Heredity (2020) 125:73–84
https://doi.org/10.1038/s41437-020-0322-z

the geneticssociety

ARTICLE

Recent introgression between Taiga Bean Goose and Tundra Bean Goose results in a largely homogeneous landscape of genetic differentiation

Jente Ottenburghs[1] · Johanna Honka[2] · Gerard J. D. M. Müskens[3] · Hans Ellegren[1]

Received: 12 December 2019 / Revised: 11 May 2020 / Accepted: 12 May 2020 / Published online: 26 May 2020

Abstract

Several studies have uncovered a highly heterogeneous landscape of genetic differentiation across the genomes of closely related species. Specifically, genetic differentiation is often concentrated in particular genomic regions ("islands of differentiation") that might contain barrier loci contributing to reproductive isolation, whereas the rest of the genome is homogenized by introgression. Alternatively, linked selection can produce differentiation islands in allopatry without introgression. We explored the influence of introgression on the landscape of genetic differentiation in two hybridizing goose taxa: the Taiga Bean Goose (*Anser fabalis*) and the Tundra Bean Goose (*A. serrirostris*). We re-sequenced the whole genomes of 18 individuals (9 of each taxon) and, using a combination of population genomic summary statistics and demographic modeling, we reconstructed the evolutionary history of these birds. Next, we quantified the impact of introgression on the build-up and maintenance of genetic differentiation. We found evidence for a scenario of allopatric divergence (about 2.5 million years ago) followed by recent secondary contact (about 60,000 years ago). Subsequent introgression events led to high levels of gene flow, mainly from the Tundra Bean Goose into the Taiga Bean Goose. This scenario resulted in a largely undifferentiated genomic landscape (genome-wide $F_{ST} = 0.033$) with a few notable differentiation peaks that were scattered across chromosomes. The summary statistics indicated that some peaks might contain barrier loci while others arose in allopatry through linked selection. Finally, based on the low genetic differentiation, considerable morphological variation and incomplete reproductive isolation, we argue that the Taiga and the Tundra Bean Goose should be treated as subspecies.

Associate Editor: Giorgio Bertorelle

Supplementary information The online version of this article (https://doi.org/10.1038/s41437-020-0322-z) contains supplementary material, which is available to authorized users.

✉ Jente Ottenburghs
jente.ottenburghs@hotmail.com

[1] Department of Evolutionary Biology, University of Uppsala, Uppsala, Sweden

[2] Department of Ecology and Genetics, University of Oulu, PO Box 3000, FI-90014 Oulu, Finland

[3] Team Animal Ecology, Wageningen Environmental Research, Wageningen University & Research, Droevendaalsesteeg 3-3A, 6708 PB Wageningen, The Netherlands

Introduction

It is increasingly appreciated that interspecific gene flow, or introgression, is a common phenomenon. Numerous species have exchanged genetic material with other species through introgressive hybridization (Barlow et al. 2018; Palkopoulou et al. 2018; Árnason et al. 2018; Wu et al. 2018; Gopalakrishnan et al. 2018), including our own species, *Homo sapiens* (Patterson et al. 2012; Vernot et al. 2016; Villanea and Schraiber 2018). This widespread genetic exchange has changed our views on the evolutionary process and the nature of species (Mallet et al. 2016; Shapiro et al. 2016; Roux et al. 2016).

A number of studies have revealed a highly heterogeneous landscape of genetic differentiation across the genomes of closely related species (Turner et al. 2005; Nadeau et al. 2012; Ellegren et al. 2012; Renaut et al. 2013). Genetic differentiation (measured for example by F_{ST}, the fixation index) between species pairs is often

concentrated in a few genomic regions, the so-called islands of differentiation (Wolf and Ellegren 2017). This finding led to the formulation of a verbal model in which such islands diverge over time (i.e. higher absolute divergence, d_{XY}) because they contain loci involved in reproductive isolation (and hence originally referred to as "genomic islands of speciation", Turner et al. 2005), whereas the rest of the genome is homogenized by interspecific gene flow (Wu 2001; Turner et al. 2005; Feder et al. 2012). This leads to small genomic regions of high divergence against a background of low divergence.

Rigorous tests of islands of differentiation have revealed that reduced diversity due to linked selection can also lead to heterogeneous genomic landscapes (Cruickshank and Hahn 2014; Wolf and Ellegren 2017). This is thought to arise from two processes: genetic hitchhiking or background selection (Cutter and Payseur 2013; Burri 2017; Rettelbach et al. 2019; Stankowski et al. 2019; Buffalo and Coop 2019). Genetic hitchhiking refers to the situation in which positive selection on a variant results in selection for the genetic region in which this advantageous variant occurs. As the advantageous variant goes toward fixation, loci linked to this variant hitchhike along and increase in frequency (Smith and Haigh 1974). Background selection involves purifying selection against recurring deleterious mutations (Charlesworth 1994). This process also reduces diversity at linked sites. Genomic regions with high levels of recombination are expected to experience less linked selection because recombination uncouples loci from the advantageous or deleterious variant under selection (Hudson and Kaplan 1995; Nordborg et al. 1996). These processes–genetic hitchhiking and background selection—can produce islands of differentiation in allopatry in the absence of gene flow.

The True Geese (genera *Anser* and *Branta*) are an excellent system to explore the consequences of introgressive hybridization on a genomic level (Ottenburghs et al. 2016a). Previous work has uncovered introgression between several goose species (Ottenburghs et al. 2016b, 2017a), but it remains to be determined when these introgression events occurred and how these species remain distinct in the face of gene flow. In this study, we focus on two Bean Goose taxa: the Taiga Bean Goose (*Anser fabalis*) and the Tundra Bean Goose (*A. serrirostris*). These taxa belong to the Bean Goose complex (which also includes the Pink-footed Goose, *A. brachyrhynchus*) and have been considered conspecific based on morphology (Delacour 1951; Sangster and Oreel 1996; Mooij and Zöckler 1999) and mitochondrial DNA (Ruokonen et al. 2008). Genomic analyses have indicated that divergence within the Bean Goose complex occurred ~2 million years ago (Ottenburghs et al. 2016b). Moreover, ecological evidence suggests that the Taiga and the Tundra Bean Goose might be distinct species since they use different breeding grounds (Burgers et al. 1991) and show differences in behavior and vocalizations (Sangster and Oreel 1996). Also, slight differences in morphology exist between the taxa in body size, shape, plumage patterns and in beak morphology and coloration: the Taiga Bean Goose has a longer beak with a broad orange marking whereas the Tundra Bean Goose has a shorter beak with a reduced orange band on the bill. However, a recent study showed that only two measurements out of total of 17 distinguished the Taiga and the Tundra Bean Goose from each other (de Jong 2019), thus considerable interspecific overlap exists. Hybrids between taxa of the Bean Goose complex have been reported (Ottenburghs et al. 2016a; Honka et al. 2017), mainly based on genetic tests because hybrids are difficult to identify due to morphological similarities with both parental species (Randler 2004). Moreover, most hybrids were reported during migration and on the wintering grounds, so it is currently not possible to pinpoint a putative hybrid zone on their breeding areas. Whether the hybrids are fertile and backcross with the parental species—and thus resulting in introgression–remains to be investigated.

In this study, we explore the evolutionary history of the Taiga and the Tundra Bean Goose using whole-genome resequencing data (on average 37× coverage with paired-end sequencing). We investigate (1) how genetic differentiation is distributed across the genome and (2) how the timing of introgression influences the structure of the genomic landscape of differentiation. We address these questions through a combination of population genomic summary statistics, including relative divergence (F_{ST}), absolute divergence (d_{XY}), nucleotide diversity (π) and Tajima's D. We also apply demographic modeling. Finally, we assess the taxonomic status of the Taiga and the Tundra Bean Goose, which has been heavily debated, by combining the genetic results with morphological and ecological information. Moreover, the Taiga Bean Goose is declining: population numbers have halved since the 1990s, but the Taiga Bean Goose is still being hunted. Current population size estimates are 53,000–57,000 individuals for the Taiga Bean Goose and 600,000 individuals for the Tundra Bean Goose (Fox and Leafloor 2018). Thus, verifying the taxonomical position of the Taiga and Tundra Bean Goose is of utmost importance for the correct management of the taxa.

Material and methods

Sequencing and quality assessment

We collected blood and tissue samples for the Taiga Bean Goose (*A. fabalis*, $n = 9$) and the Tundra Bean Goose (*A. serrirostris*, $n = 9$), migrating within Europe (Fig. 1a,

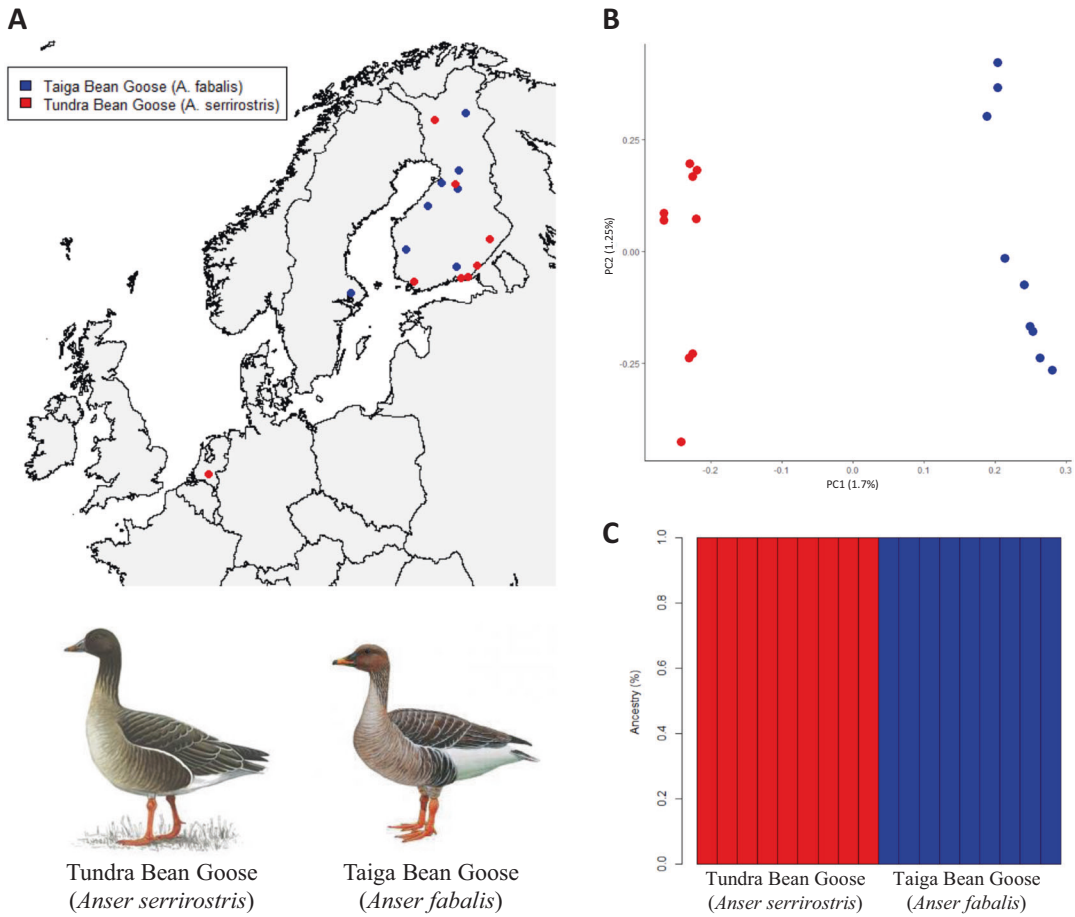


Fig. 1 Population genetic structure of Taiga and Tundra Bean Goose. a Map of sampling locations of migrating Taiga and Tundra Bean Goose. **b** Principal component analysis (PCA) and **c** *ADMIXTURE*-analyses show clear genetic differentiation between Taiga and Tundra Bean Goose. Drawings used with permission of the Handbook of the Birds of the World (del Hoyo et al. 2018).

Supplementary Table S1). Due to elusive nature of the species, especially during the breeding time, and remote breeding sites (mires and tundra-like habitats), the samples were collected from legally hunted geese during their migration. The tissue samples were collected in years 2010–2013 and stored frozen in absolute ethanol. Genomic DNA was isolated from the blood and tissue samples using the Qiagen Gentra kit (Qiagen Inc.). Quality and quantity of the DNA was measured using the Qubit (Invitrogen, Life Technologies).

Sequencing libraries were prepared from 100 ng DNA using the TruSeq Nano DNA sample preparation kit (cat# FC-121-4001/4002, Illumina Inc.), targeting an insert size of 350 bp and a target coverage of 30×. Whole-genome paired-end sequencing (150 bp) was performed on an Illumina HiSeqX following standard procedures. Sequencing reads were mapped to the reference genome of a closely related goose species with the highest quality, namely Swan Goose (*Anser cygnoides*) genome version 1.0 (Gao et al. 2016), using Burrows–Wheeler Aligner (BWA) version 0.7.17 (Li and Durbin 2009). The resulting BAM-files were sorted with Samtools version 1.6 (Li et al. 2009) and duplicates were marked with Picard version 2.10.3 (http://broadinstitute.github.io/picard/). Next, local realignment was performed using GATK version 3.7 (McKenna et al. 2010).

For each individual, a first round of variant calling was performed with GATK HaplotypeCaller. The resulting list of variants was filtered on mapping quality (MQRankSum < 0.22) and read depth (DP > 10). The variants passing these filters were then used as a reference set for base quality score recalibration (BQSR) following a bootstrapping approach in GATK (following Kardos et al. 2018). Next, we applied a hard filter in line with the GATK best practices pipeline (Van der Auwera et al. 2013), applying the following filtering criteria: QD < 2.0 | | FS > 60.0 | | MQ < 40.0 | | MQRankSum < −12.5 | | ReadPosRankSum < −8.0. The final dataset contained 13,890,330 SNPs. Different filtering steps were applied in the consequent analyses.

Population structure and differentiation

Using VCFtools version 0.1.15 (Danecek et al. 2011), we removed loci for which the *p*-value was smaller than 0.01 in

a test for excess of heterozygotes relative to Hardy–Weinberg genotype proportions. Moreover, we retained only loci with a minor allele frequency ≥ 0.05. Finally, the SNPs were filtered on linkage disequilibrium along windows of 50 markers with a R^2-threshold of 0.5. The resulting dataset of 6,221,883 SNPs provided the input for the principal component analysis (PCA) using the pca-function in Plink version 1.07 (Purcell et al. 2007). These analyses were repeated with different settings for the Hardy–Weinberg test and linkage disequilibrium to assess the robustness of the patterns.

The same dataset of 6,221,883 SNPs was used to assess the ancestry composition of each individual in *ADMIXTURE* version 1.3.0 (Alexander et al. 2009). All SNPs were formatted for the *ADMIXTURE*-analyses (i.e. converted to BED-format) using Plink version 1.07 (Purcell et al. 2007). We ran analyses with the number of clusters set from $K = 1$ to 4, and performed 10-fold cross-validation to assess the optimal number of clusters. The final admixture proportions per individual (Q-estimates representing the log-likelihood of cluster assignment) were visualized with R version 3.5.0 (R Core Team 2018).

The filtered dataset of 13,890,330 SNPs was used to construct the genomic landscape of differentiation. Summary statistics were calculated across non-overlapping windows of 200,000 nucleotides (200 kb). To assess the genome-wide heterogeneity in genetic differentiation, we calculated relative divergence (F_{ST}). However, this statistic is a relative measure of differentiation that is dependent on the underlying genetic diversity within the population (Ottenburghs et al. 2017b; Wolf and Ellegren 2017). Therefore, we also estimated absolute divergence (d_{XY}) and nucleotide diversity (π) to rule out any effects of local reductions in genetic diversity on patterns of genetic differentiation. Finally, to infer whether these regions of reduced genetic diversity are the result of (linked) selection, we calculated Tajima's D. Negative values of this statistic suggest purifying selection or population expansion (Tajima 1989). Moreover, divergent selection is expected result in higher absolute divergence (d_{XY}) and lower nucleotide diversity (π) in particular genomic regions. Hence, we correlated F_{ST} with d_{XY} and π. Relative divergence (F_{ST}) and Tajima's D were calculated using VCFtools version 0.1.15 (Danecek et al. 2011), whereas absolute divergence (d_{XY}) and nucleotide diversity (π) were calculated with the popgenWindows.py script from Martin et al. (2015) which is available here: https://github.com/simonhmartin/genomics_general. The analyses were repeated for different window sizes (10, 20, 50 and 100 kb) to rule out any effects of window size.

Because the Swan Goose genome has not been assembled on a chromosome level, we aligned scaffolds to the highest quality bird genome currently available, namely the Chicken (*Gallus gallus*) genome assembly Galgal6 (Hillier et al. 2004), with LASTZ version 1.04.00 (Harris 2007). The scaffolds were ordered and orientated based on the coordinates from the Chicken genome and consequently merged into pseudo-chromosomes. The resulting alignment was visualized with R version 3.5.0 (R Core Team 2018) using the package ggplot2 (Wickham 2016).

Demographic analyses

Demographic inference was performed using the software package DADI (Gutenkunst et al. 2009). Because demographic analyses can be biased by selection (Ragsdale et al. 2018), we only used non-coding loci (5,397,934 SNPs). These loci were selected using snpEff version 4.3T (Cingolani et al. 2012), which annotates SNPs into several functional classes, such as protein-coding, intronic, and intergenic regions. Due to the lack of an outgroup to establish the ancestral state for each SNP, we used a folded frequency spectrum. We tested several demographic models with increasing complexity to estimate the timing of gene flow between the Taiga and the Tundra Bean Goose, ranging from strict isolation to secondary contact with asymmetrical gene flow. For each scenario, ten simulations were run with different starting values to ensure proper exploration of the likelihood landscape. After convergence of parameters, the simulation with the highest likelihood was retained. The final set of parameters was converted into absolute time and population size estimates using a mutation rate of 1×10^{-9} per nucleotide per generation (Pujolar et al. 2018) and a generation time of two years (Ottenburghs et al. 2017a). Confidence intervals for parameters were generated using a bootstrap approach (10 iterations) in which 1 million SNPs were randomly selected and a demographic model was tested with DADI using the parameter values from the most likely model as a starting point.

Results

Sequencing and quality assessment

We re-sequenced the genomes of nine Taiga Bean Geese and nine Tundra Bean Geese. All 18 samples were mapped to the Swan Goose genome (Supplementary Table S2), with an average mapping percentage of 92.6% (range: 83.7–97.6) and an average sequencing depth of 37× (range: 31–44). SNP calling, following the GATK best practices guidelines (Material and methods), resulted in a final dataset of 13,890,330 SNPs.

Population structure and differentiation

The PCAs indicated that the Taiga and the Tundra Bean Goose can be separated using genomic data. The first

principal component discriminated between the two taxa and the second principal component indicated some intraspecific population structure within both taxa (Fig. 1b). This intraspecific population structure might relate to the distribution of breeding areas, but unfortunately we do not have information about sites of origin because the birds were sampled during migration. The principal components explained little genetic variance, suggesting that only a subset of genetic loci drive the genetic differences between the taxa. However, the PCA-patterns were robust to different filtering settings (Supplementary Fig. S1). In contrast, the individual ancestries estimated by *ADMIXTURE* pointed to one population ($K = 1$ had the lowest CV-error, Supplementary Fig. S2) although the analyses with $K = 2$ confidently discriminated between two genetically distinct populations under particular filtering criteria (Fig. 1c). Moreover, relaxing the thresholds for linkage disequilibrium and minor allele frequency in filtering the SNPs highlighted a more admixed pattern (Supplementary Fig. S1), suggesting that a large proportion of genetic variation is shared between the taxa. This observation is confirmed by the genomic window analyses which show that genetic divergence was concentrated in a small number of differentiated loci. The majority of genomic windows showed low levels of F_{ST} (genome-wide $F_{ST} = 0.033$, Fig. 2a) and intermediate values of d_{XY} (Fig. 2b) and π (Fig. 2c). Most genomic windows showed a negative value for Tajima's D (Fig. 2d), which can be due to purifying selection or population expansion (Tajima 1989). High F_{ST}-windows were characterized by slightly higher levels of absolute divergence (d_{XY}, Spearman correlation, $\rho = 0.14$, $p < 0.01$, Fig. 2e) and lower levels of nucleotide diversity (Spearman correlation, $\rho = -0.16$, $p < 0.01$, Fig. 2f). These results were robust against different window sizes (Supplementary Table S3).

The results from Fig. 2 were visualized in the genomic landscape of differentiation (Fig. 3, Supplementary Fig. S2). The F_{ST}-landscape was largely flat with a few notable peaks that were scattered across chromosomes (82 F_{ST}-windows above 0.25). Peaks in F_{ST} were often accompanied by lower levels of d_{XY} and a drop in nucleotide diversity in one or both taxa (e.g., highlighted regions on chromosomes 1, 2 and 3 in Fig. 3). However, in some cases, a peak in F_{ST} corresponded to an increase in d_{XY} (e.g., highlighted regions

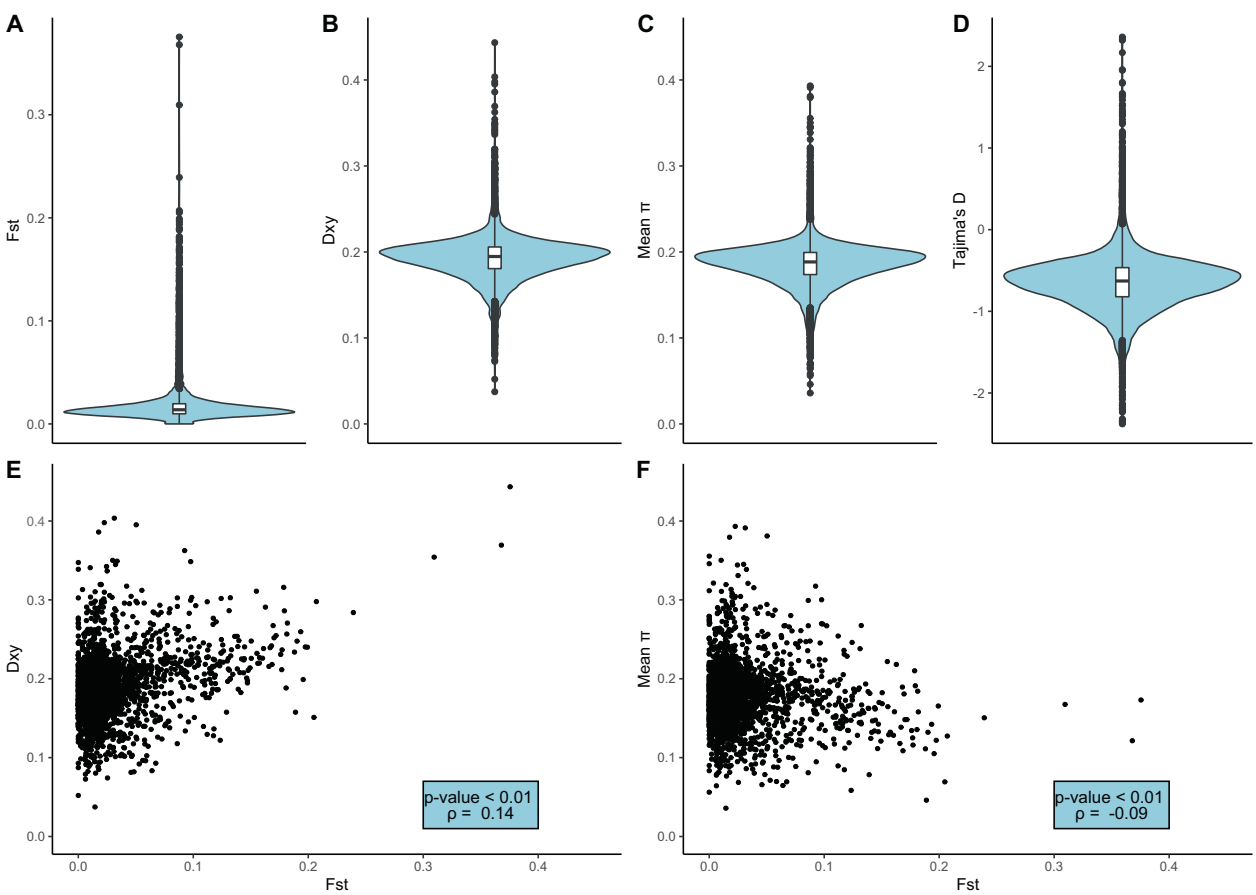


Fig. 2 Distribution of summary statistics. a Relative divergence (F_{ST}), **b** absolute divergence (d_{XY}), **c** nucleotide diversity (π) and **d** Tajima's D. Correlations between **e** F_{ST} and d_{XY} and between **f** F_{ST} and nucleotide diversity.

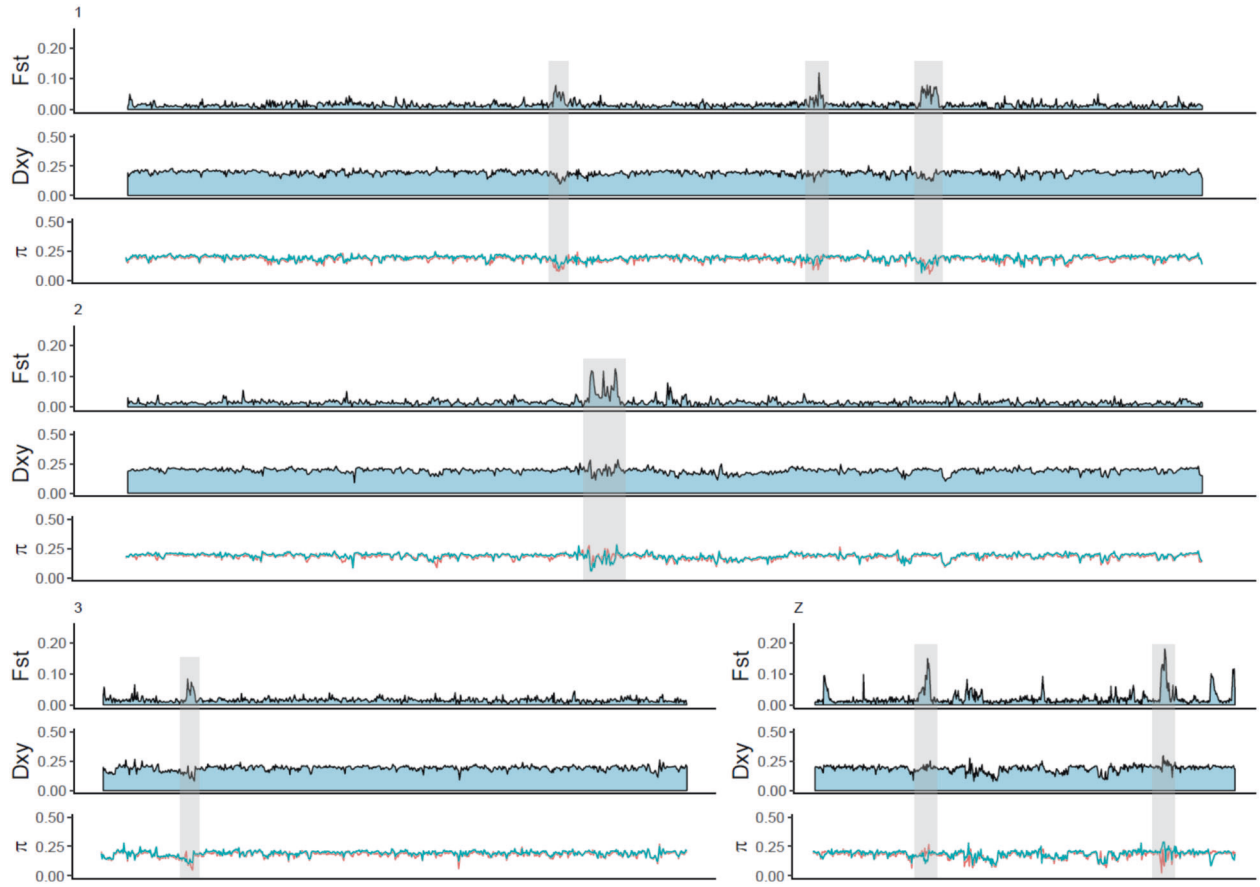


Fig. 3 The genomic landscape of Taiga and Tundra Bean Goose, with sequences aligned to chicken chromosomes, for relative divergence (F_{ST}), absolute divergence (d_{XY}) and nucleotide diversity (π). The colors in the nucleotide diversity tracks correspond to the Taiga (blue) and the Tundra Bean Goose (red). Only the first three chromosomes and the Z-chromosome are depicted, a complete picture is provided in Supplementary Fig. S3. The highlighted areas (gray boxes) show examples of differentiated islands.

on the Z-chromosome in Fig. 3). Although there were a few notable F_{ST}-peaks on the Z-chromosome (Fig. 3), the mean F_{ST} between windows on the autosomes and the Z-chromosome was not significantly different (two-sample t-test, $t = 0.49$, $p = 0.62$).

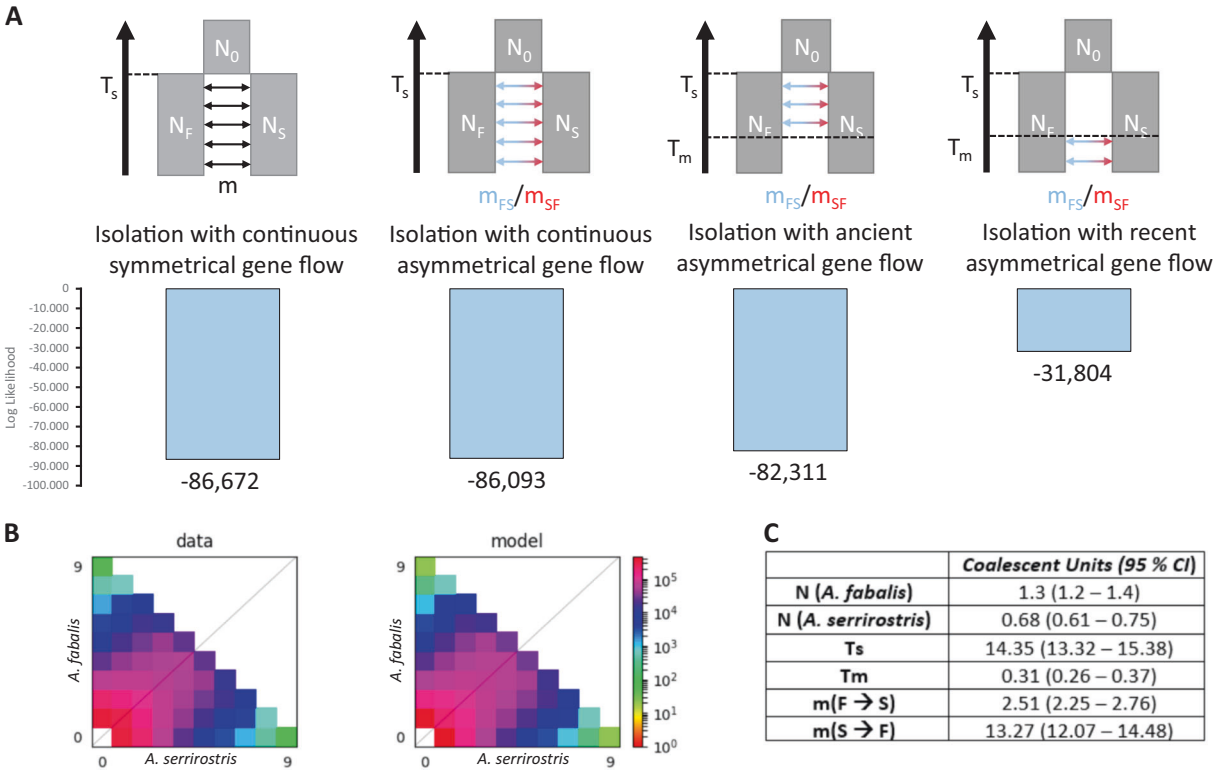
Demographic analyses

Demographic modeling indicated that a model of strict isolation was highly unlikely (log-likelihood = −472,672). The inclusion of gene flow markedly improved the likelihood estimation, as exemplified by the log-likelihood of a model with continuous, symmetrical gene flow was −86,672. Exploration of more sophisticated models with asymmetrical gene flow indicated that the most likely model (log-likelihood = −31,804) entails a scenario of secondary contact with gene flow mainly from the Tundra into the Taiga Bean Goose (Fig. 4a, b, Supplementary Table S4). Including population expansions for one of both taxa, did not improve the likelihood scores (Supplementary Fig. S4, Supplementary Table S4). Transforming the coalescent units (Fig. 4c, Supplementary Table S5) into absolute time showed that the taxa diverged ~2.66 million years ago (95% CI: 2.47–2.81 million years) and that secondary contact occurred around 58,285 years ago (95% CI: 48,658–67,918 years). Effective population sizes after the initial split were 102,508 (95% CI: 110,954–130,061) and 62,855 (95% CI: 56,102–69,608) for the Taiga Bean Goose and the Tundra Bean Goose, respectively.

Discussion

The evolutionary history of the Bean Geese

Our genomic analyses indicated that the Taiga Bean Goose and the Tundra Bean Goose can be genetically separated despite overlapping values in most morphological traits and gene flow (Fig. 1). Moreover, the demographic modeling revealed that the taxa diverged ca. 2.66 million years ago (Fig. 4), in line with previous estimates (Ruokonen et al. 2000; Ottenburghs et al. 2016b). This divergence time



	Coalescent Units (95 % CI)
N (*A. fabalis*)	1.3 (1.2 – 1.4)
N (*A. serrirostris*)	0.68 (0.61 – 0.75)
Ts	14.35 (13.32 – 15.38)
Tm	0.31 (0.26 – 0.37)
m(F → S)	2.51 (2.25 – 2.76)
m(S → F)	13.27 (12.07 – 14.48)

Fig. 4 Demographic analyses point to recent asymmetrical gene flow. a Four different demographic scenarios and their likelihood scores that were tested with DADI. The most likely model concerns isolation with recent asymmetrical gene flow (log-likelihood=−31,804). **b** Comparison between the actual data and the simulated model and **c** the parameters (in 2N-units) estimated in the most likely demographic model.

coincides with a fast global cooling trend that resulted in a circumpolar tundra belt and expansion of temperate grasslands (Zachos et al. 2001), the ideal habitats for geese to thrive (Owen 1980). After a period of allopatry, the Taiga and the Tundra Bean Goose established secondary contact about 60,000 years ago which culminated in bidirectional gene flow, though mostly from the Tundra into the Taiga Bean Goose.

The period of introgression occurred during the Weichselian Glaciation (between 75,000 and 11,000 years ago) when a cooling trend introduced tundra vegetation in the Northern hemisphere (Mangerud et al. 2011; Otvos 2015). During this period, the geese probably resided in different refugia: the Taiga Bean Goose was driven to southwestern Europe (specifically Spain) whereas the Tundra Bean Goose occurred on the tundra in western Siberia (Ploeger 1968). The warm interstadials during the Weichselian cooling period might have brought these populations in secondary contact. On the basis of the current distributions, we can assume that the Taiga Bean Goose moved northwards into the range of the Tundra Bean Goose. Initially, the moving Taiga Bean Goose might have been outnumbered by the Tundra Bean Goose in certain areas, leading to hybridization. As the range shift proceeded, the Tundra Bean Goose and previously produced hybrids were probably incorporated into the Taiga Bean Goose population, thereby overturning the numerical imbalance. Consequently, hybrids might have had a higher chance of backcrossing with the Taiga Bean Goose, resulting in the observed pattern of asymmetric gene flow from Tundra into Taiga Bean Goose (Currat et al. 2008). These findings support the widespread occurrence of introgressive hybridization between bird species in general (Rheindt and Edwards 2011; Ottenburghs et al. 2017b), and geese in particular (Ottenburghs et al. 2017a).

Islands of differentiation

Although it is possible to discriminate between the Taiga and the Tundra Bean Goose using genetic data (Fig. 1), it does not automatically follow that the taxa are genetically distinct. Indeed, PCAs tend to overemphasize differences (Björklund 2019) and *ADMIXTURE*-analyses are sensitive to filtering criteria applied to the SNPs (Lawson et al. 2018). These biases were also apparent in our analyses. Regardless of the filtering thresholds, PCAs clearly discriminated between both taxa. In the *ADMIXTURE*-analyses, on the other hand, more stringent filtering criteria uncovered varying levels of shared ancestry between the Taiga and the Tundra Bean Goose (Supplementary Fig. S2). These

findings indicate the potential issues of solely relying on PCAs and genetic ancestry analyses when assessing the genetic make-up of populations. Therefore, it is important to investigate genetic patterns in more detail, for example by exploring the genomic landscape of differentiation.

In line with the *ADMIXTURE*-analyses, the genetic divergence between the taxa seems to be driven by few genomic regions that are scattered throughout the genome, so-called islands of differentiation (Figs. 2 and 3). This pattern has been observed in other bird species, such as crows (Poelstra et al. 2014), woodpeckers (Grossen et al. 2016), warblers (Toews et al. 2016; Irwin et al. 2018), flycatchers (Ellegren et al. 2012; Burri et al. 2015), thrushes (Ruegg et al. 2014; Delmore et al. 2015), stonechats (Van Doren et al. 2017), and nightingales (Mořkovský et al. 2018). In line with previous studies, we found no significant difference in the level of genetic differentiation between islands on autosomes and on the Z-chromosome (e.g., Ellegren et al. 2012; Bay and Ruegg 2017; Mořkovský et al. 2018). Some patterns suggest that some of the islands of differentiation uncovered in this study might contribute to reproductive isolation, whereas the remainder of the genome can freely flow between the species. The positive correlation between F_{ST} and d_{XY} indicated increased genetic divergence in particular genomic regions, whereas the rest of the genome showed divergence levels close to the genome-wide average (Fig. 2e). In addition, the demographic model uncovered high levels of recent gene flow between the Taiga and the Tundra Bean Goose (Fig. 4). The islands of differentiation contained some interesting candidate genes, such as *KCNU1*, which is involved in spermatogenesis and might thus play a role in prezygotic post-mating isolation (Buffone et al. 2012). However, more detailed analyses are needed to validate these candidate genes (see Supplementary Table S6 for a list of candidate genes).

Several patterns indicated that the genomic landscape of the Bean Geese was at least in part shaped by linked selection. The negative correlation between F_{ST} and nucleotide diversity suggests that selection reduced genetic diversity in certain genomic regions (Fig. 2f). Negative values of Tajima's D across the majority of genomic windows (Fig. 2d) point to purifying selection or population expansion (Tajima 1989). Also, some of the differentiation islands did not show elevated d_{XY} values, indicating linked selection (Fig. 3). Few of the high F_{ST} islands were accompanied by a decrease in absolute divergence d_{XY} (Fig. 3). Instead of extant linked selection that does not cause a drop in d_{XY}, this result can be explained by recurrent selection (i.e. selection in a common ancestor and in the daughter species, Cruickshank and Hahn 2014; Irwin et al. 2018). Clearly, more detailed analyses are needed to determine the relative contributions of reproductive isolation and linked selection in shaping the genomic landscape of the Bean Geese. Such analyses include quantifying the relationship between levels of diversity and local recombination rate, and comparing the genomic landscapes of related goose species (Burri et al. 2015; Ravinet et al. 2017; Stankowski et al. 2019). The islands of differentiation may evolve at the same genomic regions at independent lineages even across broad taxonomical range due to linked selection at conserved genetic elements such as areas of low recombination (Burri et al. 2015; Dutoit et al. 2017; Delmore et al. 2018).

Taxonomic recommendations

The degree and character of genomic differences between the Taiga and the Tundra Bean Goose raise the question whether they should be considered separate species. Specifically, do a few differentiated regions in the genome provide enough evidence to consider them as distinct species? As a single criterion, genomic differentiation might be considered too low to justify a species rank. But, in combination with other species criteria, such as morphology, behavior and ecology, genomics could provide an extra line of evidence in species classification (Ottenburghs 2019). Indeed, avian taxonomy has become more pluralistic (Sangster 2018), combining different species criteria to justify taxonomic decisions (Alström et al. 2008; Gohli et al. 2015; Oswald et al. 2016).

Furthermore, linking islands of differentiation to other species criteria, such as morphology or reproductive isolation, can strengthen a taxonomic decision. This is nicely illustrated by the genomic analyses of Hooded Crow (*Corvus cornix*) and Carrion Crow (*C. corone*), which uncovered a single differentiated genomic region that harbored several genes involved in pigmentation and visual perception (Poelstra et al. 2014). These genetic variants have been shown to underlie the different plumage patterns (black or gray-coated) in these species (Wu et al. 2019; Knief et al. 2019). In addition, several behavioral studies uncovered assortative mating according to plumage phenotypes (Saino and Villa 1992; Risch and Andersen 1998; Haas et al. 2010). Such detailed investigations have not been performed for the Bean Goose complex. The genomic islands of differentiation uncovered in this study might be associated with morphological and behavioral differences between the Taiga and the Tundra Bean Goose, but this remains to be determined by denser sampling across the range of these taxa and experimental work on their social behavior.

On the basis of the evidence from different species criteria (e.g., genetic differentiation, reproductive isolation and morphology), one can thus assess the taxonomic status of particular taxa (Ottenburghs 2019). The first criterion to

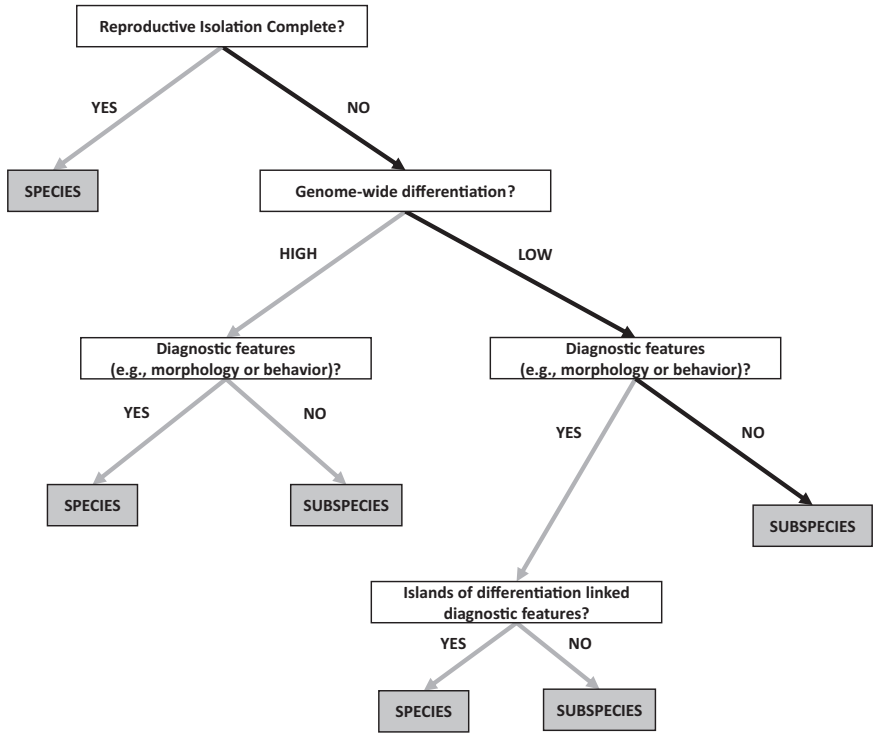


Fig. 5 Decision tree for classification of species and subspecies based on reproductive isolation, genetic differentiation and morphology. The black arrows indicate the route followed to determine the taxonomic position of the Taiga and the Tundra Bean Goose.

consider is the level of reproductive isolation between taxa. If reproductive isolation is complete, the two taxa should be considered separate species. If reproductive isolation is incomplete, the level of genomic differentiation and diagnosability (e.g., differences in behavior or morphology) can be taken into account. Here, different scenarios are possible. For example, a high level of genomic differentiation in combination with several diagnostic features suggests a species status, whereas a low level of genomic differentiation in combination with no diagnostic features indicates that the taxa should be treated as subspecies. A special situation concerns the combination of low genomic differentiation and several diagnostic features. To reach a taxonomic decision, genomic islands of differentiation can be taken into account. If the diagnostic features can be linked to particular genomic islands of differentiation (thus providing a genetic basis for these features), the taxa can be considered distinct species. If not, a subspecies status is more appropriate.

To visualize this taxonomic decision process, we constructed a decision tree which we illustrate with the information on the Taiga and the Tundra Bean Goose (Fig. 5). First, reproductive isolation between the Taiga and the Tundra Bean Goose is incomplete: both taxa are known to hybridize (Ottenburghs et al. 2016a; Honka et al. 2017) and this study uncovered high levels of recent introgression. Second, although this study shows that they are genetically distinct, the degree of genetic differentiation is very low (genome-wide $F_{ST} = 0.033$). This level of genome-wide differentiation is lower compared to other bird systems that are considered subspecies, such as *Catharus* thrushes ($F_{ST} = 0.1$; Delmore et al. 2015) and some members of the Yellow-rumped Warbler (*Setophaga coronata*) complex ($F_{ST} = 0.06$; Irwin et al. 2018). One notable exception concerns the Golden-winged (*Vermivora chrysoptera*) and Blue-winged Warblers (*Vermivora cyanoptera*) that, despite a genome-wide F_{ST} of only 0.0045, are considered distinct species (Toews et al. 2016). Third, there are no clear diagnostic features to discriminate between the Taiga and the Tundra Bean Goose (de Jong 2019). Moreover, there is considerable morphological variation within both taxa (Burgers et al. 1991). Possibly, there is clinal variation in certain traits, such as beak size, across the range of the Bean Goose complex, similar to the Greater White-fronted Goose (*A. albifrons*, Ely et al. 2005). However, the morphology of the eastern Bean Goose taxa (*A. s. serrirostris* and *A. f. middendorfii*) will need to be assessed to obtain a complete picture of morphological variation within the Bean Goose complex. On the basis of the low genetic differentiation, considerable morphological variation and incomplete reproductive isolation, we argue that the Taiga and the Tundra Bean Goose should be treated as subspecies.

Data archiving

The genome re-sequencing data are freely available in EMBL-EBI European Nucleotide Archive (http://www.ebi.ac.uk/ena) under accession number PRJEB35788. The

scripts and workflow for the analyses can be found on the following Github-page: https://github.com/JenteOttie/Goose_Genomics/tree/master/BeanGoose.

Acknowledgements We thank members of the Ellegren lab and the Suh lab for insightful discussions. We are indebted to the Naturhistoriska Riksmuseet in Stockholm and the Finnish Game and Fisheries Research Institute (now the Natural Resources Institute Finland) for providing several samples. This research was made possible by grants from the Swedish Research Council (contract 2013-8271) and the Knut and Alice Wallenberg foundation (contract 2014.0044). Sequencing was performed by the SNP&SEQ Technology Platform in Uppsala. The facility is part of the National Genomics Infrastructure (NGI) Sweden and Science for Life Laboratory. The SNP&SEQ Platform is also supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation. Computations were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome Res 19:1655–64

Alström P, Rasmussen PC, Olsson U, Sundberg P (2008) Species delimitation based on multiple criteria: The Spotted Bush Warbler *Bradypterus* thoracicus complex (Aves: Megaluridae). Zool J Linn Soc 154:291–307

Árnason Ú, Lammers F, Kumar V, Nilsson MA, Janke A (2018) Whole-genome sequencing of the blue whale and other rorquals finds signatures for introgressive gene flow. Sci Adv 4:eaap9873

Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A et al. (2013) From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. Curr Protoc Bioinformatics 43:11.10.1–11.10.33.

Barlow A, Cahill JA, Hartmann S, Theunert C, Xenikoudakis G, Fortes GG et al. (2018) Partial genomic survival of cave bears in living brown bears. Nat Ecol Evol 2:1563–1570

Bay RA, Ruegg K (2017) Genomic islands of divergence or opportunities for introgression? Proc R Soc B Biol Sci 284:20162414

Björklund M (2019) Be careful with your principal components. Evolution 73:2151–2158

Buffalo V, Coop G (2019) The linked selection signature of rapid adaptation in temporal genomic data. Genetics 213:1007–1045

Buffone MG, Ijiri TW, Cao W, Merdiushev T, Aghajanian HK, Gerton GL (2012) Heads or tails? Structural events and molecular mechanisms that promote mammalian sperm acrosomal exocytosis and motility. Mol Reprod Dev 79:4–18

Burgers J, Smit J, Vandervoet H (1991) Origins and systematics of two types of Bean Goose *Anser fabalis* (Latham 1787) wintering in the Netherlands. Ardea 79:307–315

Burri R (2017) Linked selection, demography and the evolution of correlated genomic landscapes in birds and beyond. Mol Ecol 26:3853–3856

Burri R, Nater A, Kawakami T, Mugal CF, Olason PI, Smeds L et al. (2015) Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. Genome Res 25:1656–1665.

Charlesworth B (1994) The effect of background selection against deleterious mutations on weakly selected, linked variants. Genet Res 63:213–227

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L et al. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly 6:80–92

Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Mol Ecol 23:3133–3157

Currat M, Ruedi M, Petit RJ, Excoffier L (2008) The hidden side of invasions: massive introgression by local genes. Evolution 62:1908–1920

Cutter AD, Payseur BA (2013) Genomic signatures of selection at linked sites: unifying the disparity among species. Nat Rev Genet 14:262–274

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA et al. (2011) The variant call format and VCFtools. Bioinformatics 27:2156–2158

Delacour J (1951) Taxonomic notes on the bean geese, *Anser fabalis* Lath. Ardea 39:135–142

Delmore KE, Hübner S, Kane NC, Schuster R, Andrew RL, Câmara F et al. (2015) Genomic analysis of a migratory divide reveals candidate genes for migration and implicates selective sweeps in generating islands of differentiation. Mol Ecol 24:1873–1888

Delmore KE, Lugo Ramos JS, Van Doren BM, Lundberg M, Bensch S, Irwin DE et al. (2018) Comparative analysis examining patterns of genomic differentiation across multiple episodes of population divergence in birds. Evol Lett 2:76–87

Van Doren BM, Campagna L, Helm B, Illera JC, Lovette IJ, Liedvogel M (2017) Correlated patterns of genetic diversity and differentiation across an avian family. Mol Ecol 26:3982–3997

Dutoit L, Vijay N, Mugal CF, Bossu CM, Burri R, Wolf J et al. (2017) Covariation in levels of nucleotide diversity in homologous regions of the avian genome long after completion of lineage sorting. Proc R Soc B Biol Sci 284:20162756

Ellegren H, Smeds L, Burri R, Olason PI, Backström N, Kawakami T et al. (2012) The genomic landscape of species divergence in *Ficedula* flycatchers. Nature 491:756–760

Ely CR, Fox AD, Alisauskas RT, Andreev A, Bromley RG, Degtyarev AG et al. (2005) Circumpolar variation in morphological characteristics of Greater White-fronted Geese *Anser albifrons*. Bird Study 52:104–119

Feder JL, Egan SP, Nosil P (2012) The genomics of speciation-with-gene-flow. Trends Genet 28:342–350

Fox A, Leafloor J (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

Gao G, Zhao X, Li Q, He C, Zhao W, Liu S et al. (2016) Genome and metagenome analyses reveal adaptive evolution of the host and interaction with the gut microbiota in the goose. Sci Rep. 6:32961

Gohli J, Leder EH, Garcia-Del-Rey E, Johannessen LE, Johnsen A, Laskemoen T et al. (2015) The evolutionary history of Afrocanarian blue tits inferred from genomewide SNPs. Mol Ecol 24:180–191

Gopalakrishnan S, Sinding M-HS, Ramos-Madrigal J, Niemann J, Samaniego Castruita JA, Vieira FG et al. (2018) Interspecific gene flow shaped the evolution of the genus *Canis*. Curr Biol 28:3441–3449.e5

Grossen C, Seneviratne SS, Croll D, Irwin DE (2016) Strong reproductive isolation and narrow genomic tracts of differentiation among three woodpecker species in secondary contact. Mol Ecol 25:4247–4266

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet 5:e1000695

Haas F, Knape J, Brodin A (2010) Habitat preferences and positive assortative mating in an avian hybrid zone. J Avian Biol 41:237–247

Harris R (2007) Improved pairwise alignment of genomic DNA. The Pennsylvania State University, University Park, PA, USA

Hillier LW, Miller W, Birney E, Warren W, Hardison RC, Ponting CP et al. (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432:695–716

Honka J, Kvist L, Heikkinen M, Helle P, Searle J, Aspi J (2017) Determining the subspecies composition of bean goose harvests in Finland using genetic methods. Eur J Wildl Res 63:19

del Hoyo J, Elliott A, Sargatal J, Christie D, de Juana E (2018) Handbook of birds of the world. Lynx Edicions, Barcelona

Hudson RR, Kaplan NL (1995) Deleterious background selection with recombination. Genetics 141:1605–1617

Irwin DE, Milá B, Toews DPL, Brelsford A, Kenyon HL, Porter AN et al. (2018) A comparison of genomic islands of differentiation across three young avian species pairs. Mol Ecol 27:4839–4855

de Jong A (2019) Less is better. Avoiding redundant measurements in studies on wild birds in accordance to the principles of the 3Rs. Front Vet Sci 6:195

Kardos M, Åkesson M, Fountain T, Flagstad Ø, Liberg O, Olason P et al. (2018) Genomic consequences of intensive inbreeding in an isolated wolf population. Nat Ecol Evol 2:124–131

Knief U, Bossu CM, Saino N, Hansson B, Poelstra J, Vijay N et al. (2019) Epistatic mutations under divergent selection govern phenotypic variation in the crow hybrid zone. Nat Ecol Evol 3:570

Lawson DJ, van Dorp L, Falush D (2018) A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. Nat Commun 9:3258

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N et al. (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079

Mallet J, Besansky N, Hahn MW (2016) How reticulated are species? BioEssays 38:140–149

Mangerud J, Gyllencreutz R, Lohne Ø, Svendsen JI (2011) Glacial history of Norway. Dev Quat Sci 15:279–298

Martin SH, Davey JW, Jiggins CD (2015) Evaluating the use of ABBA–BABA statistics to locate introgressed loci. Mol Biol Evol 32:244–257

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A et al. (2010) The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–1303

Mooij J, Zöckler C (1999) Reflections on the systematics, distribution and status of *Anser fabalis*. Casarca 5:103–120

Mořkovský L, Janoušek V, Reif J, Rídl J, Pačes J, Choleva L et al. (2018) Genomic islands of differentiation in two songbird species reveal candidate genes for hybrid female sterility. Mol Ecol 27:949–958

Nadeau NJ, Whibley A, Jones RT, Davey JW, Dasmahapatra KK, Baxter SW et al. (2012) Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. Philos Trans R Soc Lond B Biol Sci 367:343–353

Nordborg M, Charlesworth B, Charlesworth D (1996) The effect of recombination on background selection. Genet Res 67:159–174

Oswald JA, Harvey MG, Remsen RC, Foxworth DU, Cardiff SW, Dittmann DL et al. (2016) Willet be one species or two? A genomic view of the evolutionary history of *Tringa semipalmata*. Auk 133:593–614

Ottenburghs J (2019) Avian species concepts in the light of genomics. In: Kraus R (ed.) Avian genomics in ecology and evolution—from the lab into the wild. Springer Nature, Cham, p 211–235

Ottenburghs J, van Hooft P, van Wieren SE, Ydenberg RC, Prins HHT (2016a) Hybridization in geese: A review. Front Zool 13:1–9

Ottenburghs J, Megens H-J, Kraus R, Madsen O, van Hooft P, van Wieren S et al. (2016b) A tree of geese: A phylogenomic perspective on the evolutionary history of True Geese. Mol Phylogenet Evol 101:303–313

Ottenburghs J, Megens H-J, Kraus R, Van Hooft P, Van Wieren S, Crooijmans R et al. (2017a) A history of hybrids? Genomic patterns of introgression in the True Geese. BMC Evol Biol 17:201

Ottenburghs J, Kraus R, van Hooft P, van Wieren S, Ydenberg R, Prins H (2017b) Avian introgression in the genomic era. Avian Res 8:30

Otvos EG (2015) The Last Interglacial Stage: Definitions and marine highstand, North America and Eurasia. Quat Int 383:158–173

Owen M (1980) Wild geese of the world: their life history and ecology. Batsford, London

Palkopoulou E, Lipson M, Mallick S, Nielsen S, Rohland N, Baleka S et al. (2018) A comprehensive genomic history of extinct and living elephants. Proc Natl Acad Sci 115:E2566–E2574

Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y et al. (2012) Ancient admixture in human history. Genetics 192:1065–1093

Ploeger P (1968) Geographical differentiation in arctic Anatidae as a result of isolation during the last glacial. PhD thesis, University of Amsterdam

Poelstra J, Vijay N, Bossu C, Lantz H, Ryll B, Muller I et al. (2014) The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. Science 344:1410–1414

Pujolar JM, Dalén L, Olsen RA, Hansen MM, Madsen J (2018) First de novo whole genome sequencing and assembly of the pink-footed goose. Genomics 110:75–79

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–575

R Core Team (2018) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Ragsdale AP, Moreau C, Gravel S (2018) Genomic inference using diffusion models and the allele frequency spectrum. Curr Opin Genet Dev 53:140–147

Randler C (2004) Frequency of bird hybrids: does detectability make all the difference? J Ornithol 145:123–128

Ravinet M, Faria R, Butlin RK, Galindo J, Bierne N, Rafajlović M et al. (2017) Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. J Evol Biol 30:1450–1477

Renaut S, Grassa CJ, Yeaman S, Moyers BT, Lai Z, Kane NC et al. (2013) Genomic islands of divergence are not affected by geography of speciation in sunflowers. Nat Commun 4:1827

Rettelbach A, Nater A, Ellegren H (2019) How linked selection shapes the diversity landscape in *Ficedula* flycatchers. Genetics 212:277–285

Rheindt FE, Edwards SV (2011) Genetic introgression: an integral but neglected component of speciation in birds. Auk 128:620–632

Risch M, Andersen L (1998) Selektive Partnerwahl der Aaskrähe (Corvus corone) in der Hybridisierungszone von Rabenkrähe (*C. c. corone*) und Nebelkrähe (*C. c. cornix*). J Ornithol 139:173–177

Roux C, Fraïsse C, Romiguier J, Anciaux Y, Galtier N, Bierne N (2016) Shedding light on the grey zone of speciation along a continuum of genomic divergence. PLoS Biol 14:e2000234

Ruegg K, Anderson EC, Boone J, Pouls J, Smith TB (2014) A role for migration-linked genes and genomic islands in divergence of a songbird. Mol Ecol 23:4757–4769

Ruokonen M, Kvist L, Lumme J (2000) Close relatedness between mitochondrial DNA from seven *Anser* goose species. J Evol Biol 13:532–540

Ruokonen M, Litvin K, Aarvak T (2008) Taxonomy of the bean goose–pink-footed goose. Mol Phylogenet Evol 48:554–562

Saino N, Villa S (1992) Pair composition and reproductive success across a hybrid zone of carrion crows and hooded crows. Auk 109:543–555

Sangster G (2018) Integrative taxonomy of birds: the nature and delimitation of species. In: Tietze, D.T. (ed.) Bird species. Springer, Cham, p 9–37

Sangster G, Oreel G (1996) Progress in taxonomy of taiga and tundra bean geese. Dutch Bird 18:310–316

Shapiro BJ, Leducq J-B, Mallet J (2016) What is speciation? PLoS Genet 12:e1005860

Smith JM, Haigh J (1974) The hitch-hiking effect of a favourable gene. Genet Res 23:23–35

Stankowski S, Chase MA, Fuiten AM, Rodrigues MF, Ralph PL, Streisfeld MA (2019) Widespread selection and gene flow shape the genomic landscape during a radiation of monkeyflowers. PLoS Biol 17:e3000391

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595

Toews DPL, Taylor SA, Vallender R, Brelsford A, Butcher BG, Messer PW et al. (2016) Plumage genes and little else distinguish the genomes of hybridizing warblers. Curr Biol 26:2313–2318

Turner TL, Hahn MW, Nuzhdin SV (2005) Genomic islands of speciation in *Anopheles gambiae*. PLoS Biol 3:e285

Vernot B, Tucci S, Kelso J, Schraiber JG, Wolf AB, Gittelman RM et al. (2016) Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals. Science 352:235–239

Villanea FA, Schraiber JG (2018) Multiple episodes of interbreeding between Neanderthal and modern humans. Nat Ecol Evol 3:39

Wickham H (2016) Elegant graphics for data analysis, 1st edn. Springer-Verlag, New York, NY

Wolf JBW, Ellegren H (2017) Making sense of genomic islands of differentiation in light of speciation. Nat Rev Genet 18:87–100

Wu C-I (2001) The genic view of the process of speciation. J Evol Biol 14:851–865

Wu D-D, Ding X-D, Wang S, Wójcik JM, Zhang Y, Tokarska M et al. (2018) Pervasive introgression facilitated domestication and adaptation in the *Bos* species complex. Nat Ecol Evol 2:1139

Wu C-C, Klaesson A, Buskas J, Ranefall P, Mirzazadeh R, Söderberg O et al. (2019) *In situ* quantification of individual mRNA transcripts in melanocytes discloses gene regulation of relevance to speciation. J Exp Biol 222:jeb194431

Zachos J, Pagani M, Sloan L, Thomas E, Billups K (2001) Trends, rhythms, and aberrations in global climate 65 Ma to present. Science 292:686–693