1-1-2016

# Panmixia Across the Range of the Assortatively Mating Lesser Snow Goose

Brent B. Horowitz
*University of Denver*

# Panmixia Across the Range of the Assortatively Mating Lesser Snow Goose


## Abstract

The Lesser Snow Goose has been shown to exhibit imprinted positive assortative mating on the basis of its color phases, blue and white, and it has been proposed that the geographic cline of the color phases reflects a historic allopatry of the phases. Using ddRAD-seq, thousands of SNPs were analyzed in both phases and in colonies across the range of the species to examine genetic evidence for separation of the color phases or along the two migratory pathways, which have also been hypothesized to be separate. These analyses did not reveal any genetic differentiation between the color phases nor by migratory pathway. These data are consistent with a single panmictic population. The Lesser Snow Goose is a poor candidate for sympatric speciation, as gene flow between the phases is high, and assortative mating is focused on a single locus.



## Document Type

Thesis

## Degree Name

M.S.

## Department

Biological Sciences

## First Advisor

Thomas W. Quinn, Ph.D.

## Second Advisor

Robin Tinghitella

## Third Advisor

Scott Nichols

## Keywords

Assortative mating, Lesser snow goose, Population genetics

## Subject Categories

Genetics and Genomics

# Panmixia across the range of the assortatively mating Lesser Snow Goose

---

A Thesis Presented to

the Faculty of Natural Sciences and Mathematics

University of Denver

---

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

---

by

Brent Horowitz

November 2016

Advisor: Tom Quinn

Author: Brent B. Horowitz
Title: Panmixia across the range of the assortatively mating Lesser Snow Goose
Advisor: Tom Quinn
Degree Date: November 2016

ABSTRACT

The Lesser Snow Goose has been shown to exhibit imprinted positive assortative mating on the basis of its color phases, blue and white, and it has been proposed that the geographic cline of the color phases reflects a historic allopatry of the phases. Using ddRAD-seq, thousands of SNPs were analyzed in both phases and in colonies across the range of the species to examine genetic evidence for separation of the color phases or along the two migratory pathways, which have also been hypothesized to be separate. These analyses did not reveal any genetic differentiation between the color phases nor by migratory pathway. These data are consistent with a single panmictic population. The Lesser Snow Goose is a poor candidate for sympatric speciation, as gene flow between the phases is high, and assortative mating is focused on a single locus.

ACKNOWLEDGEMENTS

This work would not be possible without the contributions of many others. My advisor Tom Quinn has provided a steady hand throughout this project and through the lows and highs of my time at the University of Denver. Robin Tinghitella and Scott Nichols, my committee members, have my thanks for all their suggestions, questions, and helpful hints. I also owe a debt to our research group, including Shannon Murphy, Whitley Lehto, and Mayra Vidal. The DU Organismal Biologists were very helpful with the writing portion of this process, both on this work and others. Ben Fotovich was a tremendous resource and helpful guide with all things computing. Martin Margittai was helpful both with assistance with the Tecan plate reader and for serving as my defense committee chair on such short notice. The David Pollack Lab and Jonathan Shortt at CU Anschutz were helpful and generous in lending use of their Pippin Prep to our group for use. Of course, without the support of my friends and family I could not be where I am today. Mom and Dad, Grandmama, Alex, Carey, Michael, Debbie, Kim, Jeff, Kellie, Rene, and River, thank you. Lastly I would like to acknowledge the late Dr. Susan Offner for setting me on a path into biology, for which I am forever grateful.

TABLE OF CONTENTS

CHAPTER ONE: INTRODUCTION ..... 1

CHAPTER TWO: MATERIALS AND METHODS ..... 7
DNA Sources ..... 7
DNA Quantification ..... 7
Sex Identification ..... 8
Double Digest RAD-Seq ..... 8
Digestion in silico ..... 8
Double Digestion of DNA samples ..... 8
Preparation of Magnetic Beads and Bead Cleaning ..... 9
Barcoding Strategy ..... 9
Adapter Annealing ..... 10
Adapter Ligation ..... 10
Size Selection by Pippin Prep ..... 11
Quantitative PCR ..... 11
PCR Amplification and Sequencing Preparation ..... 11
Sequencing ..... 12
Computing Resources ..... 12
Demultiplexing ..... 12
Sequence Alignment ..... 12
Loci and Population Analysis ..... 13

CHAPTER THREE: RESULTS ..... 15
Sequencing Data ..... 15
Loci ..... 15
Population Analysis ..... 16
Principle Component Analysis ..... 17

CHAPTER FOUR: DISCUSSION ..... 18

BIBLIOGRAPHY ..... 32

LIST OF FIGURES

**Figure 1**. Sample locations. Pie charts show proportion of blue birds between 1969 and 1971 (Cooke et al. 1975).....23
**Figure 2**. Sequencing reads by sample tissue origin. Samples obtained from blood (n=215, $\bar{x}$=530625) obtained higher read counts than liver samples (n=92, $\bar{x}$=263145) (ANOVA, df=1, F=9.272, p=0.003). Data are pooled from multiple runs; individuals sampled multiple times are counted independently for this analysis.....24
**Figure 3.** AMOVA $F_{ST}$ values for individual loci between blue and white phases for SNPs mapped by chromosome. Colors alternate by chromosome. Not shown: microchromosomes; no microchromosome had more than one SNP present.....26
**Figure 4.** Structure analysis of snow geese by color. Individuals are arranged as vertical bars, with populations found by Structure represented as different colors. The relative size of the color indicates the probability the individual should be assigned to the population. Structures are the average of 5 replicate runs of the analysis.....28
**Figure 5.** ΔK values by K (populations) value from Structure analysis. ΔK for K=1 cannot be calculated.....29
**Figure 6.** Raw log likelihood of each Structure output K. Error bars are 95% CI.....30
**Figure 7.** Principal Coordinate Analysis of SNP data by population location. Black circles = AR, Red = CHM, Green = LPB, Blue = WI.....31

# CHAPTER ONE: INTRODUCTION

One of the basic assumptions of random mating is that individuals mate without consideration of their own phenotype. However, exceptions to this assumption are common, often causing profound impacts on the evolutionary trajectory of a species. One such violation is positive assortative mating, where individuals seek out a mate whose phenotype matches their own; this can contribute to sympatric speciation when coupled with disruptive natural selection (Barluenga et al. 2006). Assortative mating can be driven by early learning of sexual preferences from parental imprinting, often within birds (Class Aves) (Irwin and Price 1999). This learning can be considered a type of behavioral plasticity, as exposure to parents who do not share the phenotype will result in mate preference for that other phenotype (Verzijden et al. 2012). Here I present a genetic study of a subspecies of goose that has been described as exhibiting assortative mating based upon early learning of sexual preferences, to determine whether the learned mate preference has led to genetic differentiation between the two color phases on which the assortative mating is based. This assortative mating has been hypothesized to have shaped the genetic structure of the modern populations. If the two color phases are actively diverging, genetic evidence of this differentiation should be found here.

The Lesser Snow Goose (*Chen caerulescens caerulescens*, syn: *Anser caerulescens caerulescens*) is a migratory species with breeding grounds in the Russian and Canadian Arctic. Two major color phases exist: blue and white (Cooke and Buckley 1987). Historically, these two phases were considered potentially different species, and at one time the blue phase was thought

to be a juvenile form of the white (Manning et al. 1956); ecological evidence led to their classification as a single species with a color dimorphism (Cooch 1961). The phases are unevenly distributed across their geographical range on both the northern breeding grounds and the southern wintering grounds. The white phase can be found across the entire range, with the western breeding colonies described as entirely white (with extremely rare blue phase birds (McLandress and McLandress 1979)), while the blue phase is not typically found in colonies west of Banks Island in the Canadian Arctic Archipelago (Manning et al. 1956). The blue phase increases in population proportion on a west-to-east cline (Figure 1) (Cooke et al. 1975). The western limit of the blue phase has changed through time; Banks Island, for example, has not always had the blue phase present (Cooke et al. 1975).

Depending on the location of the breeding ground, flocks migrate south to different wintering grounds. There are two main migratory paths: the Pacific Flyway, and the Central and Mississippi Flyway (Ogilvie 1978). The Pacific Flyway populations winter in New Mexico, California, and Mexico; these populations breed between Wrangel Island in Russia and the Western edges of the Canadian Arctic Archipelago (Ogilvie 1978). There has been indication that some of the Wrangel Island population, termed the Northern Wrangel Island population, winters further north, in Southwestern Canada (Shorey et al. 2011). The Central and Mississippi Flyways populations winter in eastern Texas, Louisiana, Mississippi, and the Gulf Coast of Mexico while breeding in the Hudson Bay colonies (Ogilvie 1978).

Adult geese choose their mates in the southern wintering ground and formed pairs migrate north to the breeding ground together (Finney and Cooke 1978). Females are typically philopatric, returning to the breeding ground where they were born in the spring, while males following their mates more often shift between colonies (Cooke and Buckley 1987). Consequently, rates of immigration differ by sex (Shorey et al. 2011); one study identifying banded immigrants only found four immigrant females in comparison with fifty-one males

(Cooke et al. 1975). Because many males and some females migrate from their natal colonies to another colony when they reach breeding age, there is a considerable level of gene flow between breeding grounds within migratory paths (Cooke and Buckley 1987). One estimate calls 50% of new breeding adults immigrants to the colony (Cooke et al. 1975). Therefore, assuming there is no mechanism used by the geese to differentiate which individuals originated in their breeding colony, there is considerable interbreeding between colonies which are sympatric in the wintering grounds (Cooke et al. 1975). Immigration is determined by shared wintering ground; colonies which do not share wintering ground will not have migration occur between them, given that mate pairing drives migration (Shorey et al. 2011). As geese migrating along the eastern and western migratory paths also occupy geographically separated wintering grounds, the two could be considered separate populations (Ogilvie 1978). Further, one would predict if the two have been separated long enough, there were be genetic differentiation between the populations. If there is no gene flow between the two pathways, it should be reflected in this study.

The Lesser Snow Goose demonstrates assortative mating between its two color phases (Cooke et al. 1975). This mating pattern has been shown to have a basis in parental imprinting; newly breeding adults will seek out mates which possess the same color phenotype as the parents without regard for the goose's own coloration (Cooke et al. 1976). There is some evidence that non-preferred mates will be selected if the preferred color is not available, but as the wintering grounds are gatherings of thousands of geese, this is a relatively small factor in pairing (Cooke et al. 1975). One seven-year study of a mixed color phase colony (with a greater proportion of white phase) found between 9 and 13% of white phase and 28 and 31% of blue phase geese were in a mixed pairing in a given year (Cooke et al. 1976). Far more pairings of same-color phases take place than mixed pairings, violating Hardy-Weinberg equilibrium assumptions of random mating (Cooke and Cooch 1968).

The color dimorphism, historically cause for declaring the phases different species, was eventually determined to be caused by a single gene or very closely linked group of genes (Cooke and Mirsky 1972). The blue allele is incompletely dominant over the white allele, with heterozygotes demonstrating blue bodies with white bellies (Cooke and Mirsky 1972). More recently, it has been shown that the dimorphism is caused by the melanocortin-1 receptor gene, or *MC1R* (Mundy et al. 2004). A single non-synonymous base transition causes the Valine to replaced Methionine at the 85$^{th}$ amino acid in the protein in the outer region of the second transmembrane pass, leading to the emergence of the blue phase from the ancestral white phase, 380,000 ± 188,800 years ago (Mundy et al. 2004). There is considerable variation in the degree of blue coloration in the *MC1R* heterozygotes of *C. c. caerulescens,* indicating that other genes may be responsible for the particular color patterning (Mundy et al. 2004); this has been found to be true in other species with *MC1R* gene variants (Steiner et al. 2007). In the other subspecies, the Greater Snow Goose (*C. c. atlanticus*), and the closely related Ross' Goose (*C. rossii*), the vast majority (>99.9%) of the birds are completely white, with most blue phases thought to be the result of interbreeding with *C. c. caerulescens* (McLandress and McLandress 1979). In blue members of both *C. c. atlanticus* and *C. rossii,* the same V85M substitution is found in the *MC1R* gene (Mundy et al. 2004).

Natural selection for or against one phase is not well established. At one point, blue geese were increasing between 1-2% per year in colonies where they were found, despite the two phases laying similar sized clutches (Cooch 1961). White phases were said to lay eggs earlier in the season and thus suffer higher rates of predation, as well as make more frequent stops during migration, subjecting them to hunting mortality (Cooch 1961). However, statistical analyses for these claims were either not performed or not reported, and the claims were based on two years of observation of a single colony; the techniques used to find these counts have also been disputed (Cooke and Buckley 1987; Cooke et al. 1988). Further, models based on the described natural

selection concluded that blues would continue to rise to more than 75% of the populations in which they were found, but this did not occur within the indicated time frame nor subsequently; the frequencies remained largely the same over the course of a 12 year study (Cooke et al. 1975). Natural selection, if it is a factor at all, is likely not alone in explaining the color phase frequency observations (Cooke and Buckley 1987).

The current population patterns of the Lesser Snow Goose have led to questions about the historical conditions of the species. One proposal holds that the two colors were once allopatric, with a blue population in the east and a white population in the west (Cooke et al. 1988). Historical evidence lends some support to the hypothesis that there was geographical separation between the color phases, with naturalists of the day only describing colonies of one color phase or the other, rather than mixed colonies (Barnston 1860; Saunders 1917). Changes in farming habits, particularly increases in rice farming in inland Texas between the 1920s and 1930s, have opened up new feeding locations in the wintering grounds, which is where pairing occurs (Cooke et al. 1988). Both white and blue geese took to these new feeding options, changing mate availability as previously isolated colonies crossed (Cooke et al. 1988). Testing of six allozymes of blue and white geese at the La Perouse Bay colony found a small (0.003) but significant difference in the $F_{ST}$ value between the two color phases, which was the genetic evidence used to support the idea that the two phases were formerly allopatric (Cooke et al. 1988).

Other previous population genetic studies of the Lesser Snow Goose have focused on mitochondrial DNA (mtDNA). While there is evidence of multiple clades in mtDNA which spread into the present geographic populations out of glacial refugia in the Pleistocene, (Avise et al. 1992; Quinn 1992) it is likely that these clades do not provide the full picture of the population structure. As mtDNA is inherited as a single unit or haplotype, all loci analyzed from mtDNA are inherited together and are subject to lineage sorting (Moore 1995). Loci randomly selected and analyzed from nuclear chromosomes, however, are independent from one another (except for

those that are linked). Depending on the effective population size of ancestral populations and patterns of gene flow, mtDNA is also subject to the phenomenon of introgression, whereby a small set of mitochondrial haplotypes from one population move to another location and expand within the receiving population. As the ancestral populations of the Lesser Snow Goose were not necessarily the large populations seen today, it is possible that there were smaller populations while in the glacial refugia, which would leave them more subject to possible introgression. In this situation, nuclear loci would be able to resolve as of yet undiscovered population structure that cannot be found by mtDNA, and would additionally allow for analysis of biparental population structure, rather than just maternally inherited DNA. To date, no studies of autosomal markers beyond the aforementioned allozymes have been performed, and none across the geographic range of the species; although one unpublished study was performed, it found no overall significant difference between the color phases, although two of thirteen loci tested were significantly different between samples taken from one colony (Quinn 1988).

Until recently, most population studies of nuclear genomes used microsatellite or amplified fragment length polymorphism (AFLP) genotyping. However, as sequencing technology development has accelerated and become less costly, new methods that sample far more loci for similar amounts of effort have been developed. These operate through a 'reduced-representation' method. One of these methods, for example, identifies Restriction-site Associated DNA tags (RAD-tags) (Baird et al. 2008). Here, a restriction enzyme is used to cut at particular sequence sites and DNA fragments near the cut site are sequenced in parallel, identifying loci near the cut site, and only near the cut site, among individuals of a species or closely related species (Baird et al. 2008). A later, more selective method was developed that used two restriction enzymes and a tight size selection technique, only carrying through fragments which were cut by both enzymes and within a particular size range; this technique is called double digest RAD-seq (ddRAD-seq) (Peterson et al. 2012). Replacing the hydroshearing in Baird's method

with a second restriction enzyme allows modulation of the loci sampled with greater specificity (Peterson et al. 2012). Both these methods, as well as other similar protocols, provide the ability to sample the same loci repeatedly across individuals and populations of a species or closely related species.

Previously RAD-seq and ddRAD-seq have been used to study a variety of different problems. The ability to discover single nucleotide polymorphisms (SNPs) and find population structure has been established (Ogden et al. 2013). Other groups have used RAD techniques with goals of resolving phylogeography (Catchen et al. 2013a), genome-wide linkage mapping (Takahashi et al. 2013), among numerous other uses (Narum et al. 2013). Here, ddRAD-seq was used to elucidate the population structure of the Lesser Snow Goose. Previous research indicated separation between the wintering grounds of the western and eastern migratory pathways. Assortative mating as a result of imprinting of color preference between the blue and white color phases has also been described. To date, analysis of nuclear genetic loci, other than phenotypic observations of the *MC1R* gene and the aforementioned six allozymes at La Perouse Bay, across the range of the Lesser Snow Goose has not been performed. If these patterns have caused any barriers to gene flow, those should be reflected in SNP data in the genomes. By analyzing the genetic structure of the populations of *C. c. caerulescens*, evidence of these barriers can be detected, which will indicate if the Lesser Snow Goose is undergoing sympatric speciation on the basis of the color phases, or if migration exists across geographical flyways.

# CHAPTER TWO: MATERIALS AND METHODS

## DNA Sources

*Chen caerulescens caerulescens* blood samples were drawn from the brachial vein of randomly selected adult geese at La Perouse Bay (LPB) located near Churchill, Manitoba, Canada. Blood samples were mixed with an anticoagulant, mixed by inversion, and frozen within 10h(Quinn and White 1987). Sex was usually, but not always, recorded at the time of blood collection. Similar procedures were used to collect samples at Anderson River in the Northwest Territories and Cape Henrietta Maria, near Hook Point, Ontario, Canada (Figure 1).

Samples from Wrangel Island, Russia, were obtained from hunter-killed birds in southern British Colombia and the Puget Sound, locations where geese from Wrangel Island only are known to migrate (Quinn 1992) (Figure 1). Wrangel Island birds were all considered white phase and sex determination was made as below.

## DNA Quantification

DNA had been previously extracted from blood samples (Quinn and White 1987) or from liver samples (Quinn 1992). Concentrations were estimated by Picogreen fluorescence (Singer et al. 1997). Briefly, 1µL of each sample was loaded into a well of a Grenier Flat Black 96 well plate (Sigma-Aldrich, St. Louis, MO) along with 1X TE and the Picogreen reagent (Thermo Scientific, Fremont, CA), at a 1/200 dilution from the stock solution, with a final volume of 200µL. Plates were read on a Tecan Infinite M1000 (Tecan Trading AG, Switzerland) with excitation at 480nm and emission at 520nm.

Sex Identification

For individual bird samples for which sex was not recorded or known at the time of capture, PCR amplification of an intron of the chromodomain-helicase-DNA-binding-protein (CHD) gene, which differs in size between the female-specific W chromosome and shared Z chromosome (Kahn et al. 1998), was used. PCR followed the parameters of (Kahn et al. 1998) and products were run on 2% agarose 3:1 (Amresco, Solon, OH) gels in 0.5X TBE at 10V/cm.

Double Digest RAD-Seq

Samples were prepared for sequencing following a modified version of the previously published protocol (Peterson et al. 2012). Each step is described below.

Digestion in silico

The R package SimRAD version 0.96 (Lepais and Weir 2014) was used to estimate restriction enzyme cut frequency for both *Msp*I and *Eco*RI and to estimate the number of fragments generated with *Msp*I-*Msp*I, *Eco*RI-*Eco*RI and *Eco*RI -*Msp*I ends. The reference genome used was *Gallus gallus* (chicken; (Hillier et al. 2004). On average, *Msp*I cut every 2227 bp and *Eco*RI cut every 3636 bp. At the range of 300 ± 24 bp, there were estimated to be 17408 fragments with *Msp*I-*Eco*RI ends, 2142 with *Eco*RI-*Eco*RI ends, and 56032 with *Msp*I-*Msp*I ends.

Double Digestion of DNA samples

Each individual genomic DNA sample (500ng) was digested with 0.5µL MspI (20,000 units/mL, New England Biolabs, Ipswitch, MA), 0.5µL *EcoRI*-HF (20,000 units/mL, New England Biolabs), 2.0µL 10X CutSmart Buffer (500mM Potassium acetate, 200mM $MgCl_2$, 100mM Magnesium acetate, 1mg/mL bovine serum albumin, pH 7.9, New England Biolabs), with a final volume of 20µL. Reactions were placed at 37°C for 3h, then 4°C overnight.

Preparation of Magnetic Beads and Bead Cleaning

Magnetic beads for removing enzymes and salts from samples were made by modifying an existing protocol (Rohland and Reich 2012). Sera-Mag Speedbeads (Thermo Scientific) (2mL) were placed on a magnet stand (Promega, Madison, WI) and washed with 1X TE pH 8.0 three times, then placed in 2mL TE. A separate solution of 9g PEG-8000 (Amresco), 2.92g NaCl, 500µL 1M Tris-HCl, 100µL 0.5M EDTA, and $ddH_2O$ to a final volume of 48mL was prepared and mixed by hand until all components were in solution. Tween 20 (25µL) was added to the solution, followed by the cleaned Speedbeads, and the whole solution was gently rocked until evenly mixed. Quality was compared with commercial Ampure XP beads (Beckman Coulter, Brea, CA) by cleaning 2µL of NEB 50 bp DNA ladder (New England Biolabs) diluted with 18µL of $ddH_2O$ with both bead solutions as described below, followed by observation on a 1.5% agarose gel, ensuring equal retention of DNA.

Prepared magnetic beads (1.8X volumes) were added to double digested samples, mixed by pipetting, and incubated 5 minutes at room temperature (21.5°C). Samples were placed on a magnet stand and beads were allowed to move to the stand for approximately 3 minutes. The supernatant was removed and 200µL of freshly prepared 70% ethanol (EtOH) was added with the tube in place on the stand. The 70% EtOH was removed and another 200µL of EtOH was added and removed. The samples were then dried on the stand at 37°C for 4 minutes. At room temperature and off the magnet stand, 30µL of $ddH_2O$ was added and mixed by flicking. The tubes were replaced on the stand and the beads were allowed to move to the stand. The liquid was removed and saved; the beads were discarded. The cleaned samples were quantified using Picogreen as described above.

Barcoding Strategy

Each individual was assigned a unique pair of identifiers: one barcode of five bases attached at the *Eco*RI side (henceforth: P1 side) and an index of six bases attached at the *Msp*I

side (henceforth: P2 side). Samples were later identified by the combination of the two sequences; no individual sample shared both a P1 barcode and P2 index in a given sequencing run.

Adapter Annealing

Oligonucleotides were designed according to the supplement of Peterson and colleagues (Peterson et al. 2012), with each oligo pairing by base complementation with another to form a double stranded adapter; adapters for the P1 side contained the barcode sequence while the P2 adapter was universal, containing no identifying information. Oligonucleotides (Integrated DNA Technologies, Coralville, IA) were diluted to a concentration of 100µM in dd$H_2O$. Annealing buffer (10X) consisted of 100mM Tris-HCl pH8, 500 mM NaCl, and 10mM EDTA. Adapters were annealed in a solution of 40µL 100µM of each oligo, 10µL of annealing buffer 10X, and 10µL dd$H_2O$. The annealing solution was placed at 97.5°C for 2.5 minutes, then allowed to cool to room temperature overnight sitting in a 25 L water bath. This 40µM solution was diluted to 4µM working stocks using dd$H_2O$ and stored at -20°C.

Adapter Ligation

Ligation reactions consisted of 4µL of 10X ligase buffer (500mM Tris-HCl, 100mM $MgCl_2$, 100mM DTT, 1mM ATP, pH 7.5, New England Biolabs), 0.2µL of T4 DNA ligase (400,000U/mL, New England Biolabs), 0.2µL 0.4 µM *Msp*I adapter, 0.3µL 0.4 µM *Eco*RI adapter, 80 or 200 ng of DNA sample (for Run One and Two, respectively), and dd$H_2O$ to a final volume of 40µL. The reaction was incubated at room temperature for 30 minutes, then placed in a 65°C water bath for 10 minutes, and then allowed to cool to room temperature overnight. Samples with unique P1 barcodes were then combined into eight or four (for the first two runs and the third, respectively) different pools and cleaned with prepared magnetic beads as above.

Size Selection by Pippin Prep

Each pool was electrophoresed through 2% agarose gels and fragments in the size range 300 ± 24bp were isolated using a Pippin Prep apparatus (Sage Science, Beverly, MA). The manufacturer's protocol was followed, using ladder B. Output pools were between 30 and 60µL.

Quantitative PCR

Each pool was quality checked by qPCR. Each pool was set up in duplicate 20µL reactions of 10µL 2X iQ SYBR Green supermix (Bio-Rad, Hercules, CA), 0.5µL of the Pippin pool, 1µL 10µM Primer with specific P2 index, 1µL 10µM Universal P1 Primer, and 7.5µL dd$H_2O$. Reactions were run on a CFX Connect Real-Time System (Bio-Rad) for the following cycles: 95°C (5 minutes), 40 cycles of 95°C (15 seconds) 72°C (30 seconds), and melt curve analysis from 65°C to 95°C.

PCR Amplification and Sequencing Preparation

Pippin Prep pools were amplified by Phusion PCR (New England Biolabs). Reactions were 4µL Phusion HF Buffer, 7.5µL Pippin prep pool, 4µL 10µM Primer with specific P2 index, 4µL 10µM Universal P1 Primer, 0.2µL Phusion Polymerase (2U/µL), and 0.4µL of 10mM dNTPs (Sigma-Aldrich). These primers can only amplify fragments which have an EcoRI-annealed adapter on an end, preventing amplification of *Msp*I-*Msp*I fragments. As many reactions as were required to utilize all the Pippin Prep output were used. The following cycles were run on an MJ Research PTC-200 DNA Engine (GMI, Ramsey, MN): 98°C (30 seconds), 12 cycles of 98°C (7 seconds), 72°C (15 seconds), with a final cycle of 72° C (10 minutes). Products with a common index primer were combined and cleaned by magnetic beads, then quantified by Picogreen. Equimolar amounts of each pool as determined by Picogreen were pooled into the final sample for sequencing, as determined by the lowest yield pool.

Sequencing

Sequencing was performed at The Genomics and Microarray Core Facility at the University of Colorado Denver (Aurora, CO) on an Illumina HiSEQ 2500 with V4 chemistry (San Diego, CA). Sequencing used 1X125 cycles with 1X7 additional cycles to read the i7 index tag. Three sequencing runs with varied sets of individuals were performed.

Computing Resources

All data were processed on the University of Denver's High Performance Cluster (Denver CO), variously on Intel® Xeon® E5420, E5645, or E5-2650 nodes.

Demultiplexing

Individual samples were sorted from the sequencer by the Stacks (Catchen et al. 2013b) version 1.39 tool *process_radtags*, based on the barcoding strategy above, with a minimum phred33 score of 30, discarding any reads with uncalled bases or ambivalent barcodes, and a sliding window of 19 bases for control of average sequencing quality. Sequences from the same individual from multiple runs were pooled. Individuals with fewer than 100,000 total reads passing quality filters were discarded from further analysis. All demultiplexed individuals which were used were stored in the NCBI Sequence Read Archive under BioProject ID PRJNA325370, held from release until the external publication of this work. Blood and liver sample read counts were compared by an ANOVA.

Sequence Alignment

Individuals were aligned to the *G. gallus* genome (Hillier et al. 2004) using the Burrows-Wheeler Alignment algorithm program BWA-mem version 0.7.10 (Li 2013). The package SAMtools version 0.1.19 (Li et al. 2009) was subsequently used to convert the files from SAM to BAM format.

Loci and Population Analysis

The Stacks version 1.39 tool *ref_map.pl* was used to process loci and search for SNPs between individual genomic samples. Three identical raw reads in an individual were required to process a haploid locus in that individual, requiring another three reads for a heterozygous locus. Two reads mapped to the W chromosome were discarded as they mapped onto the W chromosome in both male and female samples; only female snow geese actually have W chromosomes. Population-level statistics were generated pairwise between each of the four locations and between the two color phases by the Stacks tool *populations*. Loci were included in the population analysis if they were sequenced in at least 20% of individuals in both populations. Only one SNP was processed per locus to avoid SNPs known to be closely linked; the SNP written was chosen at random. Equivalent parameters were used with raw sequence without mapping to a reference genome in the *denovo_map.pl* pipeline for confirmation of population analyses. AMOVA-corrected $F_{ST}$ values for each locus were collected for downstream analysis.

SNP data were imported to Arlequin version 3.5 (Excoffier and Lischer 2010), using the PGDSpider tool (Lischer and Excoffier 2012) and pairwise $F_{ST}$ values between populations were found, with significance tested by permuting individuals between populations (Excoffier and Lischer 2010). $F_{IS}$ values for each population were also calculated.

Structure analysis (Pritchard et al. 2000) (version 2.3.4) was performed on SNPs determined from the Stacks pipeline. The value of K, or number of populations, was varied from 1 to 6 with three replications of each K value. Structure was set with 10,000 burn-ins followed by 50,000 Markov Chain Monte Carlo repetitions. Outputs were sent to Clumpak (Kopelman et al. 2015) for visualization and Structure Harvester (Earl and vonHoldt 2012) to determine the optimal value of K (Evanno et al. 2005).

Principle Component Analysis (PCA) was done with the R package SNPRelate (Zheng et al. 2012). AMOVA $F_{ST}$ values for each locus obtained from the Stacks pipeline were mapped in a Manhattan plot by the R package qqman (Turner 2014).

# CHAPTER THREE: RESULTS

## Sequencing Data

A total of 294,780,458 reads were obtained from the sequencing runs. 146,830,949 reads were retained; the others were discarded as they were missing an *Eco*RI cut site, had an ambiguous barcode, or had an uncalled base. Samples obtained from blood had higher numbers of reads than samples obtained from liver (Figure 2). Bird species, including the Lesser Snow Goose, have nucleated erythrocytes with low gene expression (Quinn 1988). Extractions performed on blood samples from birds give high concentrations of DNA with relatively low shearing. Liver samples do not provide the same DNA extraction quality, and this was shown to carry through to sequences obtained from those samples. After discarding 16 individuals with fewer than 100,000 total reads, 119 individuals were carried forward (Table 1).

## Loci

To assess which sequence reads belongs to the same loci, the Stacks pipeline (Catchen et al. 2011) was used. Stacks identifies identical reads within individuals and assembles them into "stacks", which are then compared between individuals and populations to identify SNPs between those. Through the Stacks pipeline, 53,894 loci were found. For comparison, the *denovo_map.pl* pipeline had 54693 loci passing the same filters. SNPs were obtained from 3854 loci shared between the blue and white phases; 4421 SNPs were obtained between the four locations, with the difference in count between the color and location analyses owing to the location parameters requiring fewer individuals to pass a SNP though the 20% population filter.

Population Analysis

In order to assess whether there is separation between the color phases or migratory path breeding grounds, differentiation between populations and inbreeding within populations were determined by $F_{ST}$ and $F_{IS}$ values, respectively. The $F_{ST}$ values between the two color phases ranged from 0 to 0.19 for individual SNPs, with an overall average $F_{ST}$ of 0.013. SNPs were located on every numbered macrochromosome, including the Z (Figure 3). Between locations, pairwise $F_{ST}$ values were all calculated as below 0 (Table 2A); this held true with the locations broken down by color phase as well (Table 2B). As $F_{ST}$ values can range from 0 to 1, all values were essentially considered equal to 0; none were significantly different from 0. $F_{Is}$ values, commonly known as inbreeding coefficients, calculated for each population were also low, with only AR samples above 0; no $F_{Is}$ values reached statistical significance either. These data do not show any indication of inbreeding or isolation of populations.

Another way to determine whether there are distinct populations existing within the Lesser Snow Goose, is to perform a Structure analysis. Structure is used to differentiate between populations in a dataset and subsequently assign individuals to the populations to which they belong without regard for the original populations. Structure is run with various values of K, which are the number of populations allowed in a particular run of the model. Structure analysis (Figure 4) showed no evidence of differentiation between the locations or color phases; between K=1 and K=6, every sequenced individual was more likely assigned to a single large population than any other outcome, indicating no barrier to gene flow by color or location. Log likelihood values are not known to properly indicate the most likely population structure (Evanno et al. 2005), so the $\Delta K$ measurement, which has been shown to be more accurate, was used. Structure Harvester output indicated K=2 was the most likely population value (Figure 5). Structure Harvester requires K±1 to calculate probability; K=0 is impossible to analyze and therefore K=1 cannot be the most likely Structure by that calculation. Based on raw log likelihood values, K=2

is also the most likely value (Figure 6). Evanno and colleagues note that the log likelihood value typically is not maximized at the correct value of K, instead continuously increasing until reaching the correct value, then briefly leveling off with a slight increase in slope before a fall in likelihood. The latter is seen in the log likelihoods of the Structure of the Lesser Snow Goose (Figure 6). In all K values tested, every individual was more likely assigned to a single large population than all other possible populations combined. These data, taken together, indicate an optimal Structure of K=1.

Principle Component Analysis

To further search for differentiation between any groupings and verify the findings of the Structure analysis, a PCA was used. PCA on the 3768 SNPs which passed population filters (Figure 7) did not reveal any differentiation between samples. No populations or colors separated along any of the first four principal components, none of which explained more than 1.79% of the total variance. In all comparisons of the first four components, one major cluster was observed. The first four principal components did not reveal any differentiation between individuals, which is in agreement with the results of the Structure analysis.

# CHAPTER FOUR: DISCUSSION

My data do not provide support for the hypothesis of genetic structure in the Lesser Snow Goose, neither between blue and white phases nor across the geographical range of the species. Gene flow among breeding populations of Lesser Snow Goose occurs due to the proximity of birds from several colonies on the wintering grounds encouraging migration and interbreeding (Cooke et al. 1975). Geographical or reproductive isolation between populations or color phases has been considered throughout the literature, yet nuclear genetic data on this scale were not previously available to test these assertions, and only the single study of six allozymes from a single population offered support of the hypothesis of former allopatry of the two color phases (Cooke et al. 1988).

Previous studies on the population dynamics of the Lesser Snow Goose have described assortative mating of the color phases caused by imprinting, and most pair bonds each year are between same-phase birds (Cooke et al. 1982). Regardless, these date show that this barrier between color phases has not been strong enough to be reflected the genetic level. Every measure tested here was consistent with panmixia of the color phases. The overall $F_{ST}$ comparisons between colors did not differ significantly from 0. The Structure and PCA results also did not show differentiation between any individuals or populations. While 122 SNPs had $F_{ST}$ values significantly different from 0 at the 5% level between the color phases, this represents only 3% of the 3854 SNPs tested, fewer than should occur randomly. Previously, a hypothesis that the two color phases were once allopatric was described (Cooke et al. 1988). With the benefit of a larger

sample size, both in loci and individuals, no such history can be detected. It is unlikely that significant population changes occurred between the prior study and this one, as the samples used for both studies were acquired within 5 years of each other. With the exception of the *MC1R* gene on which the dimorphism is based, where the $F_{ST}$ values were 0.78 and 0.56 respectively between the two phases at two different colonies (Mundy et al. 2004), every measure tested here was consistent with panmixia of the color phases.

The finding that the Lesser Snow Goose is panmictic between its two color phases is unexpected, yet not inexplicable. Families of geese in the breeding grounds lack complete fidelity. Nest parasitism, when a female lays eggs in a nest which another goose subsequently incubates, is prevalent (Rockwell et al. 1983). The nest parasitism affects an average of 22% of nests in a given year, varying with the amount of quality nesting space available; years with less quality space yielded greater degrees of parasitism (Lank et al. 1989). Additionally, the existence of mating outside the pairing system causes goslings to imprinting on geese other than their own genetic parents (Mineau and Cooke 1979). Rarely, this phenomenon extends to nests with two females and no males present, with eggs from both mothers present, potentially from different males (Quinn et al. 1989). Disturbances to the nesting area, such as a predator or a wandering researcher, cause the geese in the area to temporarily scatter, and within the first few days of life, before imprinting has occurred, the goslings can get rearranged as they return to family units without rejection by the potentially new parents (Prevett and MacInnes 1980). There is a considerable rate of extra-pair fertilization and intraspecific brood parasitism which contributes to the mixing of the color phases within a given population (Quinn et al. 1987). While lower than the predictions of Hardy-Weinberg equilibrium, rates of mixed-phase pairing are higher than would be expected in a species that assortatively mates based on an individual's own coloration, rather than imprinted from the parents, as is the Lesser Snow Goose (Cooke et al. 1982). The finding of panmixia in the Lesser Snow Goose is therefore understandable in the

context of a parental imprinting system. There is reason to doubt that the color phases are strongly isolated on a timescale crossing many generations.

There are two migratory pathways used by the Lesser Snow Goose, the Pacific Flyway, and the Central and Mississippi Flyways, and they are considered geographically separate from each other, as are the wintering grounds to which they lead (Ogilvie 1978). The most obvious consequence of the stated geographical barrier is the lack of blue geese in the Pacific Flyway populations, although there have been blue phase sightings on a small scale in the Californian wintering grounds (<1%) (McLandress and McLandress 1979). However, the eastern and western locations were also found to be panmictic here, casting doubt on the efficacy of any barriers between pathways.

When mtDNA was compared, two major clades emerged (Avise et al. 1992; Quinn 1992). Both of the clades were distributed in both color phases (Avise et al. 1992; Quinn 1992). The different mtDNA clades were also distributed across locations from Wrangel Island to La Perouse Bay (Quinn 1992). Based on these studies, a hypothesis was formed which states that there were two major populations in separate glacial refugia in the Pleistocene (350,000 years ago), one of which spread across the continent and was subsequently separated into the two current geographic populations (east and west) while the other spread into the populations more recently (Quinn 1992).

When the evidence in this study is taken with the two clades which encompass the same geographic range in mtDNA, little support is lent to any extant geographical barriers to gene flow (Avise et al. 1992; Quinn 1992). Other than at the *MC1R* gene, on which the assortative mating is based, there is no evidence of any hindrance of gene flow. The dearth of blue individuals in the west may be a result of the sexual imprinting, where a lack of blues in effect self-replicates, as few to no western geese would have a blue preference imprinted upon them. This in turn would reduce the movement of blue phase birds into the western grounds, as western geese would not

choose any available blue mates, and blue geese would not enter the colonies. While preventing the expansion of the blue phase into the west, gene flow has not been prevented at other loci, indicating that geography is not a significant barrier.

As new feeding opportunities have opened up in the southern United States, the historical wintering grounds have shifted, providing new mating opportunities to previously isolated Lesser Snow Goose populations (Cooke et al. 1988). Whether or not the populations were isolated prior to the increased farmland presence is another question. If there were once strong geographical barriers separating the two flyways, the data presented here and in other work (Geramita et al. 1982) indicate that there is migration between the two. Prior studies into the mitochondrial genetics of the Lesser Snow Goose found two clades spread over the entire geographic range, further indicating that the geographical barrier is likely not absolute (Avise et al. 1992; Quinn 1992). These two clades are thought to reflect a historical, rather than present-day, geographical barrier which existed in the Pleistocene era (Avise et al. 1992; Quinn 1992).

A previous study described genetic isolation between two populations occupying Wrangel Island, termed the Northern Wrangel Island (NWI) and Southern Wrangel Island (SWI) populations. The WI samples studied here were obtained from the livers of hunter-killed geese, and whether they had the facial staining indicative of the NWI population or not is unknown; the samples could be from either population based on their collection locations. If any samples from WI sequenced here were in the NWI population, no second population was detected by Structure analysis nor by PCA. Ascertaining which WI samples are from NWI or SWI is not feasible; therefore, evidence for or against the prior hypothesis cannot be properly determined here.

The Lesser Snow Goose has become an environmental troublemaker due to overpopulation, causing issues on the breeding grounds for other species. For instance, populations of savannah sparrows declined 77% in LPB over the course of 36 years due to the foraging of the Lesser Snow Goose (Peterson et al. 2014). Despite attempts to increase hunter

kills to reduce the population of geese, as well as an increase in predation by polar bears forced ashore by climate change, the populations continue to grow (Alisauskas et al. 2011; Rockwell and Gormezano 2008). New feeding opportunities in the wintering grounds allowed for a population expansion by allowing a greater proportion of geese to survive the winter (Rockwell et al. 2003). This has resulted in considerable damage to the northern breeding grounds, where the same increase in food supply did not occur to match the increased demand. The population patterns outlined here may indicate that attempts to limit the numbers of the Lesser Snow Goose are hampered by additional migrants from other colonies. This case is reminiscent of the European eels of the Sargasso Sea, which also exhibit panmixia across their range (Als et al. 2011). As a panmictic species, the European eels require an international effort toward their conservation, given that actions in one region affect the entire range (Als et al. 2011), and the same can likely be said toward the management of the Lesser Snow Goose; the destruction of habitat occurs in the Canadian and Russian Arctic but is ultimately caused by farming in the United States. Control efforts for *C. c. caerulescens* will require efforts from numerous agencies and countries, if the destructive effects of the overabundance are to be reversed.

My findings indicate that the Lesser Snow Goose is a panmictic species, and do not support previous hypotheses that the two color phases were allopatric nor that the eastern and western migratory pathways are reproductively isolated. Given that positive assortative mating is not typically sufficient to cause speciation absent natural selection or another supporting driving force, this finding is not surprising. Between the eastern and western migratory pathways, previous research has been mixed on whether the two are isolated, and this study strongly supports movement and gene flow between the two.


**Figure 1**. Sample locations. Pie charts show proportion of blue birds between 1969 and 1971 (Cooke et al. 1975).

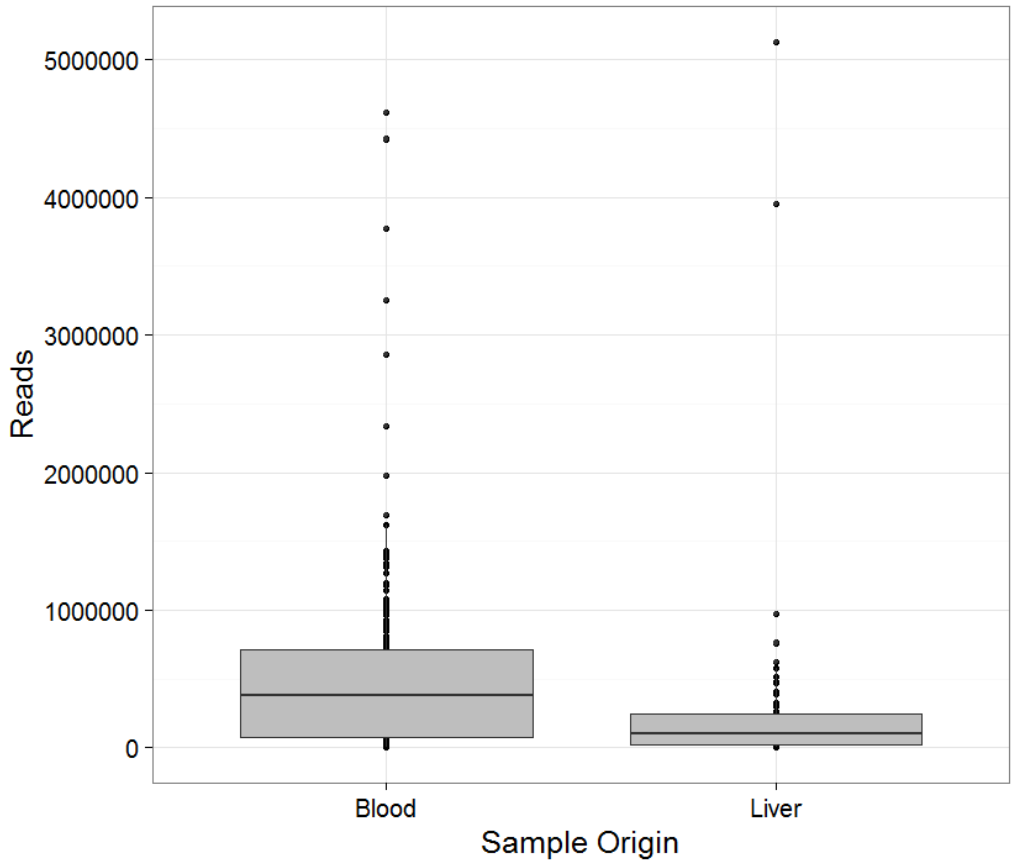


**Figure 2**. Sequencing reads by sample tissue origin. Samples obtained from blood (n=215, $\bar{x}$=530625) obtained higher read counts than liver samples (n=92, $\bar{x}$=263145) (ANOVA, df=1, F=9.272, p=0.003). Data are pooled from multiple runs; individuals sampled multiple times are counted independently for this analysis.

**Table 1**. Number of individuals included in analysis, broken down by color phase and breeding location.

| | White Individuals | Blue Individuals | Total Individuals |
|---|---|---|---|
| Anderson River | 21 | 0 | 21 |
| Cape Henrietta Maria | 7 | 16 | 23 |
| La Perouse Bay | 26 | 19 | 45 |
| Wrangel Island | 30 | 0 | 30 |
| Total | 84 | 35 | 119 |

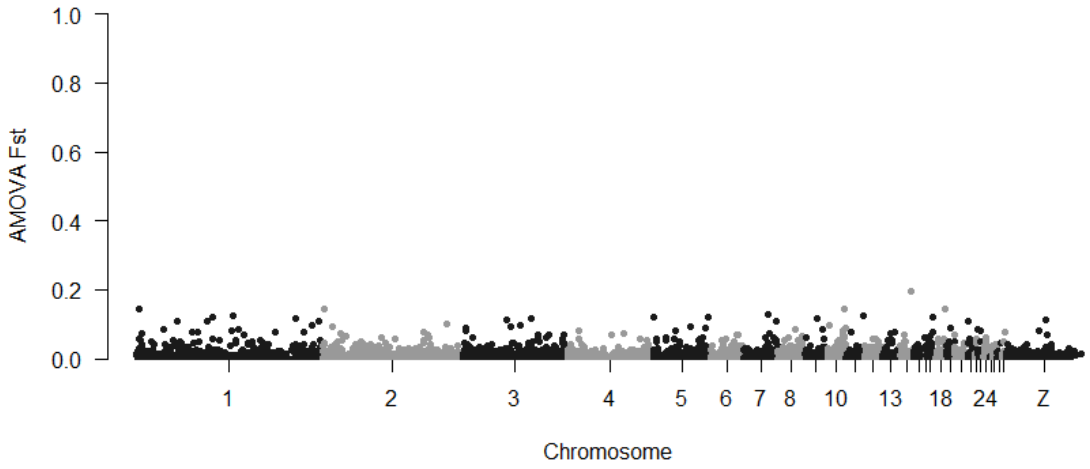


**Figure 3.** AMOVA $F_{ST}$ values for individual loci between blue and white phases for SNPs mapped by chromosome. Colors alternate by chromosome. Not shown: microchromosomes; no microchromosome had more than one SNP present.

**Table 2**. (A) Population-wide $F_{ST}$ and $F_{IS}$ values. Below diagonal are pairwise $F_{ST}$ comparisons between populations. Values below zero are the result of higher variation within populations than between them, and should be considered equal to zero. Grey values on the diagonal are $F_{IS}$ values for each population. Parenthetical values are P values. (B) $F_{ST}$ and $F_{IS}$ values with mixed populations separated by color.

| **A** | Anderson River | Cape Henrietta Maria | La Perouse Bay | Wrangel Island |
|---|---|---|---|---|
| Anderson River | 0.04061 (.09) | | | |
| Cape Henrietta Maria | -0.04153 (.99) | -0.08337 (.97) | | |
| La Perouse Bay | -0.00661 (.99) | -0.02461 (.99) | -0.10992 (1) | |
| Wrangel Island | -0.14899 (.85) | -0.07328 (.98) | -0.12062 (.99) | -0.08538 (1) |

| **B** | AR White | CHM White | CHM Blue | LPB White | LPB Blue | WI White |
|---|---|---|---|---|---|---|
| AR White | 0.0406 (.08) | | | | | |
| CHM White | -0.078 (.99) | -0.0044 (.54) | | | | |
| CHM Blue | -0.030 (.97) | -0.009 (.68) | -0.1107 (.97) | | | |
| LPB White | -0.014 (.97) | -0.032 (.97) | -0.010 (.87) | -0.0517 (.98) | | |
| LPB Blue | -0.018 (.92) | -0.113 (.97) | -0.039 (.92) | -0.038 (.97) | -0.1270 (.99) | |
| WI White | -0.149 (.99) | -0.036 (.98) | -0.091 (.99) | -0.068 (.99) | -0.247 (.99) | -0.0854 (1) |

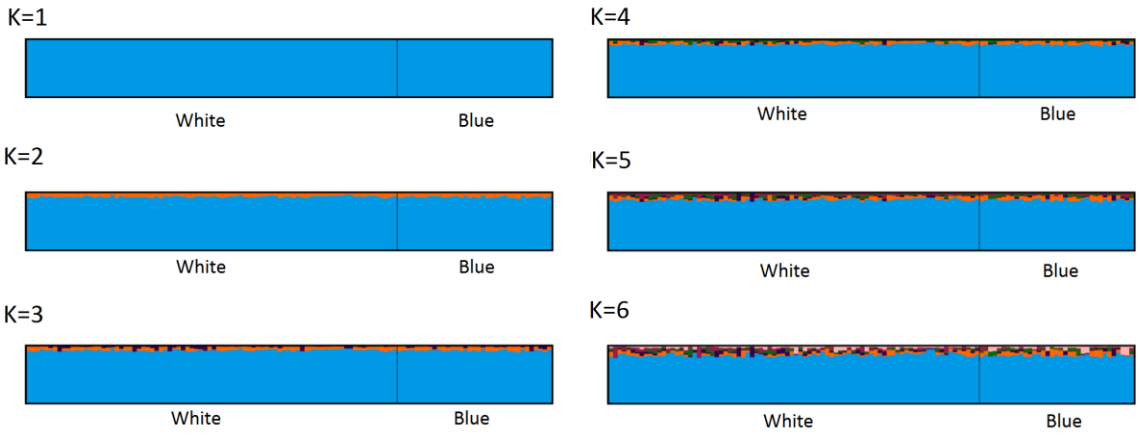


**Figure 4.** Structure analysis of snow geese by color. Individuals are arranged as vertical bars, with populations found by Structure represented as different colors. The relative size of the color indicates the probability the individual should be assigned to the population. Structures are the average of 5 replicate runs of the analysis.

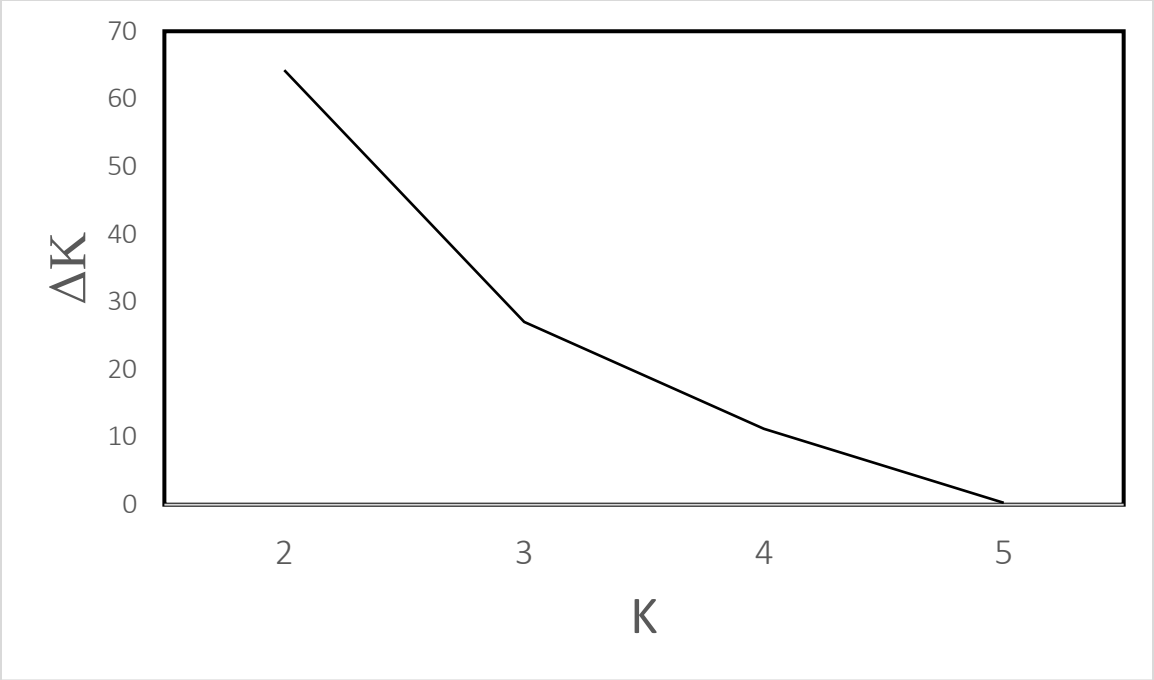


**Figure 5.** ΔK values by K (populations) value from Structure analysis. ΔK for K=1 cannot be calculated.

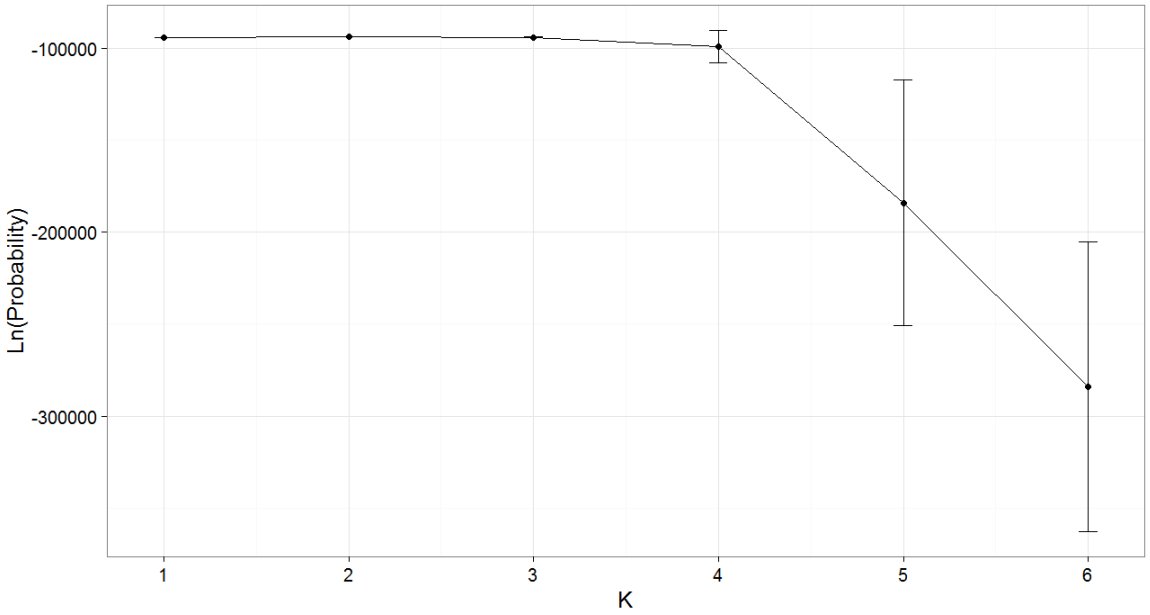


**Figure 6.** Raw log likelihood of each Structure output K. Error bars are 95% CI.

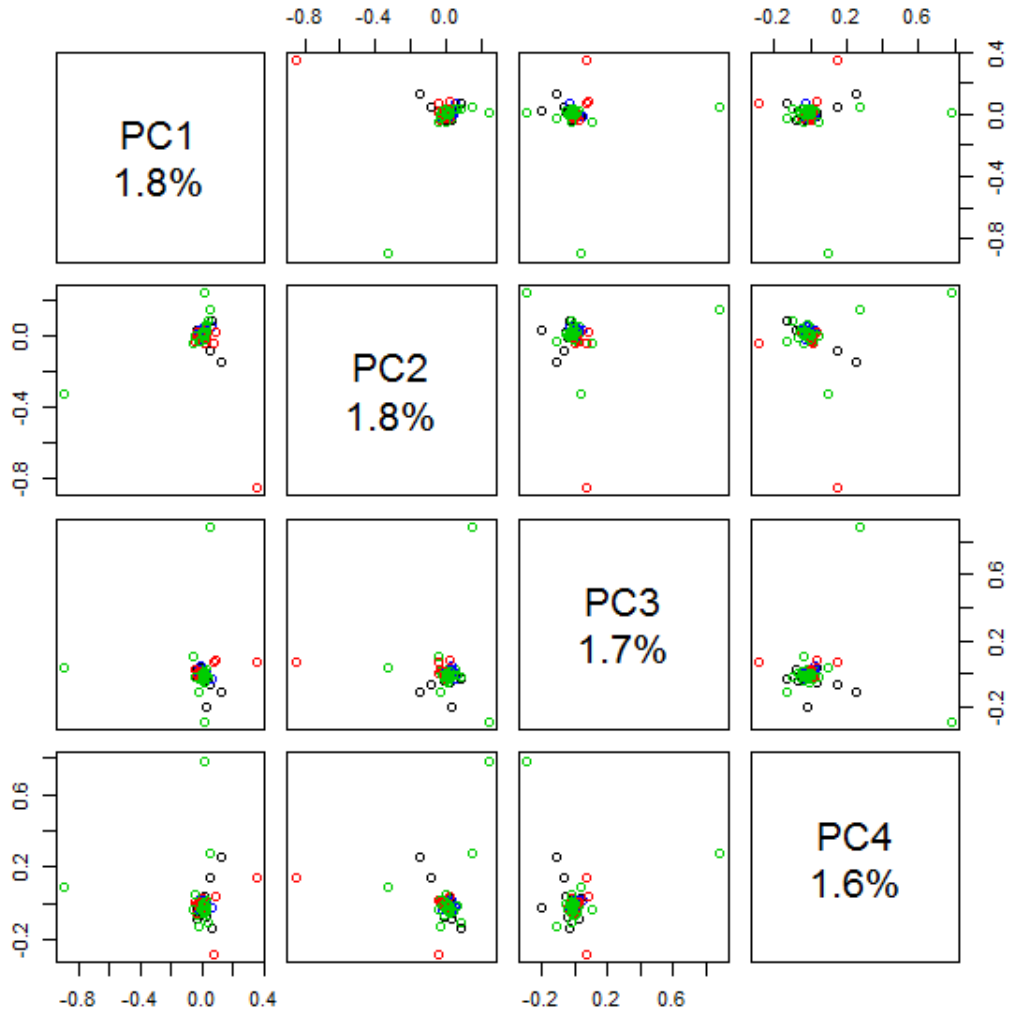


**Figure 7.** Principal Coordinate Analysis of SNP data by population location. Black circles = AR, Red = CHM, Green = LPB, Blue = WI.

# BIBLIOGRAPHY


Alisauskas RT, Rockwell RF, Dufour KW, Cooch EG, Zimmerman G, Drake KL, Leafloor JO, Moser TJ, Reed ET. 2011. Harvest, survival, and abundance of midcontinent Lesser Snow Geese relative to population reduction efforts. Wildlife Monographs 179(1):1-42.

Als TD, Hansen MM, Maes GE, Castonguay M, Riemann L, Aarestrup K, Munk P, Sparholt H, Hanel R, Bernatchez L. 2011. All roads lead to home: panmixia of European eel in the Sargasso Sea. Molecular Ecology 20(7):1333-1346.

Avise JC, Alisauskas RT, Nelson WS, Ankney CD. 1992. Matriarchal population genetic structure in an avian species with female natal philopatry. Evolution 46(4):1084-1096.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PloS One 3(10):e3376.

Barluenga M, Stölting KN, Salzburger W, Muschick M, Meyer A. 2006. Sympatric speciation in Nicaraguan crater Lake Cichlid fish. Nature 439(7077):719-723.

Barnston G. 1860. Recollections of the swans and geese of Hudson's Bay. Ibis 2(3):253-259.

Catchen J, Bassham S, Wilson T, Currey M, O'Brien C, Yeates Q, Cresko WA. 2013a. The population structure and recent colonization history of Oregon threespine stickleback determined using restriction-site associated DNA-sequencing. Molecular Ecology 22(11):2864-2883.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013b. Stacks: an analysis tool set for population genomics. Molecular Ecology 22(11):3124-3140.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. 2011. Stacks: Building and genotyping loci *de novo* from short-read sequences. G3: Genes, Genomes, Genetics 1(3):171-182.

Cooch G. 1961. Ecological aspects of the blue-snow goose complex. Auk 78(1):72-89.

Cooke F, Abraham KF, Davies JC, Findlay CS, Healey RF, Sadura A, Seguin RJ. 1982. The La Perouse Bay Snow Goose Project - A 13-Year Report. Kingston, Ontario, Canada: Department of Biology, Queen's University.

Cooke F, Buckley PA. 1987. Avian Genetics: a Population and Ecological Approach. London: Academic Press. p. 488.

Cooke F, Cooch FG. 1968. The genetics of polymorphism in the goose *Anser caerulescens*. Evolution 22(2):289-300.

Cooke F, Finney G, Rockwell R. 1976. Assortative mating in Lesser Snow Geese (*Anser caerulescens*). Behavior Genetics 6(2):127-140.

Cooke F, MacInnes C, Prevett J. 1975. Gene flow between breeding populations of Lesser Snow Geese. Auk 92(3):493-510.

Cooke F, Mirsky PJ. 1972. A genetic analysis of Lesser Snow Goose families. Auk 89(4):863-871.

Cooke F, Parkin D, Rockwell R. 1988. Evidence of former allopatry of the two color phases of Lesser Snow Geese (*Chen caerulescens caerulescens*). Auk 105(3):467-479.

Earl DA, vonHoldt BM. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4(2):359-361.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14(8):2611-2620.
Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources 10(3):564-567.
Finney G, Cooke F. 1978. Reproductive habits in the Snow Goose: The influence of female age. Condor 80(2):147-158.
Geramita JM, Cooke F, Rockwell RF. 1982. Assortative mating and gene flow in the Lesser Snow Goose: a modelling approach. Theoretical Population Biology 22(2):177-203.
Hillier LW, Miller W, Birney E, Warren W, Hardison RC, Ponting CP, Bork P, Burt DW, Groenen MA, Delany ME. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432(7018):695-716.
Irwin DE, Price T. 1999. Sexual imprinting, learning and speciation. Heredity 82(4):347-354.
Kahn NW, St. John J, Quinn TW. 1998. Chromosome-specific intron size differences in the avian CHD gene provide an efficient method for sex identification in birds. Auk 115(4):1074-1078.
Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I. 2015. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Molecular Ecology Resources 15(5):1179-1191.
Lank DB, Cooch EG, Rockwell RF, Cooke F. 1989. Environmental and demographic correlates of intraspecific nest parasitism in Lesser Snow Goose *Chen caerulescens caerulescens*. Journal of Animal Ecology 58(1):29-44.
Lepais O, Weir JT. 2014. SimRAD: an R package for simulation-based prediction of the number of loci expected in RADseq and similar genotyping by sequencing approaches. Molecular Ecology Resources 14(6):1314-1321.
Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:1303.3997 00(00):1-3.
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup GPDP. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25(16):2078-2079.
Lischer HEL, Excoffier L. 2012. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics 28(2):298-299.
Manning TH, Höhn EO, Macpherson AH. 1956. The Birds of Banks Island. Ottawa: Le Ministere.
McLandress MR, McLandress I. 1979. Blue-phase Ross' Geese and other blue-phase geese in western North America. Auk 96(3):544-550.
Mineau P, Cooke F. 1979. Rape in the Lesser Snow Goose. Behaviour 70(3/4):280-291.
Moore WS. 1995. Inferring phylogenies from mtDNA bariation: mitochondrial-gene trees versus nuclear-gene trees. Evolution 49(4):718-726.
Mundy NI, Badcock NS, Hart T, Scribner K, Janssen K, Nadeau NJ. 2004. Conserved genetic basis of a quantitative plumage trait involved in mate choice. Science 303(5665):1870-1873.
Narum SR, Buerkle CA, Davey JW, Miller MR, Hohenlohe PA. 2013. Genotyping-by-sequencing in ecological and conservation genomics. Molecular Ecology 22(11):2841-2847.
Ogden R, Gharbi K, Mugue N, Martinsohn J, Senn H, Davey J, Pourkazemi M, McEwing R, Eland C, Vidotto M. 2013. Sturgeon conservation genomics: SNP discovery and validation using RAD sequencing. Molecular Ecology 22(11):3112-3123.
Ogilvie MA. 1978. Wild geese. Vermillion, South Dakota: T. & A. D. Poyser Limited. p. 350.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. PloS One 7(5):e37135.
Peterson SL, Rockwell RF, Witte CR, Koons DN. 2014. Legacy effects of habitat degradation by Lesser Snow Geese on nesting savannah sparrows. Condor 116(4):527-537.
Prevett J, MacInnes CD. 1980. Family and other social groups in Snow Geese. Wildlife Monographs 71:3-46.
Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics 155(2):945-959.
Quinn TW. 1988. DNA sequence variation in the Lesser Snow Goose, *Anser caerulescens caerulescens*. [Kingston, Ontario, Canada]: Queen's University.
Quinn TW. 1992. The genetic legacy of mother goose–phylogeographic patterns of Lesser Snow Goose *Chen caerulescens caerulescens* maternal lineages. Molecular Ecology 1(2):105-117.
Quinn TW, Davies JC, Cooke F, White BN. 1989. Genetic analysis of offspring of a female-female pair in the Lesser Snow Goose (*Chen c. caerulescens*). Auk 106(2):177-184.
Quinn TW, Quinn JS, Cooke F, White BN. 1987. DNA marker analysis detects multiple maternity and paternity in single broods of the Lesser Snow Goose. Nature 326(6111):392-394.
Quinn TW, White BN. 1987. Identification of Restriction-Fragment-Length Polymorphisms in genomic DNA of the Lesser Snow Goose (*Anser caerulescens caerulescens*). Molecular Biology and Evolution 4(2):126-143.
Rockwell RF, Findlay CS, Cooke F. 1983. Life history studies of the Lesser Snow Goose (*Anser caerulescens caerulescens*). I. The influence of age and time on fecundity. Oecologia 56(2/3):318-322.
Rockwell RF, Gormezano LJ. 2008. The early bear gets the goose: climate change, polar bears and Lesser Snow Geese in western Hudson Bay. Polar Biology 32(4):539-547.
Rockwell RF, Witte CR, Jefferies RL, Weatherhead PJ. 2003. Response of nesting savannah sparrows to 25 years of habitat change in a snow goose colony. Ecoscience 10(1):33-37.
Rohland N, Reich D. 2012. Cost-effective, high-throughput DNA sequencing libraries for multiplexed tarrget capture. Genome Research 22(5):939-946.
Saunders W. 1917. Wild geese at Moose Factory. Auk 34(3):334-335.
Shorey RI, Scribner KT, Kanefsky J, Samuel MD, Libants SV. 2011. Intercontinental gene flow among western arctic populations of Lesser Snow Geese. Condor 113(4):735-746.
Singer VL, Jones LJ, Yue ST, Haugland RP. 1997. Characterization of PicoGreen reagent and development of a fluorescence-based solution assay for double-stranded DNA quantitation. Analytical Biochemistry 249(2):228-238.
Steiner CC, Weber JN, Hoekstra HE. 2007. Adaptive variation in beach mice produced by two interacting pigmentation genes. PLoS Biol 5(9):e219.
Takahashi T, Sota T, Hori M. 2013. Genetic basis of male colour dimorphism in a Lake Tanganyika cichlid fish. Molecular Ecology 22(11):3049-3060.
Turner SD. qqman: an R package for visualizing GWAS results using QQ and manhattan plots. bioRxiv [Internet].005165.
Verzijden MN, Ten Cate C, Servedio MR, Kozak GM, Boughman JW, Svensson EI. 2012. The impact of learning on sexual selection and speciation. Trends in Ecology & Evolution 27(9):511-519.
Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics 28(24):3326-3328.