Johanna Honka, Serena Baini, Jeremy B. Searle, Laura Kvist and Jouni Aspi

Genetic assessment reveals inbreeding, possible hybridization and low levels of genetic structure in a declining goose population

J. Honka, Ecology and Genetics Research Unit, University of Oulu, Oulu 90014, Finland

ORCID: https://orcid.org/0000-0003-2536-5869

Corresponding author

email. [johanna.honka@oulu.fi](mailto:johanna.honka@oulu.fi)

S. Baini, Department of Biology, University of Rome “Tor Vergata", Rome 00133, Italy

ORCID: https://orcid.org/0000-0002-9938-5937

J.B. Searle, Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853, USA

ORCID: https://orcid.org/0000-0001-7710-5204

L. Kvist, Ecology and Genetics Research Unit, University of Oulu, Oulu 90014, Finland

ORCID: https://orcid.org/0000-0002-2108-017

J. Aspi, Ecology and Genetics Research Unit, University of Oulu, Oulu 90014, Finland

ORCID: http://orcid.org/0000-0002-2451-3201

# Abstract

The population numbers of taiga bean goose (*Anser fabalis fabalis*) have halved during recent decades. Since this subspecies is hunted throughout most of its range, the decline is of management concern. Knowledge of the genetic population structure and diversity is important for guiding management and conservation efforts. Genetically unique subpopulations might be hunted to extinction if not managed separately, and any inbreeding depression or lack of genetic diversity may affect the ability to adapt to changing environments and increase the extinction risk. We used microsatellite and mitochondrial DNA markers to study the genetic population structure and diversity among taiga bean geese breeding within the Central flyway management unit using non-invasively collected feathers. We found some genetic structuring with the maternally inherited mitochondrial DNA between four geographic regions (*ɸ*ST = 0.11-0.20) but none with the nuclear microsatellite markers (all pairwise *F*ST-values 0.002- 0.005). These results could be explained by female natal philopatry and male-biased dispersal, which completely homogenizes the nuclear genome. Therefore, the population could be managed as a single unit. Genetic diversity was still at a moderate level (average *H*E = 0.69) and there were no signs of past population size reductions, although significantly positive inbreeding coefficients in all sampling sites (*F*IS = 0.05-0.10) and high relatedness values (*r* = 0.60-0.86) between some individuals could indicate inbreeding. In addition, there was evidence of either incomplete lineage sorting or introgression from the pink-footed goose (*A. brachyrhynchus*). The current population is not under threat by genetic impoverishment but monitoring in the future is desirable.

# Keywords: *Anser fabalis*, microsatellites, mitochondrial DNA, control region, non-invasive sampling, hybridization, citizen science

# Introduction

Knowledge of population genetic structure is essential for guiding management and conservation of species. The presence of highly divergent subpopulations with low levels of gene flow could warrant a status of separate management units (MUs; Palsbøll, Bérubé and Allendorf, 2007). Loss of genetic diversity and inbreeding depression, processes that especially affect small and fragmented populations, may lead to a loss of evolutionary potential and contribute to a higher extinction risk (Frankham, 2005). Harvesting can also lead to adverse genetic changes such as alteration of population subdivision (extirpation of local populations), loss of genetic variation and selective genetic changes (reduction in certain phenotypes targeted by hunting) that may further compromise population viability (Allendorf *et al*., 2008). Thus, knowledge of genetic structure and diversity is especially important for managing harvested species and subspecies to ensure sustainable hunting. This has been our motivation to study the taiga bean goose (*Anser fabalis fabalis*), which is hunted over most of its range but has suffered a marked decline during recent decades (Fox *et al*., 2010). However, very little is currently known about the genetic population structure and diversity of the taiga bean geese in their breeding area.

Most Holarctic goose populations are on the increase (Fox and Leafloor, 2018) to the extent that this is causing conflicts with humans, especially in agriculture (Fox and Madsen, 2017). However, a few goose populations are declining, one of them being the taiga bean goose (Fox *et al*., 2010; Fox and Leafloor, 2018) breeding in Northern Europe and Asia (Fig. 1). The population size of the taiga bean goose has nearly halved from the 90,000-100,000 individuals in the 1990s (Nilsson, van den Bergh and Madsen, 1999) to the current estimate of 52,000 individuals for the total wintering population size (Fox and Leafloor, 2018). The taiga bean goose is a quarry species, hunted throughout its range, except in the UK, Norway, the Netherlands and Belgium. Although the exact cause of decline is uncertain, overharvesting is one plausible explanation. Since the taiga bean goose population is still open to hunting, sustainable management of the population is of crucial importance. Within this framework, an International Single Species Action Plan (ISSAP) was developed by AEWA (The African-Eurasian Migratory Waterbird Agreement) to conserve the taiga bean goose (Marjakangas *et al*., 2015). Hunting of the North-East/North-West European population of the taiga bean goose can still be continued within the limits of agreed sustainable use within the ISSAP framework (Marjakangas *et al*., 2015).

The ISSAP for the taiga bean goose recognizes four subpopulations or flyway management units: Western, Central, Eastern 1 and Eastern 2 subpopulations (Marjakangas *et al*., 2015; Fig. 1). These flyway units are distinguished by isotopic composition of feathers collected from bean geese from Western, Central and Eastern 1 flyway units (Fox *et al*., 2017) but the units have not been confirmed genetically and it is not known if these units should be further subdivided into smaller management units. For effective management of the taiga bean goose, population genetic structure should be assessed in order to preserve genetic diversity. Previously, the population genetic structure of the taiga bean goose has only been studied within a limited area in Central Scandinavia belonging to the Western flyway management unit (de Jong *et al*., 2019).

The taiga bean geese are elusive, especially during the breeding period, and observing even neck-banded geese in breeding areas is rarely possible due to a long escape distance of geese (Pirkola and Kalinainen, 1984a). Thus, catching geese and sampling blood would be very difficult and stressful for the birds. Developments in non-invasive genetic sampling have allowed sampling of elusive and endangered species without the need to handle or even observe the animals (Taberlet, Luikart and Waits, 1999). For birds, nest material and moulted feathers provide a valuable source of DNA (Pearce, Fields and Scribner, 1997; Segelbacher, 2002). The taiga bean geese perform a moult from the middle of June to the middle of August and during this flightless period, they spend time in the wettest part of mires or in the vicinity of ponds, where they leave abundant cues of their presence such as tracks, signs of grazing, faeces and moulted feathers (Pirkola and Kalinainen, 1984b). In our study, we utilized this source of non-invasive feather samples collected by volunteers (citizen-science) to study the genetic population structure of the breeding taiga bean geese.

Our aim was 1) to assess large-scale genetic structure among breeding taiga bean geese in Finland (consisting of a large part of the Central Management Unit), 2) to estimate genetic diversity within this population and 3) to evaluate current effective population size and demographic fluctuations. For this, we used highly variable microsatellites first to identify individuals and parentage and then to study variation at the population level. In addition, we defined the genetic structure of maternal lineages by sequencing the most variable part of the mitochondrial control region. This region was also used to verify the subspecies (see Honka et al. 2017).

# Material and methods

Study species

The taxonomy of the bean goose and the related pink-footed goose (*A. brachyrhynchus*) has been controversial (Ruokonen and Aarvak, 2011). Currently, the bean goose is either split to two species (*A. fabalis* and *A. serrirostris*; Sangster and Oreel, 1996) or treated as one species (*A. fabalis*) with four subspecies (*A. f. fabalis*, *A. f. rossicus*, *A. f. middendorffii* and *A. f. serrirostris*; Mooij and Zöckler, 1999). Based on a recent study using genome-wide data, *A. f. fabalis* and *A. f. rossicus* should be classified as subspecies (Ottenburghs *et al*., 2020) and thus we follow here the classification of one species with four subspecies. From here on, we focus on the western subspecies and refer to *A. f. fabalis* as the taiga bean goose and *A. f. rossicus* as the tundra bean goose.

The taiga bean goose breeding distribution covers the forested taiga area from Scandinavia to Western Siberia (Scott and Rose, 1996; Fig. 1) and in Finland the ‘aapa’ mire zone (see Laitinen *et al*., 2007) with core breeding area extending from Lapland to Northern Ostrobothnia (Pirkola and Kalinainen, 1984b; Fig. 2). In Finland, the taiga bean goose is listed as vulnerable (the Red List of Finnish species; Lehikoinen *et al*., 2019) with the breeding population size estimated to be 1700-2500 pairs (i.e. 3400-5000 breeding individuals; Valkama, Vepsäläinen and Lehikoinen, 2011). Subadults and failed breeders perform a moult migration to Novaya Zemlya (Nilsson *et al*., 2010; Paasivaara, 2013; Laaksonen, Piironen and Mäntyniemi, 2019) and thus are not counted. In addition to the taiga bean goose, a very small number of tundra bean geese may breed in Finland in the most northernmost Lapland adjacent to their breeding range in Norwegian Finnmark (Aarvak and Øien, 2009; Fig. 1). This subspecies is listed as near threatened in Finland (The Red List of Finnish species; Lehikoinen *et al*., 2019).

## Sampling and DNA extraction

Bean goose feathers were collected from nests or brood-rearing/moulting sites in Finland or close to Finnish border (Sør-Varanger municipality, Norway) during years 2006-2014 (*n* = 14) and 2016-2018 (*n* = 2127) or from taiga bean geese handled for ringing (with appropriate permits for ringing) in years 2017-2018 (*n* = 20) (Fig. 2). These samples are here on referred to as the ‘Finnish population’. We used mostly a citizen science approach in which the public is involved in scientific research to collect feather samples. We also created outgroups for our Finnish population by sampling 1. Swedish breeding taiga bean geese (2014; *n* = 7), 2. migrating Russian taiga bean geese hunted from Finland (Honka *et al*., 2017) or Estonia (2010-2012, 2017; *n* = 7), 3. Norwegian breeding tundra bean geese (2002, 2006; *n* = 7) and 4. Iceland breeding pink-footed goose (Ruokonen, Aarvak and Madsen, 2005, the Natural History Museum of Reykjavik; *n* = 7) (Fig. 2). All feathers were stored in paper envelopes at room temperature prior to DNA extraction. Sampling of feathers was performed in a laboratory in which no PCR-products are handled.

DNA from a calamus and from a blood clot (when visible; Horváth *et al*., 2005) was extracted using QuickExtract DNA Extraction Solution (Lucigen) according to the manufacturer’s protocol, except for a 15 min incubation at 65 ⁰C instead of 6 min. Feathers showing very poor preservation and clear environmental exposure (brittle and discoloured calamus) were omitted from the DNA extraction. Also, feathers belonging to other bird species based on morphology (e.g. feather shape typical of common crane *Grus grus*) were omitted from DNA extraction. The extracted DNA was stored in -20 ⁰C.

Microsatellite genotyping

We genotyped the samples for 28 microsatellite loci (Noreikiene *et al*., 2012; Kleven, Kroglund and Østnes, 2016) in four multiplex reactions (A-D) (Supplementary table 1) using Multiplex PCR Kit (Qiagen) in 6 µl reaction volumes according to manufacturer’s protocol. The forward primers were fluorescently labelled with either 6FAM, NED, PET or VIC dyes (Supplementary table 1). One microliter of template DNA was used in each reaction. The thermal profile consisted of 95 °C for 15 min, followed by 40 cycles of 94°C for 30 s, 57 °C (panel A and B) or 60 °C (panel C and D) for 90 s and 72 °C for 30 s with a final extension of 60 °C for 30 min. We performed fragment analysis with an ABI 3730 and scored the alleles using GeneMapper 5 (Applied Biosystems).

We used a stepwise approach to amplify the microsatellite panels. First, we amplified the samples with panel B (five loci) to screen for quality and to reduce the number of duplicate individuals, as follows. We performed individual identification by visual inspection of the genotypes and by using the ‘Regroup genotypes’ option in the program Gimlet v.1.3.3. (Valière, 2002). If one sampling site included several identical genotypes, these were assumed to be replicates of the same individual and all except one was excluded. Next, we amplified panel A (seven loci) from the apparently good-quality samples selected in the first step. We performed the individual identification similarly as after the panel B, as genotyping errors could have been interpreted as different individuals in the first step and again excluded all but one identical genotype (taking into account the genotyping errors). Panel C and D were amplified with the samples selected in the second step and the individual identification was performed once more in order to test for any remaining errors in the data and again kept only one of the identical genotypes. Due to the inherent presence of allele drop-outs and false alleles in microsatellite data (Taberlet, Luikart and Waits, 1999), we regarded multilocus genotypes with up to three mismatches to belong to the same individual. Mismatches in 4-5 loci between multilocus genotypes were carefully checked and if the pattern was consistent with allele drop-out due to poor quality template DNA, the samples were excluded. Consensus genotypes for each individual were created manually and individuals with more than 25% missing data were excluded, resulting with 491 individuals. Three individuals were excluded from the analysis of the taiga bean geese because these samples had a mtDNA sequence of a different subspecies or were possibly still in spring migration (see below). Thus, 488 individuals were included in analyses unless otherwise stated. We calculated the unbiased probability of identity (*P*ID) and probability of identity of siblings (*P*ID SIB) (Taberlet and Luikart, 1999) from the identified individuals using the Gimlet program.

Molecular sexing

Sexing of the individuals was based on fluorescently labelled (6FAM) forward primer ASW12-D3 (Guzzetti *et al*., 2008) and reverse primer HZW278 (Gravley *et al*., 2017) targeting the HINTZ/W gametologs. We performed the PCR reactions and fragment analyses similarly as for the microsatellites except the annealing temperature was 50 °C and the final extension was performed at 72 °C. These markers produce a 287 bp (base pair) fragment in females and a 297 bp fragment in males (including the primer sequences). The sexing results were verified using seven individuals from which the sex was inspected by cloacal examination during goose ringing. In addition, we used a *χ*2-test to test if the number of females and males differed from the expectation of equal numbers.

Microsatellite analyses

The data were divided into four geographical regions (Fig. 2) and all analyses were performed on these groups. We used the program Micro-Checker v. 2.2.3 (van Oosterhout *et al*., 2004) to assess the accuracy of the microsatellite typing and the program FreeNA (Chapuis and Estoup, 2007) to calculate the null-allele rate. A custom program (Microsat\_errcalc; Honka and Merikanto, 2020) was used to calculate allele drop-out (ADO) and false allele (FA) rates from different feather samples belonging to the same individual. Number of alleles (*A*), number of private alleles (*PA*), observed heterozygosity (*H*O), expected heterozygosity (*H*E) and inbreeding coefficient (*F*IS) were calculated using the program GenAlEx 6.503 (Peakall and Smouse, 2006, 2012) and allelic richness (*A*R) was calculated using the program FSTAT 2.9.4 (Goudet, 1995). Departure from the Hardy-Weinberg equilibrium (HWE) and the degree of linkage disequilibrium (LD) were determined using the program Genepop 4.7.0 (Rousset, 2008) with a sequential Bonferroni correction (Rice 1989) applied to these tests.

We inferred parentage and sibships using the program Colony 2.0.6.5 (Jones and Wang, 2010) by setting ‘monogamy’ for both females and males (as geese are known to form stable pair bonds) and performed three iterations of the long run with inbreeding model with a full-likelihood method. All taiga bean goose individuals (*n* = 488) were placed in the candidate offspring category, females (*n* = 237) and geese of unknown sex (*n* = 62) (based on molecular sexing) were placed in the candidate mother category (total *n* = 299) and males (*n* = 189) and geese with unknown sex (*n =* 62) (based on molecular sexing) were placed in the candidate father category (total *n* = 251). We excluded each individual from being the mother or father of itself. We used the calculated ADO and FA rates (Supplementary table 1) as the marker error rates. We calculated relatedness (*r*) between goose dyads using the program ML-relate (Kalinowski, Wagner and Taper, 2006) taking account the null alleles. As it has been shown that sampling of close relatives biases genetic structure analyses, especially in the program Structure (Anderson and Dunham, 2008; Rodrígues-Ramilo and Wang, 2012), we created a dataset from which the inferred parents, all but one sibling and individuals with *r >* 0.55 were removed resulting in a subset of data with 376 non-kin individuals. All analyses were performed on this non-kin dataset except for *F*ST (fixation index), *ɸ*ST and *N*e (effective population size) that were calculated with the full dataset because the precision of these genetic estimates may suffer with the purging of all siblings (Waples and Anderson, 2017).

We inferred population structure using the program Structure 2.3.4. (Pritchard, Stephens and Donnelly, 2000; Falush, Stephens and Pritchard, 2003; Hubisz *et al.*, 2009) with the LOCPRIOR option, with individuals within 16 km treated as coming from one location. We used an admixture ancestry model and correlated allele frequencies with a burn-in of 100,000 and a run length of 1,000,000. Five iterations were performed with possible number of clusters (*K*) set from 1 to 10. The *ad hoc* approach of Evanno, Regnaut and Goudet (2005) was used to infer the most likely number of *K* clusters in the data as implemented in the program Structure Harvester (Earl and vonHoldt, 2012). We used the program Clumpak 1.1. (Kopelman *et al*., 2015) to visualize the Structure results and to create consensus among the different iterations. In addition, we used DAPC (Discriminant Analysis of Principal Components; Jombart, Devillard and Balloux, 2010) implemented in the R package (R Core Team, 2018) ‘adegenet’ 2.1.1 (Jombart, 2008; Jombart and Ahmed, 2011) to assess population structure. We randomly chose an equal number of samples in each geographical region (*n* = 75 in each population, total *n* = 300) as the DAPC analysis is not sensitive to underlying family structure but could be biased due to unequal samples sizes. DAPC was also used to compare the outgroup samples and the Finnish population by performing DAPC first on the outgroups and then importing the ‘unknown’ Finnish population into this same framework using the ‘predict’ function. Pairwise *F*ST-values were calculated using the ENA-correction implemented in the program FreeNA due to the possible presence of null alleles in the data. To test for isolation-by-distance, we performed a Mantel test implemented in the R package (R core team, 2018) ‘ade4’ 1.7.13 (Chessel, Dufour and Thioulouse, 2004; Dray, Dufour and Chessel, 2007; Dray and Dufour, 2007; Bougeard and Dray, 2018) using Euclidean distances. We also performed a spatial autocorrelation analysis (Smouse and Peakall, 1999) using the program GenAlEx to test for fine-scale geographic patterns. The significance of the analysis was tested using a heterogeneity test (Smouse, Peakall and Gonzalez, 2008). In addition, the autocorrelation analysis was performed on females and males separately to estimate sex-biases in relatedness (Banks and Peakall, 2012).

The effective population size (*N*e) was estimated with the linkage disequilibrium model (Hill, 1981; Waples, 2006; Waples and Do, 2010), assuming monogamy and using 0.05 as the critical value for allele frequency as implemented in the program NeEstimator v2.1 (Do *et al*., 2014). The *N*e estimate was compared with the estimate based on sibship assignment (Wang, 2009) by the program Colony. The presence of historical bottlenecks was evaluated with the program Migraine 0.5.4 (Leblois *et al.*, 2014) using a single population with a past population size change (OnePopVarSize) option and GSM (Generalized stepwise mutation; Pritchard *et al*., 1999) model. Migraine was run with 500 points and 2000 runs per point with seven iterations. Ancestral population size (*N*anc), current population size (*N*) and time of demographic change in generations (*T*) were resolved from scaled parameters produced by Migraine (θanc = 2*N*anc*µ*,θ = 2*Nµ* and *D* = *T*/2*N*) by assuming a microsatellite mutation rate of *µ* = 5 x 10-4 per locus per generation (Dib *et al*., 1996; Ellegren, 2000; Sun *et al*., 2012) and a generation time of 5-7.5 years (Dillingham, 2010). The demographic change was evaluated using a parameter called *N*ratio (*N/N*anc), with *N*ratio > 1 indicating a population growth and *N*ratio < 1 indicating a population bottleneck. The statistical significance was evaluated using 95% confidence intervals. If the 95% confidence intervals do not span 1, the results are statistically significant (Leblois *et al.*, 2014). For other tests of possible bottlenecks, see Supplementary text 1.

We studied hybridization between the taiga bean goose and either the pink-footed goose (taiga bean goose x pink-footed goose) or the tundra bean goose (taiga bean goose x tundra bean goose) by using a simulation study. First, we simulated 100 pure parental individuals (selected by *q >* 0.99 by a preliminary NewHybrids run) and 100 individuals in different hybrid classes (F1, F2 and backcrosses) using the program HybridLab1.0 (Nielsen, Bach and Kotlicki, 2006). We used these simulated individuals as an input for the program NewHybrids 1.1 (Anderson and Thompson, 2002) with Jeffreys-like priors for both mixing proportions and allele frequencies, setting burn-in to 20,000 sweeps and the chain length to 100,000 MCMC sweeps with the z option. We compared these results to the Structure program run with the admixture ancestry model and correlated allele frequencies with burn-in set to 10,000 and run length to 100,000. Five iterations were performed with the number of *K* set to 2. As the Structure program performed better than the NewHybrids program (Supplementary Figures 1-4), we secondly run Structure analysis with the simulated parentals and the Finnish population including also the two tundra bean geese and the plausible pink-footed goose individual (*n* = 491). In this second analysis, the aim was to search for admixture with other subspecies or species in the Finnish population.

Mitochondrial DNA analyses

The hypervariable portion of the mitochondrial control region domain I (210 bp) was amplified using primers AdCR1F and AdCR2R (Honka *et al*., 2018) from the individuals identified with microsatellite genotyping (*n =* 491). The primers were designed to contain mismatches to Numts (nuclear sequences of mitochondrial origin; Lopez *et al*., 1994), which are problematic in genetic studies if not accounted for (Sorenson and Quinn, 1998; Bensasson *et al*., 2001). The PCR reactions were performed in 10 µl volumes using 1 x Phusion HF-buffer (Thermo Fisher Scientific), 0.2 mM of each dNTPs, 0.5 µM of forward and reverse primers, 0.02 U/µl of Phusion DNA Polymerase (Thermo Fisher Scientific) and 1 µl of template DNA. The thermocycling conditions were 98 °C for 4 min, followed by 40 cycles of 98 °C for 30 s, 57 °C for 30 s and 72 °C 40 s with a final extension of 72 °C for 7 min. Double-stranded sequencing with the PCR primers was performed using BigDye Terminator v.3.1 (Applied Biosystems) and the reactions were run on an ABI3730.

The sequences were manually edited using the program CodonCode Aligner v.4.0.4. (CodonCode Corporation) and aligned with GenBank sequences of bean geese (accession numbers: EU186807-EU186812, EU186827, AF159951 and MH491806-MH491819; Ruokonen, Kvist and Lumme, 2000; Ruokonen, Litvin and Aarvak, 2008; Honka *et al*., 2017), pink-footed geese (AF159952-AF159953; Ruokonen, Kvist and Lumme, 2000) and greylag goose (*A. anser*) as an outgroup (AF159961; Ruokonen, Kvist and Lumme, 2000) using the program BioEdit 7.2.5 (Hall, 1999). A median joining-network (Bandelt, Forster and Röhl, 1999) was constructed using the program PopART (Leigh and Bryant, 2015) and number of haplotypes (*H*), haplotype (*h*) and nucleotide (*π*) diversities, Tajima’s D (*D*) and Fu’s Fs (*Fs*) were calculated using program DnaSP v. 6.12 (Rozas *et al*., 2017). The presence of genetic structure on individual and region levels was tested by an analysis of variance framework using analysis of molecular variance (AMOVA), which is based basically on hierarchical variance of gene frequencies. We calculated pairwise *ɸ*ST-values between regions and performed AMOVA analysis using the program Arlequin 3.5.2.2. (Excoffier and Lischer, 2010) with substitution model of Jukes-Cantor (Jukes and Cantor, 1969). This was the best substitution model based on AIC (Akaike Information Criteria) and BIC (Bayesian Information Criterion) values in the program MEGA X (Kumar *et al*., 2018) and is supported in the Arlequin program. A sequential Bonferroni correction (Rice, 1989) was applied to the *ɸ*ST–statistics.

# Results

## Microsatellite genotyping success

Microsatellite genotyping was successful for 1805 feather samples (at least one multiplex panel successful) with a success rate of 84%. After omitting possible redundancy, this amounted to a total of 491 individuals in our dataset. One individual that may have been on spring migration and had pink-footed goose mtDNA (individual found dead, see Fig.2) was omitted from the Finnish breeding population analyses but included in the mtDNA haplotype network, maps and hybridization analyses. Similarly, we also omitted two tundra bean goose individuals, leaving us with a final total of 488 taiga bean goose individuals. The error rate in the whole dataset (28 loci) was 0.019 per allele over all loci and 0.039 per locus over all loci. The mean ADO rate was 0.041, the mean FA rate was 0.009 and the mean null allele rate was 0.034 (Supplementary table 1). The cumulative unbiased probability of identity and the cumulative probability of identity of siblings were low (*P*ID = 5e-25, *P*ID SIB = 4e-10; Supplementary table 1) indicating that our markers can separate different individuals with high confidence.

No evidence of large allele dropout was present in the loci, but the program MicroChecker suggested that some loci show stuttering. However, the stuttering was not consistent over the geographical regions (see Fig. 2 for the regions). We removed the loci Abra14 and Abra29 due to low polymorphism, Abra9 due to high rate of missing data (43%) and Abra15 and Abra43 due to high frequency of null alleles (> 10%) (Supplementary table 1). Locus Afa18 was not in Hardy-Weinberg equilibrium in any of the four geographical regions and was removed. Therefore, a final total of 22 loci were kept for further analyses. Several other loci showed deviations from Hardy-Weinberg equilibrium and some were in linkage disequilibrium after Bonferroni correction but these were not consistent between regions, so these loci were kept in the analyses.

## Molecular sexing and relatedness

Molecular sexing using the HINTZ/W gametologs showed congruent results with the cloacal examination except for one goose out of seven. Based on the cloaca, this goose was assigned as a female while based on molecular sexing it was a male. This sample was replicated and the same result was obtained by molecular sexing. We identified 237 females and 189 males in the Finnish population which was significantly different from equality (*χ*2 = 5.40, *P* = 0.02). The sex could not be determined for 62 individuals.

Most of the geese dyads studied were unrelated (80,576; relatedness value 0.00-0.04), 38,049 dyads had relatedness values between 0.05-0.39 and 131 geese dyads had relatedness values between 0.40-0.59 indicating potentially parent-offspring or sibling relationships according to the ML-relate program. Relatedness values of approximately 0.50 were found for all known parent-offspring and sibling dyads of ringed geese (two sibling dyads and one parent offspring dyad), except for one gosling that could not have been offspring of the parents it was captured with (unrelated to the mother, *r =* 0.25 between the father and the gosling). We identified 23 candidate fathers and 28 candidate mothers with probabilities over 0.90.

We identified two male geese harbouring mtDNA of the pink-footed goose to have sired offspring with females having taiga bean goose mtDNA. Fifty-two pairs of full-siblings were identified with probabilities of 1.00, these also shared mtDNA haplotypes. Seventy-two geese dyads showed relatedness values between 0.60-0.86 suggesting potential inbreeding. Based on these results, we created a subset of the data that did not contain parent-offspring or sibling relationships or highly related individuals (*r* > 0.55), producing a final dataset of 376 non-kin individuals.

## Genetic diversity

The number of alleles varied between 171 and 194 with a mean of 184 over the geographical regions (Table 1). Allelic richness, which takes into account the sample size, was almost the same in all regions (*A*R = 7.4-7.6; Table 1). The Eastern Finland/Kainuu region had the largest number of private alleles (*PA* =9) followed by Lapland (*PA* = 7; Table 1). Observed and expected heterozygosities were similar over the regions (mean *H*O = 0.64 and mean *H*E = 0.69) with all regions showing higher expected than observed heterozygosities (Table 1). Inbreeding coefficients (*F*IS) ranged between 0.05 and 0.10 (Table 1). All regions showed deviation from the Hardy-Weinberg equilibrium when summed over loci (Fisher’s exact test; Table 1).

## Population structure

The Structure analyses indicated little difference between the regions (mapped in Fig. 2) or indeed between any individuals within the Finnish breeding taiga bean goose population (Fig. 3). The optimal *K* was 4 based on the *ad hoc* statistics Δ*K*, but the log likelihood values did not reach a clear plateau (Supplementary figure 5), instead all samples were assigned to all clusters with high admixture proportions indicating that the *K* would actually be 1.

The DAPC analyses showed that slight genetic structuring might be present between Western Finland, Eastern Finland/Kainuu and Lapland regions, as the samples mostly did not overlap (Fig. 4a). Samples from Northern Ostrobothnia/Southern Lapland region, on the other hand, overlapped with the other three regions (Fig. 4a). The DAPC analysis performed to the outgroups showed that all outgroups (Swedish taiga bean geese, Russian taiga bean geese, tundra bean geese and pink-footed geese) clustered to their own groups except for minor overlap between Swedish and Russian taiga bean geese (Fig. 4b).

When the Finnish samples were fitted to this framework, most of the samples clustered with the Russian taiga bean geese. However, some samples showed genetic affinity to Swedish taiga bean geese, tundra bean geese and, to a smaller extent, to the pink-footed geese (Fig. 4b).

Pairwise *F*ST-values were always very low, with mean values between regions of less than 0.005 (Table 2). The Western Finland region was the most differentiated from all of the other regions although none of the values were statistically significant.

We found no isolation-by-distance pattern using the Mantel test (*r =* 0.019, *P* = 0.18; Supplementary figure 7) or the spatial autocorrelation analysis (Fig. 5a). No sex-specific differences were evident either in the spatial autocorrelation analysis (Fig. 5b).

## Effective population size and bottlenecks

The effective population size (*N*e) of the full dataset including kin was estimated to be 1127.6 individuals (95% confidence intervals CI: 937.2-1392.9) using the linkage disequilibrium method. The *N*e estimate based on sibship using the full likelihood method was 1134 individuals (95% CI: 927-1327). No signs of recent population bottlenecks were detected, instead, the Migraine analysis indicated population growth as the *N*ratio was 7.8 (CI: 1.746-21.24) and the 95% confidence did not overlap with one. The Migraine analysis indicated a population growth starting from 21,766-32,648 years ago (depending on generation time 5-7.5 years; *T* = 4353 generations, 95% CI:1138-13,511 ) from 601 individuals (*N*anc, 95% CI: 255-2490) to the current census population size of 4701 individuals (*N*, 95% CI: 3790-5876) (Supplementary figure 6).

## Hybridization

The NewHybrids and the Structure analysis based on simulated data showed that identification of different hybrid categories (F1, F2, backcrosses) between the taiga bean and the pink-footed goose and between the taiga bean and the tundra bean goose is difficult using microsatellite markers due to the extent of variation within each hybrid category (Supplementary Figures 1-4). However, the pure simulated parentals were readily identifiable. We also run a Structure analysis using the simulated parental populations and the full dataset of the Finnish population including the three individuals omitted from other analyses (two Finnish tundra bean geese and the goose found dead (see Fig. 2); *n =* 491). This analysis showed that the simulated pure taiga bean geese and pink-footed geese could be separated and that some Finnish taiga bean geese showed admixture with the pink-footed goose (Fig. 6a). No pure pink-footed geese were present in our dataset according to this analysis. The analysis for the simulated taiga and tundra bean geese showed similar results. Some of the Finnish taiga bean geese showed admixture with the tundra bean goose (Fig. 6b). The two tundra bean geese sampled from Finland in this study were identified as pure tundra bean geese with > 0.99 probability (Fig. 6b, arrows).

## Mitochondrial DNA

We successfully sequenced a 210 bp long fragment of the mtDNA control region from 446 samples. Thirteen samples were identified as Numts and excluded. The vast majority of the studied individuals possessed *A. f. fabalis* mtDNA haplotypes (*n* = 432; Fig. 7). The most common haplotype among the Finnish bean geese was Fa3 (*n =* 261) followed by the haplotype FAB1a/FAB1b/Fa2 (*n* = 159). The slashes between haplotype names denote identical haplotypes based on the sequenced region. These haplotypes, however, differ based on the whole control region (Ruokonen, Litvin and Aarvak, 2008; Honka *et al.*, 2017). The two most common haplotypes were distributed throughout Finland (Fig. 8a,b,c). Four rarer haplotypes were also found: FAB3 (*n =* 2), Fa1 (*n =* 3), Fa4 (*n =* 4) and Fa7 (*n =* 3; GenBank accession number: MT023340). The rarer haplotypes were more localized with haplotype Fa1 only found in Northern Ostrobothnia, Fa4 only in Southern Lapland, Fa7 in only males in two areas in Northern Ostrobothnia and Lapland and FAB3 only in females from Ostrobothnia (Fig. 8a,b,c). Haplotype FAB3 was also common among the Russian geese, which were sampled along their migration route from south-eastern Finland (Fig. 8a). Haplotypes Fa3, FAB1a/FAB1b/Fa2 and Fa8 (GenBank accession number: MT023341) were found among the Swedish taiga bean geese (Fig. 8a).

Surprisingly, 16 individuals possessed a mtDNA haplotype typical of the pink-footed goose (Fig. 7) and these individuals were distributed throughout Finland (Fig. 8a,b,c). Two individuals had a mtDNA belonging to the tundra bean goose *A. f. rossicus* (Fig. 7). One of these was found from the Helsinki metropolitan area, which is far outside the natural breeding range of either subspecies (Fig. 8a,c). The other was found in northernmost Lapland (Fig. 8a). The Norwegian outgroup consisted solely of the tundra bean geese (Fig. 8a).

The number of haplotypes was the highest in the Northern Ostrobothnia/Southern Lapland region and the lowest in the Western and the Eastern Finland/Kainuu regions (Table 3). Haplotype and nucleotide diversities were the lowest in the Western Finland region and the highest in the Northern Ostrobothnia/Southern Lapland region (Table 3). Tajima’s Dvalue was negative in all of the regions while Fu’s Fs was negative only in the Northern Ostrobothnia/Southern Lapland and the Lapland regions (Table 3). Pairwise *ɸ*ST-values were either zero or very close to zero in all comparisons except between the Northern Ostrobothnia/Southern Lapland region and all the other regions (Table 2). The pairwise *ɸ*ST-values were even lower if the non-kin dataset was used, with some values even below zero (Table 2). Thus, the Northern Ostrobothnia/Southern Lapland region was moderately genetically differentiated (*ɸ*ST = 0.11-0.20). The pairwise *ɸ*ST-values were higher for females (*ɸ*ST-values between 0.00-0.27) than in males (*ɸ*ST-values between 0.00-0.13) (Table 4). In addition, 93.8% of the total variation was within region variation and 6.2% was among region variation (*ɸ*ST = 0.062; P < 0.001) according to AMOVA.

# Discussion

We did not detect clear population structure within the Finnish breeding taiga bean geese using microsatellite markers. All analyses suggested close to a panmictic population, except the DAPC that indicated slight structuring between Western Finland, Eastern Finland/Kainuu and Lapland (see Fig. 2). Presence of geographically localized mtDNA haplotypes and higher *ɸ*ST-values for females than in males, however, suggested at least some maternal genetic structure. It was unforeseen to find such little genetic structure within such a large geographic area, but the pairing system of geese, in which pair formation occurs already in common wintering or spring staging areas, can explain these results. We found moderate genetic diversity and signs of inbreeding within the Finnish taiga goose population.

Surprisingly, we also found that a pink-footed goose mtDNA haplotype is widespread (although at low frequency) among the taiga bean goose. This could indicate hybridization between the taiga bean goose and the pink-footed goose and admixture was also evident in the microsatellite data. In addition, we confirmed breeding of tundra bean geese in the northernmost Lapland and the presence of a vagrant tundra bean goose in Southern Finland. The microsatellite data suggested introgression between the taiga and the tundra bean goose as well.

## Genotyping success, relatedness and genetic diversity

This study proved that a citizen-science approach to feather collection was highly efficient for the elusive bean goose, with more than 2100 feathers received during years 2016-2018. We initially screened a large number of feathers as some of the samples had to be discarded due to low quality and quantity of DNA. The geographical coverage of the sampling was good, spanning the whole distribution area of the bean goose in Finland (Fig. 2). Even though bean goose feathers, especially the down feathers, can be mixed with other large birds in similar habitats, such as the common crane (*Grus grus*) and the whooper swan (*Cygnus cygnus*), only a few feathers had to be excluded as belonging to other species. This was determined either based on feather morphology or because no PCR-product amplified from pristine feathers not identifiable based on morphology.

The microsatellite genotyping success rate was rather high (84%) when compared to other studies utilising non-invasive feather samples (50-74%; Segelbacher, 2002; Hogan *et al*., 2008), probably because we did not attempt to extract DNA from the most poorly preserved feathers. The overall error rate was 2% per allele over all loci and 4% per locus over all loci. This level of error could have an effect on the analyses, but we used consensus genotypes created from several different feathers belonging to the same individual, which should reduce the amount of error. The use of the stepwise amplification protocol for microsatellites should also reduce the error as only high quality feathers that showed the least genotyping errors and easily scored alleles were selected for the next amplification step. In addition, we used programs for analysis that can take the genotyping errors into account.

Molecular sexing based on HINTZ/W gametologs identified 237 females and 189 males, deviating significantly from an equal sex ratio (*P =* 0.02). Bean geese form stable pair bonds and both parents participate in the rearing of goslings, so an equal number of females and males was expected; the reason for the deviation is unknown. Based on our molecular sexing results, one cloacally examined goose out of seven was misidentified as a female. The cloacal examination is not a 100% accurate method and the male organ can be difficult to expose if the goose was immature or not producing semen (Buckland and Guy, 2002).

Several parent-offspring and full-sib dyads were identified using the microsatellite markers. We also identified one family in which one of the goslings was not the offspring of the social parents, thus representing a possible case of gosling adoption or an intraspecific nest parasitism. Although these behaviours have not been observed in the bean goose, gosling adoption is not uncommon among geese species (Zicus, 1981; Choudbury *et al*., 1993; Larsson, Tegelström and Forslund, 1995; Nilsson and Persson, 2003; for review, see Kalmbach, 2006) and likewise intraspecific nest parasitism (Weigmann and Lamprecht, 1991; Larsson, Tegelström and Forslund, 1995; Anderholm *et al*., 2009).

Overall, a large majority of the individual goose dyads were unrelated or related to a very low degree. However, a few of the goose dyads showed high relatedness values (*r* = 0.60-0.86) indicating potential inbreeding. Even though genotyping errors could also explain the high relatedness, most of these individuals had genotypes not consistent with allele dropout, had different mtDNA haplotypes, were of different sex or came from feathers of different ringed individuals. The finding of several potentially inbred individuals was surprising as the detrimental effects of inbreeding to individual fitness, known as the inbreeding depression, are widely documented (Keller and Waller, 2002) and usually animals avoid inbreeding through several mechanisms (Pusey and Wolf, 1996). However, a single case of sibling pairing has been observed in the Canada goose (*Branta canadensis*;Lebeuf and Giroux, 2013). Positive inbreeding coefficients (*F*IS) were also observed within the Finnish population as all geographic regions deviated from the Hardy-Weinberg equilibrium (Table 1). Heterozygote deficiency could be due to Wahlund effect, which is caused by the merging of populations with different allele frequencies (Wahlund, 1928). Even though we detected only a very low level of population structure, it is possible that the population structure is very fine-scaled or there is population structure in the wintering sites instead of breeding sites. We were unable to discern exact family relationships beyond parent-offspring and siblings, which might affect the Hardy-Weinberg equilibrium and cause positive *F*IS-values.

All geographic regions showed moderate heterozygosity levels (*H*O = 0.62-0.66).Similar heterozygosity values have been observed for example in the greater white-fronted goose (*A. albifrons*; Ely, Wilson and Talbot, 2017), while the Scandinavian taiga bean geese had slightly higher heterozygosity levels (*H*O = 0.72-0.80; de Jong *et al*., 2019). Thus, genetic diversity seemed not to be reduced within the Finnish taiga bean goose population despite the recent decline in population numbers.

## Population structure and demography

We did not observe any population structure when using microsatellite markers with the Structure program as all individuals showed admixture and no genetic clustering was evident (Fig. 3) for any value of *K*. We found genetic separation between Western Finland, Eastern Finland/Kainuu and Lapland (see Fig. 2) in the DAPC analysis; however, this genetic separation was not strong (Fig. 4a). All pairwise *FST*-values were low (< 0.005) showing no clear nuclear differentiation between the geographic regions (Table 2). We did not observe isolation-by-distance either, further indicating lack of spatial genetic structuring (Fig. 5a). These findings indicate that the taiga bean goose population is close to panmictic.

On the other hand, some of the rarer mtDNA haplotypes were localised to certain areas (Fig. 8a,b,c) and the pairwise *ɸ*ST–values showed genetic differentiation (0.11-0.20), between the Northern Ostrobothnia/Southern Lapland region and all other regions (Table 2), implying at least some level of female philopatry. In addition, the pairwise *ɸ*ST–values were higher for females than in males (Table 4), indicating stronger genetic structuring in females than in males. Also, we found a difference in haplotype composition between females and males as FAB3 haplotype was only found in females and haplotype Fa7 only in males. AMOVA analysis indicated that only about 6.2% of the total variation was between regions, thus mtDNA genetic structuring is still rather limited.

However, nuclear spatial genetic patterns did not indicate sex-specific differences (Fig. 5b). As opposed to male philopatry observed in most other birds, geese show female philopatry to natal areas (Greenwood, 1980; van der Jeugd, van der Veen and Larsson, 2002). Therefore, genetic structure may be promoted especially in the maternally inherited mitochondrial DNA or the female specific W chromosome, as was seen here as higher *ɸ*ST–values in females than in males. It seems, however, that the dispersal of male bean geese is so high that it completely homogenizes the nuclear genome. Geese pair in their wintering grounds or in the spring staging areas and males follow females to the female’s natal area (Rohwer and Anderson, 1988), which allows geese from even distant breeding areas to pair and thus mediate gene flow in the biparentally-inherited nuclear DNA. Long distance dispersal of a few males could lead to panmixia as only one migrant per generation in an ideal population is enough to prevent population differentiation due to drift (Mills and Allendorf, 1996; Wang 2004, but see Vucetich and Waite, 2000). For example, in greylag and brent geese (*B. bernicla hrota*) most males breed close to their natal site but a minority of the geese undergo long-distance dispersal (Nilsson and Persson, 2001; Harrison *et al.*, 2010). On the other hand, a lack of sex-specific dispersal differences has been observed in Asian breeding swan goose (*A. cygnoid*; Zhu *et al*., 2020). The taiga bean geese winter in gregarious flocks mainly in Southern Sweden and Denmark and to a lesser extent in the Netherlands, western Germany, Poland and Britain (Nilsson, van den Bergh and Madsen, 1999; Fig. 1) providing ample opportunities for geese breeding in different areas to mix. Only the taiga bean geese belonging to the Eastern 2 population overwinter in a separate area in Central Asia (Heinicke, 2009; Fig. 1), therefore this population could show genetic differentiation.

Studies in other goose species have shown varying levels of genetic structure ranging from a lack of genetic structure (Avise *et al*., 1992; Harrison *et al*., 2010; Pellegrino *et al*., 2015), to phylogeographic clustering (Ruokonen *et al*., 2004; Pujolar *et al*., 2017) and strong genetic structuring in brood-rearing sites (Lecomte *et al*., 2009). For example, microsatellite studies in the greater white-fronted goose have discovered a panmictic population with the exception of the Greenland white-fronted goose (*A. a. flavirostris*) and the Tule goose (*A. a. elgasi*) (Ely, Wilson and Talbot, 2017; Wilson, Ely and Talbot, 2018). Also, the different flyway populations of the barnacle goose (*B. leucopsis*) studied with SNP (single nucleotide polymorphism) markers showed genetic structuring but also genetic exchange between all flyways (Jonker *et al*., 2013). Contrary to our findings in the Finnish breeding bean geese, genetic structuring was observed with microsatellite markers within the taiga bean geese breeding in Central Scandinavia (belonging to the Western flyway management unit, see Fig.1; de Jong *et al.*, 2019). However, de Jong *et al*. (2019) studied fine-scale genetic patterns in a geographically restricted area, thus the different scale (family-level structure in de Jong *et al.*, 2019) of our study could explain the contrasting results.

Both the linkage disequilibrium and the sibship method produced similar estimates of the effective population size with 1128 and 1134 individuals, respectively. The Finnish population size is estimated to be 1700-2500 breeding pairs (i.e. 3400-5000 individuals) based on survey data (Valkama, Vepsäläinen and Lehikoinen, 2011), thus excluding non-breeders and juveniles. The ratio between the *N*e estimated based on microsatellite markers (this study) and the estimate of the number of breeding individuals is 0.23–0.33. Accordingly, *N*e is just about a quarter or a third of the estimated breeding population. Low *N*e/*N* ratios (0.11-0.14) are commonly reported among animals (Frankham, 1995; Palstra and Ruzzante, 2008) and based on *N*e = 1128, the total population size (*N*) would be thus 8,057-10,255 individuals, a surprisingly large estimate. However, our estimate of the *N*e from only Finnish samples is perhaps somewhat misleading due to a continuous population over the borders of countries.

We found no indication of past population bottlenecks, instead, we inferred population expansion starting around 22,000-32,000 years ago. This was unexpected since the taiga bean goose population is in decline (Fox *et al.*, 2010; Fox and Leafloor, 2018). Probably the current population decline is too recent or not severe enough to be detected in genetic bottleneck test. The population expansion starting around 27,000 years ago coincidences with or precedes the Last Glacial Maximum (around 26,500-20,000 years ago; Clark *et al.*, 2009).

## Mitochondrial DNA

The number of haplotypes, haplotype diversity and nucleotide diversity were the highest among the Northern Ostrobothnia/Southern Lapland region followed by the Lapland region (Table 3). The mitochondrial diversity is thus higher in the northern areas, which are the core breeding regions of the taiga bean goose in Finland. The haplotype diversities were low (*h* = 0.40-0.58) compared to other bean goose subspecies (*h* = 0.68-0.86; Honka *et al*., 2017) or, for example, to the greater white-fronted goose (*h* = 0.68-0.96; Ely, Wilson and Talbot, 2017), but similar in range to the endangered population of lesser white-fronted goose (*h* = 0.37–0.53; *A. erythropus*; Ruokonen *et al*., 2004). The nucleotide diversities were over a similar range as reported for the aforementioned species.

All samples had taiga bean goose mtDNA haplotypes, except two individuals that had tundra bean goose mtDNA haplotypes and 16 individuals that had pink-footed goose mtDNA haplotypes (Fig. 7). One of the tundra bean geese was sampled from a metropolitan area in Helsinki (Fig. 8a,c), outside of the breeding range of either subspecies and was thus a vagrant bird. This individual spent the summer in parkland and did not appear to be injured, therefore it is uncertain why this individual did not migrate to its normal breeding grounds in Russian tundra. The other tundra bean goose was sampled from northernmost Finland, confirming that this subspecies breeds in Finland (Fig. 8a). Our results also genetically confirmed the findings of Aarvak and Øien (2009) that the tundra bean goose breeds in the most of Finnmark with the taiga bean goose only restricted to the east, near the Pasvik area (Fig. 8a,b,c).

## Hybridization

We also compared the Finnish population with other taiga bean geese (Russian and Swedish), the tundra bean goose and the pink-footed goose using DAPC. As expected, the Finnish geese mostly grouped with the Russian geese (Fig. 4b), which should belong to the same flyway management unit (Central; see Fig. 1). However, some Swedish geese (Western unit) showed genetic affinity to the Russian and Finnish geese and thus gene flow between the different flyways could be present. Some of the Finnish bean geese clustered closely with either the tundra bean goose or the pink-footed goose (Fig. 6). This could be due to a lack of resolution in our microsatellite markers to discriminate between different populations or an indication of possible hybridization between these populations. Whole-genome re-sequencing of the taiga and the tundra bean geese has shown that the genomes of these subspecies are homogenous expect for a few “islands of differentiation”, due to extensive gene flow 60,000 years ago (Ottenburghs *et al*., 2020). Thus, the close affinity of the subspecies is probably due to the past hybridization.

In this study, we showed that the pink-footed goose mtDNA is widespread (although at low frequency, 4% of the studied population) in the Finnish taiga bean goose population (Fig. 8a,b,c). This was unexpected because the pink-footed goose breeds in Greenland, Iceland and Svalbard and no breeding attempts have been recorded in Finland. Although the pink-footed goose can be found as a vagrant bird in Finland, and has started to regularly migrate through the Western Finland in recent decades (Heldbjerg *et al*., 2019), the finding of this many pure pink-footed geese does not seem plausible. The microsatellite data supported admixture (Fig. 6), indicating possible hybridization and introgression between the taiga bean goose and the pink-footed goose. Geese show a high propensity for hybridization (for review, see Ottenburghs *et al*., 2016b) and in the genomics era, increasing number of studies have identified that ancient hybridization, adaptive introgression and hybrid speciation are much more common than previously thought (Ottenburghs *et al.*, 2017; Taylor and Larson, 2019). Interestingly, the mtDNA of the tundra bean goose had not introgressed into the taiga bean goose according to our results. The mtDNA of the pink-footed goose may convey adaptive benefits as hybridization often leads to the introgression of adaptive genetic variation (Arnold and Kunte, 2017), but also incomplete lineage sorting can explain the presence of the pink-footed goose mtDNA (Degnan and Rosenberg, 2009). The pink-footed goose has been treated as a separate species as it has formed a monophyletic group based on mitochondrial DNA (Ruokonen, Litvin and Aarvak, 2008), although more recently a sister species relationship was suggested between the tundra bean goose and the pink-footed goose (Ottenburghs *et al*., 2016a). Further studies are needed to elucidate the phylogenetic position of the pink-footed goose and the possible admixture scenario.

## Management implications

We did not find evidence to divide the Finnish bean goose population into smaller management units or subpopulations as there was no strong genetic structuring within Finland. Therefore, the flyway management units outlined in the International Single Species Action Plan (ISSAP; Marjakangas *et al.*, 2015) seem to be justified based on our study (see Fig. 1). The genetic diversity was found to be moderate and effective population size fairly large, thus the Finnish breeding taiga bean geese are not under immediate threat by genetic impoverishment. However, as the genetic diversity was lower compared to other bean goose subspecies and widespread goose species, further reductions in genetic diversity should be avoided to maintain the evolutionary potential of this subspecies, especially since inbreeding was detected in the population. Even though we found evidence of possible hybridization and introgression, naturally occurring hybridization does not pose a threat to populations as it is natural part of species evolution (Allendorf *et al*., 2001; Taylor and Larson, 2019). Thus, natural hybridization should not disqualify species from conservation programs and protection (vonHoldt *et al*., 2018).

Non-invasive genetic sampling could be used for genetic monitoring of the taiga bean goose in the future to provide estimates of local population census sizes, survival, recruitment and temporal variation in genetic diversity in order to ensure the genetic viability of the population (Schwartz, Luikart and Waples, 2007). Besides microsatellites and mtDNA, SNP markers could be used in genetic monitoring, as SNPs do not require laborious calibration between different laboratories enabling flyway-wide monitoring, and are well suited for non-invasive samples as amplicon lengths are short and SNPs are less error prone than microsatellites (Carroll *et al*., 2018). A SNP panel has been already developed for subspecies identification for the greater white-fronted goose (Wilson *et al*., 2019) and similarly a SNP panel could provide a feasible alternative for the genetic monitoring for the taiga bean geese.

# Acknowledgments

We are grateful to the numerous volunteers who collected feathers and enabled this study. We also thank Risto Karvonen, Ingar Jostein Øien, Allan Hamari and the Natural History Museum of Reykjavik for samples. Special thanks to Antti Piironen and the Pohjois-Karjalan metsähanhimiehet for feather samples from ringed geese. We are grateful for the help with field sampling from Toni Laaksonen, Mervi Kunnasranta, Petri Timonen and Tuomas Seimola (The Natural Resource Institute of Finland) and the crew of HF Helicopters. We also thank everyone who helped advertise the feather collection with special thanks to Mikko Alhainen from the Finnish Wildlife Agency. We thank Kari Merikanto for coding the error rate calculation program. In addition, we thank the Finnish Cultural Foundation’s North Ostrobothnia Regional Fund (grant number: 60182022) and the Finnish Cultural Foundation (grant number: 00200356) for providing funding for JH.

**Data accessibility**

The program code generated during the current study is available in the GitHub repository, https://github.com/karimerikanto/microsat\_errcalc. The executable file is available on request from the corresponding author. The sequence data has been deposited in GenBank with accession numbers: MT023340-MT023341. Upon acceptance, we will upload the microsatellite genotypes in Dryad repository.

**Conflict of interest**

None declared.

**Author contributions**

**Johanna Honka**: conceptualization (equal); formal analysis (lead); investigation (lead); methodology (lead); visualization (lead); writing – original draft preparation (lead); writing – review & editing (equal). **Serena Baini**: investigation (equal); writing – review & editing (equal). **Jeremy B. Searle**: supervision (equal); writing – review & editing (equal). **Laura Kvist**: conceptualization (equal); supervision (equal); writing – review & editing (equal). **Jouni Aspi**: conceptualization (equal); supervision (lead); writing – review & editing (equal).

# References

Aarvak, T. & Øien, I.J. 2009, *Monitoring of bean goose in Finnmark county, Norway–results from 2008. NOF rapportserie report No 2‐2009*, Norsk Ornitologisk Forening, Norway.

Allendorf, F.W., England, P.R., Luikart, G., Ritchie, P.A. & Ryman, N. 2008, "Genetic effects of harvest on wild animal populations", *Trends in Ecology and Evolution,*vol. 23, no. 6, pp. 327-337.

Allendorf, F.W., Leary, R.F., Spruell, P. & Wenburg, J.K. 2001, "The problems with hybrids: Setting conservation guidelines", *Trends in Ecology and Evolution,*vol. 16, no. 11, pp. 613-622.

Anderholm, S., Marshall, R.C., van der Jeugd, H. P., Waldeck, P., Larsson, K. & Andersson, M. 2009, "Nest parasitism in the barnacle goose: Evidence from protein fingerprinting and microsatellites", *Animal Behaviour,*vol. 78, no. 1, pp. 167-174.

Anderson, E.C. & Dunham, K.K. 2008, "The influence of family groups on inferences made with the program Structure", *Molecular Ecology Resources,*vol. 8, no. 6, pp. 1219-1229.

Anderson, E.C. & Thompson, E.A. 2002, "A model-based method for identifying species hybrids using multilocus genetic data", *Genetics,*vol. 160, no. 3, pp. 1217-1229.

Arnold, M.L. & Kunte, K. 2017, "Adaptive genetic exchange: A tangled history of admixture and evolutionary innovation", *Trends in Ecology and Evolution,*vol. 32, no. 8, pp. 601-611.

Avise, J.C., Alisauskas, R.T., Nelson, W.S. & Ankney, C.D. 1992, "Matriarchal population genetic structure in an avian species with female natal philopatry", *Evolution,*vol. 46, no. 4, pp. 1084-1096.

Bandelt, H.-J., Forster, P. & Röhl, A. 1999, "Median-joining networks for inferring intraspecific phylogenies", *Molecular Biology and Evolution,*vol. 16, no. 1, pp. 37-48.

Banks, S.C. & Peakall, R. 2012, "Genetic spatial autocorrelation can readily detect sex-biased dispersal", *Molecular Ecology,*vol. 21, no. 9, pp. 2092-2105.

Bensasson, D., Zhang, D.-X., Hartl, D.L. & Hewitt, G.M. 2001, "Mitochondrial pseudogenes: Evolution's misplaced witnesses", *Trends in Ecology and Evolution,*vol. 16, no. 6, pp. 314-321.

BirdLife International 2018,*Anser fabalis. The IUCN Red List of Threatened Species 2018:  
e.T22679875A132302864*. Available: http://dx.doi.org/10.2305/IUCN.UK.2018- 2.RLTS.T22679875A132302864.en [2020, 20.01.].

Bougeard, S. & Dray, S. 2018, "Supervised multiblock analysis in R with the ade4 package", *Journal of Statistical Software,*vol. 86, no. 1, pp. 1-17.

Buckland, R.B. & Guy, G. 2002, *Goose production*, Food and Agriculture Organization of the United Nations, Rome.

Carroll, E.L., Bruford, M.W., DeWoody, J.A., Leroy, G., Strand, A., Waits, L. & Wang, J. 2018, "Genetic and genomic monitoring with minimally invasive sampling methods", *Evolutionary Applications,*vol. 11, no. 7, pp. 1094-1119.

Chapuis, M.-P. & Estoup, A. 2007, "Microsatellite null alleles and estimation of population differentiation", *Molecular Biology and Evolution,*vol. 24, no. 3, pp. 621-631.

Chessel, D., Dufour, A.B. & Thioulouse, J. 2004, "The ade4 package-I-One-table methods", *R News,*vol. 4, no. 1, pp. 5-10.

Choudhury, S., Jones, C.S., Black, J.M. & Prop, J. 1993, "Adoption of young and intraspecific nest parasitism in barnacle geese", *Condor,*vol. 95, no. 4, pp. 860-868.

Clark, P.U., Dyke, A.S., Shakun, J.D., Carlson, A.E., Clark, J., Wohlfarth, B., Mitrovica, J.X., Hostetler, S.W. & McCabe, A.M. 2009, "The Last Glacial Maximum", *Science*, vol. 325, no. 5941, pp. 710-714.

De Jong, A., Kleven, O., Ostnes, J.E., Kroglund, R.T., Vahlström, I., Nilsson, J.A.N. & Spong, G. 2019, "Birds of different feather flock together - genetic structure of taiga bean goose in Central Scandinavia", *Bird Conservation International,*vol. 29, no. 2, pp. 249-262.

Degnan, J.H. & Rosenberg, N.A. 2009, "Gene tree discordance, phylogenetic inference and the multispecies coalescent", *Trends in Ecology and Evolution,*vol. 24, no. 6, pp. 332-340.

Dib, C., Fauré, S., Fizames, C., Samson, D., Drouot, N., Vignal, A., Millasseau, P., Marc, S., Hazan, J., Seboun, E., Lathrop, M., Gyapay, G., Morissette, J. & Weissenbach, J. 1996, "A comprehensive genetic map of the human genome based on 5,264 microsatellites", *Nature,*vol. 380, no. 6570, pp. 152-154.

Dillingham, P.W. 2010, "Generation time and the maximum growth rate for populations with age-specific fecundities and unknown juvenile survival", *Ecological Modelling,*vol. 221, no. 6, pp. 895-899.

Do, C., Waples, R.S., Peel, D., Macbeth, G.M., Tillett, B.J. & Ovenden, J.R. 2014, "NeEstimator v2: Re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data", *Molecular Ecology Resources,*vol. 14, no. 1, pp. 209-214.

Dray, S., Dufour, A.B. & Chessel, D. 2007, "The ade4 package-II: Two-table and K-table methods", *R News,*vol. 7, no. 2, pp. 47-52.

Dray, S. & Dufour, A. 2007, "The ade4 package: implementing the duality diagram for ecologists", *Journal of Statistical Software,*vol. 22, no. 4, pp. 1-20.

Earl, D.A. & vonHoldt, B.M. 2012, "STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method", *Conservation Genetics Resources,*vol. 4, no. 2, pp. 359-361.

Ellegren, H. 2000, "Microsatellite mutations in the germline: Implications for evolutionary inference", *Trends in Genetics,*vol. 16, no. 12, pp. 551-558.

Ely, C.R., Wilson, R.E. & Talbot, S.L. 2017, "Genetic structure among greater white-fronted goose populations of the Pacific Flyway", *Ecology and Evolution,*vol. 7, no. 9, pp. 2956-2968.

Evanno, G., Regnaut, S. & Goudet, J. 2005, "Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study", *Molecular Ecology,*vol. 14, no. 8, pp. 2611-2620.

Excoffier, L. & Lischer, H.E.L. 2010, "Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows", *Molecular Ecology Resources,*vol. 10, no. 3, pp. 564-567.

Falush, D., Stephens, M. & Pritchard, J.K. 2003, "Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies", *Genetics,*vol. 164, no. 4, pp. 1567-1587.

Fox, A.D., Ebbinge, B.S., Mitchell, C., Heinicke, T., Aarvak, T., Colhoun, K., Clausen, P., Dereliev, S., Faragö, S., Koffijberg, K., Kruckenberg, H., Loonen, M. J. J. E., Madsen, J., Mooij, J., Musil, P., Nilsson, L., Pihl, S. & Van Der Jeugd, H. 2010, "Current estimates of goose population sizes in western Europe, a gap analysis and an assessment of trends", *Ornis Svecica,*vol. 20, no. 3-4, pp. 115-127.

Fox, A.D., Hobson, K.A., de Jong, A., Kardynal, K.J., Koehler, G. & Heinicke, T. 2017, "Flyway population delineation in taiga bean geese *Anser fabalis fabalis* revealed by multi-element feather stable isotope analysis", *Ibis,*vol. 159, no. 1, pp. 66-75.

Fox, A.D. & Madsen, J. 2017, "Threatened species to super-abundance: The unexpected international implications of successful goose conservation", *Ambio,*vol. 46, no. 2, pp. 179-187.

Fox, A.D. & Leafloor, J.O. (eds) 2018, *A global audit of the status and trends of Arctic and Northern Hemisphere goose populations*, Conservation of Arctic Flora and Fauna International Secretariat, Akureyri, Iceland.

Frankham, R. 2005, "Genetics and extinction", *Biological Conservation,*vol. 126, no. 2, pp. 131-140.

Frankham, R. 1995, "Effective population size/adult population size ratios in wildlife: A review", *Genetical Research,*vol. 66, no. 2, pp. 95-107.

Goudet, J. 1995, "FSTAT (Version 1.2): A computer program to calculate F-statistics", *Journal of Heredity,*vol. 86, no. 6, pp. 485-486.

Gravley, M.C., Sage, G.K., Schmutz, J.A. & Talbot, S.L. 2017, "Development of microsatellite loci exhibiting reverse ascertainment bias and a sexing marker for use in emperor geese (*Chen canagica*)", *Avian Biology Research,*vol. 10, no. 4, pp. 201-210.

Greenwood, P.J. 1980, "Mating systems, philopatry and dispersal in birds and mammals", *Animal Behaviour,*vol. 28, no. 4, pp. 1140-1162.

Guzzetti, B.M., Talbot, S.L., Tessler, D.F., Gill, V.A. & Murphy, E.C. 2008, "Secrets in the eyes of black oystercatchers: A new sexing technique", *Journal of Field Ornithology,*vol. 79, no. 2, pp. 215-223.

Hall, T. 1999, "BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT", *Nucleic Acids Symposium Series,*vol. 41, pp. 95-98.

Harrison, X.A., Tregenza, T., Inger, R., Colhoun, K., Dawson, D.A., Gudmundsson, G.A., Hodgson, D.J., Horsburgh, G.J., McElwaine, G. & Bearhop, S. 2010, "Cultural inheritance drives site fidelity and migratory connectivity in a long-distance migrant", *Molecular Ecology,*vol. 19, no. 24, pp. 5484-5496.

Heinicke, T. 2009, "Status of the Bean Goose *Anser fabalis* wintering in central Asia", *Wildfowl,*vol. 59, pp. 77-99.

Heldbjerg, H., Madsen, J., Amstrup, O., Bakken, J., Balsby, T., Christensen, T.K., Clausen, K.K., Cottaar, F., Frikke, J., Gundersen, O.M., Kjeldsen, J.P., Koffijberg, K., Kuijken, E., Månsson, J., Nicolaisen, P.I., Nielsen, H.H., Nilsson, L., Reinsborg, T., Pessa, J., Shimmings, P., Tombre, I., Verhaeghe, F., Verscheure, C. & Westerbring, M.J. 2019, *Pink-footed goose Svalbard population: Status report 2018-2019*, AEWA European Goose Management Platform, Perth, United Kingdom.

Hill, W.G. 1981, "Estimation of effective population size from data on linkage disequilibrium", *Genetical Research,*vol. 38, no. 3, pp. 209-216.

Hogan, F.E., Cooke, R., Burridge, C.P. & Norman, J.A. 2008, "Optimizing the use of shed feathers for genetic analysis", *Molecular Ecology Resources,*vol. 8, no. 3, pp. 561-567.

Honka, J., Heino, M.T., Kvist, L., Askeyev, I.V., Shaymuratova, D.N., Askeyev, O.V., Askeyev, A.O., Heikkinen, M.E., Searle, J.B. & Aspi, J. 2018, "Over a thousand years of evolutionary history of domestic geese from Russian archaeological sites, analysed using ancient DNA", *Genes,*vol. 9, no. 7. DOI 10.3390/genes9070367

Honka, J., Kvist, L., Heikkinen, M.E., Helle, P., Searle, J.B. & Aspi, J. 2017, "Determining the subspecies composition of bean goose harvests in Finland using genetic methods", *European Journal of Wildlife Research,*vol. 63, no. 1. DOI 10.1007/s10344-017-1077-6

Honka, J. & Merikanto, K. 2020, *Microsat\_errcalc. GitHub repository*. Available: https://github.com/karimerikanto/microsat\_errcalc [2020, 27.01.].

Horváth, M.B., Martínez-Cruz, B., Negro, J.J., Kalmár, L. & Godoy, J.A. 2005, "An overlooked DNA source for non-invasive genetic analysis in birds", *Journal of Avian Biology,*vol. 36, no. 1, pp. 84-88.

Hubisz, M.J., Falush, D., Stephens, M. & Pritchard, J.K. 2009, "Inferring weak population structure with the assistance of sample group information", *Molecular Ecology Resources,*vol. 9, no. 5, pp. 1322-1332.

Jombart, T., Devillard, S. & Balloux, F. 2010, "Discriminant analysis of principal components: A new method for the analysis of genetically structured populations", *BMC Genetics,*vol. 11, no. 1. DOI 10.1186/1471-2156-11-94

Jombart, T. 2008, "adegenet: A R package for the multivariate analysis of genetic markers", *Bioinformatics,*vol. 24, no. 11, pp. 1403-1405.

Jombart, T. & Ahmed, I. 2011, "adegenet 1.3-1: New tools for the analysis of genome-wide SNP data", *Bioinformatics,*vol. 27, no. 21, pp. 3070-3071.

Jones, O.R. & Wang, J. 2010, "COLONY: A program for parentage and sibship inference from multilocus genotype data", *Molecular Ecology Resources,*vol. 10, no. 3, pp. 551-555.

Jonker, R.M., Kraus, R.H.S., Zhang, Q., Van Hooft, P., Larsson, K., Van Der Jeugd, H. P., Kurvers, R. H. J. M., Van Wieren, S.E., Loonen, M. J. J. E., Crooijmans, R. P. M. A., Ydenberg, R.C., Groenen, M.A.M. & Prins, H.H.T. 2013, "Genetic consequences of breaking migratory traditions in barnacle geese *Branta leucopsis*", *Molecular Ecology,*vol. 22, no. 23, pp. 5835-5847.

Jukes, T.H. & Cantor, C.R. 1969, "Evolution of protein molecules" in *Mammalian Protein Metabolism*, ed. H.N. Munro, Academic Press, New York, NY, pp. 21-132.

Kalinowski, S.T., Wagner, A.P. & Taper, M.L. 2006, "ML-RELATE: A computer program for maximum likelihood estimation of relatedness and relationship", *Molecular Ecology Notes,*vol. 6, no. 2, pp. 576-579.

Kalmbach, E. 2006, "Why do goose parents adopt unrelated goslings? A review of hypotheses and empirical evidence, and new research questions", *Ibis,*vol. 148, no. 1, pp. 66-78.

Keller, L.F. & Waller, D.M. 2002, "Inbreeding effects in wild populations", *Trends in Ecology and Evolution,*vol. 17, no. 5, pp. 230-241.

Kleven, O., Kroglund, R.T. & Østnes, J.E. 2016, "Isolation, characterization and multiplex PCR development of bean goose (*Anser fabalis*) microsatellite loci", *Journal of Ornithology,*vol. 157, no. 2, pp. 641-646.

Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A. & Mayrose, I. 2015, "Clumpak: A program for identifying clustering modes and packaging population structure inferences across K", *Molecular Ecology Resources,*vol. 15, no. 5, pp. 1179-1191.

Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. 2018, "MEGA X: Molecular evolutionary genetics analysis across computing platforms", *Molecular Biology and Evolution,*vol. 35, no. 6, pp. 1547-1549.

Laaksonen, T., Piironen, A. & Mäntyniemi, S. 2019, "Metsähanhien epäonninen kesä ja sulkasatomuutto Novaja Zemljalle [Unlucky summer for taiga bean geese and moult migration to Novaya Zemlya]", *Metsästäjä,*vol. 6, pp. 34-35.

Laitinen, J., Rehell, S., Huttunen, A., Tahvanainen, T., Heikkilä, R. & Lindholm, T. 2007, "Mire systems in Finland–special view to aapa mires and their water-flow pattern", *Suo,*vol. 58, no. 1, pp. 1-26.

Larsson, K., Tegelström, H. & Forslund, P. 1995, "Intraspecific nest parasitism and adoption of young in the barnacle goose: Effects on survival and reproductive performance", *Animal Behaviour,*vol. 50, no. 5, pp. 1349-1360.

Lebeuf, A.P. & Giroux, J. 2013, "Sibling pairing in temperate-nesting Canada geese", *The Wilson Journal of Ornithology,*vol. 125, no. 2, pp. 398-401.

Leblois, R., Pudlo, P., Néron, J., Bertaux, F., Reddy Beeravolu, C., Vitalis, R. & Rousset, F. 2014, "Maximum-likelihood inference of population size contractions from microsatellite data", *Molecular Biology and Evolution,*vol. 31, no. 10, pp. 2805-2823.

Lecomte, N., Gauthier, G., Giroux, J.-F., Milot, E. & Bernatchez, L. 2009, "Tug of war between continental gene flow and rearing site philopatry in a migratory bird: The sex-biased dispersal paradigm reconsidered", *Molecular Ecology,*vol. 18, no. 4, pp. 593-602.

Lehikoinen, A., Jukarainen, A., Mikkola-Roos, M., Below, A., Lehtiniemi, T., Pessa, J., Rajasärkkä, A., Rintala, J., Rusanen, P., Sirkiä, P., Tiainen, J. & Valkama, J. 2019, "Linnut [Birds]" in *Suomen lajien uhanalaisuus–Punainen kirja 2019 [The 2019 Red List of Finnish Species]*, eds. E. Hyvärinen, A. Juslén, E. Kemppainen, A. Uddström & U.M. Liukko, Ympäristöministeriö & Suomen ympäristökeskus.

Leigh, J.W. & Bryant, D. 2015, "POPART: Full-feature software for haplotype network construction", *Methods in Ecology and Evolution,*vol. 6, no. 9, pp. 1110-1116.

Lopez, J.V., Yuhki, N., Masuda, R., Modi, W. & O'Brien, S.J. 1994, "Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat", *Journal of Molecular Evolution,*vol. 39, no. 2, pp. 174-190.

Marjakangas, A., Alhainen, M., Fox, A.D., Heinicke, T., Madsen, J., Nilsson, L., Rozenfeld, S. & (Compilers) 2015, *International Single Species Action Plan for the conservation of the taiga bean goose Anser fabalis fabalis*, AEWA Technical Series No. XX, Bonn, Germany.

Mills, L.S. & Allendorf, F.W. 1996, "The one-migrant-per-generation rule in conservation and management", *Conservation Biology,*vol. 10, no. 6, pp. 1509-1518.

Mooij, J.H. & Zöckler, C. 1999, "Reflections on the systematics, distribution and status of *Anser fabalis* (Latham, 1787)", *Casarca,*vol. 5, pp. 103-120.

Nielsen, E.E., Bach, L.A. & Kotlicki, P. 2006, "HYBRIDLAB (version 1.0): A program for generating simulated hybrids from population samples", *Molecular Ecology Notes,*vol. 6, no. 4, pp. 971-973.

Nilsson, L., De Jong, A., Heinicke, T. & Sjöberg, K. 2010, "Satellite tracking of bean geese *Anser fabalis fabalis* and *A. f. rossicus*from spring staging areas in northern Sweden to breeding and moulting areas", *Ornis Svecica,*vol. 20, no. 3-4, pp. 184-189.

Nilsson, L. & Kampe-Persson, H. 2003, "Why should greylag goose *Anser anser* parents rear offspring of others?", *Wildfowl,*vol. 54, pp. 25-37.

Nilsson, L. & Persson, H. 2001, "Natal and breeding dispersal in the Baltic greylag goose *Anser anser*", *Wildfowl,*vol. 52, pp. 21-30.

Nilsson, L., van den Bergh, L. & Madsen, J. 1999, "Taiga Bean Goose *Anser fabalis fabalis*" in *Goose populations in the Western Palearctic. A review of status and distribution. Wetlands International Publication No. 48.*, eds. J. Madsen, G. Cracknell & T. Fox, Wetlands International, National Environmental Research Institute, Rönde, pp. 23-39.

Noreikiene, K., Teacher, A.G.F., Madsen, J. & Gienapp, P. 2012, "Isolation and characterization of 55 novel microsatellite markers for the pink-footed goose (*Anser brachyrhynchus*)", *Conservation Genetics Resources,*vol. 4, no. 2, pp. 423-428.

Ottenburghs, J., Honka, J., Müskens, G. J. D. M. & Ellegren, H. 2020, "Recent introgression between taiga bean goose and tundra bean goose results in a largely homogeneous landscape of genetic differentiation", *Heredity,*vol. 125, no. 1-2, pp. 73-84.

Ottenburghs, J., Megens, H.-J., Kraus, R.H.S., Van Hooft, P., Van Wieren, S.E., Crooijmans, R. P. M. A., Ydenberg, R.C., Groenen, M.A.M. & Prins, H.H.T. 2017, "A history of hybrids? Genomic patterns of introgression in the true geese", *BMC Evolutionary Biology,*vol. 17, no. 1. DOI 10.1186/s12862-017-1048-2

Ottenburghs, J., Megens, H.-J., Kraus, R.H.S., Madsen, O., van Hooft, P., van Wieren, S.E., Crooijmans, R P M A, Ydenberg, R.C., Groenen, M.A.M. & Prins, H.H.T. 2016a, "A tree of geese: A phylogenomic perspective on the evolutionary history of true geese", *Molecular Phylogenetics and Evolution,*vol. 101, pp. 303-313.

Ottenburghs, J., van Hooft, P., van Wieren, S.E., Ydenberg, R.C. & Prins, H.H.T. 2016b, "Hybridization in geese: A review", *Frontiers in Zoology,*vol. 13, no. 1. DOI 10.1186/s12983-016-0153-1

Paasivaara, A. 2013, "Mukana muutolla: Satelliittilähetin antaa uutta tietoa taigametsähanhesta [Along with migration: Satellite transmitters provide new information about taiga bean goose]", *Metsästäjä,*vol. 1, pp. 10-13.

Palsbøll, P.J., Bérubé, M. & Allendorf, F.W. 2007, "Identification of management units using population genetic data", *Trends in Ecology and Evolution,*vol. 22, no. 1, pp. 11-16.

Palstra, F.P. & Ruzzante, D.E. 2008, "Genetic estimates of contemporary effective population size: What can they tell us about the importance of genetic stochasticity for wild population persistence?", *Molecular Ecology,*vol. 17, no. 15, pp. 3428-3447.

Peakall, R. & Smouse, P.E. 2012, "GenALEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update", *Bioinformatics,*vol. 28, no. 19, pp. 2537-2539.

Peakall, R. & Smouse, P.E. 2006, "GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research", *Molecular Ecology Notes,*vol. 6, no. 1, pp. 288-295.

Pearce, J.M., Fields, R.L. & Scribner, K.T. 1997, "Nest materials as a source of genetic data for avian ecological studies", *Journal of Field Ornithology,*vol. 68, no. 3, pp. 471-481.

Pellegrino, I., Cucco, M., Follestad, A. & Boos, M. 2015, "Lack of genetic structure in greylag goose (*Anser anser*) populations along the European Atlantic flyway", *PeerJ,*vol. 2015, no. 8. DOI 10.7717/peerj.1161

Pirkola, M.K. & Kalinainen, P. 1984a, "Use of neckbands in studying the movements and ecology of the bean goose *Anser fabalis*", *Annales Zoologici Fennici,*vol. 21, pp. 259-263.

Pirkola, M.K. & Kalinainen, P. 1984b, "The status, habitats and productivity of breeding populations of bean goose, *Anser fabalis fabalis*", *Swedish Wildlife Research,*vol. 13, pp. 9-48.

Pritchard, J.K., Seielstad, M.T., Perez-Lezaun, A. & Feldman, M.W. 1999, "Population growth of human Y chromosomes: A study of Y chromosome microsatellites", *Molecular Biology and Evolution,*vol. 16, no. 12, pp. 1791-1798.

Pritchard, J.K., Stephens, M. & Donnelly, P. 2000, "Inference of population structure using multilocus genotype data", *Genetics,*vol. 155, no. 2, pp. 945-959.

Pujolar, J.M., Dalén, L., Hansen, M.M. & Madsen, J. 2017, "Demographic inference from whole-genome and RAD sequencing data suggests alternating human impacts on goose populations since the last ice age", *Molecular Ecology,*vol. 26, no. 22, pp. 6270-6283.

Pusey, A. & Wolf, M. 1996, "Inbreeding avoidance in animals", *Trends in Ecology and Evolution,*vol. 11, no. 5, pp. 201-206.

R Core Team 2018, *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.* Available: https://www.R-project.org/.

Rice, W.R. 1989, "Analyzing tables of statistical tests", *Evolution,*vol. 43, no. 1, pp. 223-225.

Rodríguez-Ramilo, S.T. & Wang, J. 2012, "The effect of close relatives on unsupervised Bayesian clustering algorithms in population genetic structure analysis", *Molecular Ecology Resources,*vol. 12, no. 5, pp. 873-884.

Rohwer, F.C. & Anderson, M.G. 1988, "Female-biased philopatry, monogamy, and the timing of pair formation in migratory waterfowl" in *Current Ornithology, vol 5*, ed. R.F. Johnston, Springer, Boston, MA, pp. 187-221.

Rousset, F. 2008, "GENEPOP'007: A complete re-implementation of the GENEPOP software for Windows and Linux", *Molecular Ecology Resources,*vol. 8, no. 1, pp. 103-106.

Rozas, J., Ferrer-Mata, A., Sanchez-DelBarrio, J.C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S.E. & Sanchez-Gracia, A. 2017, "DnaSP 6: DNA sequence polymorphism analysis of large data sets", *Molecular Biology and Evolution,*vol. 34, no. 12, pp. 3299-3302.

Ruokonen, M., Aarvak, T. & Madsen, J. 2005, "Colonization history of the high-arctic pink-footed goose *Anser brachyrhynchus*", *Molecular Ecology,*vol. 14, no. 1, pp. 171-178.

Ruokonen, M. & Aarvak, T. 2011, "Typology revisited: Historical taxa of the bean goose - pink-footed goose complex", *Ardea,*vol. 99, no. 1, pp. 103-112.

Ruokonen, M., Kvist, L., Aarvak, T., Markkola, J., Morozov, V.V., Øien, I.J., Syroechkovsky Jr., E.E., Tolvanen, P. & Lumme, J. 2004, "Population genetic structure and conservation of the lesser white-fronted goose *Anser erythropus*", *Conservation Genetics,*vol. 5, no. 4, pp. 501-512.

Ruokonen, M., Kvist, L. & Lumme, J. 2000, "Close relatedness between mitochondrial DNA from seven *Anser* goose species", *Journal of Evolutionary Biology,*vol. 13, no. 3, pp. 532-540.

Ruokonen, M., Litvin, K. & Aarvak, T. 2008, "Taxonomy of the bean goose-pink-footed goose", *Molecular Phylogenetics and Evolution,*vol. 48, no. 2, pp. 554-562.

Sangster, G. & Oreel, G.J. 1996, "Progress in taxonomy of taiga and tundra bean geese", *Dutch Birding,*vol. 18, pp. 310-316.

Schwartz, M.K., Luikart, G. & Waples, R.S. 2007, "Genetic monitoring as a promising tool for conservation and management", *Trends in Ecology and Evolution,*vol. 22, no. 1, pp. 25-33.

Scott, D.A. & Rose, P.M. 1996, *Atlas of Anatidae populations in Africa and western Eurasia.*Wetlands International, Wageningen.

Segelbacher, G. 2002, "Noninvasive genetic analysis in birds: Testing reliability of feather samples", *Molecular Ecology Notes,*vol. 2, no. 3, pp. 367-369.

Smouse, P.E. & Peakall, R. 1999, "Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure", *Heredity,*vol. 82, no. 5, pp. 561-573.

Smouse, P.E., Peakall, R. & Gonzales, E. 2008, "A heterogeneity test for fine-scale genetic structure", *Molecular Ecology,*vol. 17, no. 14, pp. 3389-3400.

Sorenson, M.D. & Quinn, T.W. 1998, "Numts: A challenge for avian systematics and population biology", *Auk,*vol. 115, no. 1, pp. 214-221.

Sun, J.X., Helgason, A., Masson, G., Ebenesersdóttir, S.S., Li, H., Mallick, S., Gnerre, S., Patterson, N., Kong, A., Reich, D. & Stefansson, K. 2012, "A direct characterization of human mutation based on microsatellites", *Nature Genetics,*vol. 44, no. 10, pp. 1161-1165.

Taberlet, P. & Luikart, G. 1999, "Non-invasive genetic sampling and individual identification", *Biological Journal of the Linnean Society,*vol. 68, no. 1-2, pp. 41-55.

Taberlet, P., Luikart, G. & Waits, L.P. 1999, "Noninvasive genetic sampling: Look before you leap", *Trends in Ecology and Evolution,*vol. 14, no. 8, pp. 323-327.

Taylor, S.A. & Larson, E.L. 2019, "Insights from genomes into the evolutionary importance and prevalence of hybridization in nature", *Nature Ecology and Evolution,*vol. 3, no. 2, pp. 170-177.

Valière, N. 2002, "GIMLET: A computer program for analysing genetic individual identification data", *Molecular Ecology Notes,*vol. 2, no. 3, pp. 377-379.

Valkama, J., Vepsäläinen, V. & Lehikoinen, A. 2011, *The third Finnish breeding bird atlas. – Finnish Museum of Natural History and Ministry of Environment* [Homepage of Finnish Museum of Natural History and Ministry of Environment], [Online]. Available: http://atlas3.lintuatlas.fi/english [2019, 29.12.].

Van der Jeugd, H. P., Van der Veen, I. T. & Larsson, K. 2002, "Kin clustering in barnacle geese: Familiarity or phenotype matching?", *Behavioral Ecology,*vol. 13, no. 6, pp. 786-790.

Van Oosterhout, C., Hutchinson, W.F., Wills, D.P.M. & Shipley, P. 2004, "MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data", *Molecular Ecology Notes,*vol. 4, no. 3, pp. 535-538.

vonHoldt, B.M., Brzeski, K.E., Wilcove, D.S. & Rutledge, L.Y. 2018, "Redefining the role of admixture and genomics in species conservation", *Conservation Letters,*vol. 11, no. 2. DOI 10.1111/conl.12371

Wahlund, S. 1928, "Zusammensetzung von Populationen und Korrelationserscheinungen vom Standpunkt der Vererbungslehre aus betrachtet [Composition of populations and correlation phenomena from the point of view of heredity]", *Hereditas,*vol. 11, no. 1, pp. 65-106.

Wang, J. 2009, "A new method for estimating effective population sizes from a single sample of multilocus genotypes", *Molecular Ecology,*vol. 18, no. 10, pp. 2148-2164.

Wang, J. 2004, "Application of the one-migrant-per-generation rule to conservation and management", *Conservation Biology,*vol. 18, no. 2, pp. 332-343.

Waples, R.S. 2006, "A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci", *Conservation Genetics,*vol. 7, no. 2, pp. 167-184.

Waples, R.S. & Anderson, E.C. 2017, "Purging putative siblings from population genetic data sets: A cautionary view", *Molecular Ecology,*vol. 26, no. 5, pp. 1211-1224.

Waples, R.S. & Do, C. 2010, "Linkage disequilibrium estimates of contemporary Ne using highly variable genetic markers: A largely untapped resource for applied conservation and evolution", *Evolutionary Applications,*vol. 3, no. 3, pp. 244-262.

Weigmann, C. & Lamprecht, J. 1991, "Intraspecific nest parasitism in bar-headed geese, *Anser indicus*", *Animal Behaviour,*vol. 41, no. 4, pp. 677-688.

Wilson, R.E., Ely, C.R. & Talbot, S.L. 2018, "Flyway structure in the circumpolar greater white-fronted goose", *Ecology and Evolution,*vol. 8, no. 16. DOI 10.1002/ece3.4345

Wilson, R.E., Sonsthagen, S.A., DaCosta, J.M., Ely, C.R., Sorenson, M.D. & Talbot, S.L. 2019, *Identification of single nucleotide polymorphisms for use in a genetic stock identification system for greater white-fronted goose (Anser albifrons) subspecies wintering in California*, US Geological Survey. Reston, VA

Vucetich, J.A. & Waite, T.A. 2000, "Is one migrant per generation sufficient for the genetic management of fluctuating populations?", *Animal Conservation,*vol. 3, no. 3, pp. 261-266.

Zicus, M.C. 1981, "Canada goose brood behavior and survival estimates at Crex Meadows, Wisconsin", *Wilson Bulletin,*vol. 93, pp. 207-217.

Zhu, Q., Damba, I., Zhao, Q., Yi, K., Batbayar, N., Natsagdorj, T., Davaasuren, B., Wang, X., Rozenfeld, S., Moriguchi, S., Zhan, A., Cao, L. & Fox, A.D. 2020, "Lack of conspicuous sex-biased dispersal patterns at different spatial scales in an Asian endemic goose species breeding in unpredictable steppe wetlands", *Ecology and Evolution,*vol. 10, no. 14, pp. 7006-7020.

Table 1. Summary statistics for the 376 non-kin taiga bean geese (*Anser fabalis fabalis*) genotyped by 22 microsatellite loci and grouped by four geographical regions (see Fig. 2). Sample size (*n*), total number of alleles (*A*), allelic richness (*A*R), number of private alleles (*PA*), observed heterozygosity (*H*O), expected heterozygosity (*H*E), inbreeding coefficient (*F*IS) and *P*-value from test for deviation from the Hardy-Weinberg equilibrium (*P*HWE). Statistically significant *P*-values after Bonferroni correction shown in bold. Standard errors are in parentheses.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Region** | ***n*** | ***A*** | ***A*R** | ***PA*** | ***H*O** | ***H*E** | ***F*IS** | ***P*HWE** |
| Western Finland | 50 | 171 | 7.5 | 3 | 0.64 (0.03) | 0.69 (0.03) | 0.07 (0.03) | **0.00** |
| Eastern Finland/Kainuu | 140 | 194 | 7.6 | 9 | 0.64 (0.03) | 0.70 (0.03) | 0.07 (0.02) | **0.00** |
| Northern Ostrobothnia/ Southern Lapland | 87 | 183 | 7.6 | 4 | 0.66 (0.04) | 0.69 (0.03) | 0.05 (0.02) | **0.00** |
| Lapland | 99 | 186 | 7.4 | 7 | 0.62 (0.04) | 0.69 (0.03) | 0.10 (0.02) | **0.00** |
| Mean over regions |  | 184 | 7.5 | 6 | 0.64 (0.02) | 0.69 (0.02) | 0.07 (0.01) | **0.00** |

Table 2. Pairwise *F*ST- and *ɸ*ST-values among Finnish taiga bean goose (*Anser fabalis fabalis*). Pairwise *F*ST-values for microsatellite data below the diagonal among 488 Finnish taiga bean goose grouped by geographical regions (see Fig. 2) with non-kin individuals in parenthesis (*n* = 375). Pairwise *ɸ*ST-values for mitochondrial data (*n* = 447) above diagonal with non-kin individuals in parenthesis (*n* = 343). Statistically significant values after Bonferroni correction indicated with an asterisk.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Western Finland | Eastern Finland/ Kainuu | Northern Ostrobothnia/ Southern Lapland | Lapland |
| Western Finland | - | 0.0228  (0.0004) | 0.1977\*  (0.1481\*) | 0.0045  (-0.0047) |
| Eastern Finland/ Kainuu | 0.0032 (0.0024) | - | 0.0779\*  (0.0855\*) | 0.0000  (-0.0081) |
| Northern Ostrobothnia/ Southern Lapland | 0.0046 (0.0028) | 0.0016  (0.0004) | - | 0.1107\*  (0.0951\*) |
| Lapland | 0.0042  (0.0022) | 0.0026  (0.0008) | 0.0025  (0.0007) | - |

Table 3. Genetic variability in the hypervariable portion of the mitochondrial control region domain I (210 base pairs) among non-kin Finnish taiga bean geese (*Anser fabalis fabalis*, *n* = 342). Number of samples (*n*), number of haplotypes (*H*), haplotype diversity (*h*), nucleotide diversity (*π*), Tajima’s D (*D*) and Fu’s Fs (*Fs*). NS: non-significant values. The different geographical regions are mapped in Figure 2.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Region** | ***n*** | ***H*** | ***h* (SD)** | ***π* (SD**) | ***D*** | ***Fs*** |
| Western Finland | 47 | 3 | 0.399 (0.069) | 0.0021 (0.0004) | -0.087 (NS) | 0.038 (NS) |
| Eastern Finland/Kainuu | 122 | 3 | 0.464 (0.036) | 0.0039 (0.0009) | -0.828 (NS) | 2.243 (NS) |
| Northern Ostrobothnia/ Southern Lapland | 84 | 6 | 0.577 (0.035) | 0.0048 (0.0011) | -0.934 (NS) | -0.527 (NS) |
| Lapland | 90 | 5 | 0.499 (0.042) | 0.0037 (0.0009) | -1.433 (NS) | -0.342 (NS) |
| Finland | 343 | 7 | 0.519 (0.018) | 0.0040 (0.0005) | -0.864 (NS) | -0.789 (NS) |

Table 4. Pairwise *ɸ*ST-values among Finnish taiga bean goose (*Anser fabalis fabalis*) for mitochondrial data. Pairwise *ɸ*ST-values for females below diagonal (*n* = 218) and for males above diagonal (*n =* 174). Statistically significant values after Bonferroni correction indicated with an asterisk.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Western Finland | Eastern Finland/ Kainuu | Northern Ostrobothnia/ Southern Lapland | Lapland |
| Western Finland | - | -0.0005 | 0.1267\* | -0.0241 |
| Eastern Finland/ Kainuu | 0.0293 | - | 0.0462\* | -0.0065 |
| Northern Ostrobothnia/ Southern Lapland | 0.2689\* | 0.08768\* | - | 0.0753\* |
| Lapland | -0.0079 | -0.0048 | 0.1664\* | - |

**Figure legends**

**Fig. 1** Breeding and wintering distributions of the taiga (*Anser fabalis fabalis*) and tundra (*A. f. rossicus*) bean goose shown with different colours and autumn migration routes shown with coloured arrows. The flyway management units used in the International Single Species Action Plan (Marjakangas *et al*. 2015) for the taiga bean goose are also shown with dashed ellipses. Map redrawn from BirdLife International (2018) and Marjakangas *et al*. (2015).

**Fig. 2** The breeding distribution of taiga bean goose (*Anser fabalis fabalis*) in Finland and the approximate location of sampling sites of the current study. The exact sampling sites are not shown in order to protect the breeding locations and the sites closer than 16 km were merged. Samples are divided here into four geographical regions shown with different symbols and colours. Samples of tundra bean goose (*A. f. rossicus*) mtDNA are indicated with a star and one dead goose that was not ascertained as breeding, is indicated with a black dot. The sampling sites of outgroups are also shown. The background breeding distribution map was created by interpolating the breeding category indexes (1 = unlikely breeding, 2 = possible breeding, 3 = probable breeding and 4 = confirmed breeding) of the Finnish Breeding Bird Atlas using the Inverse Distance Weighted method in ArcGIS-software. The breeding index data are from: Results of the 3rd Finnish bird atlas. Finnish Museum of Natural History, University of Helsinki Luomus. Used with Creative Commons Attribution 4.0 license.

**Fig. 3** Structure assignment plots for *K* = 4 using the LOCPRIOR option for non-kin Finnish breeding taiga bean geese (*Anser anser fabalis*) (*n* = 376). Sampling locations within 16 km from each other were combined into single locations for the LOCPRIOR option. Each vertical bar represents one individual.

**Fig. 4** DAPC (discriminant analysis of principal components) for Finnish breeding taiga bean geese (*Anser fabalis fabalis*) and outgroups (Russian and Swedish taiga bean goose, Norwegian tundra bean goose (*A. f. rossicus*) and Icelandic pink-footed goose (*A. brachyrhynchus*)). a) DAPC for Finnish breeding taiga bean geese with an equalized number of samples per geographic region (*n* = 75 per region, chosen randomly; as mapped in Fig. 2). Number of PCs = 100. b) DAPC for outgroups: Russian and Swedish taiga bean goose, Norwegian tundra bean goose and Icelandic pink-footed goose. The Finnish breeding taiga bean goose samples (Finland) were fitted to this DAPC analysis as ‘unknown’ samples in order to identify the clustering of the Finnish samples without assuming any prior genetic group. Number of PCs = 12. Inertia ellipses represent graphical summaries of a cloud of points.

**Fig. 5** Correlogram from spatial autocorrelation analysis for a) non-related taiga bean geese (*Anser fabalis fabalis; n* = 376) and b) females and males separately (*n =* 328). The autocorrelation coefficient (*r*) was plotted against the function of distance class (50 km). 95% confidence interval (dashed lines, U = upper limit, L = lower limit) was determined by 999 permutations with the null hypothesis of no population structure and 95% error bars were determined by bootstrap resampling of 1000 replicates.

**Fig. 6** Structure plot with *K* = 2 for: a) simulated taiga bean goose (*Anser fabalis fabalis*), simulated pink-footed goose (*A. brachyrhynchus*) and bean geese breeding in Finland; b) simulated taiga bean goose, simulated tundra bean goose (*A. f. rossicus*) and bean geese breeding in Finland. Two tundra bean goose individuals identified using mitochondrial DNA indicated in the Finnish dataset with an arrow. These appear as tundra bean geese based on their microsatellites as well.

**Fig. 7** Median joining haplotype network for the Finnish breeding bean geese, different bean goose subspecies (*A. f. fabalis*, *A. f. rossicus*, *A. f. serrirostris* and *A. f. middendorffii*) and pink-footed goose (*A. brachyrhynchus*). A mtDNA sequence of greylag goose (*A. anser*) was used to root the haplotype network. The sizes of the circles are proportional to the frequency of each haplotype and tick marks across branches indicate the number of mutational differences. Forward slashes between haplotype names denote identical haplotypes based on the sequenced fragment but differing based on the whole control region sequence.

**Fig. 8** Mitochondrial haplotypes of all the Finnish bean geese (*Anser fabalis*), only males or females and outgroup samples (indicated with an asterisk) showed on a map as pie charts. a) All Finnish taiga bean geese and outgroups. The outgroup in south-eastern Finland and Estonia consists of Russian taiga bean geese (*A. f. fabalis*), hunted along their migration route. The outgroup in Sweden consists of breeding or moulting taiga bean geese, the outgroup in Norway consists of breeding or moulting tundra bean geese (*A. f. rossicus*) and the outgroup in Iceland consists of museum feathers of pink-footed goose (*A. brachyrhynchus*). b) Male taiga bean geese. c) Female taiga bean geese. The size of the circles corresponds to the frequency of the haplotype. Close sample sites were merged for better visualization of the data.