGS to assay genome variation in a species of interest is to perform high-throughput Illumina short-read sequencing on a population sample. In order to identify and analyse genetic variation, the sequencing reads generated by this approach must be mapped to a reference genome assembly. It is therefore important that a reference assembly of the species of interest, or a closely related species, is already available. There are now major international efforts to generate reference genome assemblies for all eukaryotic life, and reference genome assemblies are available for rapidly increasing numbers of species (Formenti et al. 2022; Lewin et al. 2022). The possibility of assaying genetic variation on the genome scale is therefore available for an increasingly large number of wild species.

Genetic monitoring of pollinators

Several recent reports have uncovered evidence for massive insect declines, which are potentially even more severe than declines observed in other animal taxa (Wagner 2020; Sánchez-Bayo & Wyckhuys 2019; Seibold et al. 2019; van Klink et al. 2020). In one key longitudinal study in northern Germany, more than 75% loss of flying insect biomass was reported in less than 30 years (Hallmann et al. 2017) and many indicators suggest that this is a global trend (Webster et al. 2023; Sánchez-Bayo & Wyckhuys 2019). Insects are critical components of ecosystems as a food source and through their roles in processes such as decomposition and pollination. Globally, some of the main factors causing insect declines are habitat loss, pesticides, and climate change (Wagner 2020).

A decline of pollinating insects has been identified in multiple studies, which is of particular concern due to their vital role to ecology and agriculture (Potts et al. 2010). An assessment report on pollinators, pollination and food production produced by IPBES highlighted the enormous economic value of insect pollination and evidence for their decline in abundance and diversity (IPBES 2016). Pollinators were considered high priority for inclusion in genetic monitoring programs in a report to SEPA (Posledovich et al. 2021b), which identified 34 pollinating insect species for inclusion in such a program. Fifteen species were identified as high priority, including ten bumblebee species.

Inbreeding can be particularly problematic for bees and bumblebees

The presence of inbreeding is damaging to small populations. Inbred individuals have increased levels of homozygosity in their genome, which causes recessive deleterious alleles to be present in the homozygous state where they cause decreased fitness and genetic abnormalities. Inbreeding is particularly damaging when it occurs in species of Hymentoptera, which include all bees, wasps and ants, because of the genetic basis of how sex is determined in these species (Lozier & Zayed 2017; Zayed & Packer 2005).

In contrast to humans and other vertebrates, in which both males and females are diploid, Hymentoptera are haplodiploid: males are haploid and females are diploid. This is governed by the complementary sex determination system, in which heterozygosity at a sex-determining locus is required to produce females (Crozier 1971; Heimpel & de Boer 2007). The production of females is assured by the presence of a large number of alleles at this locus, so that in large populations, diploid individuals almost always inherit two different alleles and become females. In rare cases, however, a diploid individual will inherit two copies of the same allele and will develop into a diploid male instead of a female.

Diploid males are unable to reproduce, and it is extremely damaging to a population if diploid males are produced at high frequencies. Diploid males are more likely to result from mating between two related individuals and thus are more common in smaller populations in which inbreeding regularly occurs. The production of diploid males therefore places an additional genetic burden on populations of Hymentoptera, and leads to an increased risk of extinction due to inbreeding (Zayed & Packer 2005). Identification of diploid males in a population is therefore an indicator that it is vulnerable to extinction.

Bumblebees are threatened by climate change

There are approximately 260 species of bumblebee in the world. They all belong to the genus *Bombus* and are separated into 15 currently recognised subgenera (Cameron et al. 2007; Williams et al. 2008). They are effective pollinators that are particularly important in temperate regions of the northern hemisphere. Declines in bumblebee species abundance and distribution have been reported globally in recent decades (Williams et al. 2009; Cameron & Sadd 2020). According to the IUCN Red List, nearly one-quarter of the bumble bee species that have been assessed are declining (De Meulemeester et al. 2016). Globally, factors such as habitat loss and degradation, pathogens and pesticides are major contributors to these declines. In addition, as bumblebees are typically cold adapted, a warming climate is predicted to have a major negative impact on their populations (Cameron & Sadd 2020).

The prognosis for bumblebee populations in Europe has been investigated in a report entitled "Climatic Risk and Distribution Atlas of European Bumblebees" (Rasmont et al. 2015) which resulted from the EU FP7 project STEP (Status and Trends of European Pollinators; http://step-project.net/). This project utilised over one million records of all bumblebee species found in Europe in order to model their climatic niche and project future suitable habitats on the basis of multiple future climate change scenarios. This study found that the majority of species are predicted to lose a significant proportion of their current ranges within this century. Up to 36% of species are predicted to lose more than 80% of their current range, with alpine and arctic species among the most severely threatened by climate change.

The distributions of both North American and European bumblebee species have shifted northwards in recent decades. However, although the southern edges of their distributions have retreated rapidly, expansions of the northern edges are occurring much more slowly.

Climate change is the most likely explanation for these changes, and in particular the increased occurrence of extreme weather events that limit survival (Soroye et al. 2020; Kerr et al. 2015). Bumblebee populations in the arctic and alpine regions are particularly threatened by climate change because it is impossible for them to shift their distributions further north (Williams & Osborne 2009; Rasmont et al. 2015; Soroye et al. 2020; Kerr et al. 2015).

Aim of study: Assessment of genetic diversity in mountain bumblebees in Sweden

The main focus of this study is to provide an assessment of genetic variation in bumblebee species found in mountain habitats in Sweden. These species are among the most threatened by climate change (Rasmont et al. 2015). We focus on six species that are restricted to arctic and mountain environments, and one species that is present in the Swedish mountains but is also widespread in multiple habitats throughout Europe (Table 1, Appendix I). Records of these seven species from Artdatabanken are shown in Fig. 1, and reflect their geographical range in Sweden. These seven species belonging to three different subgenera and were all regarded as high priority for inclusion in genetic monitoring of pollinators by SEPA (Posledovich et al. 2021b).

The six mountain-restricted species belong to two subgenera: Alpinobombus and Pyrobombus. *Bombus balteatus*, *Bombus hyperboreus*, *Bombus pyrrhopygus*, and *Bombus alpinus* are classified as Alpinobombus. *B. balteatus* is relatively common in the Swedish mountains, whereas the other three have a more restricted range, inhabiting higher latitudes and elevations. *B. pyrrhopygus* was previously known as *B. polaris* but this species has been split in two based on phylogenetic analysis. This revision means that *B. polaris* is considered to be restricted to North America, whereas *B. pyrrhopygus* is found in Eurasia (Williams et al. 2019, 2015). In contrast to the other species, *B. hyperboreus* is a parasitic species, laying eggs in other bumblebee nests rather than establishing its own nest (Ødegaard et al. 2015). *Bombus lapponicus* and *Bombus monticola* are classified as Pyrobombus. Both of these species are found throughout the Swedish mountains.

In addition to their range in Scandinavia, two of these species are also present in mountain environments further south in Europe. *B. monticola* is found in high elevations at many locations across Europe including the British Isles, Pyrenees, Alps and Balkans. *B. alpinus* occurs at the highest elevations in the Alps in addition to the Scandinavian mountains. The other mountain species *B. lapponicus*, *B. pyrrhopygus*, *B. balteatus*, *B. hyperboreus* are not present south of Scandinavia. All of these species apart from *B. alpinus* and *B. monticola* are also found further east in Siberia (Rasmont et al. 2015).

The final species *Bombus pascuorum* is a member of the subgenus Thoracobombus. In contrast to the other species, it is common all over Europe, present in a range of habitats from the Arctic to the Mediterranean. It is the most abundant bumblebee species in Europe, and is relatively unthreatened by climate change compared to other species (Rasmont et al. 2015). It therefore serves as a very useful comparison to other species as it is expected to have relatively high levels of genetic variation.

All of the mountain species are threatened by climate change, and those that are most at risk are the ones restricted to the highest elevations and latitudes. *B. pyrrhopygus*, *B. alpinus* and *B. hyperboreus*. In particular, *B. alpinus* and *B. hyperboreus* are classified as vulnerable in the species red list whereas all the other species studied here are considered least concern as

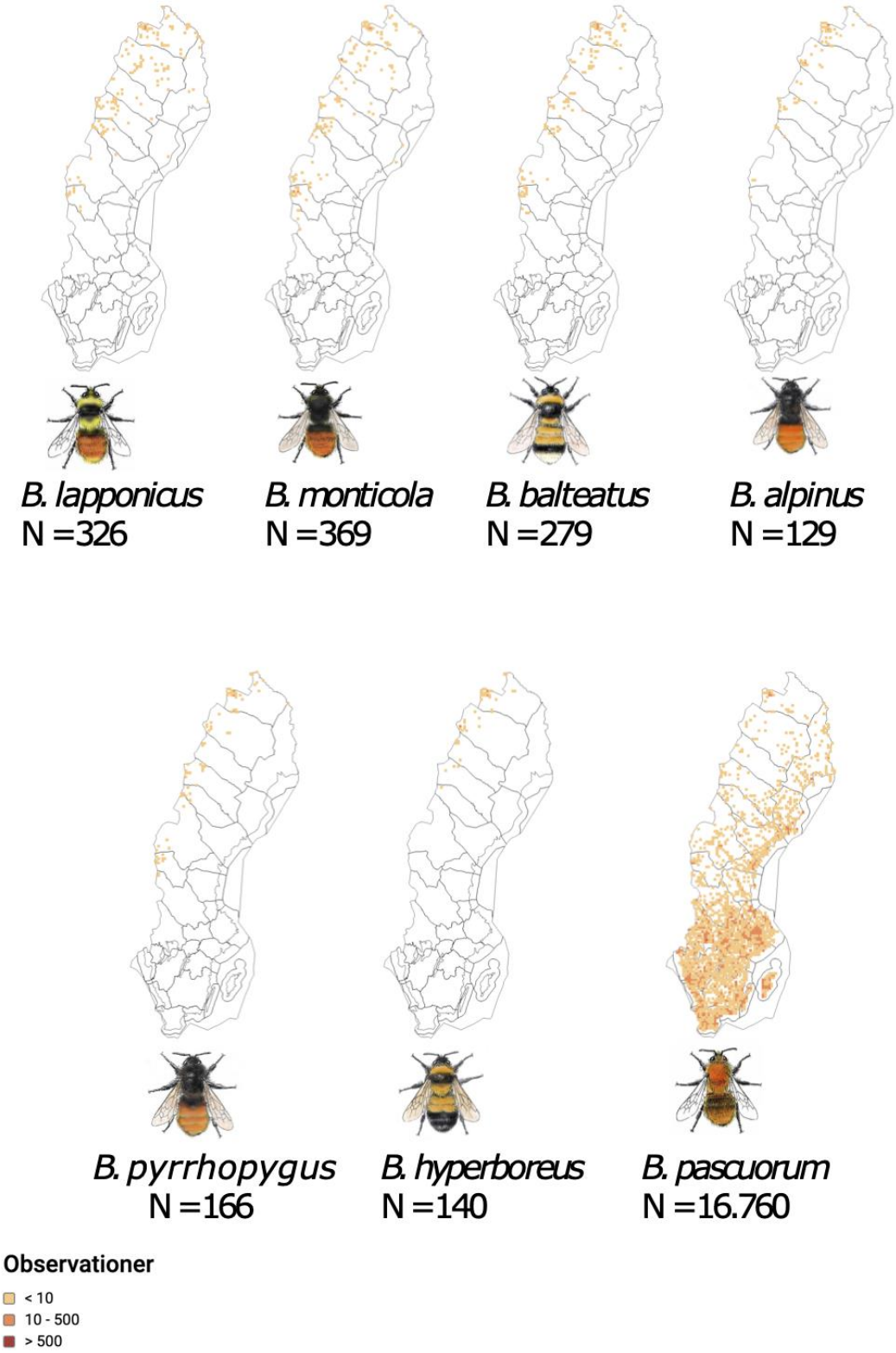


Figure 1. *Distribution of study species in Sweden. Points on the map represent observations since 1997 reported in Artdatabanken (N is the number of observations). B. lapponicus, B. monticola and B. balteatus are found widely in the Swedish mountains and in the Arctic. B. alpinus, B. pyrrhopygus and B. hyperboreus are restricted to higher elevations and latitudes in these areas. B. pascuorum is a common bumblebee in the whole of Europe, distributed from the Arctic to the Mediterranean. Maps from artfakta.se. Drawings © Björn Cederberg.*

they are not immediately threatened (Table 1). In addition to red list classifications, species have been classified into climate risk categories based on species distribution modelling (Rasmont et al. 2015). Two of the species under study here, *B. pyrrhopygus*, and *B. hyperboreus*, are assigned the most extreme risk class, HHHR, indicating that >95% of their inhabitable range is predicted to be lost by 2100 under a realistic future climate change scenario. The other four mountain species, *B. alpinus*, *B. balteatus*, *B. lapponicus*, and *B. monticola* are assigned the category HHR, indicating a loss of >85%.

Bombus pascuorum is a highly polytypic species with 24 described subspecies found throughout Europe (Lecocq et al. 2015). Four of these subspecies are found in Sweden. *B. p. smithianus* north of the Arctic circle, *B. p. sparreanus* in Norrland south of the Arctic circle, *B. p. pallidofacies* in Svealand and Götaland, *B. p. gotlandicus* on Gotland. Subspecies of the other species of interest have not been described. *B. pascuorum* is the most widespread and abundant bumblebee species in Europe and can survive in urban and suburban areas where other bumblebee species are scarce (Rasmont et al. 2015). *B. pascuorum* is assigned the lower climate risk category R, indicating >50% predicted loss in a similar time. It is important to note that these estimates do not take the possibility of migration into account. *B. pascuorum* may be able to account for loss of suitable habitat by expanding to higher latitudes and elevations. In contrast, the mountain species, are already at the northern limits of their potential distributions.

The goal of the project is to provide a first estimate of genetic variation in these seven bumblebee species using whole-genome sequencing by assaying populations throughout their range in Sweden. This is used to investigate population genetic parameters such as N_E and determine levels of population subdivision and inbreeding. The results are used to evaluate the priority and feasibility to include these species in a in genetic monitoring program and provide first estimates of population genetic parameter for such a program.

Genomic resources available for focal species

We chose to assay genetic variation with whole genome sequencing as it is cost-effective and provides the maximum amount of information available to accurately estimate population genetic parameters. To analyse genetic variation in a species using genome sequencing, one needs a reference genome assembly of the species of interest or closely-related one to use for mapping sequencing reads.

As the Pyrobombus and Alpinobombus species form two distinct clusters, it is possible to use a single representative species' genome from each subgenus to map reads from all of the other species of that subgenus. As part of previous investigations into these groups, we have produced genome assemblies of *B. lapponicus sylvicola* (Pyrobombus) and *B. kirbiellus* (Alpinobombus) using samples from the Rocky Mountains, USA (Christmas et al. 2021, 2022). *B. lapponicus sylvicola* in USA is an extremely close relative of *B. lapponicus* in Sweden. The most recent taxonomic revision considers them be the same species (*B. lapponicus*) although they were previously considered to be two species, *B. sylvicola* (North America) and *B. lapponicus* (Eurasia) (Martinet et al. 2019). A similar situation exists for *B. kirbiellus* in North America and *B. balteatus* in Eurasia. In this case, these two species were previously both considered to be *B. balteatus*, but the most recent taxonomic revision has split *B. balteatus* into two species *B. kirbiellus* in North America and *B. balteatus* in Eurasia (Williams et al. 2019). The genome assemblies that are available are therefore from species currently known as *B. lapponicus sylvicola* (previously *B. sylvicola*) and *B. kirbiellus*

(previously *B. balteatus*), which are closely related to the study species *B. lapponicus* and *B. balteatus* found in Sweden. *Bombus pascuorum* is distantly related to the other study species and its genome assembly is publicly available.

Previous studies of focal species

As mentioned above, the mountain bumblebee species studied here have sister species in North America and their subgenera have circumpolar distributions. We have previously used whole genome sequencing to assess genetic variation and population structure in *B. lapponicus sylvicola* (sister species of *B. lapponicus*) and *B. kirbiellus* (sister species of *B. balteatus*) in several high-elevation locations in Colorado, USA (Christmas et al. 2021, 2022). One significant finding was the detection of the presence of a previously unknown cryptic species, which was indistinguishable from *B. lapponicus sylvicola* on the basis of standard morphological characters. Out of 281 samples identified as *B. lapponicus sylvicola*, 67 formed a distinct genetic cluster based on whole genome sequence data, representing this cryptic species, which we named "incognitus" (Christmas et al. 2021). Whether this species is present in Sweden is unknown. Furthermore, it is possible that surveys of genetic variation using whole-genome sequencing in other species may uncover more hidden cryptic species.

We have also analysed levels of genetic variation and population fragmentation in the Colorado populations of *B. lapponicus sylvicola* and *B. kirbiellus*. We found no evidence that populations were structured across a distance of over 150 km in the Rocky Mountains, indicating that there are no barriers to gene flow between mountains (Christmas et al. 2022). Similarly low levels of population structure have been observed for other mountain bumblebee species (Ghisbain et al. 2020; Lozier et al. 2011), although species of bumblebee with low abundance can exhibit high population differentiation (Darvill et al. 2006). Levels of population structure in the Swedish mountain bumblebees are unknown.

Summary of specific aims

Here we aimed to perform an assessment of levels of intraspecific genetic variation in six species of mountain bumblebee and one common bumblebee. This was achieved by collecting 333 bumblebee samples from throughout each species' range in Sweden and performing whole-genome sequencing. The data was used to estimate a range of population genetic parameters for each species to inform about their conservation status, and to evaluate the importance and feasibility of including the species in a long-term genetic monitoring program. Specific aims were as follows:

1) To generate phylogenetic trees of all the samples in order to a) confirm evolutionary relationship between species, b) confirm the identity of each sample, c) identify samples that were misidentified and confirm their identity, and d) search for the presence of any previously-unknown cryptic species that genetically cluster independently to known species.

2) To perform an analysis of population structure in each species to detect the presence of subpopulations and barriers to gene flow across their ranges. This analysis was carried out by analysing phylogenetic trees, principle-components analysis (PCA) and using the clustering algorithm Admixture (Alexander et al. 2009).

3) To estimate levels of genetic variation across the genome and use these values to estimate effective population size (N_E) for each species. We used Watterson's theta (θ_W) (Watterson

1975) to estimate genetic variation. This is an estimate of the harmonic mean of effective population size over evolutionary time that resulted in levels of genetic variation seen in the population in the present day. We also measured observed heterozygosity (H_{OBS}).

4) To perform a reconstruction of historical variation in N_E. We used the program SMC++, which analyses variation in coalescent time and the site frequency spectrum of genetic variation across the genome (Terhorst et al. 2017). This method to reconstructs historical fluctuations in N_E through evolutionary time and can provide insight into whether recent population declines have occurred.

5) To analyse the length distribution of runs of homozygosity (ROH) in samples of each species. The distribution of ROH is affected by population dynamics and reflects the prevalence of population bottlenecks and inbreeding (Ceballos et al. 2018).

RESULTS SUMMARY

- We collected and whole-genome-sequenced 333 samples of seven species of bumblebee in Sweden.
- Six of the species are restricted to arctic and alpine environments, and were collected in Lappland, Jämtland and Härjedalen. The seventh species is widespread and was collected from locations throughout Sweden.
- Estimates of N_E from genetic variation (θ_W) were in the range of 55,000 – 225,000 and species with more restricted distributions at higher latitudes and altitudes tend to be at the lower end of this range.
- No cryptic species were detected and species relationships reconstructed using phylogenetic trees matched expectations from previous studies.
- We did not uncover evidence for geographical population structure in the mountain bumblebee species using PCA and Admixture, indicating the absence of major barriers to gene flow between extant populations in the locations we surveyed.
- We identified a diploid male of the species *Bombus hyperboreus*, collected near the southern limit of its range. The production of diploid males, which are sterile, indicates that this population is suffering negative effects of inbreeding.
- We used SMC++ to infer past fluctuations in N_E. Two of the species that are currently rarer and have restricted distributions (*B. hyperboreus* and *B. pyrrhopygus*) showed signals of population decline since the last ice age.
- Analysis of runs of homozygosity (ROH) in each diploid sample revealed that the diploid male *B. hyperboreus* was homozygous at more than half of its genome, but that signs of inbreeding in other samples were much weaker or absent.
- We compared estimates of N_E in these species with their sister species in North America and found that N_E of the sister species of *B. balteatus* was substantially higher in North America than in Europe, indicative of higher abundance.

RESULTS

Field sampling of seven bumblebee species

Samples of the study species were collected during the summer of 2021. A large collecting effort was conducted through a field trip to Abisko and the vicinity, which yielded samples of all seven species (263 samples in total). Additional excursions to collect mountain species were made to Jämtland (19 samples) and Härjedalen (23 samples). Sampling efforts in these mountain localities further south were less successful. It is unclear whether this lack of success reflects declining populations, temporary population fluctuations, or other effects such as poor weather conditions. Samples of the widespread species *B. pascuorum* were collected from throughout its range in Sweden, including Gotland and Skåne. This incorporated six main geographic areas in which four different subspecies and hybrids between them were located.

The final datasets consisted of 333 samples. Three of the species - *B. alpinus*, *B. pyrrhopygus* and *B. hyperboreus* - have a more restricted distribution at higher elevations and latitudes. These species were found in lower numbers than the other, more widespread, mountain species and were only present in a subset of localities. All of the sampling locations for all

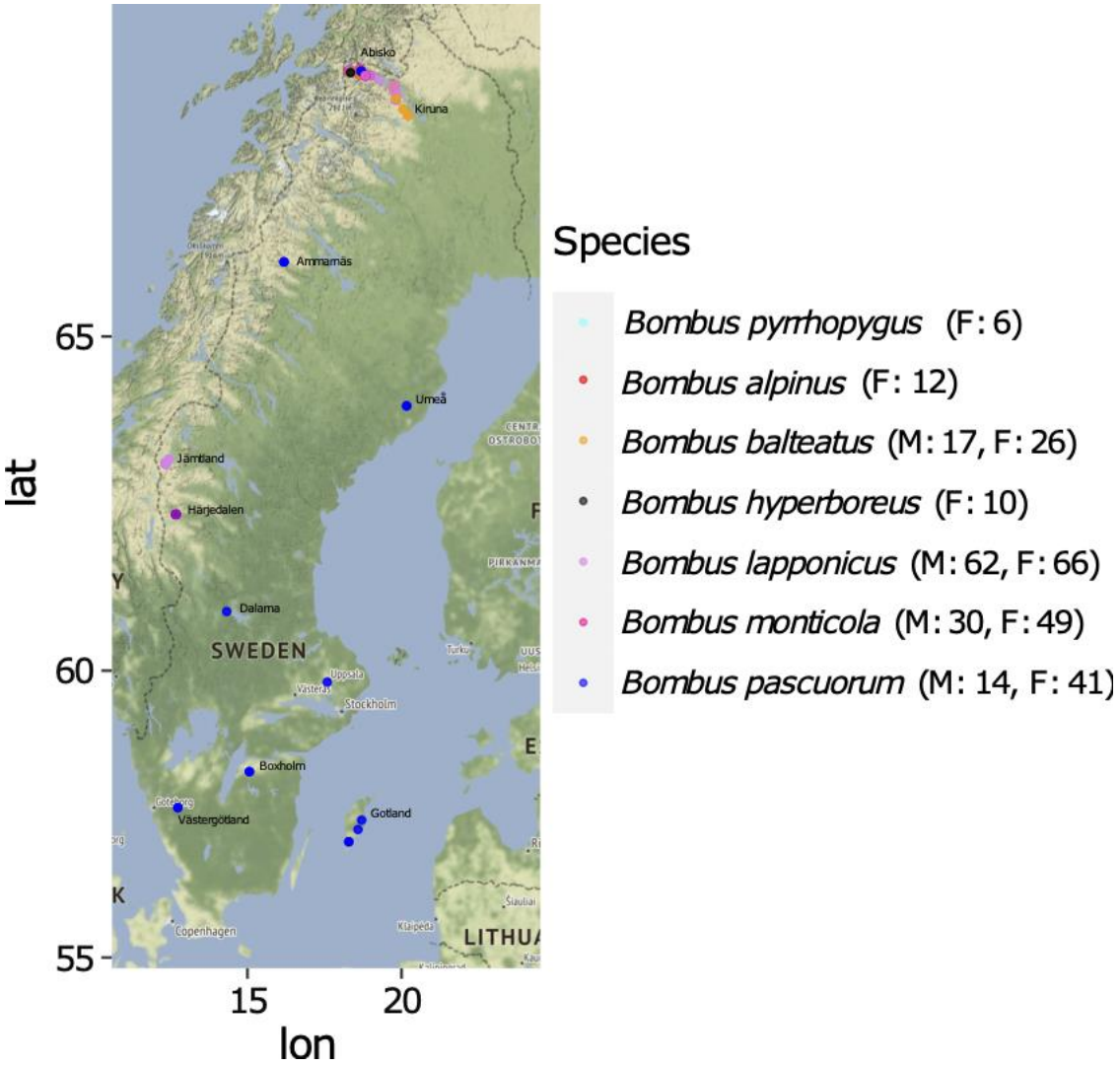


Figure 2. *Map of bumblebee sample collection for this study. The solid dots in different colours denote the sampling sites for each species across Sweden. Number of males (M) and females (F; workers and queens) collected are given.*

seven species are displayed in Fig. 2. A summary of samples collected in each location is presented in Table 2. A full list of samples that were sequenced and associated metadata is presented in Supplementary Table S1.

Whole-genome sequencing and bioinformatic analysis

All samples were subjected to whole-genome sequencing using the Illumina NovaSeq instrument at National Genomics Infrastructure, SciLifeLab, Stockholm. Raw sequence reads from each sample were aligned to the appropriate reference genome to generate three separate multi-species alignments: 1) The 207 Pyrobombus samples (*B. lapponicus* and *B. monticola*) were aligned to *B. lapponicus sylvicola* (ASM1967717v1)(Christmas et al. 2021). 2) The 71 Alpinobombus samples (*B. alpinus, B. pyrrhopygus, B. balteatus, B. hyperboreus*) were aligned to the *B. kirbiellus* reference assembly (ASM1920181v1)(Christmas et al. 2022). 3) The 55 *B. pascuorum* samples aligned to the *B. pascuorum* reference genome assembly (iyBomPasc1.1).

We first determined ploidy of each sample by detecting the presence of heterozygous sites. This analysis confirmed whether samples were male (haploid) or female (diploid). One

sample of *B. hyperboreus* was identified morphologically as a male, but was found to be diploid on the basis of the genetic analysis, indicating that it is a diploid male (see below for additional analysis). We updated sample classifications in cases of incorrect classification in the field. Species identity was then confirmed by constructing neighbour-joining (NJ) trees for each of the three multi-species alignments. The seven different species were clearly distinguishable from NJ trees. A small number of samples that had been incorrectly classified were identified from NJ trees and correctly classified. In addition, the identity of the two species *B. alpinus* and *B. pyrrhopygus*, which are difficult to identify due to their convergence in morphology, could be confirmed by a combination of genetic clustering in NJ trees and detailed morphological analysis.

Levels of genetic variation

We estimated levels of genetic variation in all species using Watterson's theta (θ_W) and nucleotide diversity (average number of pairwise differences), which are both estimates of the population mutation parameter theta (Watterson 1975)(Table 3). Estimates of θ_W are 0.06% - 0.25%, which is slightly lower than the range observed in populations of the honeybee *Apis mellifera* (Wallberg et al. 2014). This parameter can be converted into an estimate of effective population size (N_E). This estimate represents a harmonic average in N_E over the time period to the coalescence of sequences in the sample tens of thousands of years ago (Nadachowska-Brzyska et al. 2022).

Levels of N_E vary between species. The highest estimates of N_E are exhibited by *B. pascuorum* (~220,000; Table 3). This is consistent with the fact that it is a widespread species in Europe, found in many habitats, which may be expanding in numbers (Rasmont et al. 2015). For comparison, populations of managed honeybees in Europe, which are also widespread have similar N_E of around ~200,000 (Wallberg et al. 2014). Two other species that are relatively widespread in the Swedish mountains, *B. lapponicus* and *B. balteatus* also exhibit relatively high N_E (~195,000 and ~180,000 respectively). *B. pyrrhopygus* also exhibited relatively high N_E (~170,000), which is surprising considering that it restricted to high elevations and was the species observed the fewest number of times during collection. Conversely, those species with the most restricted distributions had significantly lower levels of genetic diversity. Estimates of N_E for *B. alpinus*, *B. hyperboreus*, were all in the vicinity of 60,000 (Table 3). Surprisingly, *B. monticola* also has a N_E in this range, despite being one of the more numerous mountain species that might be expected to have higher N_E.

In general, therefore differences in levels of long-term N_E between species were consistent with expectations based on species abundance and range size, with some exceptions. None of the estimates of N_E are close to 500, below which a population is considered to be vulnerable from the negative effects of genetic drift (Allendorf et al. 2022). These analyses suggest that N_E has been historically large in these species and they do not show signs of recent drastic declines. However, these estimates of N_E reflect it is important to sample at multiple recent time points in order to accurately determine N_E and how it changes in contemporary populations.

Population structure

We estimated population structure and species relationships of all species using three methods: Neighbour-joining (NJ) trees (Saitou & Nei 1987), principle-components analysis (PCA) and Admixture analyses (Alexander et al. 2009).

The NJ tree of *B. lapponicus* and *B. monticola* reveals that the samples cluster by species with no additional structuring apparent (Fig. 3). It is also interesting to note that no additional cryptic species were observed. A similar study of the sister species *B. lapponicus sylvicola* in Colorado revealed that 67 out of 281 samples originally identified as *B. l. sylvicola* belonged to a previously undescribed species "incognitus" (Christmas et al. 2021). We did not observe cryptic species related to *B. lapponicus* in our dataset of 129 samples. This is consistent with the hypothesis that "incognitus" is restricted to North America, and whether it occurs outside of Colorado is unknown.

Samples of *B. lapponicus* from Lappland, Jämtland and Härjedalen showed no evidence of population substructure or clustering by geography in the PCA (Fig. 4). This was further supported by the Admixture analysis, in which a single population was the best supported model by cross-validation (Fig. 5). Samples of *B. monticola* from Abisko formed clusters according to PCA (Fig. 4). However, the three samples that form a separate cluster were collected in exactly the same location and the most likely reason for the clustering is therefore that they are related. No geographical structure was detected in the PCA and the Admixture analysis for *B. monticola* between Lappland, Jämtland and Härjedalan, which also supports a single panmictic population (Fig. 5). These results, therefore, do not provide any evidence for barriers to gene flow between populations of either *B. lapponicus* or *B. monticola* in the Swedish mountains, which is consistent with the hypothesis that they are breeding freely across their range in Sweden.

NJ trees for the species of subgenus Alpinobumbus - *B. balteatus*, *B. alpinus*, *B. pyrrhopygus*, and *B. hyperboreus* - reveal the evolutionary relationships between these species (Fig. 6). These species relationships are similar to those published based on analyses of phenotypic characters and molecular barcodes (Williams et al. 2019), which place *B. alpinus* and *B. pyrrhopygus* as sister taxa. The species *B. alpinus* and *B. pyrrhopygus* are extremely similar in morphology, and the phylogenetic tree from genome data was an important aid to confirm the species identity of each sample of these species.

The presence of genetic structure among populations of Alpinobombus species was investigated using PCA and Admixture. No significant evidence of clustering either within or among geographical regions was observed for any of these four species using PCA (Fig. 7). A single population was the best-supported model by cross-validation using Admixture for all species (Fig. 8). However, it should be noted that the sample sizes for *B. alpinus*, *B. pyrrhopygus* and *B. hyperboreus* were all relatively low (6 – 12 individuals per species), which results in low power to detect population structure. *B. alpinus* was only found in Abisko, which means it is impossible to identify genetic structuring among geographic regions for this species. For *B. hyperboreus*, a single sample collected in Jämtland did not cluster with the other samples from Abisko, which could indicate the presence of geographical structure in the species although more samples would be required to confirm or refute this.

In contrast to the species already discussed, *B. pascuorum* is a polytypic species found in a wide range of habitats across Europe (Lecocq et al. 2015). The sampling strategy for this species incorporated the four subspecies present in Sweden, in addition to samples identified

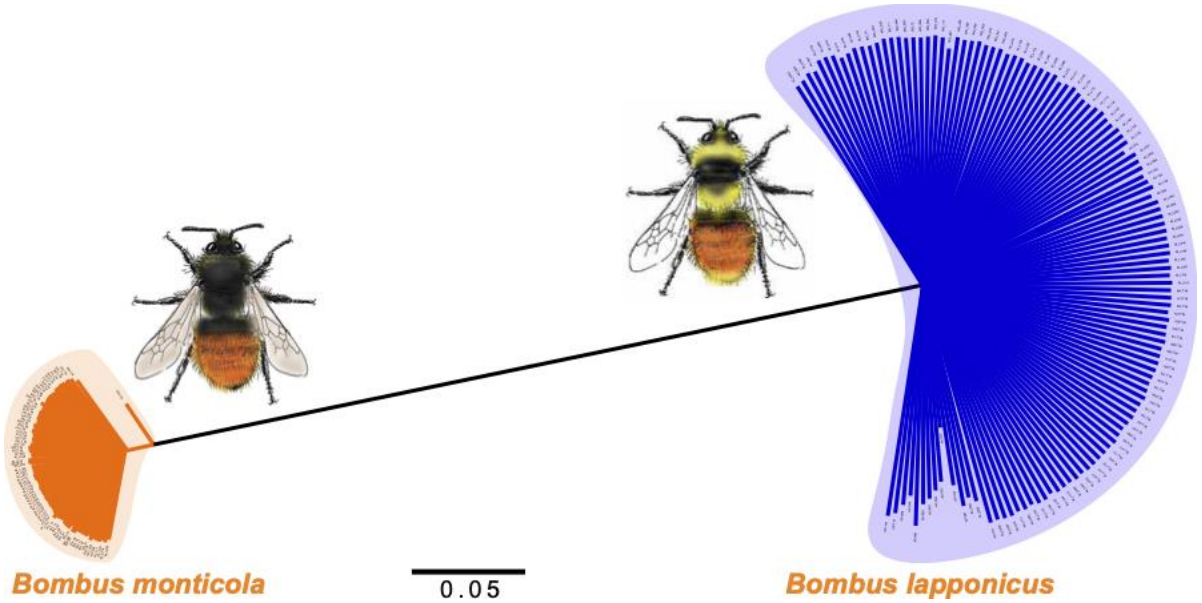


Figure 3. *Neighbour-joining tree for samples of B. lapponicus and B. monticola generated using whole-genome sequence data. Both species are in the subgenus Pyrobombus. Each external branch represents an individual sample.*

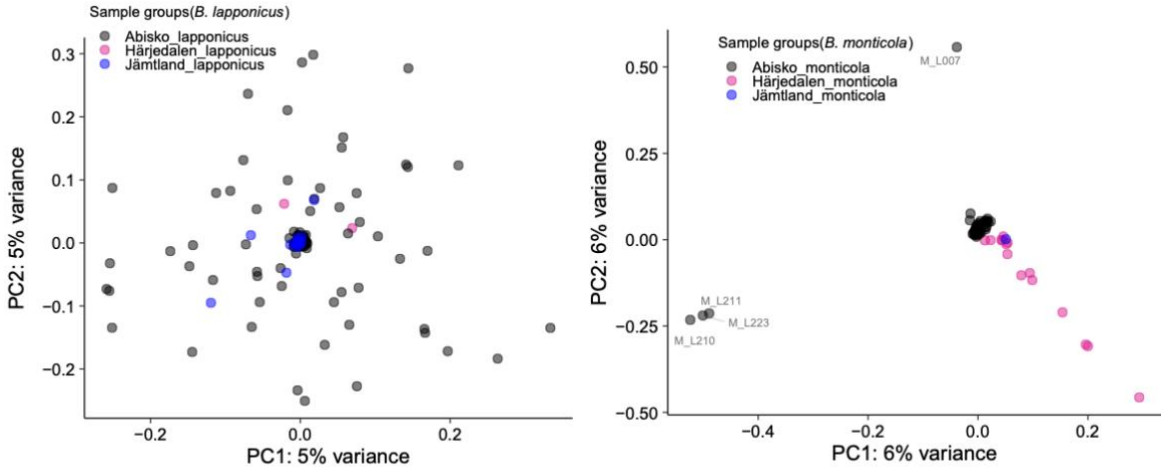


Figure 4. *Principal-component analyses for all geographical populations of B. lapponicus (left) and B. monticola (right). Sampling locations of each sample are marked.*

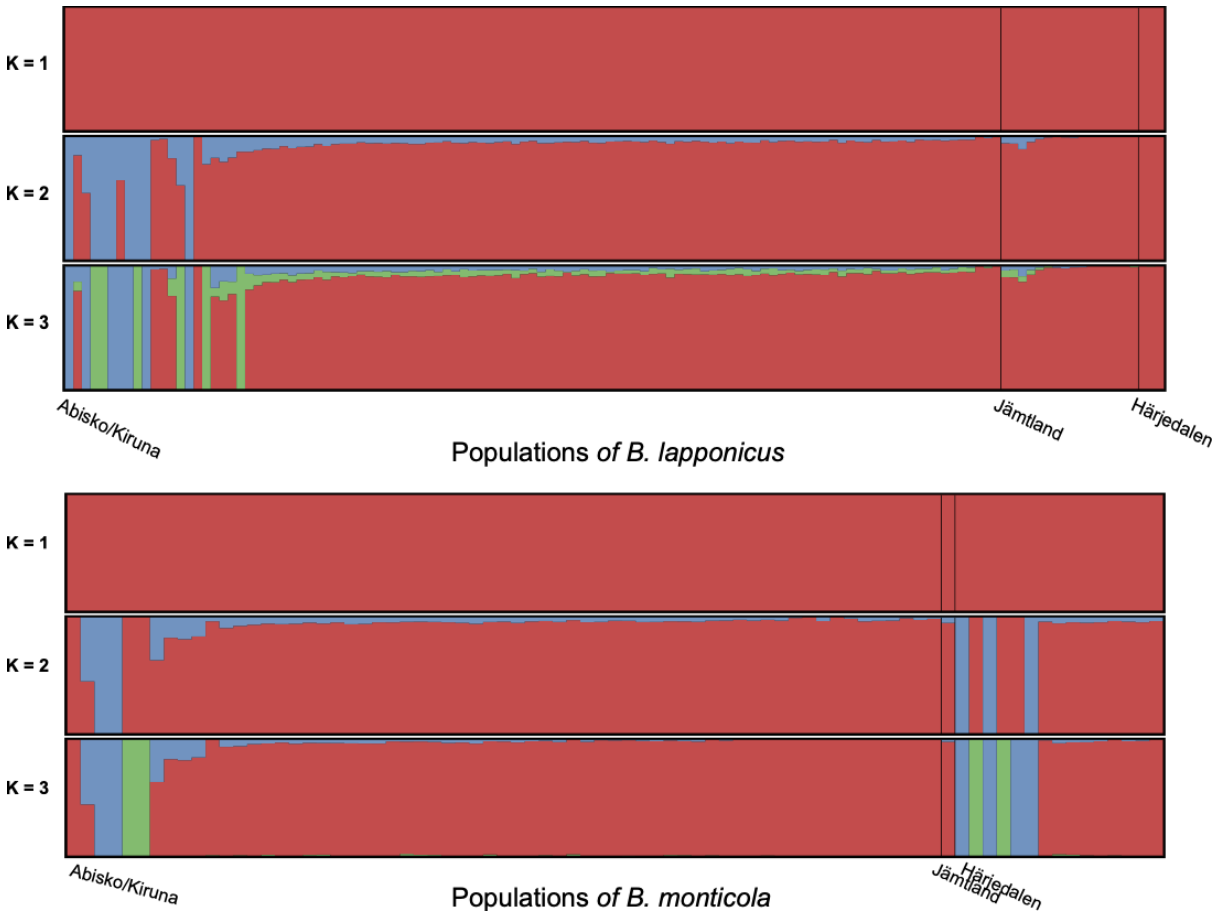


Figure 5. *Admixture analyses for all geographical populations of B. lapponicus (top) and B. monticola (bottom). No significant structure was detected for either species and K=1 was the best supported parameter value for both species.*

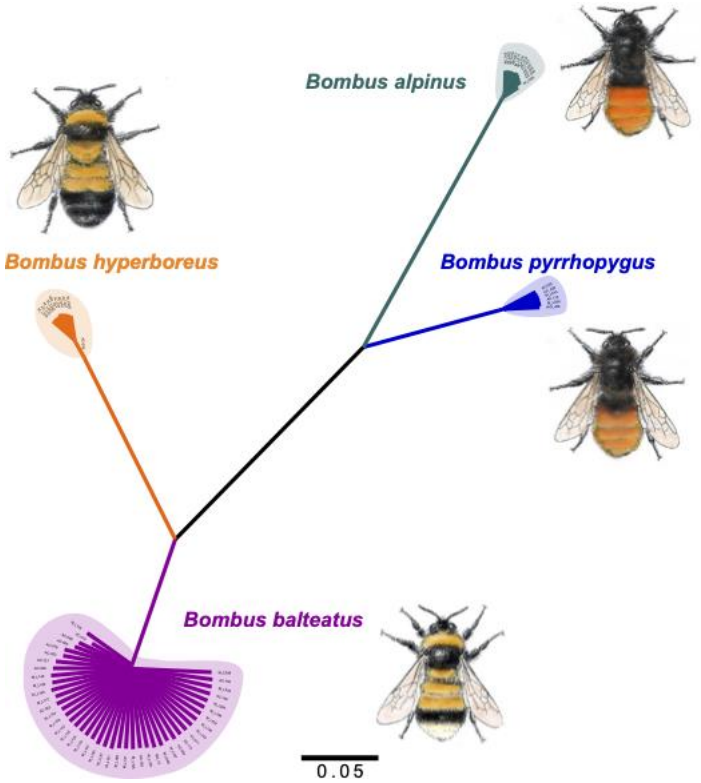


Figure 6. *Neighbour-joining tree for B. alpinus, B. pyrrhopygus, B. balteatus, and B. hyperboreus. All species are in the subgenus Alpinobombus. Each external branch represents an individual sample.*

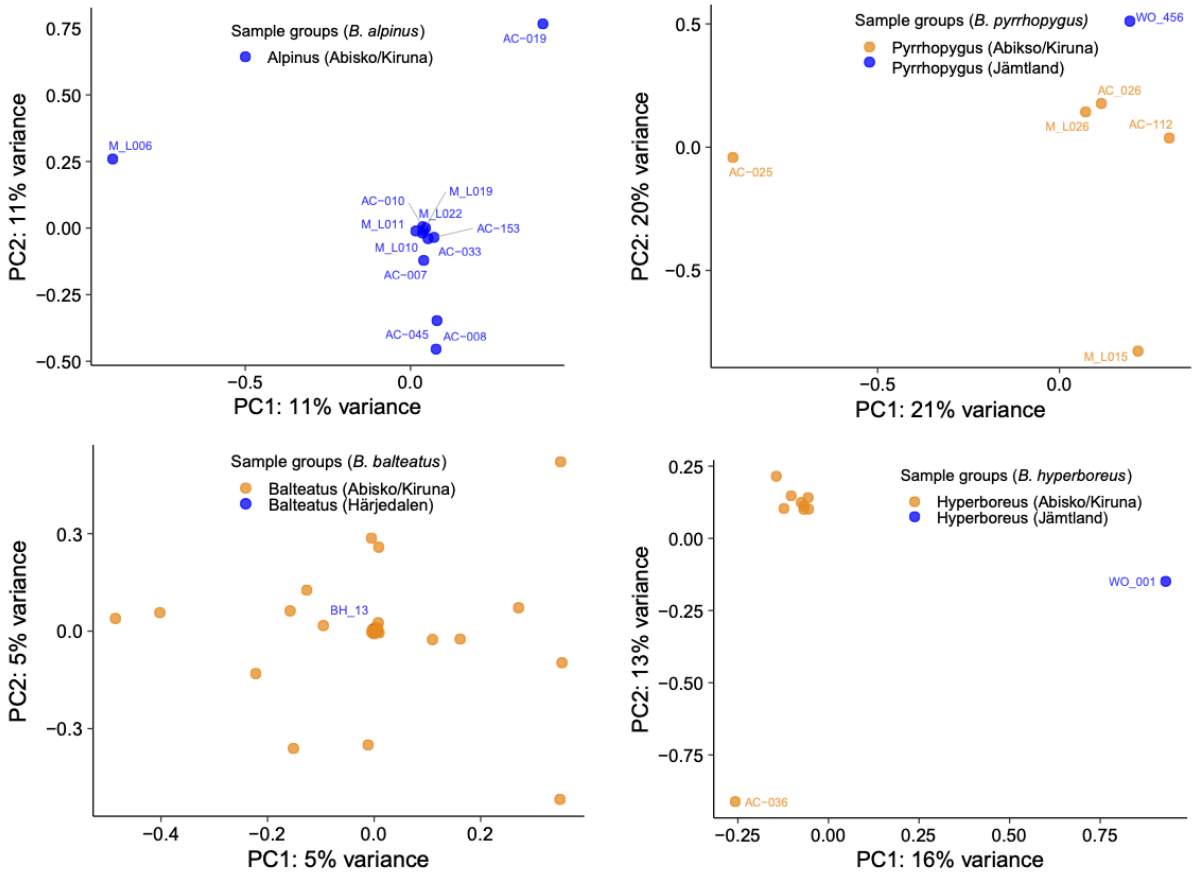


Figure 7. *Principal-component analyses for all geographical populations of B. alpinus, B. pyrrhopygus, B. balteatus, and B. hyperboreus. Sampling locations of each sample are marked.*

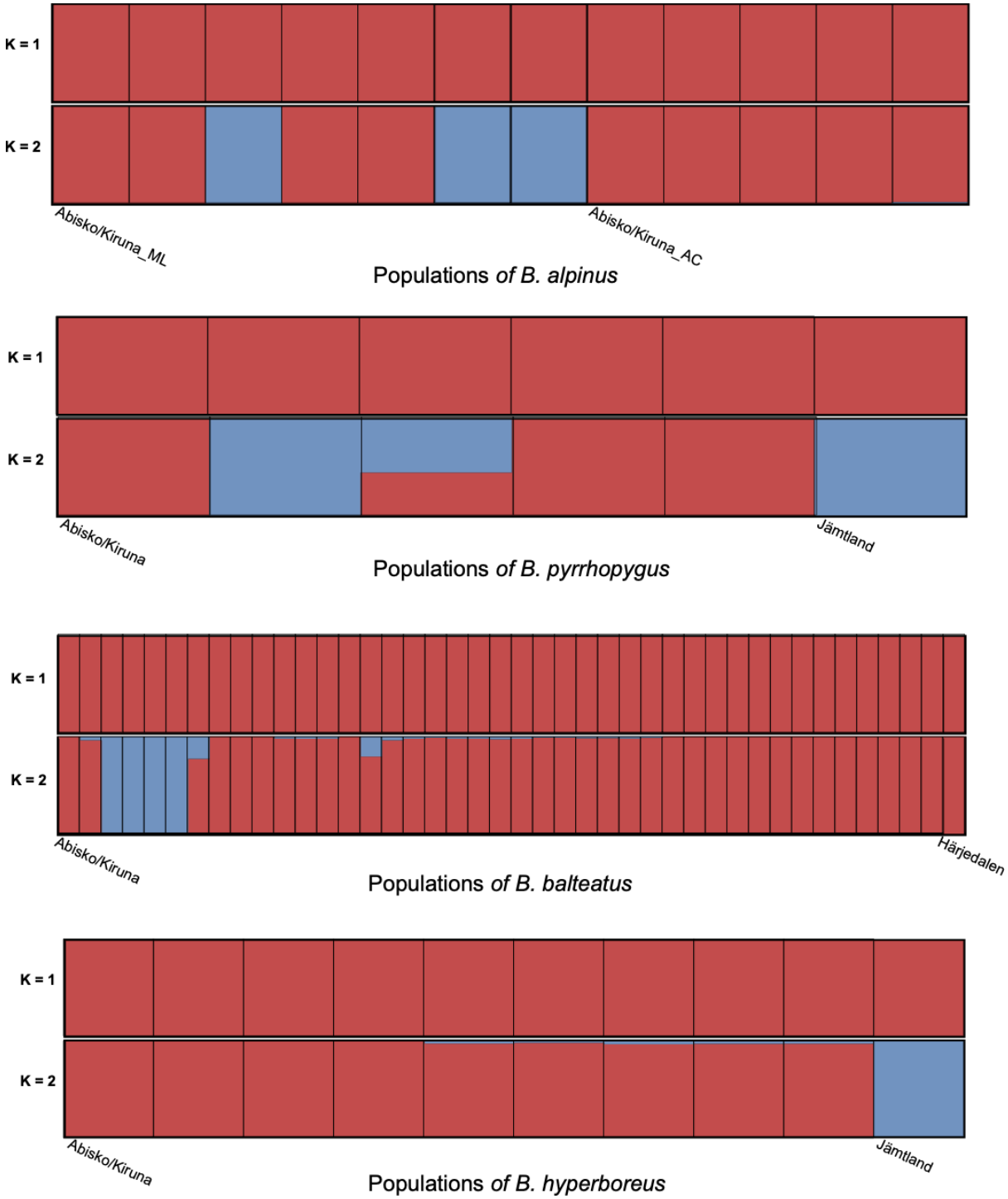


Figure 8. Admixture analyses for all geographical populations of B. alpinus, B. pyrrhopygus, B. balteatus, and B. hyperboreus. No significant structure was detected for any of the species and K=1 was the best supported parameter value for all species.

as hybrids between pairs of subspecies collected in hybrid zones. The NJ tree of these samples revealed limited levels of clustering (Fig. 9). The samples collected from Gotland, from the subspecies, *B. p. gotlandicus*, were clearly separated from the other samples. A limited amount of separation was also observed for samples from Abisko, consisting of the subspecies *B. p. smithianus*. The remaining samples, consisting of *B. p. sparreanus*, *B. p. pallidofacies*, and hybrids between subspecies, collected over a wide geographical area, do not show evidence of substructure. The PCA plot reveals a similar pattern, with samples from Gotland and Abisko forming separate clusters from the remaining samples from mainland Sweden. These remaining samples do not show evidence for geographic structuring (Fig. 10). The Admixture analysis also revealed that samples from Abisko and Gotland formed separate clusters from the remaining samples (Fig 11). Cross-validation analysis indicated that K=2 had best support, implying two major populations, one on Gotland and one on mainland Sweden, although the Admixture plots also show some degree of distinctness of the Abisko population.

Taken together, these analyses indicate that the six mountain bumblebee species studied here do not exhibit obvious population substructure across their ranges in Sweden. By contrast, the widespread polytypic species *B. pascuorum* exhibits a degree of substructure. A population on Gotland appears to be isolated from the mainland population. Furthermore, samples collected in Abisko appear to have a degree of isolation from other mainland populations further south, which display different colouration patterns and are described as belonging to different subspecies (Lecocq et al. 2015).

Inference of historical fluctuations in N_E using SMC++

We used the program SMC++ to infer historical fluctuations in N_E in all seven of the study species (Terhorst et al. 2017). This method uses the inferred distribution of coalescence times at loci across the genome and the site frequency distribution of SNPs to infer temporal variation in N_E. Hence, whereas standard estimators such as θ_W provide a harmonic average of N_E over evolutionary history, SMC++ is able to estimate fluctuations in N_E over evolutionary time.

Results of the SMC++ analysis for all species are shown in Fig. 12. The period since ~10^4 years ago represents time since the end of the last ice age. During this time period, populations of *B. hyperboreus* are inferred to have undergone a rather drastic decline. This analysis suggests that their current N_E could be less than 5,000, although historically it has been more than 10 times higher (Table 3, ~70,000). The species *B. pyrrhopygus* also shows signs of decline in this period, which could indicate a contemporary N_E around 25,000 compared to a value of ~75,000 estimated from θ_W. Both of these species have restricted arctic distributions, but may have been more widespread earlier in the current post-glacial period when conditions were more favourable in a greater part of Europe. It is therefore possible that their populations have declined in the current post-glacial period. The other mountain species under investigation do not show signals of major population expansions or declines and have been relative stable during the Holocene.

In contrast to the arctic and mountain species, there are signs that *B. pascuorum* has undergone a population expansion during this period. *B. pascuorum* is currently the most widespread and abundant bumblebee in Europe and is able to survive in a range of modern habitats (Lecocq et al. 2015). It is therefore not unlikely that this species has expanded in the current postglacial period.

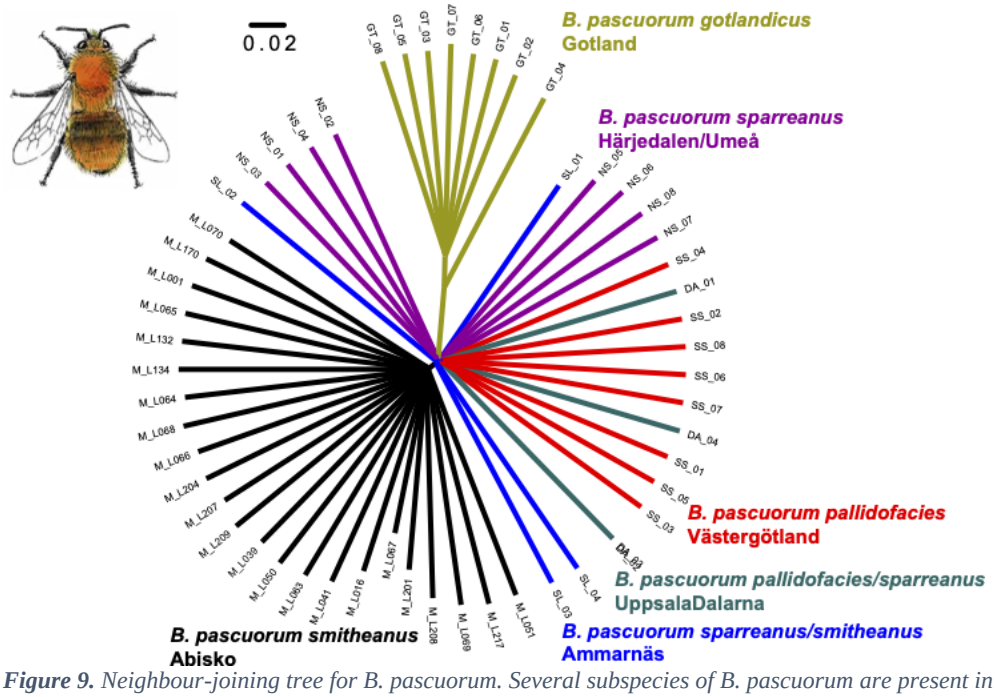


Figure 9. *Neighbour-joining tree for B. pascuorum. Several subspecies of B. pascuorum are present in Sweden. Each external branch represents an individual sample. Sampling locations and subspecies are indicated, including several from hybrid zones.*

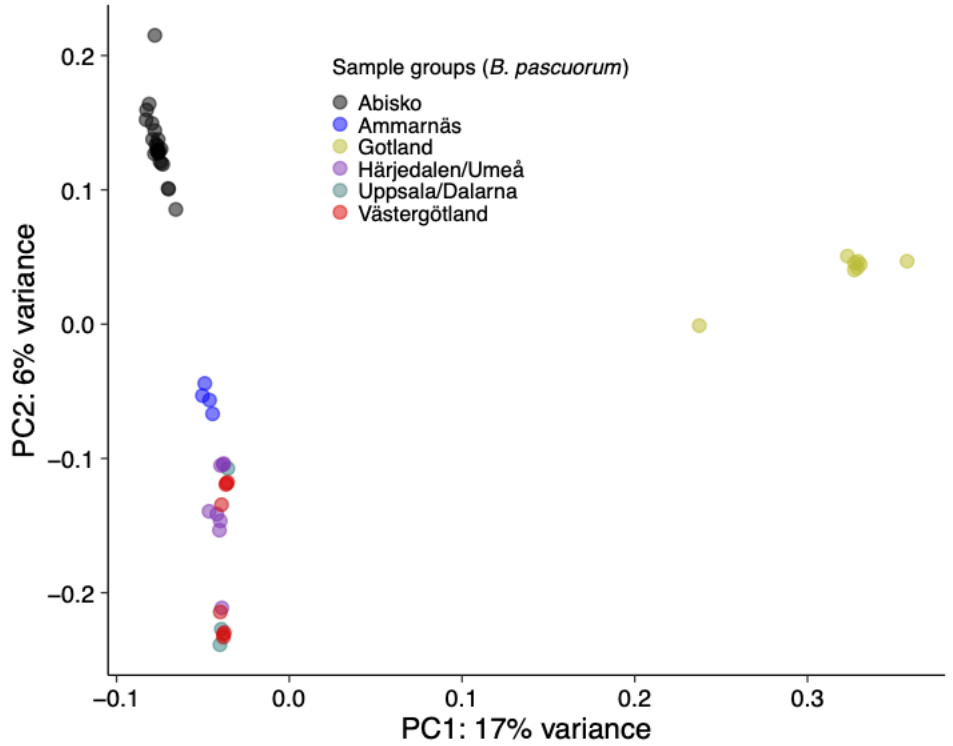


Figure 10. *Principal-component analyses for all geographical populations of B. pascuorum. Sampling locations are indicated.*

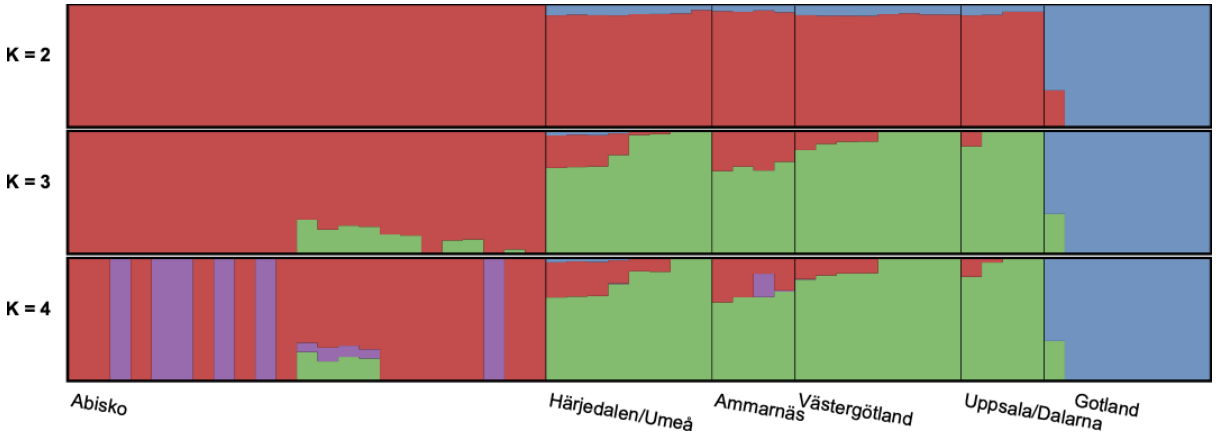


Figure 11. *Admixture runs for all geographical populations of B. pascuorum. The best supported parameter value is K=2, which indicates the presence of a distinct subpopulation on Gotland (B. pascuorum gotlandicus).*

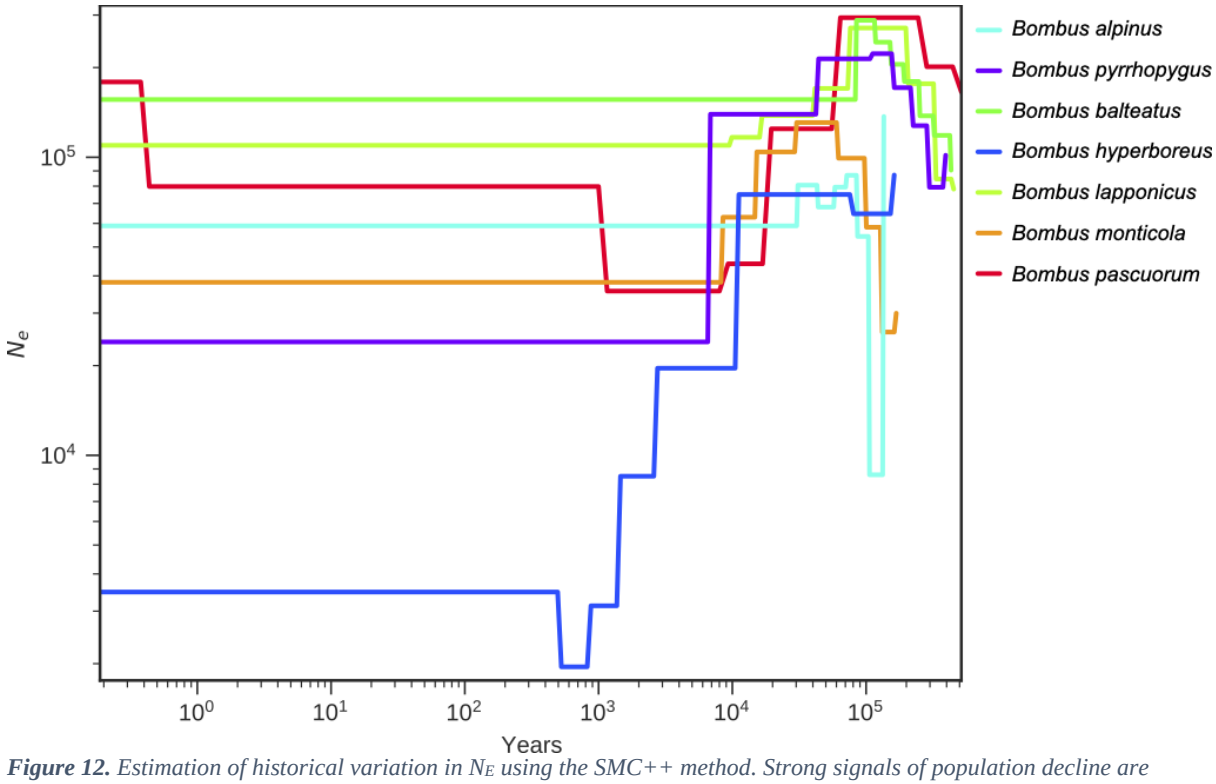


Figure 12. *Estimation of historical variation in N_E using the SMC++ method. Strong signals of population decline are inferred for B. hyperboreus, whereas population size increases are inferred for B. pascuorum.*

Runs of homozygosity

We scanned the genomes of all diploid samples (typically females) for runs of homozygosity (ROH). Only diploid samples were used because haploid samples cannot contain ROH. In general, inbreeding is expected to generate long ROH blocks in the genome. In contrast, the genomes of outbred populations can contain many shorter ROH blocks (Ceballos et al. 2018). The distribution of ROH in populations can therefore be indicative of inbreeding in a population, which could indicate that a population is at risk.

A summary of the ROH analysis is presented in Table 4 and a full analysis per sample is shown in Supplementary Figs. S1-4. Fig. 13 shows the proportion of the genome of each sample that falls into ROH >0.1 Mb, grouped by species. The most striking finding was a single individual of *B. hyperboreus* (WO_001) with >55% of the genome lying in ROH longer than 0.1 Mb (Fig. 13). This pattern is predicted to result from a sib-mating. This sample was also classified as a diploid male, which indicates that it is homozygous at the sex-determining locus. The sample is the only *B. hyperboreus* collected from Jämtland. All of the other samples of this species were found in the vicinity of Abisko. The observation of this highly inbred individual suggests that inbreeding could potentially be common in *B. hyperboreus*, and particularly so closer to the southern limits of its distribution in Jämtland. The observation of a diploid male indicates that this population is suffering negative effects of inbreeding.

A small number of individuals of *B. pyrrhopygus*, *B. pascuorum* and *B. monticola* were also identified with 5-20% of the genome covered by ROH blocks longer than 0.1 Mb indicating that inbreeding occurs sporadically in these species as well (Fig. 13). Outliers with long ROH tracts were not detected in the other species, but may still occur at low frequencies. Only a small number of samples contained any ROHs longer than 1 Mb and for the majority of species, none of samples had any such long ROHs (Table 4). The mean coverage of the genome by ROH is <5% for all species apart from *B. hyperboreus*, where the presence of the one highly inbred diploid male individual strongly influences this measure.

Average levels of ROH in *B. pascuorum* are high compared to the other species, with ROH covering ~5% of the genome on average (Table 4). As *B. pascuorum* is an extremely widespread bumblebee, it is unlikely that it is experiencing inbreeding. It is therefore unlikely that the average levels of ROH observed in any of the species here reflect a strong impact of inbreeding on their populations as they do not differ markedly from those observed in *B. pascuorum*.

Comparison of levels of genetic variation with North American mountain bumblebee species

Sister species of *B. lapponicus* and *B. balteatus* are also found in mountain and arctic locations in North America. These species are highly similar in morphology and recently evolutionarily diverged from their European counterparts. The sister species of *B. lapponicus* in North America is also known as *B. lapponicus* (subspecies *sylvicola*). The sister species of *B. balteatus* in North America is *B. kirbiellus*. For both *B. lapponicus* and *B. balteatus*, there is discussion regarding whether the sister species in North America and Europe should be considered the same or a separate species (Williams et al. 2019; Martinet et al. 2019). As the sister species are highly similar in lifestyle and behaviour it is interesting to determine whether levels of genetic variation differ between the European and North American populations.

We previously collected 214 samples of *B. lapponicus* and 299 samples of *B. kirbiellus* from Rocky Mountains, Colorado, USA (Christmas et al. 2021, 2022). The dataset generated here also permitted comparisons of genetic diversity in this population (Table 5). To do this we jointly called genetic variation in both datasets simultaneously in order to directly compare levels of genetic variation. For the two populations of *B. lapponicus* (USA and Sweden), we find that levels of nucleotide diversity and effective population size are extremely similar. We find that *B. kirbiellus* (USA) has significantly higher levels of genetic variation compared to *B. balteatus* (Sweden). This indicates that this sister species may be relatively more abundant in North America, which is consistent with observations from field collections (Christmas et al. 2022).

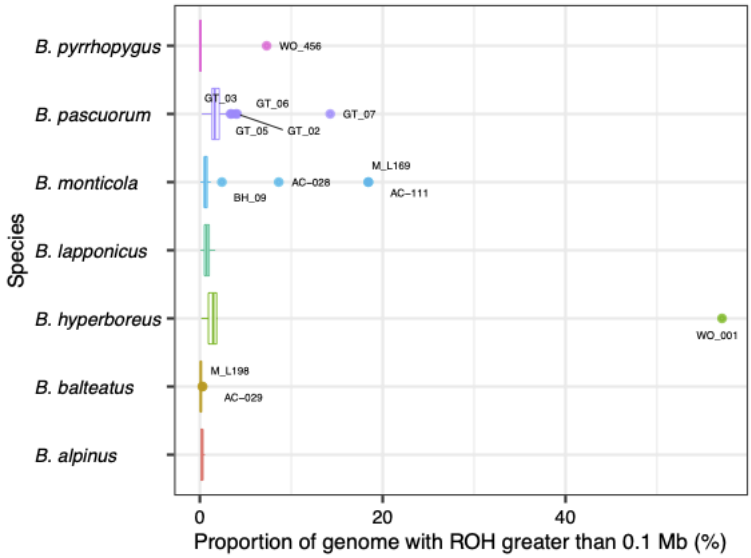


Figure 13. *Proportion of genome of each sample comprising ROH > 0.1 Mbp. Outliers are labelled.*

DISCUSSION SUMMARY

- Estimates of N_E for all species are much higher than commonly seen in endangered species, indicating that none of the species are at immediate risk of reduced fitness inbreeding.
- However, our estimates of N_E represent harmonic means over evolutionary time, and recent declines could be obscured. It is therefore important to obtain further measurements separated in time in order to accurately infer population trends.
- The lack of population structure observed in the Swedish mountains mirrors what is observed in sister species in North America. Detection of population structure in future population surveys would be indicative of declining populations.
- We detect some signals of population decline in our data:
    - We observe a highly homozygous diploid male *B. hyperboreus* in the southern extremes of its range, indicating that the population is suffering the effects of inbreeding.
    - We infer that both *B. hyperboreus* and *B. pyrrhopygus* have experienced population declines since the last ice age using SMC++ analysis. These species both have restricted arctic/alpine distributions.
- The six mountain species studied here are all classified in the "very high risk" or "extremely high risk" categories of risk from climate change, which suggests that it is important to accurately estimate population trends and monitor them in future.
- We propose two main avenues for future assessment of these species:
    - Genetic monitoring of populations in future can be used to reveal population trends in future and assess the impact of climate change
    - Extensive museum collections of the species including specimens >100 years old can be used to estimate genetic variation in the past and assess evidence for population declines

DISCUSSION

We used whole genome sequencing to survey genetic variation and population structure of six bumblebee species native to the Swedish mountains (n=278), and one widespread bumblebee species (n=55). The results are useful to evaluate the conservation status of each species, to design future genetic monitoring programs and to inform management decisions. This study also provides baseline data for use in future genetic monitoring. Below we discuss the main implications of our findings. We also outline proposals for future genetic monitoring and for comparisons with genetic variation in the past by sequencing museum specimens.

Relatively high N_E in mountain bumblebees

A value of N_E for a population less than 500 is generally considered to be cause for concern, suggesting it may be at risk of decline or extinction (Allendorf et al. 2022). Populations with low N_E can suffer from the negative effects of genetic drift and inbreeding, which cause deleterious mutations to build up and be expressed in homozygotes, leading to greater mortality (Willi et al. 2022). Small populations are also vulnerable to extinction from other causes. For all of the species under study here, we estimate N_E to be greater than 50,000. This suggests that the species are unlikely to face immediate threats due to drift or inbreeding. However, it is important to note that this estimate of N_E from θ_W reflects the harmonic mean over evolutionary time, rather than being an estimate specifically of contemporary N_E.

Additional studies, such measuring how genetic variation changes over time in a population, are required to determine whether present-day N_E differs from this estimate based on average levels of genetic variation.

Variation in estimates of N_E among species reflect differences in their range and abundance. *B. pyrrhopygus*, *B. balteatus*, and *B. lapponicus* have N_E close to that estimated for the widespread species *B. pascuorum* (~200,000), which indicates their range and abundance are also relatively large. Levels of N_E are also similar to other widespread social bees such as honeybees (Wallberg et al. 2014). By contrast, *B. alpinus*, *B. monticola* and *B. hyperboreus* have lower N_E than the other mountain bumblebee species. *B. alpinus* and *B. hyperboreus* have a more restricted distribution in at higher elevations and latitudes, and are therefore expected to have smaller population sizes. Lower levels of N_E in *B. monticola* are surprising considering that this species is fairly common in the Swedish mountains.

Low levels of population structure in mountain bumblebees

The six mountain species of bumblebees studied here did not show evidence for population structure among the mountain locations where they were collected in Lapland, Jämtland and Härjedalen. This indicates that there are no substantial barriers to gene flow and that the populations have historically been continuous across the Swedish mountain range. This provides an important baseline for future genetic monitoring, because detection of population differentiation in future would indicate declines.

Several previous studies have examined levels of population structure in bumblebee populations. Studies of genetic structure of mountain bumblebee species from Colorado, USA including *B. kirbiellus* and *B. lapponicus sylvicola* that are sister species of two of the species studied here do not show evidence for geographical substructure within the Rocky Mountains (Christmas et al. 2022). Other studies have however found evidence for genetic differentiation of these populations on different mountain ranges in North America (Koch et al. 2017; Whitley 2018). In general, studies in both Europe and North America have found that relatively abundant bumblebee species such as *B. terrestris* do not exhibit population structure over thousands of kilometres (Colgan et al. 2022; Ghisbain et al. 2020; Lozier et al. 2011), whereas rare and declining species (e.g. *B. muscorum*) may show structure on a scale of only 10 km (Darvill et al. 2006). Genetic differentiation of populations can occur in mountain ranges if species are restricted to high elevations, as with *B. bifarius* in the southern part of its range in the western USA (Jackson et al. 2018). The results presented here are therefore consistent with observations from other studies indicating that bumblebees do not exhibit population structure in the absence geographical barriers unless they have particularly low abundance.

Some species show evidence of inbreeding and population declines

Haplodiploid organisms such as bumblebees and other Hymenoptera are particularly vulnerable to inbreeding because it can lead to homozygosity at the sex-locus (see Background). This produces diploid males, which are unable to successfully fertilise females. The presence of diploid males can therefore severely reduce the fitness of a population. When populations become small and inbreeding is common, this effect can lead to a particularly extreme extinction vortex (Zayed & Packer 2005).

In our analysis we identified one example of a diploid male – a *B. hyperboreus* collected in Jämtland. This sample also had the highest degree of inbreeding detected in any sample: >50% of its genome consisted of ROH greater than 100 kbp in length. It was the only *B. hyperboreus* collected in Jämtland. It is noteworthy that this is the only observation of *B. hyperboreus* this far south reported in Artdatabanken, which has records over the last 25 years, with all of the other 143 reports restricted to Lapland (the majority in the vicinity of Abisko). This indicates that this sample was found at or close to the southern limit of the species distribution in the Swedish mountains. This finding indicates that inbreeding may be a large threat to the population of *B. hyperboreus*, particularly on the edges of its range.

A small number of samples of *B. pyrrhopygus* and *B. pascuorum* were also found to have extensive ROH. This suggests that inbred individuals occur at detectable frequencies in these populations and it may be important to monitor the populations in future for increased levels of inbreeding.

We also performed a SMC++ analysis to estimate fluctuations in N_E over the evolutionary history of the populations (Terhorst et al. 2017). This analyses variation in coalescent time across the genome can only be performed using genome-scale population sequencing data. The ability to perform such analyses is a major advantage of using whole-genome sequencing to survey genetic diversity. A significant finding of this analysis was that *B. hyperboreus* and to a lesser extent *B. pyrrhopygus* showed signals of population decline during the current postglacial period in the last ~10,000 years. This could suggest that the range of these species has shrunk and that current N_E are much smaller than predicted from overall levels of genetic variation. The observation of inbred individuals in these species from the ROH analysis also supports this suggestion.

Assessment of species vulnerability

Three of the mountain species studied here, *B. lapponicus*, *B. monticola* and *B. balteatus*, are found relatively commonly in the Swedish mountains. Of these, *B. monticola* is also found at higher elevations in the UK and in mountain ranges in mainland Europe, whereas the other species are restricted to the Scandinavian mountains. According to the IUCN Red List, all these species are categorised as "least concern" (Rasmont et al. 2015). However, they also fall in the climate risk category HHR (very high risk), indicating they are predicted to lose >85% of their current range by 2100. This implies that all of these species would occupy a significantly reduced range in the Scandinavian mountains, and *B. monticola* would additionally lose much of its range elsewhere in Europe except the high Alps.

Our analysis of genome variation in these three species did not uncover any evidence for depleted levels of genetic variation that would suggest they lie in the immediate risk zone for experiencing negative genetic effects due to inbreeding or genetic drift. We find levels of variation to be lower than expected in *B. monticola* based on its relative abundance, but no indications of population decline or inbreeding. This suggests that the populations are currently healthy, although will be strongly threatened by climate change in future.

The other three mountain species, *B. alpinus*, *B. pyrrhopygus*, and *B. hyperboreus*, have a much more restricted distribution in the Swedish mountains, occurring at higher elevations and latitudes. Of these, *B. alpinus* is also present at higher altitudes in the Alps, whereas the others are restricted to Scandinavia. *B. alpinus* and *B. hyperboreus* are categorised as "vulnerable" according to the red list, whereas *B. pyrrhopygus* is considered "least concern".

B. hyperboreus and *B. pyrrhopygus* are placed in the HHHR climate risk category (extremely high risk), indicating they are predicted to lose >95% of their range by 2100 whereas *B. alpinus* is placed in category HHR (very high risk).

Among these three species, we find evidence of population declines or inbreeding over evolutionary time. *B. hyperboreus* is a less abundant species with a very restricted range and, in contrast to the other species studied here, it has a specialised parasitic lifestyle. Our analysis indicated that this species has become much less abundant in the current postglacial period, and we uncovered one highly inbred diploid male individual among the ten samples we collected. This species is considered to be one of the most vulnerable to extinction in Europe (Rasmont et al. 2015) and our analysis suggests it may be suffering negative genetic effects of population declines, particularly in the southern limit of its range. We also find evidence that *B. alpinus* and *B. pyrrhopygus* exhibit lower or declining N_E relative to the other species, consistent with their relatively restricted abundance and distribution.

Climate change is the main threat to the six mountain bumblebee species studied here. Our analysis suggests their current ranges in the Swedish mountains would be sufficient to support healthy populations in future, although there are some indications of inbreeding and population declines. However, these ranges are predicted to shrink drastically under future scenarios of climate change, which could lead to population bottlenecks and fragmentation and eventually loss of all suitable habitat. Monitoring of these populations using both genetic and conventional methods would be advisable in order to understand how their populations are affected by changing climate in future.

Recommendations for future studies of mountain bumblebee populations

Conducting inventories of species distribution and abundance through observations in the field would be a powerful way to monitor the effects of future climate change. Currently, observations are recorded in Artportalen, but there are only a few hundred observations of each mountain species present in this database from a period of over 20 years. This does not accurately reflect abundance and is insufficient to detect changes over time. This is mainly due to the lack of organised field trips to the Swedish mountains to identify bumblebees. During this project, we conducted a small number of field trips to the mountains. We noted that we were less successful at collecting mountain bumblebees in the southern parts of the mountain range, but whether this is due to chance or reflects declining populations is unknown. All of the mountain species studied here are predicted to lose the southern extents of their ranges and become more restricted to higher elevations and latitudes due to climate change in future. Long-term coordinated efforts to record observations of these species throughout the Swedish mountains are required in order to monitor changes in their distributions.

The data presented here can be used as a baseline for future studies. Temporal change in several statistics with relevance to population viability can be investigated (Hoban et al. 2022). First, changes in levels of genetic variation can be assayed using observed and expected heterozygosity (H_{OBS} and H_{EXP}) or by θ_W. Secondly, changes in N_E can be monitored using estimates based on measures of genetic variation such as θ_W and using newly-developed methods based on linkage disequilibrium (LD) that are applicable to species with higher N_E such as most insects (Nadachowska-Brzyska et al. 2022; Hollenbeck et al. 2016; Browning & Browning 2015; Santiago et al. 2020; Webster et al. 2023). Thirdly, the extent of ROH can be monitored over time. If the extent of ROHs in a population increases over

time, this is a sign of increased inbreeding (Ceballos et al. 2018). The occurrence of highly inbred individuals can also be detected and monitored by this statistic. Fourthly, population structure can be detected using algorithms such as Admixture (Alexander et al. 2009) and genetic distance between populations can be estimated using F_{ST} statistics.

All of the statistics mentioned above can be used to detect changes in genetic variation by comparing values at multiple time points. Several studies have suggested threshold values of these statistics that lead to classification of populations as vulnerable or threatened (Schwartz et al. 2007). For example, a loss of heterozygosity >0.1% per year over a long time period would lead to a substantial loss of genetic variation. A standard assessment is that 95% of genetic variation should be predicted to remain over 100 years in order to consider that a population is not vulnerable (Allendorf et al. 2022).

We propose two main follow-up studies: Firstly, a genetic monitoring program can be used to determine future changes in the statistics mentioned above compared to the initial values presented here. We propose sampling all species studied here at approximately 5-year intervals in future for repeated genetic analysis. A sample size of no more than 50 in total per species should be sufficient, from up to 10 sampling localities spread across a north-south transect incorporating their entire range in the Swedish mountains. Secondly, contemporary changes in genetic variation can be uncovered by comparison with historical populations, which can be assayed using genome sequencing of museum specimens. We have contacted several museums in Norway and Sweden that house extensive collections, including many specimens collected >100 years ago. Whole-genome sequencing of these specimens is feasible by removing a small piece of tissue. Detailed proposals for these follow-up studies are available in Appendix II.

This study has demonstrated the utility of whole-genome sequencing for assaying genetic variation, which is both cost-effective and assays all available genetic information in an individual. It is particularly advantageous in bumblebees because of their small genome size, which means that several hundred samples can be sequenced simultaneously on the same sequencing lane and provide sufficient data. It would therefore be feasible to conduct genetic monitoring of mountain bumblebees in future using whole-genome sequencing. Whole-genome sequencing is also the optimal technique for recovering DNA sequences from museum specimens (Dabney et al. 2013).

An important recommendation for future efforts to assay genetic variation using genetic monitoring would be to conduct non-lethal sampling, for example by removing part of a back tarsus, rather than sampling the whole individual. This is particularly important for declining species such as *B. hyperboreus* whose populations could potentially be damaged by excessive sampling of whole individuals. These sampling techniques are also applicable to museum specimens, which should remain intact after sampling.

Outlook for management

Habitat loss and pesticides are strongly implicated in bumblebee declines worldwide (Cameron & Sadd 2020; Webster et al. 2023). However, these factors are unlikely to be as important for the populations under study here as the Swedish mountains are relatively unaffected by agriculture and urbanisation compared to the rest of Europe. There is however, a strong risk that these areas will become uninhabitable to the mountain species studied here due to a warming climate and its indirect effects on habitat and ecology.

Management options are available if severe declines occur in these species in future. One possibility is to actively translocate species to regions with suitable climates (Thomas 2011), although this approach is controversial (Webber et al. 2011). The most threatened species studied here (*B. alpinus*, *B. pyrrhopygus*, *B. hyperboreus*, *B. balteatus*) are predicted to disappear from the European mainland within a century. Three of these species (or their sister species) are reported to occur in the extreme north of the Siberian Arctic, Alaska, northern Canada and/or Greenland, where they are likely to be able to survive longer than in Europe due to a colder climate. *B. alpinus* is not reported outside of Europe and therefore would be extinct if it disappeared from this continent. A potential way to conserve threatened Scandinavian species could therefore be to introduce them to suitable habitats in other continents or northern archipelagos such as Svalbard. Translocations from populations in these locations could also be used to support the Scandinavian ones. However, more research is needed to confirm whether the sister species found on other continents are the same or different species to those studied here and whether they are able to successfully interbreed before conducting translocations into areas where these sister species exist (Williams et al. 2019; Martinet et al. 2019).

Conclusion

In summary, due to the extreme loss of habitat predicted for mountain bumblebee species due to future climate change, it is important to monitor their populations for signs of range contraction, population declines and loss of genetic diversity. Genetic monitoring could be an important component of this monitoring and would permit early detection of signs of population fragmentation or inbreeding. This study shows that such monitoring is feasible and provides the first set of data that could be used for a monitoring program.

METHODS

Sampling and DNA extraction

Specimens of seven bumblebee species were collected during the summer of 2021 from several locations in Sweden. This involved field trips to high elevation locations in Lapland, Härjedalen and Jämtland to collect mountain species and additional collections of the widespread species *B. pascuorum* in locations throughout Sweden. A large number of collaborators participated in the sampling effort. Sampling locations are displayed in Fig. 2, a summary of collections is found in Table 2, and a full list of samples and associated data is presented in Table S1. Samples were captured with sweeping hand nets and kept temporarily in Falcon tubes accompanied with ice packs. Species and sex/caste of each sample was identified using a guide (Ødegaard et al. 2015). Samples were stored in 95% ethanol at -20°C. The DNA was extracted using the Qiagen DNeasy Blood & Tissue Kit from a middle leg and thorax muscle taken from the coxal cavity of the middle leg.

Sequencing and variant calling

Paired-end sequencing libraries were prepared with the Nextera DNA Flex Library Preparation Kit and samples were sequenced on an Illumina NovaSeq6000 with 2 x 150 bp reads. The read data for each species were mapped to their corresponding reference genomes. *B. monticola* and *B. lapponicus* were mapped to *B. lapponicus sylvicola* (Pyrobombus) (NCBI Project ID: PRJNA646847). *B. balteatus*, *B. alpinus*, *B. hyperboreas* and *B.*

pyrrhopygus were mapped to *B. kirbiellus* (previously *B. balteatus*, Alpinobombus) (PRJNA704506), and *B. pascuorum* was mapped to *B. pascuorum* (*Thoracobombus*) (PRJEB43481). Mapping was performed using the BWA-MEM algorithm (Li 2013). Mappings were sorted and written to bam files using samtools. Picard was used to add read groups and mark duplicate reads in the bam files that were then indexed in samtools. The variants were called following the GATK best practices for non-model species (https://gatk.broadinstitute.org/hc/en-us/articles/360035535932-Germline-short-variant-discovery-SNPs-Indels-, last access March 1, 2022). HaplotypeCaller was run with each bam files with flag "-ploidy" to generate individual-specific gVCF files. All gVCF files were merged on a per contig basis by GenomicsDBImport, and called variants by GenotypeGVCFs. The resulting variant datasets were hard-filtered following the GATK recommended thresholds (https://gatk.broadinstitute.org/hc/en-us/articles/360035890471-Hard-filtering-germline-short-variants, last access March 3, 2022) to keep only high-quality and high-confidence variants.

Additional SNP filters were applied with bcftools (Li 2011): (1) only including biallelic SNPs with minimum minor allele count of 2 or greater and excluding SNPs with read depth and genotype quality per site in each sample under 3 and 20, respectively, i.e. (-v snps -m 2 -M 2 --min-ac 2:minor -e 'FMT/DP < 3 | FMT/GQ < 20'); (2) excluding sites with missing data exceeding 0.5, i.e. 'F_MISSING > 0.5'; (3) excluding sites with high mean depth across all samples and excessive heterozygosity (INFO/DP > X | INFO/ExcessHet > 25), where X = 1000 was for both *B. pascuorum* and *B. balteatus* datasets and X = 3134 for *B. sylvicola*. The additional filters INFO/DP and INFO/ExcessHet allowed us to remove sites due to alignments to multiple regions and variant calling errors. Heterozygous variants that were detected when HaplotypeCaller was run with "-ploidy 2" for haploid samples were false and filtered out for each cohort dataset. Three SNP datasets were generated, corresponding to sets of samples mapping to each of the three reference genomes, ready for the downstream analyses.

In order to directly compare levels of genetic variation in *B. lapponicus* and *B. balteatus* with their sister species in North America, the gVCF files generated by HaplotypeCaller were combined with a similar dataset of samples collected previously in the US Rocky mountains (Christmas et al. 2021, 2022). The variants were then genotyped and filtered for the two combined gVCF files following the same steps and filters except for -e 'INFO/DP > 25000' used for them. This cut-off 25000 was chosen on the basis of INFO/DP distributions in each combined dataset.

Population structure

Reconstruction of NJ trees (Saitou & Nei 1987), principal component analyses (PCA), and Admixture (Alexander et al. 2009) analyses were performed to reveal the genetic diversity of each geographical population using the SNP datasets. For these analyses only SNPs with minor allele frequency greater than 5% were included. Scripts provided by Simon Martin were used to calculate distance matrix for NJ trees taking into account mixed ploidy of samples (https://github.com/simonhmartin/genomics_general). The trees were made and visualized with FastMe (Guindon & Gascuel 2003) and FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/), respectively. The NJ trees were used to identify samples where species identification in the field was ambiguous and the morphology of these samples was studied by an expert (Björn Cederberg). This provided a confident species identification for all samples corroborated by both genetics and morphology. Samples were

initially all called as diploids, which allowed us to confidently confirm the sex of each sample by the absence of heterozygous sites in haploid males.

PCA analyses for each species were implemented in PLINK v1.9 (Chang et al. 2015). To better present the geographical clustering, one sample (DA_2) of *B. pascuorum* overlapped with sample DA_03 was removed due to their close relatedness, probably from the same nest. The Admixture program was used for 8 runs of population structure by specifying the level of ploidy for each sample (Alexander et al. 2009). The best-supported values of K were checked using "cross-validation" by adding "--cv" flag in this program.

Estimation of effective population size

Effective population size (N_E) for each species was estimated using an estimate of population mutation rate, Watterson's estimator (θ_W), involving two equations: $\theta_W = \frac{K}{a_n \text{ (genome size)}}$, where K is the number of segregating sites in the species and a_n is the (n-1)th harmonic number, n denoting the number of chromosome sets. N_E can be obtained using an estimate of θ_W and an estimate of the mutation rate per base μ based on $\theta = 2pN_E\mu$, where p is the ploidy of the species (p = 1.5 for haplodiploid bumblebees) (Fu 1994; Ferretti & Ramos-Onsins 2015). For these analyses, we used a direct estimate of mutation rate from *B. terrestris* ($\mu = 3.6 \times 10^{-9}$) (Liu et al. 2017).

Detection of runs of homozygosity (ROH)

Different algorithms can be implemented to identify ROH discussed in the review (Ceballos et al. 2018) such as PLINK that scans each chromosome by moving a window of fixed size (Shi et al. 2020) and HMM methods in BCFtools/ROH (Grossen et al. 2020).
With fine-scale genomic data mapped to a reference genome, we can identify homozygous segments of the genome that have few or no heterozygous nucleotide sites because two haplotypes inherited from the parents are identical. The distribution of ROH tract lengths can reveal details of demographic history and the timescale of inbreeding. For instance, smaller ROHs are inherited from distant ancestors and they reflect older inbreeding, whereas long blocks of ROH reflect more recent inbreeding (Ceballos et al. 2018; Grossen et al. 2020)

Segments of ROHs for females from each species using "--homozyg" flag in PLINK v1.9 (Chang et al. 2015), which uses a sliding window approach to scan each individual's genotype at each marker position across the genome. The following criteria were defined to present the length of ROHs: (1) a minimum ROH length of 10 kb containing at least 50 SNPs; (2) a maximum gap between consecutive SNPs of 1 Mb; (3) a minimum density of one SNP in 50 kb; (4) a sliding window of 50 SNPs across the genome with one SNP step; (5) no tolerance of heterozygous sites and a maximum three missing genotypes; (6) a default window threshold of 0.05. The lengths of ROH were divided into three classes: 0.01~0.5Mb, 0.5~1 Mb, 1~1.5 (detected maximum length) Mb.

Inference of population demographic history

SMC++ (Terhorst et al. 2017) was used to infer past fluctuations in N_E. To ensure reliability of SNP calling, only female samples with read coverage greater than 13X~ were selected for running SMC++. The mean depth per individual was estimated using the "--depth" flag in VCFtools (Danecek et al. 2011). The previously filtered VCF files were used as inputs by

keeping only these samples. Eventually, 49 samples were used for the demographic inference, including seven *B. pascuorum*, 10 *B. monticola* and 11 *B. lapponicus*, as well as 12 *B. balteatus* and three for *B. alpinus*, *B. hyperboreus*, and *B. pyrrhopygus*, respectively. The "composite likelihood (-d)" option was used as it incorporates genealogical information from additional distinguished individuals into the analysis, potentially leading to improved estimates (Terhorst et al. 2017). Estimations from SMC++ were scaled using a generation time (g) of 1 year and a mutation rate (μ) 3.6×10^{-9} substitutions per site per year (see above).

Storage of samples and data archiving

All specimens collected during this study are stored in 95% ethanol at low temperature (-20°C) and are largely intact except for the removal of a small amount of tissue for DNA extraction. All sequence data generated by the study have been deposited at the National Center for Biotechnology Information (NCBI) together with associated metadata under BioProject PRJNA890771. Scripts for bioinformatic analysis are available at GitHub: https://github.com/yzliu01/Swedish_mountain_bumblebee.

CONTRIBUTIONS AND ACKNOWLEDGEMENTS

The following people participated in sample collection: Matthew Webster (Lappland, Gotland), Aoife Cantwell-Jones (Lappland), Johan Lindell (Lappland), Björn Cederberg (Härjedalen, Uppland, Dalarna), Willhelm Osterman (Jämtland), Peter Nolbrant (Västergötland), Niklas Johansson (Östergötland), Sven Hellqvist (Västerbotten).

Yuanzhen Liu, Anna Olsson, and Tuuli Larva performed DNA extractions. Yuanzhen Liu performed the data analysis. Björn Cederberg provided illustrations, and performed morphological classification of some samples. Göran Holmström provided photographs of the study species. Turid Everitt and Matthew Christmas assisted with data analysis.

The project made use of facilities at the Abisko Scientific Research Station, which is connected to SITES (Swedish Infrastructure for Ecosystem Science), INTERACT (International Network for Terrestrial Research and Monitoring in the Arctic) and SLU (Swedish University of Agricultural Sciences).

The computational analysis was enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) at UPPMAX partially funded by the Swedish Research Council through grant agreement no. 2018-05973. We also acknowledge support from the National Genomics Infrastructure in Stockholm funded by Science for Life Laboratory, the Knut and Alice Wallenberg Foundation and the Swedish Research Council.

REFERENCES

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19:1655–1664. doi: 10.1101/gr.094052.109.

Allendorf FW, Funk WC, Aitken SN, Byrne M, Luikart G. 2022. *Conservation and the Genomics of Populations*. Oxford University Press.

Browning SR, Browning BL. 2015. Accurate Non-parametric Estimation of Recent Effective Population Size from Segments of Identity by Descent. Am J Hum Genet. 97:404–418. doi: 10.1016/j.ajhg.2015.07.012.

Cameron SA, Hines HM, Williams PH. 2007. A comprehensive phylogeny of the bumble bees (Bombus). Biological Journal of the Linnean Society. 91:161–188. doi: 10.1111/j.1095-8312.2007.00784.x.

Cameron SA, Sadd BM. 2020. Global Trends in Bumble Bee Health. Annu. Rev. Entomol. 65:209–232. doi: 10.1146/annurev-ento-011118-111847.

Ceballos FC, Joshi PK, Clark DW, Ramsay M, Wilson JF. 2018. Runs of homozygosity: windows into population history and trait architecture. Nat Rev Genet. 19:220–234. doi: 10.1038/nrg.2017.109.

Ceballos G, Ehrlich PR, Dirzo R. 2017. Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. Proc. Natl. Acad. Sci. U.S.A. 114. doi: 10.1073/pnas.1704949114.

Chang CC et al. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 4:s13742-015-0047–8.

Christmas MJ et al. 2022. A genomic and morphometric analysis of alpine bumblebees: Ongoing reductions in tongue length but no clear genetic component. Mol Ecol. 31:1111–1127. doi: 10.1111/mec.16291.

Christmas MJ et al. 2021. Genetic Barriers to Historical Gene Flow between Cryptic Species of Alpine Bumblebees Revealed by Comparative Population Genomics. Mol Biol Evol. doi: 10.1093/molbev/msab086.

Colgan TJ et al. 2022. Genomic Signatures of Recent Adaptation in a Wild Bumblebee. Molecular Biology and Evolution. 39:msab366. doi: 10.1093/molbev/msab366.

Crozier RH. 1971. Heterozygosity and Sex Determination in Haplo-Diploidy. The American Naturalist. 105:399–412. doi: 10.1086/282733.

Dabney J et al. 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. PNAS. 110:15758–15763. doi: 10.1073/pnas.1314445110.

Danecek P et al. 2011. The variant call format and VCFtools. Bioinformatics. 27:2156–2158. doi: 10.1093/bioinformatics/btr330.

Darvill B, Ellis JS, Lye GC, Goulson D. 2006. Population structure and inbreeding in a rare and declining bumblebee, *Bombus muscorum* (Hymenoptera: Apidae). Molecular Ecology. 15:601–611. doi: 10.1111/j.1365-294X.2006.02797.x.

De Meulemeester T et al. 2016. *European Red List of Bees*. Publications Office of the European Union doi: doi/10.2779/51181.

Des Roches S, Pendleton LH, Shapiro B, Palkovacs EP. 2021. Conserving intraspecific variation for nature's contributions to people. Nat Ecol Evol. 5:574–582. doi: 10.1038/s41559-021-01403-5.

Díez-del-Molino D, Sánchez-Barreiro F, Barnes I, Gilbert MTP, Dalén L. 2018. Quantifying Temporal Genomic Erosion in Endangered Species. Trends in Ecology & Evolution. 33:176–185. doi: 10.1016/j.tree.2017.12.002.

Ellegren H, Galtier N. 2016. Determinants of genetic diversity. Nature Reviews Genetics. 17:422–433. doi: 10.1038/nrg.2016.58.

Ferretti L, Ramos-Onsins SE. 2015. A generalized Watterson estimator for next-generation sequencing: From trios to autopolyploids. Theoretical Population Biology. 100:79–87. doi: 10.1016/j.tpb.2015.01.001.

Formenti G et al. 2022. The era of reference genomes in conservation genomics. Trends in Ecology & Evolution. 37:197–202. doi: 10.1016/j.tree.2021.11.008.

Fu YX. 1994. Estimating effective population size or mutation rate using the frequencies of mutations of various classes in a sample of DNA sequences. Genetics. 138:1375–1386. doi: 10.1093/genetics/138.4.1375.

Ghisbain G et al. 2020. Substantial genetic divergence and lack of recent gene flow support cryptic speciation in a colour polymorphic bumble bee (*Bombus bifarius*) species complex. Systematic Entomology. 45:635–652. doi: 10.1111/syen.12419.

Grossen C, Guillaume F, Keller LF, Croll D. 2020. Purging of highly deleterious mutations through severe bottlenecks in Alpine ibex. Nat Commun. 11:1001. doi: 10.1038/s41467-020-14803-1.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic biology. 52:696–704.

Hallmann CA et al. 2017. More than 75 percent decline over 27 years in total flying insect biomass in protected areas. PLOS ONE. 12:e0185809. doi: 10.1371/journal.pone.0185809.

Heimpel GE, de Boer JG. 2007. Sex Determination in the Hymenoptera. Annu Rev Entomol. 53:209–230. doi: 10.1146/annurev.ento.53.103106.093441.

Hoban S et al. 2021. Global Commitments to Conserving and Monitoring Genetic Diversity Are Now Necessary and Feasible. Bioscience. 71:964–976. doi: 10.1093/biosci/biab054.

Hoban S et al. 2022. Global genetic diversity status and trends: towards a suite of Essential Biodiversity Variables (EBVs) for genetic composition. Biological Reviews. 97:1511–1538. doi: 10.1111/brv.12852.

Hoffmann AA et al. 2021. An endangered flightless grasshopper with strong genetic structure maintains population genetic variation despite extensive habitat loss. Ecology and Evolution. 11:5364–5380. doi: 10.1002/ece3.7428.

Hohenlohe PA, Funk WC, Rajora OP. 2021. Population genomics for wildlife conservation and management. Molecular Ecology. 30:62–82. doi: 10.1111/mec.15720.

Hollenbeck CM, Portnoy DS, Gold JR. 2016. A method for detecting recent changes in contemporary effective population size from linkage disequilibrium at linked and unlinked loci. Heredity. 117:207–216. doi: 10.1038/hdy.2016.30.

IPBES. 2019. *Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services*. Zenodo doi: 10.5281/zenodo.6417333.

IPBES. 2016. *Summary for policymakers of the assessment report of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services on pollinators, pollination and food production.* Zenodo doi: 10.5281/zenodo.2616458.

Jackson JM et al. 2018. Distance, elevation and environment as drivers of diversity and divergence in bumble bees across latitude and altitude. Molecular Ecology. 27:2926–2942. doi: 10.1111/mec.14735.

Kerr JT et al. 2015. CLIMATE CHANGE. Climate change impacts on bumblebees converge across continents. Science. 349:177–180. doi: 10.1126/science.aaa7031.

van Klink R et al. 2020. Meta-analysis reveals declines in terrestrial but increases in freshwater insect abundances. Science. 368:417–420. doi: 10.1126/science.aax9931.

Koch JB, Looney C, Sheppard WS, Strange JP. 2017. Patterns of population genetic structure and diversity across bumble bee communities in the Pacific Northwest. Conserv Genet. 18:507–520. doi: 10.1007/s10592-017-0944-8.

Laikre L et al. 2020. Post-2020 goals overlook genetic diversity. Science. 367:1083–1085. doi: 10.1126/science.abb2748.

Lecocq T, Brasero N, Martinet B, Valterovà I, Rasmont P. 2015. Highly polytypic taxon complex: interspecific and intraspecific integrative taxonomic assessment of the widespread pollinator Bombus pascuorum Scopoli 1763 (Hymenoptera: Apidae). Syst Entomol. 40:881–890. doi: 10.1111/syen.12137.

Lewin HA et al. 2022. The Earth BioGenome Project 2020: Starting the clock. Proceedings of the National Academy of Sciences. 119:e2115635118. doi: 10.1073/pnas.2115635118.

Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 27:2987–2993.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:1303.3997.

Liu H et al. 2017. Direct Determination of the Mutation Rate in the Bumblebee Reveals Evidence for Weak Recombination-Associated Mutation and an Approximate Rate Constancy in Insects. Mol Biol Evol. 34:119–130. doi: 10.1093/molbev/msw226.

Lozier JD, Strange JP, Stewart IJ, Cameron SA. 2011. Patterns of range-wide genetic variation in six North American bumble bee (Apidae: Bombus) species. Mol Ecol. 20:4870–4888. doi: 10.1111/j.1365-294X.2011.05314.x.

Lozier JD, Zayed A. 2017. Bee conservation in the age of genomics. Conserv Genet. 18:713–729. doi: 10.1007/s10592-016-0893-7.

Martinet B et al. 2019. Integrative taxonomy of an arctic bumblebee species complex highlights a new cryptic species (Apidae: Bombus). Zool J Linn Soc. 187:599–621. doi: 10.1093/zoolinnean/zlz041.

Nadachowska-Brzyska K, Konczal M, Babik W. 2022. Navigating the temporal continuum of effective population size. Methods in Ecology and Evolution. 13:22–41. doi: 10.1111/2041-210X.13740.

Ødegaard F, Staverløkk A, Gjershaug JO, Bengtson R, Mjelde A. 2015. *Humler i Norge*. Norsk institutt for naturforskning: Trondheim.

Posledovich D, Ekblom R, Laikre L. 2021a. *Mapping and monitoring genetic diversity in Sweden: A proposal for species, methods and costs*. Naturvårdsverket: Stockholm.

Posledovich D, Ekblom R, Laikre L. 2021b. *Mapping and monitoring genetic diversity in Sweden: Suggestions for pollinating species*. Naturvårdsverket: Stockholm.

Potts SG et al. 2010. Global pollinator declines: trends, impacts and drivers. Trends Ecol Evol. 25:345–53. doi: 10.1016/j.tree.2010.01.007.

Rasmont P et al. 2015. Climatic Risk and Distribution Atlas of European Bumblebees. BR. 10:1–236. doi: 10.3897/biorisk.10.4749.

Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution. 4:406–25.

Sánchez-Bayo F, Wyckhuys KAG. 2019. Worldwide decline of the entomofauna: A review of its drivers. Biological Conservation. 232:8–27. doi: 10.1016/j.biocon.2019.01.020.

Santiago E et al. 2020. Recent Demographic History Inferred by High-Resolution Analysis of Linkage Disequilibrium. Molecular Biology and Evolution. 37:3642–3653. doi: 10.1093/molbev/msaa169.

Schwartz MK, Luikart G, Waples RS. 2007. Genetic monitoring as a promising tool for conservation and management. Trends in Ecology & Evolution. 22:25–33. doi: 10.1016/j.tree.2006.08.009.

Seibold S et al. 2019. Arthropod decline in grasslands and forests is associated with landscape-level drivers. Nature. 574:671–674. doi: 10.1038/s41586-019-1684-3.

Shi L et al. 2020. Estimation of inbreeding and identification of regions under heavy selection based on runs of homozygosity in a Large White pig population. J Animal Sci Biotechnol. 11:46. doi: 10.1186/s40104-020-00447-0.

Soroye P, Newbold T, Kerr J. 2020. Climate change contributes to widespread declines among bumble bees across continents. Science. 367:685–688. doi: 10.1126/science.aax8591.

Terhorst J, Kamm JA, Song YS. 2017. Robust and scalable inference of population history from hundreds of unphased whole genomes. Nat Genet. 49:303–309. doi: 10.1038/ng.3748.

Thomas CD. 2011. Translocation of species, climate change, and the end of trying to recreate past ecological communities. Trends in Ecology & Evolution. 26:216–221. doi: 10.1016/j.tree.2011.02.006.

Wagner DL. 2020. Insect Declines in the Anthropocene. Annu Rev Entomol. 65:457–480. doi: 10.1146/annurev-ento-011019-025151.

Wallberg A et al. 2014. A worldwide survey of genome sequence variation provides insight into the evolutionary history of the honeybee *Apis mellifera*. Nat Genet. 46:1081–1088. doi: 10.1038/ng.3077.

Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. Theoretical Population Biology. 7:256–276.

Webber BL, Scott JK, Didham RK. 2011. Translocation or bust! A new acclimatization agenda for the 21st century? Trends in Ecology & Evolution. 26:495–496. doi: 10.1016/j.tree.2011.06.007.

Webster MT, Beaurepaire A, Neumann P, Stolle E. 2023. Population genomics for insect conservation. Annu Rev Anim Biosci. 11.

Whitley KM. 2018. Genetic structure and gene flow barriers among populations of an alpine bumble bee (Bombus balteatus) in the central Rocky Mountains. M.Sc., Appalachian State University.

Willi Y et al. 2022. Conservation genetics as a management tool: The five best-supported paradigms to assist the management of threatened species. Proceedings of the National Academy of Sciences. 119:e2105076119. doi: 10.1073/pnas.2105076119.

Williams P, Colla S, Xie Z. 2009. Bumblebee Vulnerability: Common Correlates of Winners and Losers across Three Continents. Conservation Biology. 23:931–940. doi: 10.1111/j.1523-1739.2009.01176.x.

Williams PH et al. 2015. Genes Suggest Ancestral Colour Polymorphisms Are Shared across Morphologically Cryptic Species in Arctic Bumblebees. PLOS ONE. 10:e0144544. doi: 10.1371/journal.pone.0144544.

Williams PH et al. 2019. The arctic and alpine bumblebees of the subgenus Alpinobombus revised from integrative assessment of species' gene coalescents and morphology (Hymenoptera, Apidae, Bombus). Zootaxa. 4625:1–68. doi: 10.11646/zootaxa.4625.1.1.

Williams PH, Cameron SA, Hines HM, Cederberg B, Rasmont P. 2008. A simplified subgeneric classification of the bumblebees (genus Bombus). Apidologie. 39:46–74. doi: 10.1051/apido:2007052.

Williams PH, Osborne JL. 2009. Bumblebee vulnerability and conservation world-wide. Apidologie. 40:367–387. doi: 10.1051/apido/2009025.

Zayed A, Packer L. 2005. Complementary sex determination substantially increases extinction proneness of haplodiploid populations. Proceedings of the National Academy of Sciences. 102:10742–10746. doi: 10.1073/pnas.0502271102.

Supplementary Figures and Tables

Table S1. Summary statistics of bumblebee samples from Sweden that were collected and sequenced as part of the study.

Figure S1. Length of ROHs (Kb) across genome for each diploid (female) sample of *B. lapponicus*.

Figure S2. Length of ROHs (Kb) across genome for each diploid (female) sample of *B. monticola*.

Figure S3. Length of ROHs (Kb) across genome for each diploid (female) sample of *B. alpinus*, *B. balteatus*, *B. hyperboreus*, and *B. pyrrhopygus*.

Figure S4. Length of ROHs (Kb) across genome for each diploid (female) sample of *B. pascuorum*.

Appendix I – Photographs of study species

Appendix II – Proposals for additional studies

Table 1 – Species of bumblebees analysed in this study

***Bombus* sp.**	**Subgenus**	**English**	**Swedish**	**IUCN red list (3.1)**	**Climate risk[1]**	**Distribution in Europe**
alpinus	Alpinobombus	Alpine bumblebee	Alphumla	Vulnerable	HHR	European highlands and Arctic
pyrrhopygus	Alpinobombus	Polar bumblebee	Polarhumla	Least concern	HHHR	Arctic
balteatus	Alpinobombus	Golden-belted bumblebee	Fjällhumla	Least concern	HHR	Scandinavian highlands and Arctic
hyperboreus	Alpinobombus	High-Arctic bumblebee	Tundrahumla	Vulnerable	HHHR	Arctic
lapponicus	Pyrobombus	Lapland bumblebee	Lapphumla	Least concern	HHR	Scandinavian highlands and Arctic
monticola	Pyrobombus	Mountain bumblebee	Berghumla	Least concern	HHR	European highlands and Arctic
pascuorum	Thoracobombus	Common carder bee	Åkerhumla	Least concern	R	Common

[1]Climate risk categories: **HHHR**, extremely high climate change risk, loss of >95% of range by 2100; **HHR**, very high climate change risk, loss of >85% of range by 2100, **HR**, high climate change risk, loss of >70% of range by 2100, **R**, climate change risk, loss of >50% of range by 2100, **LR**, lower climate change risk, <50% loss of range by 2100 (Rasmont et al. 2015).

Table 2 – Bumblebee samples used in this study.

	Landskap									
Bombus sp.	Lappland	Jämtland	Härjedalen	Västerbotten	Uppland	Dalarna	Östergötland	Västergötland	Gotland	**Total**
alpinus	12	0	0	0	0	0	0	0	0	12
pyrrhopygus	5	1	0	0	0	0	0	0	0	6
balteatus	42	0	1	0	0	0	0	0	0	43
hyperboreus	9	1	0	0	0	0	0	0	0	10
lapponicus	109	16	3	0	0	0	0	0	0	128
monticola	63	1	15	0	0	0	0	0	0	79
pascuorum	23	0	4	8	2	2	4	4	8	55
Totals	**263**	**19**	**23**	**8**	**2**	**2**	**4**	**4**	**8**	**333**

Table 3 – Genetic variation in seven Swedish bumblebee species

***Bombus* sp.**	**Male**	**Female**	**Haploid genomes sampled**	**No. of SNPs**	**Watterson's theta per base (θ_W) (%)**	**Nucleotide diversity (π) (%)**	**Effective population size (N_E)**	**Observed heterozygosity (H_{OBS})**
alpinus		12	24	570,149	0.061	0.08	56,534	0.333
pyrrhopygus		6	12	1,390,927	0.184	0.236	170,545	0.371
balteatus	17	26	69	2,370,054	0.197	0.24	182,673	0.283
hyperboreus		10	20	660,899	0.074	0.099	68,978	0.306
lapponicus	62	66	194	3,095,046	0.21	0.227	194,583	0.280
monticola	28	48	124	1,085,476	0.08	0.091	73,814	0.298
pascuorum	14	41	96	3,800,079	0.241	0.273	222,742	0.218

Table 4 – Summary statistics of runs of homozygosity

		No. of ROH				
***Bombus* sp.**	**Female No.**	**0.01~0.1 Mb**	**0.1~1 Mb**	**>1 Mb**	**Mean total length per individual (Mb)**	**Mean genome coverage by total ROH (%)**
alpinus	12	301	48	0	1.72	0.69%
pyrrhopygus	6	290	87	0	5.17	2.07%
balteatus	26	4797	55	0	4.51	1.80%
hyperboreus	10	1844	804	1	25.47	10.18%
lapponicus	53	7511	771	0	5.45	2.16%
monticola	49	3799	971	2	6.99	2.77%
pascuorum	41	12329	1325	3	14.42	4.69%

Table 5 – Comparison of genetic variation in related species from Northern Europe and North America.

Species	Haploid genomes sampled	No. of SNPs	Watterson's theta per base (θ_w) (%)	Nucleotide diversity (π) (%)	Effective population size (N_E)
lapponicus cluster					
B. lapponicus (SWE)	194	2,514,341	0.171	0.161	158,075
B. lapponicus[1] (USA)	428	3,010,424	0.18	0.219	166,652
balteatus cluster					
B. balteatus (SWE)	69	2,250,496	0.187	0.207	173,458
B. kirbiellus[2] (USA)	598	5,824,609	0.334	0.271	309,427

[1]The species in North America currently named *Bombus lapponicus* (subspecies *sylvicola*), was previously considered to be a separate species named *Bombus sylvicola* (Martinet et al. 2019).
[2]The species *Bombus kirbiellus* in North America has recently been renamed and was previously known as *Bombus balteatus* (Williams et al. 2019).

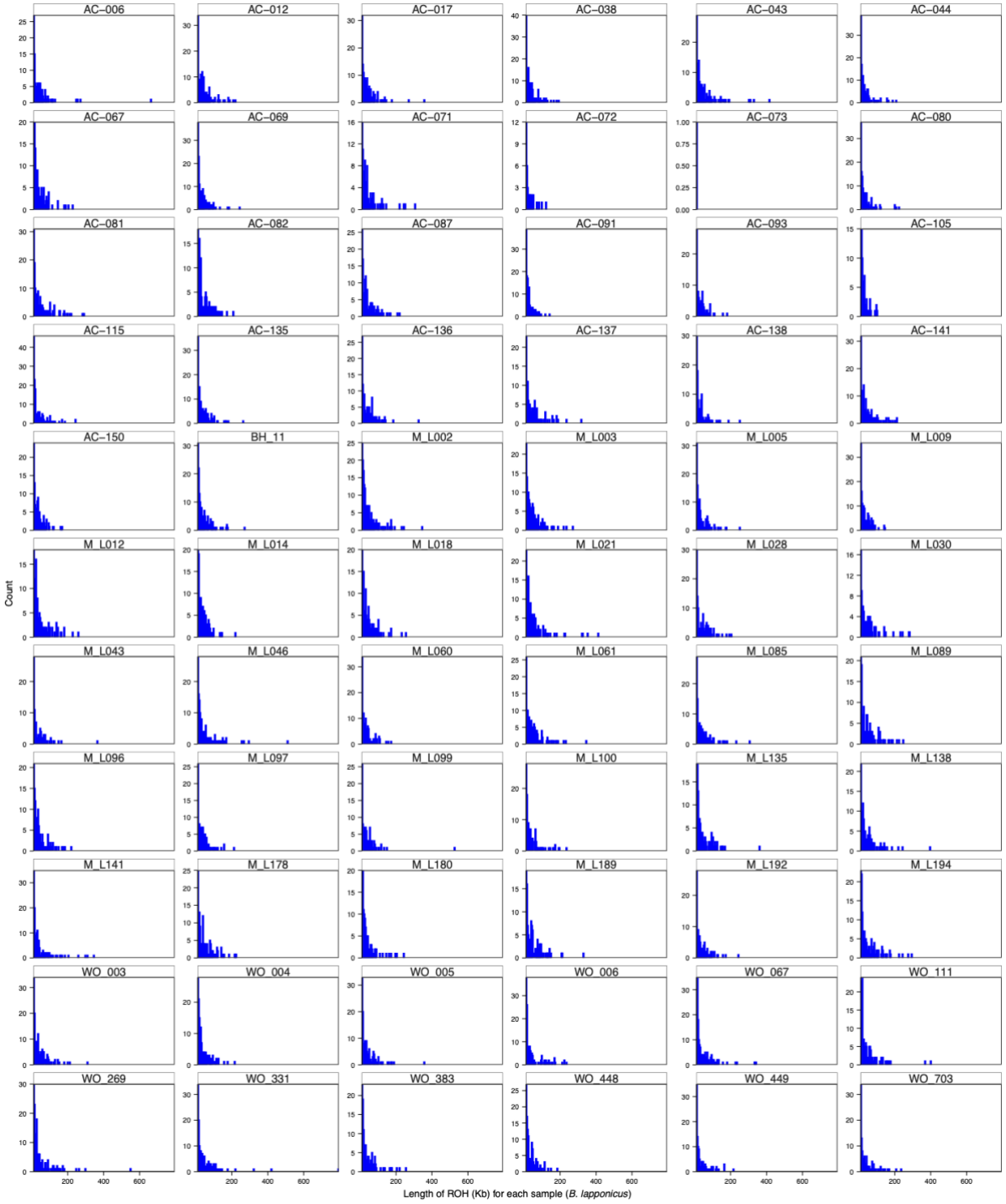


Figure S1. *Length of ROHs (Kb) across genome for each diploid (female) sample of B. lapponicus.*

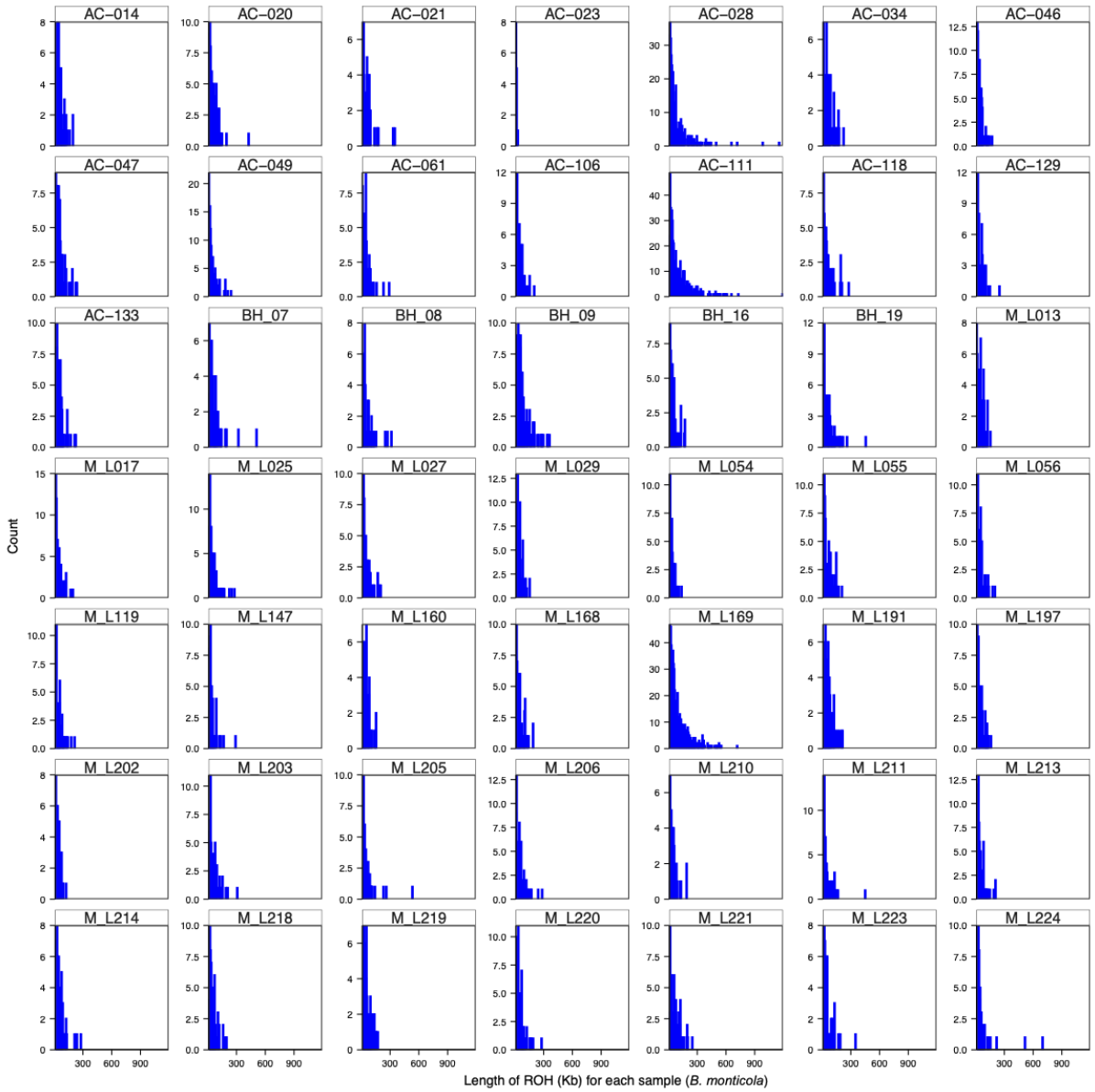


Figure S2. *Length of ROHs (Kb) across genome for each diploid (female) sample of B. monticola.*

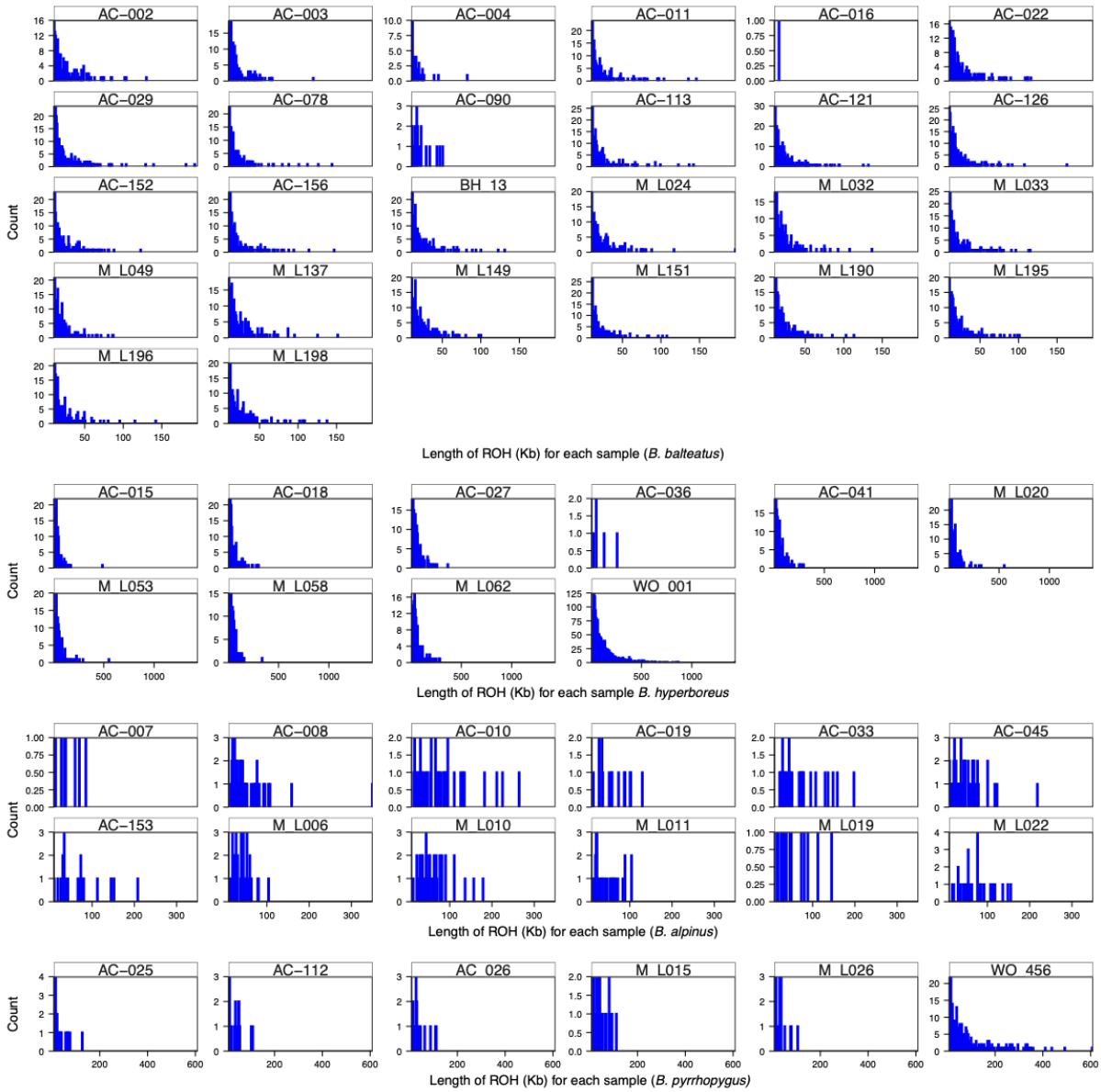


Figure S3. *Length of ROHs (Kb) across genome for each diploid (female) sample of B. alpinus, B. balteatus, B. hyperboreus, and B. pyrrhopygus.*

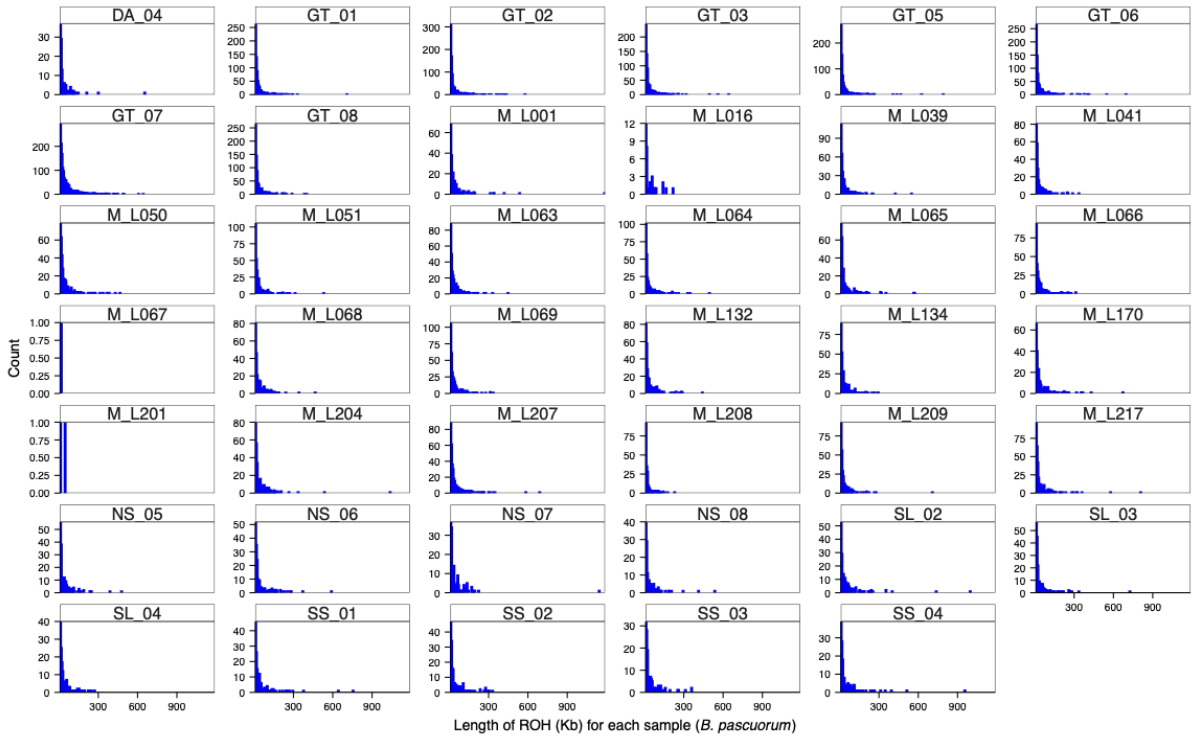


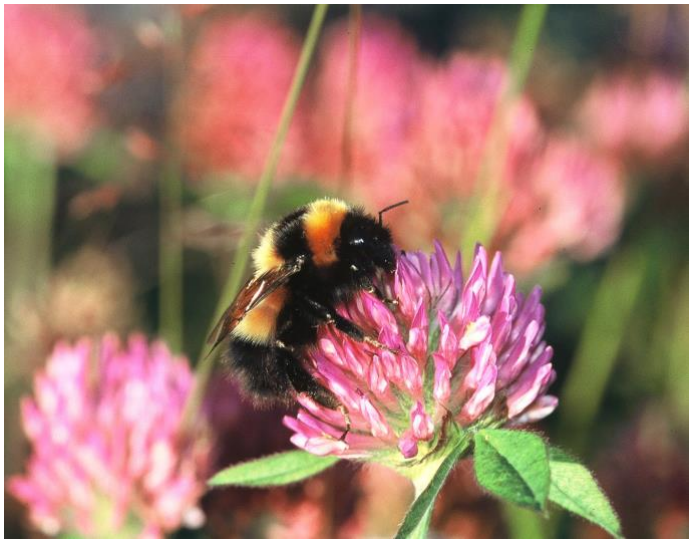
Figure S4. *Length of ROHs (Kb) across genome for each diploid (female) sample of B. pascuorum.*

APPENDIX I

PHOTOGRAPHS OF STUDY SPECIES

Bombus alpinus

Bombus balteatus



Bombus hyperboreus

Bombus lapponicus

Bombus monticola

Bombus pascuorum
(subsp. *pallidofacies*)

All photographs © Göran Holmström

APPENDIX II

SUGGESTIONS FOR ADDITIONAL STUDIES

To learn more about population trends and conservation status of the species studied, we can compare the patterns of genetic variation presented in this report to populations in the future or in the past. This can be done by: 1) Establishing a genetic monitoring programme and 2) Analysing genetic variation using historical museum specimens. Here we propose project plans and potential budgets for these two research programmes.

1) Genetic monitoring

A genetic monitoring programme could be conducted by repeating the present study at regular intervals in the future. A reasonable time frame would be every five years. For each monitoring study, the aim would be to collect up to 20 samples of each of the seven species from the three localities already assayed (Härjedalen, Jämtland, Abisko – Lapland). These study locations are suitable to track the effects of climate change on bumblebee populations because they are likely to be more stable in the northern localities (Abisko) but more vulnerable in the southern localities. The rarer species that are more specialised to arctic climates currently have fairly stable populations at higher elevations in Abisko but are extremely rare in the southern localities. We advocate non-lethal sampling of the rarer species to prevent damaging their populations. Bioinformatic and population genetic analysis will follow the methodology used in the current report.

Below we provide a potential budget for each five-year implementation of genetic monitoring:

	Total Budget	SEK
1	Salaries including social costs	
	Postdoc 10 months	550,250
	Lab assistant 6 months	330,150
2	Total running costs	585,600
3	Localities costs	73,000
4	**Subtotal**	1,539,000
5	Indirect costs	353,970
6	**Total project cost**	**1,892,970**
	Specification of running costs	
1	Library prep of population samples	257,600
2	Illumina DNA sequencing	168,000
3	Molecular genetics consumables	40,000
4	Computing	20,000
5	Field work	100,000
6	**Total running costs**	585,600

Budget items:
1) Salaries including social costs
A postdoc for 10 months, whose responsibilities include a) sample collection b) project management c) sequence data analysis d) preparing report.
A lab assistant for 6 months, whose responsibilities will include a) sample organisation b) DNA extractions c) coordination with sequencing platform: National Genomics Infrastructure (NGI), SciLifeLab.

2) Running costs
2a) Library prep of samples
The budget includes sequencing library preparation for 368 samples at a cost of 700 SEK per sample. This uses the Illumina Nextera Flex library prep with user QC to reduce cost. This service is provided by NGI.
2b) Illumina DNA sequencing
Sequencing of samples on 4 lanes of Illumina NovaSeq with an S4 flow cell. Performed at NGI.
2c) Molecular genetics consumables
Cost for DNA extraction and purification kits and other lab consumables (tubes, pipette tips etc.).
2d) Computing
Computing equipment for research assistant and postdoc, which are needed for project management and data analysis.
2e) Field work
Trips to the Swedish mountains for project members, field assistants and for other collaborators to collect bumblebee samples.

3) Localities costs
Localities costs associated with employing a postdoc and lab assistant in my department.

4) Indirect costs
23% indirect costs.

2) Analysis of genetic variation in museum specimens.

We propose analysing genetic variation in museum specimens from all seven species studied here collected at several timepoints, in order to determine whether the contemporary samples demonstrate a reduction in genetic variation. An inventory of specimens of two common mountain species from major collections in Scandinavia is presented in Figure A1 below. We aim to collect samples from at least 3 different timepoints: >100 years old, 50-100 years old and 20-50 years old. We will endeavour to use samples collected as close as possible to the contemporary sampling locations.

DNA extraction from historical samples will be done in collaboration with the Ancient DNA Facility at SciLifeLab, using non-destructive methods. A single leg will be used from each sample to minimise damage to the museum specimen. A bioinformatic pipeline GenErode is available at SciLifeLab specifically designed for analysis of genetic variation in sequence from historical specimens and comparison of samples of different ages. In total we aim to sequence at least 10 samples from each locality and timepoint. As the DNA extraction and

bioinformatic analysis is more challenging than for modern samples, we have budgeted for a 2-year postdoc contract.

Date/Location	*B. balteatus*	*B. lapponicus*
University Museum of Bergen		
pre 1900	7	15
1900-1950	197	272
1950-1980	343	1180
All samples	559	1503
Natural History Museum, Oslo		
pre 1900	5	39
1900-1950	7	56
1950-1980	1	27
All samples	65	261
The Arctic University Museum, Tromsø		
pre 1900	64	107
1900-1950	72	236
1950-1980	0	0
All samples	149	448
Museum of Natural History, Stockholm		
pre 1900	3	5
1900-1950	45	22
1950-1980	88	28
All samples	170	108

Figure A1 – Inventory of historical mountain bumblebee samples in Scandinavia

Below we provide a potential budget for a study of genetic variation in museum samples

	Total Budget	**SEK**
1	Salaries including social costs	
	Postdoc 24 months	1,320,600
	Lab assistant 6 months	330,150
2	Total running costs	530,600
3	Localities costs	146,000
4	**Subtotal**	2,326,750
5	Indirect costs	535,153
6	**Total project cost**	**2,861,903**
	Specification of running costs	
1	Library prep of population samples	252,00
2	Illumina DNA sequencing	168,000
3	Molecular genetics consumables	40,000
4	Computing	20,000
5	Travel	50,000
6	**Total running costs**	530,000

Budget items:
1) Salaries including social costs
A postdoc for 24 months, whose responsibilities include a) sample collection from museum specimens b) project management c) sequence data analysis d) preparing report.
A lab assistant for 6 months, whose responsibilities will include a) sample organisation b) DNA extractions c) coordination with sequencing platform: National Genomics Infrastructure (NGI), SciLifeLab.

2) Running costs
2a) Library prep of samples
The budget includes sequencing library preparation for 210 samples at a cost of 1200 SEK per sample. This uses the SMARTer ThruPLEX DNA-seq library prep method suitable for analysis of historical samples.
2b) Illumina DNA sequencing
Sequencing of samples on 4 lanes of Illumina NovaSeq with an S4 flow cell. Performed at NGI.
2c) Molecular genetics consumables
Cost for DNA extraction and purification kits and other lab consumables (tubes, pipette tips etc.).
2d) Computing
Computing equipment for research assistant and postdoc, which are needed for project management and data analysis.
2e) Travel
The postdoc will need to travel to museums in order to collect samples for sequencing

3) Localities costs

Localities costs associated with employing a postdoc and lab assistant in my department.

4) Indirect costs
23% indirect costs.