INVESTIGATING THE ROLE OF DISPERSAL ON THE GENETIC STRUCTURE OF WILD POPULATIONS

by

John Henry Powell III

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Fish and Wildlife Biology

MONTANA STATE UNIVERSITY
Bozeman, Montana

March 2014

©COPYRIGHT

by

John Henry Powell III

2014

All Rights Reserved

DEDICATION

I dedicate this dissertation to my wonderful wife Cris, for whom this journey has included multiple moves and many long days, and, who through it all has provided the love and support I needed to reach this point.

ACKNOWLEDGEMENTS

I would like to thank my advisors, Steven Kalinowski and Bob Gresswell, who brought me to Montana State University and fostered both my development as a geneticist and my understanding of what it means to be a scientist. In addition I would like to thank my committee members Megan Higgs, Robb Leary and Ed Adams. Megan has helped me realize my interest in applied statistics and patiently discussed the many data analysis ideas, of varying degrees of quality I have had along the way. Robb gave me my first paying job in science and has helped shape my understanding of hybridization more than anyone. And Ed has generously given his time to make sure that my comprehensive examination and defense were rigorous but reasonable.

Because my dissertation covers a broad range of species, thanks are due to my collaborators, Paul Cross, Bob Garrott, Clint Muhlfeld and Jay Rotella who afforded me the opportunity to explore diverse questions in population genetics. The National Science Foundation provided most of the financial support though a graduate research fellowship. Thanks also to Carol Johnson for helping me develop as an educator as well as a scientist.

And finally, I would like to thank my wife Cris and daughter Vida Eloise who have provided their love and support for me throughout my graduate education. My parents and brother helped instill a love for the outdoors and a curiosity of nature that provided the impetus for a career in science. And for my friends, Matt Corsi, Ann Marie Reinhold, John Syslo and Joe Naughton, who frequently served as the voice of reason for many of my ideas and were instrumental in helping me enjoy life in Bozeman.

TABLE OF CONTENTS

1. GENERAL INTRODUCTION 1

Literature Cited 6

2. ESTIMATING INBREEDING DEPRESSION IN A NON-PEDIGREED POPULATION OF WEDDELL SEALS *LEPTONYCHOTES WEDDELLII* 8

Contribution of Authors and Co-Authors 8
Manuscript Information Page 9
Abstract 10
Introduction 11
Methods 13
Study Area and Organism 13
Sample Selection and Genotyping 14
Two-Step Maximum Likelihood Analysis 16
Estimating Individual Inbreeding Coefficients 16
Modeling the Association Between Inbreeding and Fitness 18
Lifetime Reproductive Success 22
Survival Across the Study 23
Age at First Reproduction 24
Frequency of Reproduction 25
Results 25
Discussion 27
Literature Cited 35

3. MICROSATELLITES INDICATE MINIMAL BARRIERS TO MULE DEER *ODOCOILEUS HEMIONUS* DISPERSAL ACROSS MONTANA, USA 40

Contribution of Authors and Co-Authors 40
Manuscript Information Page 41
Abstract 42
Introduction 43
Material and Methods 44
Results 50
Discussion 53
Acknowledgements 55
References 61

TABLE OF CONTENTS – CONTINUED

4. HYBRIDIZATION DID NOT APPEAR TO INCREASE DISPERSAL IN A WESTSLOPE CUTTHROAT TROUT METAPOPULATION 66

Contribution of Authors and Co-Authors 66
Manuscript Information Page 67
Abstract 68
Introduction 69
Methods 72
Study Area 72
Molecular Methods 73
Preliminary Genetic Data Analysis 73
Statistical Analysis of Dispersal Patterns 75
F_{ST} Estimation in Hybridized Samples 77
Modified Algorithm 78
Mixture Model Performance 80
Results 82
Preliminary Genetic Data Analysis 82
Statistical Analysis of Dispersal Patterns 84
Discussion 85
Literature Cited 98

5. GENERAL CONCLUSIONS 102

Literature Cited 106

APPENDICES 107

APPENDIX A: A Test for Gene Tree Discordance at Diagnostic Loci 108
APPENDIX B: Landscape Genetic Analysis of Westslope Cutthroat Trout *Oncorhynchus clarkii lewisi* Populations in Glacier National Park 113

LITERATURE CITED 129

LIST OF TABLES

Table Page

2.1. Primer sequences and marker information for the 12 novel microsatellite loci used in this study. 32

2.2. Candidate regression models for each of the individual fitness surrogates ranked based on their relationship to the model with the minimum AICc value. Only those models that converged are shown. Explanatory variables included individual inbreeding coefficient (*INB*), birth location (*COL*) and a variable in the mean wait-time between pupping intervals models indicating whether the interval was following an individual seal's first pup (*FIRST*)....33

3.1. Estimated expected heterozygosity and allelic richness for the 14 loci included in the final analyses. Allelic richness was standardized to a sample size of 20 genes using rarefaction (Kalinowski 2005). Values are separated based on broad-scale sampling regions of Montana/Wyoming (MT/WY), Utah (UT), and Colorado (CO)....56

3.2. Haplotype frequencies from 76 mitochondrial control region sequences. Group 1 included 19 samples from northeastern Montana, Group 2 included 19 samples from the northwestern corner of Montana, Group 3 included 6 samples from west central Montana, Group 4 included 13 samples from southwestern Montana, and Group 5 included 19 samples from southeastern Montana (Figure 3.1)....57

3.3. Nucleotide diversity across 489 bp of the mitochondrial control region, with associated 95% confidence intervals, present in the five sampling groups located within Montana. Group 1 included 19 samples from northeastern Montana, Group 2 included 19 samples from the northwestern corner of Montana, Group 3 included 6 samples from west central Montana, Group 4 included 13 samples from southwestern Montana, and Group 5 included 19 samples from southeastern Montana (Figure 3.1)....58

LIST OF TABLES – CONTINUED

Table Page

3.4. Pairwise F_{ST} for mitochondrial sequences (above diagonal) and microsatellite genotypes (below diagonal) with values less than the Benjamini-Hochberg (1995) corrected alpha-levels in bold. Group 1 included 19 samples from northeastern Montana, Group 2 included 19 samples from the northwestern corner of Montana, Group 3 included 6 samples from west central Montana, Group 4 included 13 samples from southwestern Montana, and Group 5 included 19 samples from southeastern Montana (Figure 3.1)........58

4.1. Site information for the 17 sample locations from the Middle Fork Flathead River, with figure codes corresponding to Figure 4.1, and 95% confidence intervals for the estimated proportion of rainbow trout genes (P(RBT)) at a site and the mixing degree statistic (m_d) reported in brackets. 90

4.2. Locus names for diagnostic and polymorphic loci used in this analysis. 91

4.3. Observed bias in estimates of Weir and Cockerham's θ calculated from the observed sample (Naïve), the posterior medians from the Bayesian mixture model (Bayesian), and the sample after removing all rainbow trout alleles (True). 92

4.4. Observed variance in estimates of Weir and Cockerham's θ calculated from the observed sample (Naïve), the posterior medians from the Bayesian mixture model (Bayesian), and the sample after removing all rainbow trout alleles (True). 92

B.1. Information for the 16 sample locations from the North Fork Flathead River and Middle Fork Flathead River, with figure codes corresponding to Figure B.1. 121

B.2. Locus names for the SNP loci used in this analysis. 121

LIST OF TABLES – CONTINUED

Table Page

B.3. Pairwise θ (Weir and Cockerham 1984) with site labels corresponding to Figure B.1 and Table B.1. 122

LIST OF FIGURES

Figure Page

2.1. Lifetime reproductive success, measured as an individual based instantaneous population growth rate given their individual inbreeding coefficients. The solid line represents the estimated association between individual inbreeding coefficient and the lifetime reproductive success of a seal based on a truncated-Normal regression model fit to censored data, with the 95% prediction band for this analysis denoted by the gray shading and the 95% confidence band denoted by the dotted lines..34

3.1. Locations for 359 mule deer samples collected at hunter check stations and used in final analyses. Open circles indicate samples genotyped at microsatellite loci only, while closed circles indicate samples genotyped at microsatellite loci and at 489 bp of the mitochondrial control region. Closed squares represent samples from Utah, and closed triangles represent samples from Colorado both of which only had microsatellite loci genotyped. Numbers indicate the geographical subgroup designations used for analyses of sex-biased dispersal................................59

3.2. Visualization of the relationship between allele sharing distance (ASD) for pairs of individual mule deer and Euclidean distance in kilometers. The regression line is plotting the function $ASD = \frac{e^{0.8921+5.722*10^{-5}(Euclidian\ distance)}}{1+e^{0.8921+5.722*10^{-5}(Euclidian\ distance)}}$..60

3.3. Plot of the first two coordinates of the principal coordinate analysis run in GENALEX version 6.41 (Peakall and Smouse 2006). Open circles represent samples from the Montana/Wyoming sampling region (n = 320), closed squares represent samples from the Utah sampling region (n = 29), and closed triangles represent samples from the Colorado sampling region (n = 10)..60

LIST OF FIGURES – CONTINUED

Figure Page

4.1. Map of the sample locations in the Middle Fork Flathead River drainage, with site labels corresponding to Table 4.1. 93

4.2. Isolation by distance for all putatively flow connected sites in the Middle Fork Flathead River drainage after removing the samples from Upper Park Creek and Autumn Creek. The dashed line plots the estimated association between genetic distance and geographic distance among hybridized sites, represented by open circles, whereas the solid line plots the estimated association among sites where at least one sample did not contain hybridized individuals. This relationship was estimated using both the pairs where a single site did not contain hybridized individuals, represented by closed circles, and pairs where neither site contained hybridized individuals, represented by gray triangles. 94

4.3. Isolation by distance for all putatively flow connected sites in the Middle Fork Flathead River drainage. The dashed line plots the estimated association between genetic distance and geographic distance among hybridized sites, represented by open circles, whereas the solid line plots the estimated association among sites where at least one sample did not contain hybridized individuals. This relationship was estimated using both the pairs where a single site did not contain hybridized individuals, represented by closed circles, and pairs where neither site contained hybridized individuals, represented by gray triangles. 95

4.4. Histograms of the number of individuals in a sample with an estimated proportion of rainbow trout genes (P(RBT)) in their genome. Individual sites are plotted in each sample, labeled with the site number that corresponds to Figure 4.1 and Table 4.1. The estimated proportion of rainbow trout genes at the site is plotted as a point along the x-axis for reference. 96

LIST OF FIGURES – CONTINUED

Figure Page

4.5. Relationship between the average proportion of rainbow trout genes (P(RBT)) and the ratio of $\frac{F_{ST}}{1-F_{ST}}$ between the pairs of sites after removing samples from Upper Park Creek and Autumn Creek from analysis. The filled circles represent comparisons between samples where at least one site did not contain hybrid individuals, whereas the open circles represent comparisons between hybridized sites. 97

B.1. Map of the sample locations in the North Fork Flathead River and Middle Fork Flathead River drainages, with site labels corresponding to Table B.1. 123

B.2. Stream tree based on pairwise θ values, with site labels corresponding to Figure B.1 and Table B.1. The approximate location of barriers is indicated by gray dashes across the channel. Multiple barriers within the same segment were condensed to a single marker to increase readability. 124

B.3. Distribution of pairwise θ between the samples collected in I) Upper Muir Creek, II) Upper Park Creek, and III) Autumn Creek and other connected sites (gray bars). The solid line plots the kernel density of pairwise θ values among connected sites, and the dashed line plots the kernel density of pairwise θ values between geologic barrier isolated sites and connected sites. 125

ABSTRACT

The movement of individuals among populations (dispersal) is an ecological process that can affect the genes of populations. Identifying the role this process plays in the wild can be difficult due to uncertainty caused by other genetic and ecological processes. Here, I present three studies investigating the role of dispersal in wild populations of Weddell seals (*Leptonychotes weddellii*), mule deer (*Odocoileus hemionus*) and westslope cutthroat trout (*Oncorhynchus clarkii lewisi*). Each of these studies presents, to my knowledge, novel statistical approaches within this discipline that account for the uncertainty caused by other genetic and ecological processes when investigating the role of dispersal. In the second chapter I present a study that uses a two-step maximum likelihood analysis, fit with the data cloning algorithm, to incorporate uncertainty in estimating the probability that an individual has multiple copies of the same gene from a given ancestor into estimates of the association between this probability and the number of pups a female Weddell seal will produce in her life. This study did not find strong support for the hypothesis that increased similarity of an individual's genes reduced the number of pups she would produce across her lifetime. In the third chapter I use individual based genetic distance measures to investigate the association between features of the landscape and the genetic similarity of mule deer. This study found no detectable barriers to dispersal of mule deer across Montana. The fourth chapter presents a study comparing the dispersal rate between samples of westslope cutthroat trout from Glacier National Park that have interbred with introduced rainbow trout (*Oncorhynchus mykiss*) to samples that have not. There was no detectable difference in the dispersal rate of westslope cutthroat trout that have interbred with introduced rainbow trout relative to those that have not. In conclusion, dispersal is an important ecological process affecting the genes of populations, but understanding the role it plays across a landscape requires adequately accounting for the uncertainty due to other genetic and ecological processes. The three studies presented highlight different ways of addressing this problem when investigating the role of dispersal in the wild.

CHAPTER 1

GENERAL INTRODUCTION

Studies in population genetics can be broadly classified as concerning one of two main topics; the effects of genetic processes on populations and the effects of ecological processes on allele frequencies (Hartl and Clark 2007). An example of a study falling into this first category is Kondrashov's (1988) examination of the effects of synergistic epistasis and the deleterious mutation rate on the evolution of sex. In contrast, Sewall Wright's investigation of the effects of restricted dispersal on genetic distance between individuals in a continuous landscape (Wright 1943) would highlight a study falling into the second category. And, while there is no way to isolate the effects if any one genetic or ecological process from the others, by and large this dissertation is an investigation of the effects of ecological processes on the allele frequencies in populations.

Specifically, I present three studies that investigate the effects of the process of migration either at the individual level (Chapter Two), population level (Chapter Three) or species level (Chapter Four). Because of differences in the ecological and genetic definition of migration, in this dissertation I will refer to this process by the ecological term dispersal to highlight that I am investigating the effects of the movement of individuals between populations that ends in successful reproduction. Focusing on the scale of individuals in a population, chapter two examines the effects of restricted dispersal, which can lead to increased relatedness within a finite population, on the lifetime reproductive success of individual Weddell seals (*Leptonychotes weddellii*) from

a population in Erebus Bay, Antarctica. At the scale of populations, chapter three investigates the association between the landscape and inter-individual genetic distance in a population of mule deer (*Odocoileus hemionus*) in Montana. At the species scale, chapter four compares dispersal rates in non-hybridized westslope cutthroat trout samples (*Oncorhynchus clarkii lewisi*) to samples that have introgressed with non-native rainbow trout (*Oncorhynchus mykiss*) collected in tributary streams and lakes to the Middle Fork Flathead River.

In addition to investigating the effects of dispersal at the individual, population, and species level, each chapter also introduces statistical methods that address current problems with accounting for uncertainty in parameter estimates. In each case these methods increase the associated uncertainty in parameter estimates. Therefore, while it may appear that each study is trying to remove all possible hope of finding a statistically significant association between dispersal and the parameter of interest, as Clopper and Pearson (1934) pointed out, there is little value in an estimate without a true idea of its uncertainty.

When dispersal is restricted, genetic drift in finite populations will lead to individuals having alleles identical by descent in their genome (Hartl and Clark 2007). This increased homozygosity due to the combination of isolation and finite population size increases the number of recessive deleterious alleles in a homozygous state, which is the major cause of inbreeding depression (Charlesworth and Charlesworth 1999). The best method for testing the association between inbreeding and fitness is to use pedigree information to estimate individual inbreeding coefficients (Pemberton 2004). However,

in wild populations pedigrees are not always available. Therefore, because of the relationship between genome-wide heterozygosity and inbreeding coefficients, researchers have developed methods of estimating inbreeding depression from correlations between multi-locus heterozygosity and the fitness surrogates (Szulkin et al. 2010). However, simulation and empirical work all demonstrate a weak correlation between multilocus heterozygosity and individual inbreeding coefficients (Balloux et al. 2004, Slate et al. 2004). This low association between multilocus heterozygosity and inbreeding coefficients highlights the need to incorporate the known error in estimating inbreeding coefficients when testing for the presence of inbreeding depression in wild populations that do not have pedigree data. Chapter two addresses this need by presenting a two-step analysis that accounts for the multilevel nature of the data. This modeling approach, fit using the data cloning algorithm (Lele et al. 2007), incorporates estimation error into tests of inbreeding depression in a wild population of Weddell seals from Erebus Bay, Antarctica.

Moving up in the biological hierarchy to the level of populations, the effects of dispersal can be analogous to mutation in its ability to introduce novel genetic variation into a population (Hartl and Clark 2007). While chapter three was not developed to specifically address the introduction of novel genetic variation into a population, the hope was to use patterns of genetic structure to infer routes of dispersal that may be important for potential chronic wasting disease transmission and management in a mule deer population in Montana. Because the sampling units of this study were individual mule deer, we wanted to perform inter-individual tests for an association between genetic

distance and landscape features across Montana. However, an underlying assumption of one of the most common inter-individual tests (isolation by distance) is that the relationship between inter-individual genetic distance and geographic distance is linear, which is not the case (Graves 2012). Graves (2012) addressed this non-linearity by applying an analogous test to those used for constructing semi-variograms in spatial statistics. Chapter three presents an alternate method for estimating inter-individual isolation by distance that uses Mantel tests to construct a confidence interval around the estimated association between the probability that two individuals share an allele at a locus and their geographic separation.

At the highest level of biological organization studied in this dissertation, chapter four investigated whether hybridization between species may affect the rate of dispersal among native westslope cutthroat trout populations. Because of widespread introgression in westlsope cutthroat trout, this subspecies of cutthroat trout is at risk for genomic extinction (Allendorf et al. 2001). In addition, hybridization appears to be spreading in the face of reduced fitness of hybrid individuals (Allendorf et al. 2004). One hypothesis for this continued spread of hybridization is that hybrid individuals have a higher dispersal rate than native westslope cutthroat trout (Allendorf et al. 2004). Chapter four presents an extension of the *Structure* algorithm (Pritchard et al. 2000) allowing for the estimation of genetic distance between hybridized and non-hybridized samples. This genetic distance is used to investigate the patterns of isolation by distance among hybridized, and non-hybridized, sites. By comparing these patterns of isolation by distance, chapter four presents a direct test to the hypothesis that low levels of

hybridization increase dispersal rates in westslope cutthroat trout populations. A small simulation study is also undertaken to investigate the bias and variance of estimates of F_{ST} in hybrid populations.

Finally, chapter five discusses the information gained from each of these studies. As a note, because each of these studies was developed for publication, one of which is already in print, I use the pronoun “we” throughout chapter two, three and four to acknowledge the contributions of the co-authors listed at the beginning of these chapters.

Literature Cited

Allendorf F. W., R. F. Leary, N. P. Hitt, K. L. Knudsen, L. L. Lundquist, and P. Spruell. 2004. Intercrosses and the U.S. Endangered Species Act: should hybridized populations be included as westslope cutthroat trout? Conserv. Biol. 18:1203-1213.

Allendorf F. W., R. F. Leary, P. Spruell, and J. K. Wenburg. 2001. The problem with hybrids: setting conservation guidelines. Trends Ecol. Evol. 16:613-622.

Balloux, F., W. Amos, and T. Coulson. 2004. Does heterozygosity estimate inbreeding in real populations? Mol. Ecol. 13:3021-3031.

Charlesworth, B., and D. Charlesworth. 1999. The genetic basis of inbreeding depression. Genet. Res. 74:329-340.

Clopper, C. J., and E. S. Pearson. 1934. The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika 26:404-413.

Graves, T. A. 2012. Spatial ecology of grizzly bears in northwestern Montana and estimating resistance to gene flow. Ph.D. Dissertation, Northern Arizona University, Flagstaff, A.Z.

Hartl, D. L., and A. G. Clark. 2007. Principles of population genetics 4th ed. Sinauer Associates Inc., Sunderland, M.A.

Kondrashov, A. 1988. Deleterious mutations and the evolution of sexual reproduction. Nature 336:435-440.

Lele, S. R, B. Dennis, and F. Lutscher. 2007. Data cloning: easy maximum likelihood estimation for complex ecological models using Bayesian Markov chain Monte Carlo methods. Ecol. Lett. 10:551-563.

Pemberton, J. 2004. Measuring inbreeding depression in the wild: the old ways are the best. Trends Ecol. Evol. 19:613-615.

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

Slate, J., P. David, K. G. Dodds, B. A. Veenvliet, B. C. Glass, T. E. Broad, and J. C. McEwan. 2004. Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: theoretical expectations and empirical data. Heredity 93:255-265.

Szulkin M., N. Bierne, and P. David. 2010. Heterozygosity-fitness correlations: a time for reappraisal. Evolution 64:1202-1217.

Wright, S. 1943. Isolation by distance. Genetics 28:114-138.

CHAPTER TWO

ESTIMATING INBREEDING DEPRESSION IN A NON-PEDIGREED POPULATION OF WEDDELL SEALS *LEPTONYCHOTES WEDDELLII*

Contribution of Authors and Co-Authors

Manuscript in Chapter 2

Author: John H. Powell

Contributions: Developed study questions and appropriate analysis methods, analyzed the data, and wrote the manuscript.

Co-Author: Steven T. Kalinowski

Contributions: Assisted with study design and development of study questions, discussed results and edited manuscript.

Co-Author: Mark L. Taper

Contributions: Helped develop appropriate analysis methods, discussed results and provided comments on the manuscript

Co-Author: Jay J. Rotella

Contributions: Conceived the study, obtained funding, assisted with developing study questions.

Co-Author: Robert A. Garrott

Contributions: Conceived the study, obtained funding, assisted with developing study questions.

Manuscript Information Page

John H. Powell, Steven T. Kalinowski, Mark L. Taper, Jay J. Rotella, R. A. Garrott
Journal Name: Evolution
Status of Manuscript:
__X_ Prepared for submission to a peer-reviewed journal
____ Officially submitted to a peer-review journal
____ Accepted by a peer-reviewed journal
____ Published in a peer-reviewed journal

Publisher: Wiley Periodicals Inc.

Abstract

Inbreeding depression can reduce the viability of wild populations. Detecting inbreeding depression in the wild is difficult, and developing accurate estimates of inbreeding can be time and labor intensive. In this study we used a two-step modeling procedure to incorporate uncertainty inherent in estimating individual inbreeding coefficients from multilocus genotypes into estimates of inbreeding depression in a population of Weddell seals (*Leptonychotes weddellii*). For this study we genotyped 154 seals from the population in Erebus Bay, Antarctica at 29 microsatellite loci, 12 of which are novel. We found suggestive, but inconclusive, evidence for inbreeding depression in lifetime reproductive success, adult survival, age at maturity, and the reproductive interval of female seals in this population. The estimated inbreeding depression in lifetime reproductive success in this population was lower than the average inbreeding depression in mortality of wild mammal populations. The two-step modeling procedure presented in this paper provides a method for estimating the magnitude of a known source of error, which is assumed absent in classic regression models, and incorporating this error into inferences about inbreeding depression. These models, therefore, allow for a better evaluation of the uncertainty surrounding the biological importance of inbreeding depression in non-pedigreed wild populations.

Introduction

The reduction in fitness of inbred individuals is one of the oldest observations of population genetics (Ives and Whitlock 2002). Inbreeding depression was first recognized among domesticated species (Darwin 1868), but also affects captive and wild populations (Ralls et al. 1979, Ralls et al. 1988, Crnokrak and Roff 1999, Hedrick and Kalinowski 2000, Keller and Waller 2002). The negative effects of inbreeding are usually stronger in the wild (Crnokrak and Roff 1999), and have the potential to reduce the viability of these populations (Gilipin and Soulé 1986, Keller and Waller 2002).

Although there are many methods available to test for inbreeding depression (reviewed in Hedrick and Kalinowski 2000), they are all based on a comparison of fitness traits between inbred and non-inbred individuals. Whether inbreeding depression is estimated using a regression model assuming additive effects of loci (Hedrick and Kalinowski 2000), or multiplicative effects of loci (Morton et al. 1956), inbreeding coefficients are included as explanatory variables and are assumed to be known quantities. While measurement error can exist in pedigree inbreeding coefficients due to errors in the pedigree (Slate et al. 2004), non-zero inbreeding coefficients of the founders (Balloux et al. 2004, Slate et al. 2004), or relatedness of the founders (Slate et al. 2004), the high correlation between shallow (4 to 5 generations) and deep (50 generations) pedigrees (Balloux et al. 2004) indicates that treating pedigree inbreeding coefficients as known is appropriate for analyses of inbreeding depression.

However, because of the difficulty of constructing pedigrees in the wild, and the relationship between genome-wide heterozygosity and inbreeding, multilocus

heterozygosity is used as a surrogate for pedigree inbreeding coefficients in studies of inbreeding depression (Slate and Pemberton 2002). Szulkin et al. (2010) derived a method to calculate inbreeding depression from the observed association between heterozygosity and fitness, and their estimator has recently been shown to be practically unbiased across a range of simulated scenarios (Kardos et al. 2013). However, although substituting multilocus heterozygosity for pedigree inbreeding coefficients avoids potentially difficult data collection, empirical (Slate et al. 2004) and simulation (Balloux et al. 2004) studies indicate that multilocus heterozygosity is only weakly associated with individual inbreeding coefficients. Therefore, when estimating inbreeding depression from the association between heterozygosity and fitness it is no longer valid to treat inbreeding coefficients as known.

Selecting an analysis method that treats individual inbreeding coefficients as random quantities will allow for a more appropriate representation of our lack of knowledge in the inbreeding depression parameter we are trying to estimate. By accounting for this uncertainty a more appropriate range of plausible values of inbreeding depression can be calculated, relative to classical analysis methods that treat explanatory variables as known. This additional uncertainty can help better assess the biological importance of this process in non-pedigreed wild populations. Here, we present a two-step maximum likelihood analysis, fit using the data cloning algorithm (Lele et al. 2007), that incorporates error in estimating inbreeding coefficients from multilocus genotypes into estimates of inbreeding depression. We use this method to examine the possible

effects of inbreeding in a population of Weddell seals (*Leptonychotes weddellii*) that have been the focus of a long term ecological study in Erebus Bay, Antarctica.

Methods

Study Area and Organism

Weddell seals in Erebus Bay, Antarctica have been the focus of a long term ecological study since 1969 (Proffitt et al. 2007, Rotella et al. 2009). In this population between 8 and 14 colonies of female Weddell seals form each austral spring along tidal cracks in the sea ice (see Cameron and Siniff 2004 for a map) to pup (Rotella et al. 2009). Individuals in this population have been uniquely marked and annually surveyed (Proffitt et al. 2007, Rotella et al. 2009), and as such there is a wealth of information available about the survival and reproductive output of female Weddell seals. Females in this population reach maturity between 4 and 14 years of age (Hadley et al. 2006, Hadley et al. 2008), and produce a single pup (Gelatt et al. 2001) every 1.5-2.2 years (Hadley et al. 2007). Individual seals exhibit strong philopatry (Cameron and Siniff 2004), indicating that all reproductive events of mature seals are captured. Within Erebus Bay, individual survival and age at first reproduction are known to be associated with the breeding colony in which an individual was born (Hastings and Testa 1998, Hadley et al. 2008), as well as the year of an individual's birth (Garrott et al. 2011). In addition, this population of seals has founded a smaller isolate on White Island, Antarctica in which inbreeding is associated with reduced pup survival (Gelatt et al. 2010).

Sample Selection and Genotyping

One hundred and fifty four individual seals, born between 1980 and 1992, were genotyped. Seal lifetime ranged from a minimum of 7 years to an individual who was observed in the population for a total of 29 years. Because the most recent year class of sampled individuals had a maximum 19 year history in the population, all life histories mere truncated to a maximum of 19 years. By tracking reproductive individuals for 19 years we were able to capture the bulk of the reproductive output of each individual, because only 9% of all pups in this population were produced by mothers greater than 17 years of age (Proffitt et al. 2008). Tissue samples were collected and stored as described in Gelatt et al. (2010) and individuals were genotyped at 29 microsatellite loci.

Twelve novel microsatellite sequences were isolated by ecogenics GmbH (Switzerland) using the high-throughput genomic sequencing approach of Abdelkrim et al. (2009; Table 1). One µg of genomic DNA was analyzed on a Roche 454 GS-FLX platform (Roche, Switzerland) using a 1/16th run and the GS FLX titanium reagents. This produced a total of 8,001 reads with an average length of 193 bp. Of these, 273 contained a microsatellite insert suitable for primer design with a tetra- or trinucleotide sequence repeated at least six times, or a dinucleotide sequence of at least 10 repeats. Primers were designed for a total of 32 microsatellite inserts, all of which were tested for polymorphism.

The 12 additional microsatellite markers produced by this screen (Table 2.1) were split into three multiplex reactions each amplifying four loci. The forward primer of each locus was labeled at the 5’ end with a specific dye (6FAM, VIC, NED and PET, Life

Technologies, CA). Loci in these multiplexes were (1) *Lew-001339*, *Lew-001859*, *Lew-001845*, *Lew-001677*, (2) *Lew-001873*, *Lew-002658*, *Lew-004467*, *Lew-002762*, and (3) *Lew-007425, Lew-006657*, *Lew-006174*, *Lew-005761*. Multiplex PCR reactions consisted of 1 µM of each primer, 2 µL 5X MyTaq reaction buffer (Bioline, MA), 1 µM M13 primer (Schuelke 2000), 0.5 µg MyTaq HS DNA polymerase (Bioline, MA), 1 µl of template DNA and enough water for a final reaction volume of 10 µl. The PCR profile consisted of one activation step at 95 °C for 15 min followed by 30 cycles (95 °C for 15s, 55 °C for 15 s and 72 °C for 10 s), 10 cycles (95 °C for 15s, 53 °C for 15 s and 72 °C for 10 s), and a final extension step at 72 °C for 30 min. Microsatellite fragments were visualized using a 3100-Avant Genetic Analyzer, with allele calls made using Genemapper v. 3.7 (Life Technologies, CA). Seals were also genotyped at *Lw-4*, *Lw-7*, *Lw-10*, *Lw-11*, *Lw-16*, *Lw-20*, *Lc-6*, *Lc-13*, *Lc-18*, *Lc-26*, *Lc-28*, *Hi-8*, *Hi-14*, *Hi-15*, *Hi-16*, *Hi-20* (Davis et al. 2002) and *G1A* (Paetkau and Strobeck 1995) as described in Davis et al. (2002).

We calculated expected heterozygosity and the mean number of alleles in this sample using GenAlEx version 6.501 (Peakall and Smouse 2006, Peakall and Smouse 2012). We tested for the presence of pairwise linkage disequilibrium and for whether loci were in Hardy-Weinberg equilibrium, using the exact test of Guo and Thompson (1992), using Genepop on the web (Raymond and Rousset 1995, Rousset 2008). For both these analyses Markov chains were run for 10,000 iterations in each of the 100 batches.

Two-Step Maximum Likelihood Analysis

Parameter estimates and associated estimates of uncertainty for all analyses in this paper were made using the data cloning algorithm (Lele et al. 2007). This method allows frequentist based inference to be made from Bayesian hierarchical models (Lele et al. 2007), and as such makes the hierarchical Bayesian framework available for any researcher regardless of their data analysis philosophy. Because the data cloning algorithm was used to fit the regression models in this study, and cloning the data involved copying both the number of individuals as well as the genotypic information for each individual, models for estimating individual inbreeding coefficients were run separate from the regression analyses. Due to the additional steps necessary to run a data cloning analysis, we present its use here. Other researchers interested in a Bayesian analysis can combine the two-steps we present into a single model with the same formulations and make posterior inference from the original data.

Estimating Individual Inbreeding Coefficients

We used the Gibbs sampling method of Vogl et al. (2002) to estimate individual inbreeding coefficients. This method relates frequencies of the mth allele at the lth locus (p_{lm}) and the inbreeding coefficient of the ith individual (F_i) , which is expressed as a proportion between zero and one, using variables indicating whether or not the alleles at a specific locus (x_{il}) came from a common ancestor (i.e. were identical by descent). The indicator variables were modeled as $\mathrm{x_{il}|F_i, \mathbf{p}_{lm}} \sim \mathrm{Bernoulli}(\pi_i)$, where $\pi_i = 1 - \frac{(1- F_i)*p_{lm}}{(1- F_i)*p_{lm} + F_i}$. To prevent WinBUGS version 1.4.3 (Lunn et al. 2000) from crashing we

constrained the probability of success for this Bernoulli distribution between 0.00001 and 0.99999 (Lesaffre and Lawson 2012).

Assuming loci are independent, individual inbreeding coefficients were modeled as $\mathrm{F_i}|\mathrm{x_{i1}}, \ldots, \mathrm{x_{il}} \sim Beta((\sum_\mathrm{l} \mathrm{x_{il}} + \mathrm{a}), (\mathrm{l} - \sum_\mathrm{l} \mathrm{x_{il}} + \mathrm{b}))$. While Vogl et al. (2002) suggest using a = b = 0.001 for the beta distribution for F_i, we selected the Bayes-Laplace prior (a = b = 1) to avoid the unintended informative nature of other common “noninformative” beta prior distributions, such as the Jeffrey’s prior (a = b = ½) or the Haldane Prior (a = b = 0), in the case where individuals have no alleles identical by descent (Tuyl et al. 2008).

Finally, the vector of indicator variables identifying if alleles were identical by descent was used to specifying a vector for each locus in the population ($\boldsymbol{Z}_l$) that was the total number of alleles identical in appearance (i.e. identical by state) that were not also identical by descent. This was simply the sum of all alleles of a given state across all individuals at the locus, where alleles identical by decent within an individual were counted only once. This vector was used to model the conditional distribution of allele frequencies given identity by descent as $\mathbf{p}_{\mathrm{lm}}|\mathrm{X} \sim Dirichlet((\mathrm{Z}_1 + \alpha), \ldots, (\mathrm{Z_m} + \alpha))$, where α = 1 for a uniform Dirichlet prior distribution (Pritchard et al. 2000) on the allele frequencies ($\boldsymbol{p}_{lm}$, Vogl 2002). We parameterized this distribution by modeling individual values $\delta_{lm}|X \sim Gamma(\mathrm{Z_m} + \alpha, 1)$, so that the vector $\mathbf{p}_{\mathrm{lm}}$ had elements $\frac{\delta_{lm}}{\sum_m \delta_{lm}}$ (Spiegelhalter et al. 2003). See Vogl et al. (2002) for a more detailed description of the Gibbs sampling algorithm used for estimating individual inbreeding coefficient.

The parameters for the beta distribution modeling individual inbreeding coefficients were estimated by fitting the model of Vogl et al. (2002) to the original data,

as well as to datasets that had a total of 5, 10, 20 and 40 copies of the original data using WinBUGS version 1.4.3 (Lunn et al. 2000) and the dclone package (Sólymos 2010) in R version 2.15.0 (R Core Development Team 2012). Each model was fit using 3 independent chains run for 10,000 iterations. We saved every 5th sample after discarding the initial 5,000 samples. Model convergence to a stationary distribution was assessed by visual inspection of the trace plots and calculating the $\hat{R}$ statistic (Gelman et al. 2004) using the coda package in R (Plummer et al. 2006). Modal values for the number of loci identical by decent ($\sum_l x_{il}$) in the ith individual were used to parameterize the beta distribution of individual inbreeding coefficients in all cloned regression models. In cases where the posterior distribution was multimodal, the average of the observed modes was calculated and used.

Modeling the Association Between Inbreeding and Fitness

We tested for the presence of inbreeding depression in mature seals in this population by selecting a fitness surrogate that incorporated survival across the 19 year duration of the study, age at first maturity, and the frequency with which a seal reproduced. This statistic is the first eigenvalue of a Leslie matrix constructed for each individual. This value is an individual based analog (r_i) to the instantaneous population growth rate (r). In estimating the relationship between inbreeding and r_i, we included explanatory variables for the birth colony and year of birth for each seal due to their association with survival or reproduction in previous studies (Hastings and Testa 1998, Hadley et al. 2008, Garrott et al. 2011). Birth colony was included as an indicator

variable specifying whether or not a seal had been born in one of the Delbridge Island colonies (see Hadley et al. 2008 for a delineation of colonies).

In addition to estimating the relationship between a composite measure of lifetime reproductive success and inbreeding, we also estimated the association between inbreeding and survival across the 19 year duration of the study, mean age at first reproduction, and the frequency of reproductive events. These three additional regression models were run in an effort to identify whether a reduction in fitness at a specific stage of an individual's life was driving the possible inbreeding depression in lifetime reproductive success.

For estimating the association between individual inbreeding coefficients and r_i, probability of survival, age at first reproduction, and probability of reproducing in a given year we selected between up to 10 nested candidate models. In the description of the analyses below we present the richest model that was fit for each fitness variable with all other subsets of variables, each of which included inbreeding coefficient, except for the intercept only model, being investigated as well. Because of known differences in the wait-time following first reproduction (Hadley et al. 2008), the probability that an individual reproduced in a given year was modeled with parallel lines model differentiating between the wait-time following an individual's first pup and the wait-time for all subsequent pups. This model was specified using a variable whose value was one if the interval was following an individual's first pup and zero otherwise. Model selection, among the nested candidate models, was completed by calculating AICc values for the models (Burnham and Anderson 2002, Ponciano et al. 2009). Models selection

was performed in an effort to reduce the potential loss of precision in estimating the regression coefficient for individual inbreeding coefficient that may occur with too many explanatory variables in the model (Ramsey and Schafer 2002). The total number of parameters was assumed to simply equal the number of regression parameters (k = 1, … , 4). Because r_i was modeled as a normally distributed variable an additional variance parameter was also included in calculating the total number of parameters for the models estimating the relationship between r_i and inbreeding coefficient. Models within a Δ_{AICc} value of less than two were considered equivalent (Burnham and Anderson 2002). AICc was selected due to ease of calculation and because it closely approximated Bengtsson and Cavanaugh's (2006) AICi for state space models. All models were sampled with WinBUGS version 1.4.3 (Lunn et al. 2000), using the dclone package (Sólymos 2010) in R version 2.15.0 (R Core development Team 2012).

The length of each chain for models investigating the association between individual inbreeding coefficient and a given fitness surrogate was set based on visual inspection of the trace plots and to ensure the value of the $\hat{R}$ statistic (Gelman et al. 2004), as calculated with the coda package (Plummer et al. 2006), was below 1.1 in the most saturated model for all parameters. For data cloning, each model was fit on 1, 5, 10, 20 and 40 total copies of the data, with individual inbreeding coefficients modeled with beta distributions parameterized using the modal values for the number of loci with alleles identical by decent in an individual as described previously in the methods. The convergence of the data cloning algorithm was determined using the statistics outlined in Lele et al. (2010) as implemented in the dclone package (Sólymos 2010). These statistics

were 1) the largest eigenvalue of the posterior variance covariance matrix, which checks the degeneracy of the posterior distributions, 2) the mean squared error and 3) a correlation-like fit statistic, which together check the sufficiency of the normal approximation (Sólymos 2010). We also plotted the logarithm of the posterior variances scaled based on the posterior variance observed using the original data to ensure they had a linear decrease with the logarithm of the number of clones, and that ultimate scaled variance was below 0.05 (Sólymos 2010). Models whose parameters failed to converge to a multivariate normal distribution based on 40 total copies of the data were removed from subsequent analysis.

Estimates of regression parameters were calculated as the mean of the posterior distribution of the model fit on 40 total copies of the data, with 95% Wald's confidence intervals calculated in the dclone package (Sólymos 2010). We also developed confidence bands and prediction bands for each model.

Confidence bands were developed by taking 10,000 random draws from a multivariate normal distribution with a vector of means and a variance covariance matrix calculated from the model fit on 40 total copies of the data. Confidence bands were calculated as the 2.5 and 97.5% quantiles of the distribution of estimated individual fitness trait values at a given inbreeding coefficient.

Prediction bands for these individual regressions were developed by calculating the variance covariance matrix of regression coefficients from the model fit on 40 total copies of the data. This matrix and the mean coefficient values were then used to parameterize a new WinBUGS model, in which posterior draws of individual based

instantaneous population growth rate, individual maximum ages, age at maturity, or wait-time between reproductive events were saved for each of 1,000 equally spaced inbreeding coefficient values on the interval (0, 0.25). This interval was chosen to span the data cloned estimates of observed individual inbreeding coefficients in the sample. Prediction bands were calculated as the 2.5 and 97.5% quantiles of the draws of individual fitness trait values at a given inbreeding coefficient from a single chain run for 10,000 iterations, saving every 5th iteration after discarding the first 5,000 samples. For more detail on the construction of prediction bands from data cloned models see Sólymos (2010).

Lifetime Reproductive Success. We modeled $r_i \sim N(\mu_i, \sigma_r^2)$, where this normal distribution was truncated below zero and censored above 0.3819509. The distribution was truncated at zero because by including only reproductive individuals in this analysis the minimum individual based instantaneous population growth rate was zero (a single pup produced). In addition, the distribution was censored above 0.3819509 because this was the maximum observable value across the first 19 years of an individual's life (corresponding to a seal maturing at four years of age and producing a single pup every year until age 19). A normal distribution was selected because instantaneous population growth rate is assumed to be normally distributed (Harris 1986), an assumption which has been found to hold for three Australian mammal populations (Hone 1999).

The richest model fit was $\mu_i = \beta_0 + \beta_1 * F_i + \beta_2 * Col_i + \beta_3 * F_i * Col_i + \gamma$, where Col_i was an indicator variable whose value was one if an individual was born in a Delbridge Island colony and zero otherwise. The k regression parameters, k = 0, …, 3, were modeled as $\beta_k \sim N(0,\ \sigma^2 = 10000)$, with the intercept parameter β_0 truncated

between 0 and 0.3819509 using the truncated normal distribution in WinBUGS (Lunn 2003). These prior distributions were selected to be flat over a much larger region than the range of expected parameter values (Gelman and Hill 2007). We modeled $\sigma_r \sim half - Cauchy(0, 0.013)$, and $\gamma \sim N(0, \sigma_\gamma^2)$ where $\sigma_\gamma \sim half - Cauchy(0, 0.004)$. The scale parameter for these half-Cauchy prior distributions was set to be four times larger than the observed standard deviations in the data (Gelman 2006). For these models the individual inbreeding coefficient was modeled as $F_i \sim Beta(A_i, B_i)$. The values of A_i and B_i were the modal values of the shape and scale parameters of the posterior distribution of individual inbreeding coefficients calculated in the first step of the analysis. Individual based instantaneous population growth rate models were run on three independent chains (starting values drawn from $\beta_0 \sim N(0.15,\ \sigma^2 = 0.0001)$, $\beta_{1,2,3} \sim N(0,\ \sigma^2 = 0.01)$, $F_i \sim Unif(0.01, 0.1)$, $\sigma_r = |\xi|/\sqrt{\chi}$, $\xi \sim N(0.1,\ \sigma^2 = 0.000225)$, $\chi \sim N(0.3,\ \sigma^2 = 0.36)$, $\sigma_\gamma = |\zeta|/\sqrt{\tau}$, $\zeta \sim N(0.002,\ \sigma^2 = 0.000025)$, $\tau \sim N(0.4,\ \sigma^2 = 0.36)$) for 20,000 iterations, saving every 5th iteration after discarding the first 15,000 samples.

Survival Across the Study. Probability of survival was assumed constant across all ages of mature seals (Hadley et al 2006). We modeled age of death $death_i \sim Geom(s_i)$, censored between 4 and 19 to account for the fact that seals were not observed before maturity, the minimum age of which is 4 (Hadley at al. 2008), and the minimum number of years for which all individuals had been tracked was 19. The richest model that was fit to the data was $\log\left(\frac{(1-s_i)}{s_i}\right) = \beta_0 + (\beta_1 + \theta_Y) * F_i + \beta_2 *$

$Col_i + \beta_3 * F_i * Col_i + \gamma_Y$, where Col_i and F_i were identical to the regression models fit to r_i. The parameters θ_Y and γ_Y were group level variables for the year (Y = 1980, …, 1992) in which an individual was born. The k regression parameters, k = 0, …, 3, were modeled as $\beta_k \sim N(0,\ 10000)$, $\theta_Y \sim N(0, \sigma_\theta^2)$ where $\sigma_\theta \sim Unif(0, 100)$, and $\gamma_Y \sim N(0, \sigma_\gamma^2)$ where $\sigma_\gamma \sim Unif(0, 100)$. Age of death models were run on three independent chains (starting values drawn from $\beta_{0,1,2,3} \sim N(0,\ \sigma^2 = 1)$, $\theta_Y \sim N(0,$ $\sigma^2 = 1)$, $\gamma_Y \sim N(0,\ \sigma^2 = 1)$, $\sigma_\theta \sim Unif(0,\ 10)$, $\sigma_\gamma \sim Unif(0, 10)$) for 20,000 iterations, saving every 5th iteration after discarding the first 15,000 samples.

Age at First Reproduction. Age at first reproduction was modeled as $rep_i \sim NBin(4, p_i)$. Four events, years of survival, needed to have been observed because it corresponds to the minimum age at maturity in the population (Hadley et al. 2008). Minimum age at maturity was subtracted from all responses prior to model fit. The richest model was $\log\left(\frac{4*(1-p_i)}{p_i}\right) = \beta_0 + (\beta_1 + \theta_Y) * F_i + \beta_2 * Col_i + \beta_3 * F_i * Col_i +$ γ_Y, where Col_i and F_i were identical to the regression models fit to r_i, and θ_Y and γ_Y were group level parameters given the year of an individual's birth (Y = 1982, …, 1992). The k regression parameters, k = 0, …, 3, were modeled as $\beta_k \sim N(0,\ \sigma^2 = 10000)$, $\theta_Y \sim N(0, \sigma_\theta^2)$ where $\sigma_\theta \sim Unif(0, 100)$, and $\gamma_Y \sim N(0, \sigma_\gamma^2)$ where $\sigma_\gamma \sim Unif(0, 100)$. Age at first reproduction models were run on three independent chains (starting values drawn from $\beta_{0,1,2,3} \sim N(0,\ \sigma^2 = 1)$, $\theta_Y \sim N(0,\ \sigma^2 = 1)$, $\gamma_Y \sim N(0,\ \sigma^2 = 1)$, $\sigma_\theta \sim Unif(0,\ 10)$, $\sigma_\gamma \sim Unif(0, 10)$) for 10,000 iterations, saving every 5th iteration after discarding the first 5,000 samples.

Frequency of Reproduction. Finally, the wait-time between pups was modeled as $wait_i \sim Geom(\pi_i)$, after subtracting a single year from each individual's response. The probability that a seal reproduced in a given year was assumed constant for all pups, except if it was following the first pup (Hadley et al. 2008). The richest model fit was $\log\left(\frac{(1-\pi_i)}{\pi_i}\right) = \beta_0 + (\beta_1 + \theta_Y) * F_i + \beta_2 * Col_i + \beta_3 * F_i * Col_i + \beta_4 * First_i + \gamma_Y$, where Col_i and F_i were identical to the regression models fit to r_i, and θ_Y and γ_Y were group level parameters given the year of an individual's birth (Y = 1982, …, 1992). The variable $First_i$ had a value of one if the wait-time was following an individual's first pup, and zero if not. The k regression parameters, k = 0, …, 4, were modeled as $\beta_k \sim N(0,\ 10000)$, $\theta_Y \sim N(0, \sigma_\theta^2)$ where $\sigma_\theta \sim Unif(0, 100)$, and $\gamma_Y \sim N(0, \sigma_\gamma^2)$ where $\sigma_\gamma \sim Unif(0, 100)$. Individual wait-time models were run on three independent chains (starting values drawn from $\beta_{0,1,2,3} \sim N(0,\ \sigma^2 = 1)$, $\theta_Y \sim N(0,\ \sigma^2 = 1)$, $\gamma_Y \sim N(0,$ $\sigma^2 = 1)$, $\sigma_\theta \sim Unif(0,\ 10)$, $\sigma_\gamma \sim Unif(0, 10)$) for 20,000 iterations, saving every 5th iteration after discarding the first 15,000 samples.

Results

Expected heterozygosity observed at the microsatellite loci was consistent with observed levels in other non-exploited placental mammals (0.677, Garner et al. 2005). The expected heterozygosity was 0.716 (95% confidence interval from 0.668 to 0.764) in this population of seals, with an average of 10.2 alleles per genotyped locus. Expected heterozygosities at our newly developed microsatellite markers were within the range observed in previously described loci (*Lc-28* = 0.297, *Lc-6* = 0.888; Table 2.1), although

these loci in general had fewer observed alleles (average observed in previously described loci = 11.2, Table 2.1). We did not observe any evidence of linkage disequilibrium between loci following a Benjamini-Hochberg false discovery rate correction (smallest P-value = 0.0012, adjusted $\alpha = 0.0001$). Of the 29 amplified loci, one had an excess of homozygotes (*Lew-001677*, P-value = 0.00, $F_{IS} = 0.0664$) based on estimated F_{IS} values (Weir and Cockerham 1984).

Based on AIC_C values, the regression model estimating the individual based instantaneous population growth rate as a function of individual inbreeding coefficient was equivalent to the model that also included an individual's birth colony (Table 2.2). In the interest of parsimony we selected the model with fewer coefficients for analysis. Based on this sample of 154 individuals we estimated that for every increase in the individual inbreeding coefficient of 0.0625 (the equivalent of a first cousin mating) there was an associated change in the individual based instantaneous population growth rate of approximately -0.0066 (95% confidence interval from -0.0154 to 0.0022). This estimated decrease corresponds to a lifetime reproductive success for an individual that was the product of a first cousin mating that is 96.6% as large as a non-inbred individual (95% confidence interval from 91.5% to 101.1%). Because AIC_C values indicated that including individual inbreeding coefficients in the model led to a better fit (Table 2.2), but the 95% confidence interval around the regression parameter for individual inbreeding coefficient included zero, we believe there is suggestive, but inconclusive, evidence that inbreeding reduces the lifetime reproductive success of individual Weddell seals in this population (Figure 2.1).

We estimated that for an increase in the individual inbreeding coefficient of 0.0625 there was approximately an associated 1.02 times increase in the mean maximum observed age (95% confidence interval from 0.86 to 1.22 times). We also estimated that an increase in the individual inbreeding coefficient of 0.0625 corresponded to approximately a 1.04 times increase in the mean age at maturity (95% confidence interval from 0.92 to 1.18 times). Finally, after accounting for whether a given pup was the mother's first, there was an approximately 1.06 times estimated increase in the wait-time between pups for every increase in the individual inbreeding coefficient of 0.0625 (95% confidence interval from 0.94 to 1.20 times). The average estimated wait-time following the first pup in this population was approximately 2.08 times longer than for subsequent pups across all inbreeding coefficient values (95% confidence interval from 1.59 to 2.73 times).

Discussion

We used molecular data to estimate the associated between inbreeding and lifetime reproductive success of mature female Weddell seals in Erebus Bay, Antarctica. The use of a two-step maximum likelihood analysis allowed us to incorporate uncertainty in estimating individual inbreeding coefficients from multilocus genotypes into our assessment of inbreeding depression in this population. We found suggestive, but inconclusive, evidence for an association between individual inbreeding coefficient and lifetime reproductive success.

Because the confidence interval around the estimated association between inbreeding and lifetime reproductive success included zero, it remains unclear whether inbreeding decreased the lifetime reproductive success of Weddell seals in this population. However, the data did suggest that if inbreeding affects lifetime reproductive success, the effect must be relatively minor. The estimated cost of inbreeding (δ *sensu* Lande and Schemske 1985) in lifetime reproductive success for mature individuals with an inbreeding coefficient of 0.25 (equivalent to a full sibling mating) was lower (0.137, 95% confidence interval from 0.032 to 0.260) than the average cost of inbreeding for mortality in wild (2.155 after standardizing for $F = 0.25$, Crnokrak and Roff 1999) or captive mammal populations (0.33, Ralls et al. 1988). It is important to note that these costs of inbreeding are calculated at an inbreeding coefficient beyond any of the median values observed in our sample. In addition, the conclusion that the effect of inbreeding depression is small is supported by the fact that the estimated 5.5% reduction in lifetime reproductive success (95% confidence interval from 4.7 to 6.5%) of mature female seals with an inbreeding coefficient of 0.1 is at the low end of the expected range of inbreeding depression based on a rule of thumb derived from agricultural systems (Frankel and Soule 1981 reported by Keller and Waller 2002). Therefore, this low cost of inbreeding in lifetime reproductive success indicates that inbreeding depression does not play as large role in determining the lifetime reproductive success of mature female Weddell seals as it does in shaping other fitness parameters in other mammalian, or agricultural, populations.

While the estimated value of inbreeding depression in the lifetime reproductive success of female Weddell seals in this population from Erebus Bay was low, there is

inbreeding depression has been detected in juvenile survival of Weddell seals in an isolated population on White Island founded by individuals from Erebus Bay (Gelatt et al. 2010). Therefore, we may not have detected a strong signature of inbreeding depression because juvenile survival was not included in our analysis. In estimating the lifetime reproductive success of female seals in this population we only included mature individuals; seals that died before reproducing were not included in this analysis. It is likely that inbreeding depression in Weddell seals may manifest itself early in the seals' life and have little effect upon adults that reach breeding age. Because our study did not cover the entire lifespan of individuals; it is also possible that in addition to juvenile survival, inbreeding depression may affect senescent individuals. By tracking individuals across 19 years we only captured a single year of the senescent period for female Weddell seals (Proffitt et al. 2007), and as such may not have been able to detect the effects of senescence in this population. Therefore, as this population is the focus of a long term study, the effects of inbreeding on senescence of seals should be revisited once all genotyped individuals have completed their life cycle.

In this study we were unable to detect previously identified associations between either the colony of birth, or the year of birth, and any of the life history parameters we investigated. This fact raises the possibility that our sample was not representative of the larger population. However, the similarity between our estimated probability of survival, age at maturity and reproductive interval and previous findings in this population indicates this is not the case. For example, we estimated the annual probability of survival for a non-inbred mature individual to be 0.945 (95% confidence interval from 0.933 to

0.954), which compares well with previous reported probabilities of survival for mature individuals (0.94 Hadley et al. 2006, 0.905-0.942 Hadley et al. 2007, 0.93-0.94 Hadley et al. 2008). Our estimated mean age at first reproduction for non-inbred seals of 7.39 years (95% confidence interval from 6.93 to 7.92 years) matched well with previously reported estimates in this population (7.67 years Hadley et al. 2006, 7.55-8.45 years Hadley et al. 2008). We estimated the probability of reproduction in the year following an individual's first pup to be 0.548 (95% confidence interval from 0.442 to 0.649), which was close to the estimated probability (0.46) of Hadley et al. (2008). We also estimated the probability that an experienced breeder produced a pup the following year to be 0.716 (95% confidence interval from 0.684 to 0.746) which was slightly higher than was previously reported (0.67, SD = 0.09) by Hadley et al. (2008).

Although we did not statistically detect an association between inbreeding and lifetime reproductive success in this population of seals, the two-step models presented in this study provide a method for incorporating some of the uncertainty inherent in estimating individual inbreeding coefficients from multilocus genotypes into estimates of inbreeding depression. This analysis allowed us to provide an appropriate context (i.e. the maximum plausible amount of inbreeding depression in this population was modest) for the observed level of inbreeding depression in this non-pedigreed population of Weddell seals.

Inbreeding is an important component to population viability, but one whose inclusion in a population viability analysis is often not advocated due to a lack of available information about its magnitude (Morris and Doak 2002). In this study we

provide a useful model for directly estimating inbreeding depression in the wild when pedigree information is unavailable. This two-step analysis of inbreeding depression incorporates uncertainty in estimating individual inbreeding coefficients from multilocus genotypes and population allele frequencies using the methods of Vogl et al. (2002). Crucially, this method provides a set of plausible values for inbreeding depression that accounts for the error associated with estimating inbreeding coefficients. As such this method may allow for the incorporation of inbreeding into population viability analyses for species where a lack of pedigree information previously prevented an accurate assessment of inbreeding depression.

Table 2.1. Primer sequences and marker information for the 12 novel microsatellite loci used in this study.

Locus		Primer sequences 5'-3'	Repeat motif	Size (bp)	Number of alleles	H_E
Lew-001339	F	GCACTCCAGTTTCCTTGGAC	$(AC)_{14}$	106-135	10	0.631
	R	AGGAGCTTAGTAGGCAATCC				
Lew-001677	F	ACAAGGGATTCTTAGGGAACTG	$(AC)_{15}$	214-231	15	0.799
	R	TCCAGTGGTAATAACTTGCAAAC				
Lew-001845	F	TGTAACCTCAAGGGTCCCAC	$(TG)_{12}$	147-153	4	0.519
	R	GCGTCTGGAGTGTGGAATTG				
Lew-001859	F	TCTCCCTGTTCATTAGATCCTG	$(AC)_{13}$	98-114	8	0.622
	R	GAGCCAACTTGCATTGTGTTC				
Lew-001873	F	TGTTTCCGGTTGGGCTATTC	$(TG)_{14}$	133-155	8	0.624
	R	ACGATAGATTGGGCCTTGTC				
Lew-002658	F	ATTCTCAGACCTCAGGGAGC	$(GT)_{16}$	122-146	8	0.709
	R	CATCCTGAGTTTGGCCTTGG				
Lew-002762	F	TGTTCCATCTCCTGCCACTC	$(AC)_{13}$	164-185	9	0.759
	R	ATCTGGGGAAAGGTGGGTTC				
Lew-004467	F	TGCACAGTATAAAACAGGATAGAGG	$(AC)_{14}$	90-111	11	0.788
	R	CCAGAGAGAGCCTGTGTACG				
Lew-005761	F	AGAGAGGGTCATTAGAGACAGC	$(AC)_{13}$	158-174	8	0.782
	R	ATGACTCTTCATGGGCGTGC				
Lew-006174	F	TGGTGAACTCAACAAGGGAAAG	$(AC)_{15}$	101-112	7	0.738
	R	TGTATTGCTCAGCCCAACTC				
Lew-006657	F	GCATGCTGGGTCATGAGTG	$(GT)_{13}$	82-103	11	0.785
	R	GCCCCACGATGTTACTAAGTTG				
Lew-007425	F	AAGTTTTATGTGGGCATCCG	$(GT)_{12}$	96-106	5	0.560
	R	GCCGTTCACATTTCTGCCTC				

Table 2.2. Candidate regression models for each of the individual fitness surrogates ranked based on their relationship to the model with the minimum AICc value. Only those models that converged are shown. Explanatory variables included individual inbreeding coefficient (*INB*), birth location (*COL*) and a variable in the mean wait-time between pupping intervals models indicating whether the interval was following an individual seal's first pup (*FIRST*).

Model Structure	K	Δ_{AICc}
r_i		
$\beta_0 + \beta_1 INB$	3	0.00
$\beta_0 + \beta_1 INB + \beta_2 COL$	4	0.09
$\beta_0 + \beta_1 INB + \beta_2 COL + \beta_3 INB * COL$	5	2.24
β_0	2	2.63
Minimum age		
β_0	1	0.00
$\beta_0 + \beta_1 INB$	2	1.55
$\beta_0 + \beta_1 INB + \beta_2 COL$	3	3.50
$\beta_0 + \beta_1 INB + \beta_2 COL + \beta_3 INB * COL$	4	5.47
Age at maturity		
β_0	1	0.00
$\beta_0 + \beta_1 INB$	2	1.57
$\beta_0 + \beta_1 INB + \beta_2 COL$	3	3.57
$\beta_0 + \beta_1 INB + \beta_2 COL + \beta_3 INB * COL$	4	5.68
Pupping interval		
$\beta_0 + \beta_4 FIRST$	1	0.00
$\beta_0 + \beta_1 INB + \beta_4 FIRST$	2	1.04
$\beta_0 + \beta_1 INB + \beta_2 COL + \beta_4 FIRST$	3	2.07
$\beta_0 + \beta_1 INB + \beta_2 COL + \beta_3 INBCOL + \beta_4 FIRST$	4	2.83

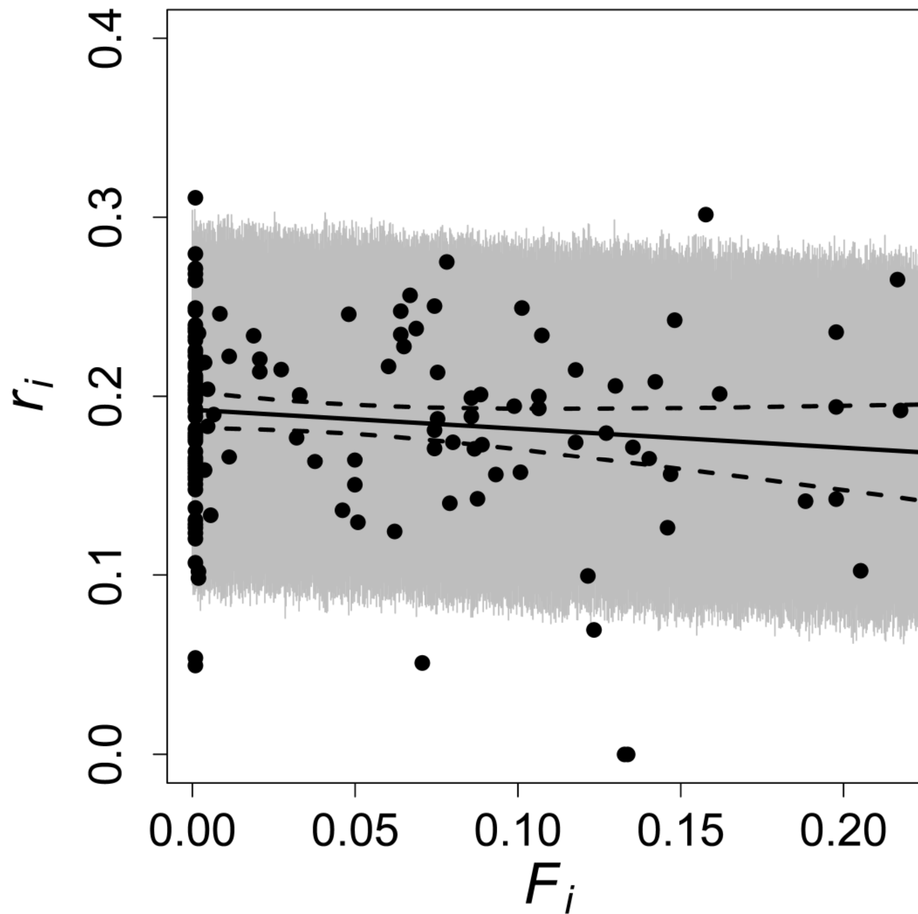

Figure 2.1. Lifetime reproductive success, measured as an individual based instantaneous population growth rate given their individual inbreeding coefficients. The solid line represents the estimated association between individual inbreeding coefficient and the lifetime reproductive success of a seal based on a truncated-Normal regression model fit to censored data, with the 95% prediction band for this analysis denoted by the gray shading and the 95% confidence band denoted by the dotted lines.

Literature Cited

Abdelkrim, J., B. C. Robertson, J.-A. L. Stanton, and N. J.Gimmel. 2009. Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. Biotechniques 46:185-191.

Balloux, F., W. Amos, and T. Coulson. 2004. Does heterozygosity estimate inbreeding in real populations? Mol. Ecol. 13:3021-3031.

Bengtsson, T. and J. E. Cavanaugh. 2006. An improved Akaike information criterion for state-space model selection. Comput. Stat. Data An. 50:2635-2654.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Stat. Soc. B 57:289-300.

Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference, 2nd ed. Springer, New York, NY.

Cameron, M. F., and D. B. Siniff. 2004. Age-specific survival, abundance, and immigration rates of a Weddell seal (*Leptonychotes weddellii*) population in McMurdo Sound, Antarctica. Can. J. Zoolog. 82:601-615.

Crnokrak, P., and D. A. Roff. 1999. Inbreeding depression in the wild. Heredity 83:260-270.

Darwin, C. R. 1868. Variation of animals and plants under domestication. John Murry, London, U.K.

Davis, C. S., T. S. Gelatt, D. Siniff, and A. Strobeck. 2002. Dinucleotide microsatellite markers from the Antarctic seals and their use in other Pinnipeds. Mol. Ecol. Notes 2:203-208.

Frankel, O. H., and M. E. Soulé. 1981. Conservation and evolution. Cambridge University Press, Cambridge, U.K.

Garner, A., J. L. Rachlow, and J. F. Hicks. 2005. Patterns of genetic diversity and its loss in mammalian populations. Conserv. Biol. 19:1215-1221.

Garrott, R. A., J. J. Rotella, D. B. Siniff, C. L. Parkinson, and G. E. Stauffer. 2011. Environmental variation and cohort effects in an Antarctic predator. Okios 120:1027-1040.

Gelatt, T.S., C. S. Davis, I. Stirling, D. B. Siniff, C. Strobeck, and I. Delisle. 2010. History and fate of a small isolated population of Weddell seals at White Island, Antarctica. Conserv. Genet. 11:721-735.

Gelatt, T. S., C. S. Davis, D. B. Siniff, and C. Strobeck. 2001. Molecular evidence for twinning in Weddell seals (*Leptonychotes weddellii*). J. Mammal. 82:491-499.

Gelman, A. 2006. Prior distributions for variance parameters in hierarchical models. Bayesian Anal. 1:515-533.

Gelman, A., and J. Hill. 2007. Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, New York, NY.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 2004. Bayesian data analysis, 2nd ed. Chapman & Hall, Boca Raton, FL.

Gilpin, M. E., and M. E. Soulé. 1986. Minimum viable populations: the process of species extinctions. In: *Conservation Biology the Science of Scarcity and Diversity* (ed Soul ME), pp. 13-34, Sinauer Associates, Sunderland.

Guo, S. W., and E. A. Thompson. 1992. Performing the exact test for Hardy-Weinberg proportion for multiple alleles. Biometrics 48:361-371.

Hadley, G. L., J. J. Rotella, and R. A. Garrott. 2008. Spatial variation in age-specific probabilities of first reproduction for Weddell seals. Oikos 117:1165-1174.

Hadley, G. L., J. J. Rotella, and R. A. Garrott. 2007. Evaluation of reproductive costs for Weddell seals in Erebus Bay, Antarctica. J. Anim. Ecol. 76:448-458.

Hadley, G. L., J. J. Rotella, R. A. Garrott, and J. D. Nichols. 2006. Variation in probability of first reproduction of Weddell seals. J. Anim. Ecol. 75:1058-1070.

Harris, R. B. 1986. Reliability of trend lines obtained from variable counts. J. Wildlife Manage. 50:165-171.

Hastings K. K., and J. W. Testa. 1998. Maternal and birth colony effects on survival of Weddell seal offspring from McMurdo Sound, Antarctica. J. Anim. Ecol. 67:722-740.

Hedrick, P. W., and S. T. Kalinowski 2000. Inbreeding depression in conservation biology. Annu. Rev. Ecol. Syst. 31:139-162.

Hone, J. 1999. On the rate of increase (r): patterns of variation in Australian mammals and the implications for wildlife management. J. Appl. Ecol. 36:709-718.

Ives, A. R., and M. C. Whitlock. 2002. Inbreeding and metapopulations. Science 295:454-455.

Kardos, M., F.W. Allendorf, and G. Luikart. 2013. Evaluating the role of inbreeding depression in heterozygosity-fitness correlations: how useful are tests for identity disequilibrium? Mol. Ecol. Resour. doi: 10.1111/1755-0998.12193

Keller L. F., and D. M. Waller. 2002. Inbreeding effects in wild populations. Trends Ecol. Evol. 17:230-241.

Lande, R., and D. W. Schemske. 1985. The evolution of self fertilization and inbreeding depression in plants. I. Genetic models. Evolution 39:24-40.

Lele, S. R., K. Nadeem, and B. Schmuland. 2010. Estimability and likelihood inference for generalized linear mixed models using data cloning. J. Am. Stat. Assoc. 105:1617-1625.

Lele, S. R, B. Dennis, and F. Lutscher. 2007. Data cloning: easy maximum likelihood estimation for complex ecological models using Bayesian Markov chain Monte Carlo methods. Ecol. Lett. 10:551-563.

Lesaffre, E., and A. Lawson. 2012. Bayesian biostatistics. John Wiles & Sons, West Sussex, UK.

Lunn, D. 2003. WBDevDJLTruncatedNormal documentation. Available at http://www.winbugs-development.org.uk/. Accessed on December 29, 2013.

Lunn, D. J., A. Thomas, N. Best, and D. Spiegelhalter. 2000. WinBUGS – a Bayesian modeling framework: concepts, structure, and extensibility. Stat. Comput. 10:325-337.

Morris, W. F., and D. F. Doak. 2002. Quantitative conservation biology. Sinauer Associates, Inc., Sunderland, MA.

Morton, N. E., J. F. Crow, and H. J. Muller. 1956. An estimate of the mutational damage in man from data on consanguineous marriages. Proc. Nat. Acad. Sci. USA 42:855-863.

Paetkau, D., and C. Strobeck. 1995. The molecular basis and evolutionary history of a microsatellite null allele in bears. Mol. Ecol. 4:519-520.

Peakall, R., and P. E. Smouse. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol. Ecol. Notes 6:288-295.

Peakall, R., and P. E. Smouse. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. Bioinformatics 28:2537-2539.

Plummer, M., N. Best, K. Cowles, and K. Vines. 2006. CODA: convergence diagnostics and output analysis for MCMC. R News 6:7-11.

Ponciano, J. M., M. L. Taper, B. Dennis, and S. R. Lele. 2009. Hierarchical models in ecology: confidence intervals, hypothesis testing, and model selection using data cloning. Ecology 90:356-362.

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

Proffitt, K. M., R. A. Garrott, and J. J. Rotella. 2008. Variation in offspring sex ratio among individual Weddell seal (*Leptonychotes weddellii*) females of different quality. Behav. Ecol. Sociobiol. 62:1679-1687.

Proffitt, K. M., R. A. Garrott, J. J. Rotella, and K. E. Wheatley. 2007. Environmental and senescent related variations in Weddell seal body mass: implications for age-specific reproductive performance. Oikos 116:1683-1690.

R Core Development Team. 2012. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org. 05/2012.

Ralls, K., J. D. Ballou, and A. Templeton. 1988. Estimates of lethal equivalents and the cost of inbreeding in mammals. Conserv. Boil. 2:185-193.

Ralls, K., K. Brugger, and J. Ballou. 1987. Inbreeding and juvenile mortality in small populations. Science 206:1101-1103.

Ramsey, F. L., and D. W. Schafer. 2002. The statistical sleuth a course in methods of data analysis, Second Edition. Duxbury, Pacific Grove, CA.

Raymond, M., and F. Rousset. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J. Hered. 86:248-249.

Rotella, J. J., W. A. Link, J. D. Nichols, G. L. Hadley, R. A. Garrott, and K. M. Proffitt. 2009. An evaluation of density-dependent and density-independent influences on population growth rates in Weddell seals. Ecology 90:975-984.

Rousset., F. 2008. Genepop '007: a complete reimplementation of the Genepop software for Windows and Linux. Mol. Ecol. Notes 4:137-138.

Schuelke, M. 2000. An economic method for the fluorescent labeling of PCR fragments. Nature Biotechnology 18:233-234.

Slate, J., and J. M. Pemberton. 2002. Comparing molecular measures for detecting inbreeding depression. J. Evol. Biol. 15:20-31.

Slate, J., P. David, K. G. Dodds, B. A. Veenvliet, B. C. Glass, T. E. Broad, and J. C. McEwan. 2004. Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: theoretical expectations and empirical data. Heredity 93:255-265.

Sólymos, P. 2010. dclone: data cloning in R. The R Journal 2:29-37.

Spiegelhalter, D., A. Thomas, N. Best, and D. Lunn. 2003. WinBUGS user manual, version 1.4, January 2003. Available at http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/manual14.pdf. Accessed December 29, 2013.

Szulkin M., N. Bierne, and P. David. 2010. Heterozygosity-fitness correlations: a time for reappraisal. Evolution 64:1202-1217.

Tuyl, F., R. Gerlach, and K. Mengersen. 2008. A comparison of Bayes-Laplace, Jeffreys, and other priors: the case of zero events. Am. Stat. 62:40-44.

Vogl, C., A. Karhu, G. Moran, and O. Savolainen. 2002. High resolution analysis of mating systems: inbreeding in natural populations of *Pinus radiata*. J. Evol. Biol. 15:433-439.

Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38:1358-1370.

CHAPTER THREE

MICROSATELLITES INDICATE MINIMAL BARRIERS TO MULE DEER *ODOCOILEUS HEMIONUS* DISPERSAL ACROSS MONTANA, USA

Contribution of Authors and Co-Authors

Manuscript in Chapter 3

Author: John H. Powell

Contributions: Helped develop study questions, developed appropriate analysis methods, analyzed the data, and wrote the manuscript.

Co-Author: Steven T. Kalinowski

Contributions: Assisted with study design and development of study questions, discussed results and edited manuscript.

Co-Author: Megan D. Higgs

Contributions: Helped develop appropriate analysis methods, discussed results and edited manuscript

Co-Author: Michael R. Ebinger

Contributions: Helped develop study questions and provided comments on the manuscript.

Co-Author: Ninh V. Vu

Contributions: Genotyped samples, aligned and submitted sequences to GENBANK, drafted molecular methods section of the manuscript.

Co-Author: Paul C. Cross

Contributions: Conceived the study, obtained funding, assisted with developing study questions, discussed results and edited manuscript.

Manuscript Information Page

John H. Powell, Steven T. Kalinowski, Megan D. Higgs, Michael R. Ebinger, Ninh V. Vu, Paul C. Cross
Journal Name: Wildlife Biology
Status of Manuscript:
____ Prepared for submission to a peer-reviewed journal
____ Officially submitted to a peer-review journal
____ Accepted by a peer-reviewed journal
__X_ Published in a peer-reviewed journal

Publisher: Nordic Board for Wildlife Research (NKV)
Issue in which manuscript appears: Volume 19, Number 1

The following chapter was published in 2013 in the journal "Wildlife biology" and appears here with permission from the Nordic Board of Wildlife Research (NKV). The full reference, which is to be used for all citations, for this work is:

Powel, J.H., S.T. Kalinowski, M.D. Higgs, M.R. Ebinger, N.V. Vu, and P.C. Cross. 2013. Microsatellites indicate minimal barriers to mule deer Odocoileus hemionus dispersal across Montana, USA. Wildlife Biology 19:102-110.

Abstract

To better understand the future spread of chronic wasting disease, we conducted a genetic assessment of mule deer population structure across the state of Montana, USA. Individual based analyses were used to test for population structure in the absence of *a priori* designations of population membership across the sampling area. Samples from the states of Wyoming, Colorado, and Utah (USA) were also included in the analysis to provide a geographic context to the levels of population structure observed within Montana. Results showed mule deer across the entire study region were characterized by weak isolation by distance and a lack of spatial autocorrelation at distances larger than 10 km. We found evidence for contemporary male bias in dispersal, with female mule deer exhibiting higher mean individual pairwise genetic distance than males. We tested for potential homogenizing effects of past translocations within Montana but were unable to detect a genetic signature of these events. Our results indicate high levels of connectivity among mule deer populations in Montana and suggest few, if any, detectable barriers to mule deer gene flow or chronic wasting disease transmission.

Introduction

The population structure of organisms across a landscape is fundamental part of the natural history of any species that wildlife managers will want to know. Because the amount of genetic divergence among subdivided populations is directly related to the movement of individuals (Hartl & Clark 2007), estimating population structure provides an indirect quantification of connectivity across the landscape. Previous investigations of genetic divergence in mule deer (*Odocoileus hemionus*) have found that population structure corresponds to subspecies boundaries (Latch et al. 2009, Pease et al. 2009).

Mule deer are distributed across western North America, ranging from northern Canada to central Mexico, and from the Pacific Ocean to east of the Rocky Mountains (Anderson and Wallmo 1984, Mackie et al. 2003). By the end of the 19th century, mule deer populations had experienced declines in both size and extent across much of their range (Mackie et al. 2003). Within the state of Montana, USA, these population declines led to the establishment of a translocation program for mule deer (Picton & Lonner 2008), which occurred concurrently with improvements in habitat quality and reductions in both predator densities and human harvest of deer (Mackie et al. 1998, Picton & Lonner 2008). Populations of mule deer peaked between the 1950s and early 1970s and now appear to be regulated by environmental conditions and hunting and predation pressure (Mackie et al. 1998, Picton & Lonner 2008).

Chronic wasting disease could greatly impact the future management of mule deer. Chronic wasting disease is expected to reduce mule deer population sizes (Dulberger et al. 2010), but might also reduce hunting pressure due to perceived health

hazards (Williams et al. 2002). As of 2012, chronic wasting disease is present in mule deer populations in the states of South Dakota and Wyoming, USA, and the Canadian provinces of Saskatchewan and Alberta, all of which border Montana. Control strategies for reducing chronic wasting disease infections focus on disrupting the routes through which the disease is spread (Miller et al. 2006). The goal of this investigation was to describe the genetic population structure of mule deer sampled across Montana, with the ultimate goal of better understanding potential chronic wasting disease transmission across the landscape, and thus helping to inform potential chronic wasting disease management actions. Additionally, this study will contribute to our understanding of mule deer genetic structure across large geographic regions.

Material and Methods

We collected samples of lymph node tissue, on a volunteer basis at hunter check stations, from 370 mule deer harvested during fall 2007 and 2008 across a study area aimed to encompass the entire state of Montana, with additional samples collected in Wyoming; Colorado; and Utah, USA (Figure 3.1). We also collected 19 samples from white-tailed deer harvested in northwest Montana. At time of collection we recorded GPS coordinates for the approximate location of harvest, along with the sex of the deer. ArcMap version 9.3.1 (Environmental Systems Research Institute, Inc., Redlands, CA) was used to reproject all coordinates into NAD83 UTM zone 12. For an indication of the size of our study area, the largest distance between two deer in our analysis was 1432 km.

We isolated DNA from lymph node tissue using Qiagen DNeasy 96 Tissue Kits and suspended DNA extracts in Buffer AE (Qiagen, MA). We performed polymerase chain reactions (PCR) on 16 microsatellite loci in a multiplex format. Multiplex PCRs amplified four unique loci, each of which was labeled with a forward primer at the 5' end for a specific dye (6FAM, VIC, NED and PET, Life Technologies, CA). Loci included in the four multiplexes were: (1) M, P, K and N (2) D, Q, O and R (3) E, BM4107, Rt30 and Rt7 and (4) G, Ovir, Rt24 and Cervid3 (Bishop et al. 1994, Wilson et al. 1997, DeWoody et al. 1995, Jones et al. 2000). A typical multiplex PCR reaction consisted of 1 µM of each primer, 5 µL 2X Qiagen Multiplex PCR Kit, approximately 50 ng DNA, and enough water for a final volume of 10 µL (Qiagen). The thermoprofile consisted of one activation step at 95 °C for 15 minutes followed by 40 cycles of 94 °C for 30 seconds, 60 °C for 90 seconds and 72 °C for 60 seconds, and a final extension step at 72 °C for 30 minutes. In addition we sequenced 489 base pairs of the D-loop, or control region, of the mitochondria using primers developed by Latch et al. (2008). PCR chemistry consisted of 1 µM of each primer, 2 µL 5X MyTaq Reaction Buffer, 0.375 Units MyTaq HS DNA polymerase, approximately 50 ng DNA, and enough water for a final volume of 10 µL (Bioline, MA). The thermoprofile consisted of one activation step at 95 °C for 4 minutes followed by 40 cycles of 94 °C for 30 seconds, 55 °C for 15 seconds and 72 °C for 30 seconds, and a final extension step at 72 °C for 1 minute. We used ExoSAP-IT for PCR cleanup (Affymetrix). Sequencing reaction and subsequent cleanup were done using Big Dye Terminator v3.1 Cycle Sequencing Kit and ethanol/EDTA/sodium acetate precipitation as suggested by the manufacturer (Life Technologies, CA). Both

microsatellite fragments and D-loop DNA sequences were visualized using a 3100-Avant Genetic Analyzer (Life Technologies, CA). Scoring of genotypes was performed using Genemapper v. 3.7 (Life Technologies, CA), and sequences were manually aligned with the aid of Sequencher v. 4.1 (Gene Codes Corporation, MI). Mitochondrial D-loop sequences were deposited in GenBank (accession numbers: JN040634–JN040649, JN040651–JN040684, JN040686–JN040687, JN040691–JN040693, JN040697–JN040716).

We screened for the presence of hybridization between mule deer and white-tailed deer using the admixture model in the program STRUCTURE version 2.3.3 (Pritchard et al. 2000, Falush et al. 3003, Falush et al. 2007). We tested for the number of genetic clusters using five independent runs of STRUCTURE for K between one and five, each run for 100,000 Markov Chain Monte Carlo (MCMC) iterations after a burn-in of 10,000 iterations. A recessive alleles model was used with missing genotypes entered as null allele homozygotes. We did not set prior membership for the 19 white-tailed deer collected in Northwest Montana. The ΔK method of Evanno et al. (2005) was used to select the appropriate K, after which we ran 10 independent runs of STRUCTURE using the same model parameters but with a fixed K. Results from these final runs were compiled into a single consensus cluster membership using the FullSearch method in CLUMPP version 1.1.2 (Jakobsson & Rosenberg 2007), with results visualized using DISTRUCT version 1.1 (Rosenberg 2004). We classified hybrids as those individuals who had a consensus estimated genetic contribution from white-tailed deer of greater than 10%. Following the removal of hybridized and misclassified individuals, we

estimated the number of genetic clusters in the study area using the same STRUCTURE parameterization as the initial analysis of the number of genetic clusters.

We compared observed genotype frequencies with Hardy-Weinberg expectations using the exact test of Guo and Thompson (1992). Population structure can affect genotype frequencies, so we analyzed three broad-scale sampling regions (Montana/Wyoming, Utah, and Colorado) independently. We used the Markov chain exact test in GENEPOP on the web (Rousset & Raymond 1995, Rousset 2008) to perform these calculations. Allele frequencies and unbiased expected heterozygosity for each sampling region were estimated using GENALEX version 6.41 (Peakall & Smouse 2006). We tested the homogeneity of mean expected heterozygosities using linear combinations. A total of three tests were run to compare the mean expected heterozygosity of each broad-scale sampling region to the common mean of the other regions. Allelic richness was calculated for a constant sample size of 20 genes in each sampling location using rarefaction with the program HP-RARE (Kalinowski 2005). Global F_{ST} was estimated, using the methodology of Weir and Cockerham (1984), among the three sampling regions with the program F-STAT version 2.9.3.2, and a 95% confidence interval was calculated using bootstrap re-sampling across loci (Goudet 2002).

To test for the presence of isolation by distance we first calculated a count of dissimilar alleles between two deer at a given locus using Equation (1) of Eding and Meuwissen (2001), except that we defined a success as an event where two alleles were different. Allele sharing distance (ASD) was then the estimated by summing these counts

of allelic dissimilarity across loci and dividing by the total number of comparisons. Binomial logistic regression was used to estimate the linear relationship between Euclidian distance and the log odds of having dissimilar alleles. A Mantel test (Mantel 1967) with 5000 random permutations was used to calculate a p-value for the test of no relationship between geographic distance and the log odds of having dissimilar alleles, and a 95% confidence interval for the estimated change in the log odds of having dissimilar alleles as a function of Euclidean distance was calculated by inverting the Mantel hypothesis test as described in Manly (2007), using 10,000 random permutations.

We further tested for a relationship between genetic and geographic distance using the multivariate spatial autocorrelation methods of Smouse and Peakall (1999), as implemented in GENALEX version 6.41 (Peakall & Smouse 2006). We tested the null hypothesis of no spatial structure within equal distance classes of 10 km using 10,000 random permutations, and corrected p-values for false discovery (Benjamini & Hochberg 1995). We ran a principal coordinates analysis for the three sampling locations using the covariance-standardized method in GENALEX version 6.41 (Peakall & Smouse 2006).

We used a re-sampling test to explore whether translocations may have reduced the genetic distance between donor and recipient locations relative to historic levels. Translocation zones were set to include all individuals within 20, 40, 60, 80, and 100 km of the centroid of Montana counties that received translocated individuals (Picton & Lonner 2008). These removal distances were selected to account for the expected increase in the geographic area affected by a translocation event due to dispersal over time. We first calculated the observed slope of the logistic regression line between the log

odds of having different alleles and Euclidean distance for the data after removing all individuals located within a specified distance of a translocation release site. We then randomly removed the same number of individuals and re-estimated the slope of the regression line. The p-values for these tests were calculated as the proportion of 5000 random removals with estimated regression slopes greater than or equal to the slope that was observed after removing individuals within the translocation zone.

We used a combination of mitochondrial DNA (mtDNA) and microsatellite markers to test for sex-biased dispersal in mule deer. We tested for historical sex-biased dispersal by comparing the genetic divergence observed between sampling groups at a sex-linked marker (mtDNA) to that observed at biparentally inherited markers (Prugnolle & de Meeus 2002). We analyzed a total of 489 bp of the mitochondrial control region in 76 samples from Montana that were pooled into 5 distinct sampling groups (Figure 3.1). Samples were pooled into these five groups because of their geographic separation within the larger sampling region of Montana. We used ARLEQUIN version 3.5 (Excoffier & Lischer 2010) to estimate F_{ST} for microsatellite genotypes and for the mitochondrial sequences following a multiple hits correction (Tamura 1992). Estimates of female specific dispersal are confounded by the fact that uniparental inheritance reduces the effective population size of this marker to one quarter the size of biparentally inherited markers. This increases the genetic divergence expected at mitochondrial loci. We calculated the expected increase in genetic divergence based on both an island model of migration and a model where populations evolve in complete isolation based on equations provided in Zink and Barrowclough (2008). Isolation by distance was

estimated for these samples using Mantel tests with 5000 random permutations (Manly 2007) comparing the ratio of genetic distance $F_{ST}/(1-F_{ST})$ to the natural logarithm of geographic distance as proposed by Rousset (1997). We tested for contemporary sex-bias in dispersal using pairwise ASD from samples located in Montana and Wyoming. Samples were split into female (n = 106) and male (n = 213) groups and the difference in mean ASD between the sexes was assessed using a permutation test, the null distribution of which was developed by randomly permuting sex and recalculating the difference in mean ASD. This test is analogous to the F_{ST} based test for sex-biased dispersal of Goudet et al. (2002). Unless otherwise stated, all analyses were performed in the statistical computing package R (R Development Core Team 2009).

Results

STRUCTURE analysis provided strong evidence that the tissue samples of putative mule deer collected in the field included white-tail deer and hybrids between mule and white tailed deer. When we included genotypes of deer known to be white-tailed in a STRUCTURE analysis, and varied the number of genetic clusters (K), the best supported number clusters was two (one for each species as shown by the mode of ΔK Evanno et al. 2005). This analysis suggested that seven samples labeled as mule deer in the field were actually white-tailed deer and that four additional deer appeared to be post first-generation hybrids between white-tailed deer and mule deer. The hybridized individuals in the dataset had admixtures of 12.5%, 14.5%, 29.8% and 40.9% white-tailed deer genetic contribution. We removed all misclassified and hybridized samples from

subsequent analyses, reducing the data set to 359 individual deer, 320 of which were located in the Montana/Wyoming sampling region, 29 in the Utah sampling region, and 10 in the Colorado sampling region. A subsequent STRUCTURE analysis indicated a single genetic cluster existed across this region. For this analysis the number of genetic clusters was selected based on the mean log posterior probabilities of the data given K.

Two of the loci we genotyped (O and Q) had an excess of homozygotes and were removed from subsequent analyses. Previous investigations have also observed an excess of homozygotes at these loci (Pease et al. 2009), which suggests there may be null alleles exist at these loci. One locus (P) was identified as out of Hardy-Weinberg equilibrium in the Montana/Wyoming sampling region from a Bonferroni corrected p-value ≈ 0.00. This locus was not removed from analysis because the F_{IS} at this locus (0.03) was within the range (0.007 to 0.078) of positive F_{IS} values that were not identified as out of Hardy-Weinberg equilibrium, and this locus was in Hardy-Weinberg equilibrium in both other sampling locations.

Estimates of heterozygosity were similar across all sampling locations (Table 3.1) (all p-values ≥ 0.47), with an average expected heterozygosity of 0.701 across the entire study region (95% confidence interval from 0.662 to 0.740). Average allelic richness was also similar among the broad-scale sampling regions (5.49 in Montana/Wyoming, 5.12 in Utah, and 5.64 in Colorado). These broad-scale sampling regions were characterized by low levels of genetic divergence with an estimated global F_{ST} of 0.012 (95% confidence interval from 0.008 to 0.016).

Data are consistent with weak isolation by distance in mule deer across this region (p-value = 0.0334, Figure 3.2). We estimated that a 100 km increase in the distance between two mule deer was associated with a 1.0057-fold increase in the odds that randomly chosen alleles at the same locus in the two deer are different (95% confidence interval from 0.9995 to 1.0118).

We found no evidence in the spatial autocorrelation analysis to support spatial structuring of mule deer genotypes within our smallest distance class of 10 km (p-value = 0.033, adjusted α-level = 0.004). There was also no apparent clustering among the broad-scale regions in the principle coordinate analysis (Figure 3.3), the first two axes of which explained 38.1% of the variation in genotypic distance.

We did not detect any evidence that historical translocations of mule deer in Montana have homogenized the population structure of this species. That is, a directed removal of samples located within a given proximity to translocation release sites did not increase the observed level of isolation by distance (smallest p-value was 0.057 for removing all individuals within 80 km of the translocation site, adjusted α = 0.01).

A total of 39 mitochondrial haplotypes were observed in the sample of 76 mitochondrial sequences (Table 3.2). Nucleotide diversity appeared to be reduced in samples collected from central and southwestern Montana relative to other sampling groups across the state (Table 3.3). We observed higher pairwise F_{ST} values for mtDNA than for nuclear DNA (Table 3.4), four comparisons of which indicated recent geographic structuring of groups (Zink & Barrowclough 2008) or historic sex-biased dispersal (Prugnolle & de Meeus 2002) based on the fact that only the mtDNA F_{ST} indicated

considerable genetic structure with estimates greater than 0.2 (Zink & Barrowclough 2008). We did not detect isolation by distance for either the microsatellite or mitochondrial genotypes in these smaller sampling groups (p-values = 0.100 and 0.197 respectively). However, a higher mean ASD at microsatellite loci was observed for female mule deer ($ASD_{FEMALE} - ASD_{MALE} = 0.013$, p-value = 0.04 from 5,000 permutations) across Montana and Wyoming.

Discussion

The weak isolation by distance we observed in Montana is consistent with previous dispersal estimates (Wright 1943, Wright 1946), which indicate that both male and female mule deer often move long distances (Anderson & Wallmo 1984, Mackie et al. 2003). The fact that there are few, if any, barriers to dispersal, whether complete or permeable, is further indicated by the lack of observed autocorrelation of genotypes across this study area. Finally, the suggestive evidence of potential sexually dimorphic patterns of genetic divergence suggests that males are probably more important than females for the long-distance spread of chronic wasting disease into new regions. This possibility was suggested in a previous study of infected mule deer populations in Colorado (Miller & Conner 2005).

Levels of genetic structure similar to our results have previously been found in studies of mule deer population structure conducted at broad spatial scales. For example, Latch et al. (2009) found no phylogenetic structure below the subspecies level when examining mtDNA. Similarly, while Pease et al. (2009) found five genetic clusters of

mule deer within the state of California, USA these clusters corresponded with previous subspecies classifications. Cullingham et al. (2011) detected spatial autocorrelation of female mule deer extending up to 2 km in Alberta and Saskatchewan, but their estimated global F_{ST} (0.008) indicated similar levels of genetic structure to our study area. However these authors did find evidence of two genetic clusters within their study area, and individuals assigned to clusters logically based on patterns of isolation by distance (Cullingham et al. 2011). Therefore, our study appears to fit well with previous work on mule deer, and in general indicates populations are characterized by low levels of genetic structure below the subspecies level.

Cross et al. (2005) showed that disease transmission across a spatially structured population is a function of migration rate and the infectious period of the pathogen. Duration of infectiousness is related to host immune responses, lifespan, and environmental persistence of the pathogen. The observed high levels of connectivity among mule deer likely represent a conservative estimate of actual dispersal because molecular methods only record successful reproduction following dispersal events (Cushman et al. 2006). Therefore, our findings of limited genetic isolation suggest high connectivity across the sampling area, which combined with the potentially long-term survival of prions in the environment (Brown & Gajdusek 1991, Seidel et al. 2007), suggest few, if any, barriers to chronic wasting disease spread.

Acknowledgements

Major funding for this project was provided by the U.S. Geological Survey, and a National Science Foundation and National Institutes of Health Ecology of Infectious Disease Grant DEB-1067129. J. Powell was supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1049562. Additional funding was provided by the National Science Foundation to S. Kalinowski under Grant No. DEB-000717456. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government. Samples provided by Montana Fish, Wildlife and Parks were paid for by the sale of hunting and fishing licenses in Montana and Federal Aid in Wildlife Restoration matching grants. We thank S. Amish, T. Andrews, J. Gude, and three anonymous reviewers for helpful comments on this manuscript.

Tables and Figures

Table 3.1. Estimated expected heterozygosity and allelic richness for the 14 loci included in the final analyses. Allelic richness was standardized to a sample size of 20 genes using rarefaction (Kalinowski 2005). Values are separated based on broad-scale sampling regions of Montana/Wyoming (MT/WY), Utah (UT), and Colorado (CO).

Locus	**Expected Heterozygosity**				**Allelic Richness**		
	MT/WY	UT	CO	Total	MT/WY	UT	CO
K	0.664	0.712	0.726	0.701	4.238	4.300	4.000
M	0.594	0.522	0.584	0.567	3.694	3.301	3.000
N	0.866	0.859	0.889	0.872	8.583	7.275	9.000
P	0.653	0.658	0.689	0.667	4.526	4.056	4.000
D	0.490	0.586	0.468	0.515	3.896	4.340	3.000
R	0.677	0.594	0.553	0.608	4.150	4.090	4.000
BM4107	0.655	0.709	0.863	0.742	5.600	6.189	9.000
E	0.515	0.584	0.489	0.529	3.374	3.532	3.000
Rt30	0.843	0.811	0.763	0.806	7.990	6.387	6.000
Rt7	0.804	0.838	0.847	0.830	6.459	6.650	8.000
Cervid3	0.734	0.697	0.737	0.723	5.918	5.186	7.000
G	0.623	0.647	0.626	0.632	3.031	2.995	3.000
OvirA	0.826	0.813	0.916	0.852	7.500	6.781	10.000
Rt24	0.854	0.752	0.700	0.769	7.881	6.648	6.000

Table 3.2. Haplotype frequencies from 76 mitochondrial control region sequences. Group 1 included 19 samples from northeastern Montana, Group 2 included 19 samples from the northwestern corner of Montana, Group 3 included 6 samples from west central Montana, Group 4 included 13 samples from southwestern Montana, and Group 5 included 19 samples from southeastern Montana (Figure 3.1).

Haplotype	Group 1	Group 2	Group 3	Group 4	Group 5
1	0.105	-	-	-	0.105
2	0.053	-	-	-	-
3	0.263	-	-	-	-
4	0.053	-	-	-	-
5	0.053	-	-	-	-
6	0.053	-	-	-	-
7	0.053	-	-	-	-
8	0.105	-	-	-	-
9	0.053	-	-	-	-
10	0.105	-	0.167	-	-
11	0.053	-	-	-	-
12	0.053	-	-	-	-
13	-	0.105	-	-	-
14	-	0.158	-	-	-
15	-	0.053	-	0.308	-
16	-	0.158	-	-	-
17	-	0.053	-	-	-
18	-	0.105	-	0.077	0.158
19	-	0.158	0.167	-	-
20	-	0.053	-	-	-
21	-	0.053	-	-	-
22	-	0.053	-	0.077	-
23	-	0.053	-	-	-
24	-	-	0.333	0.077	-
25	-	-	0.167	-	-
26	-	-	0.167	-	-
27	-	-	-	0.231	-
28	-	-	-	0.077	-
29	-	-	-	0.077	-
30	-	-	-	0.077	-
31	-	-	-	-	0.053
32	-	-	-	-	0.211
33	-	-	-	-	0.053
34	-	-	-	-	0.053
35	-	-	-	-	0.053
36	-	-	-	-	0.105
37	-	-	-	-	0.053
38	-	-	-	-	0.053
39	-	-	-	-	0.105

Table 3.3. Nucleotide diversity across 489 bp of the mitochondrial control region, with associated 95% confidence intervals, present in the five sampling groups located within Montana. Group 1 included 19 samples from northeastern Montana, Group 2 included 19 samples from the northwestern corner of Montana, Group 3 included 6 samples from west central Montana, Group 4 included 13 samples from southwestern Montana, and Group 5 included 19 samples from southeastern Montana (Figure 3.1).

	Nucleotide Diversity	**95% Confidence Interval**
Group 1	0.024	[0.011, 0.036]
Group 2	0.028	[0.013, 0.042]
Group 3	0.019	[0.007, 0.030]
Group 4	0.013	[0.006, 0.020]
Group 5	0.029	[0.014, 0.043]

Table 3.4. Pairwise F_{ST} for mitochondrial sequences (above diagonal) and microsatellite genotypes (below diagonal) with values less than the Benjamini-Hochberg (1995) corrected alpha-levels in bold. Group 1 included 19 samples from northeastern Montana, Group 2 included 19 samples from the northwestern corner of Montana, Group 3 included 6 samples from west central Montana, Group 4 included 13 samples from southwestern Montana, and Group 5 included 19 samples from southeastern Montana (Figure 3.1).

	Group 1	**Group 2**	**Group 3**	**Group 4**	**Group 5**
Group 1	-	**0.137**	**0.357**	**0.511**	**0.133**
Group 2	**0.020**	-	0.013	**0.212**	-0.017
Group 3	0.015	-0.008	-	0.122	0.054
Group 4	**0.032**	0.014	-0.022	-	**0.203**
Group 5	**0.016**	**0.030**	0.004	0.008	-

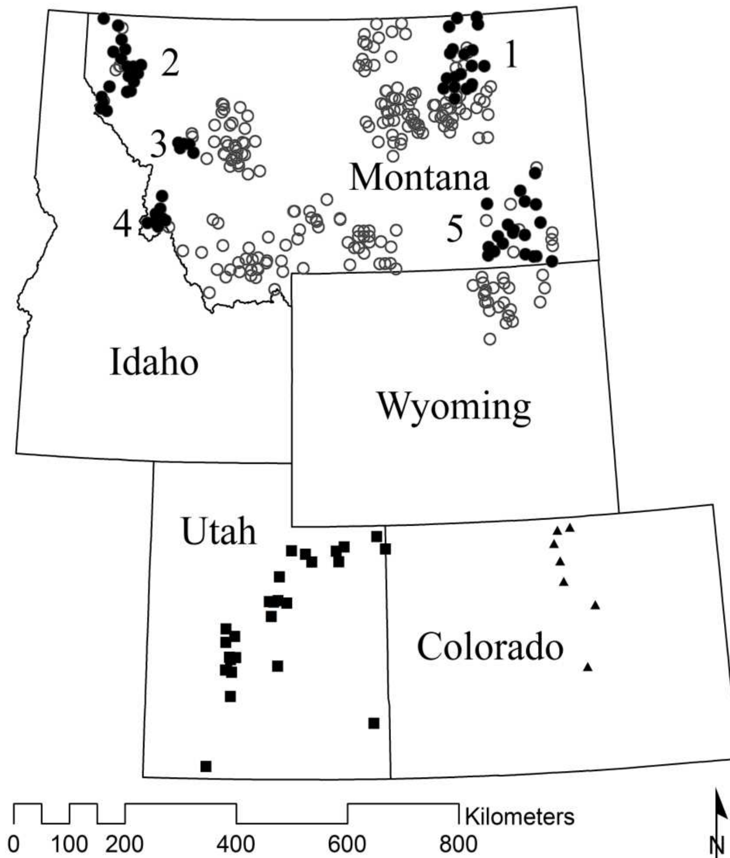

Figure 3.1. Locations for 359 mule deer samples collected at hunter check stations and used in final analyses. Open circles indicate samples genotyped at microsatellite loci only, while closed circles indicate samples genotyped at microsatellite loci and at 489 bp of the mitochondrial control region. Closed squares represent samples from Utah, and closed triangles represent samples from Colorado both of which only had microsatellite loci genotyped. Numbers indicate the geographical subgroup designations used for analyses of sex-biased dispersal.

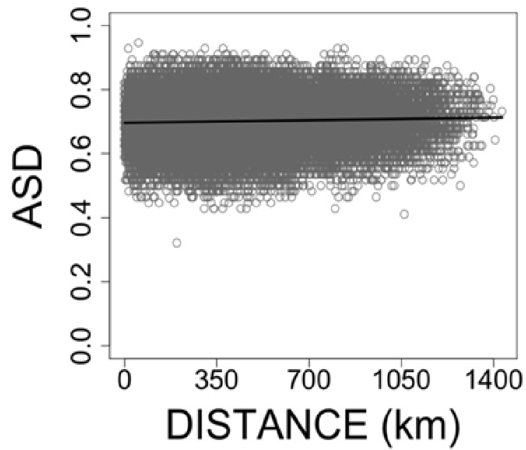

Figure 3.2. Visualization of the relationship between allele sharing distance (ASD) for pairs of individual mule deer and Euclidean distance in kilometers. The regression line is plotting the function ________________.

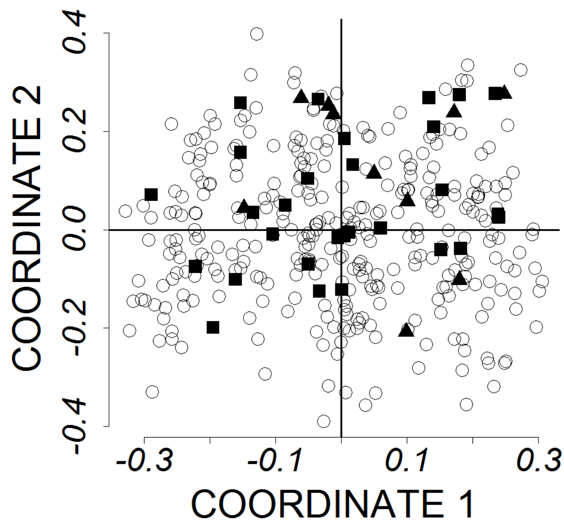

Figure 3.3. Plot of the first two coordinates of the principal coordinate analysis run in GENALEX version 6.41 (Peakall and Smouse 2006). Open circles represent samples from the Montana/Wyoming sampling region (n = 320), closed squares represent samples from the Utah sampling region (n = 29), and closed triangles represent samples from the Colorado sampling region (n = 10).

References

Anderson, A., Wallmo, O. 1984: *Odocoileus hemionus*. - Mammalian Species 219:1-9.

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

Bishop, M.D., Kappes, S.M., Keele, J.W., Stone, R.T., Sunden, S.L.F., & Hawkins, G.A. 1994: A genetic linkage map for cattle. – Genetics 136:619–639.

Brown, P., & Gajdusek, D.C. 1991: Survival of scrapie virus after 3 years' interment. – The Lancet 337:269-270.

Cross, P.C., Lloyd-Smith, J.O., Johnson, P.L.F., & Getz, W.M. 2005: Duelling timescales of host movement and disease recovery determine invasion of disease in structured populations. – Ecology Letters 8:587-595.

Cullingham, C.I., Nakada, S.M., Merrill, E.H., Bollinger, T.K., Pybus, M.J., & Coltman, D.W. 2011: Multiscale population genetic analysis of mule deer (*Odocoileus hemionus hemionus*) in western Canada sheds new light on the spread of chronic wasting disease. – Canadian Journal of Zoology 89:134-147.

Cushman, S.A., McKelvey, K.S., Hayden, J., & Schwartz, M.K. 2006: Gene flow in complex landscapes: testing multiple hypotheses with causal modeling. – The American Naturalist 168:486-499.

DeWoody J.A., Honeycutt R.L., & Skow L.C. 1995: Microsatellite markers in white-tailed deer. – Journal of Heredity 86:317-319.

Dulberger, J., Hobbs, N.T., Swanson, H.M., Bishop, C.J., & Miller, M.W. 2010: Estimating chronic wasting disease effects on mule deer recruitment and population growth. – Journal of Wildlife Diseases 46:1086-1095.

Eding, H., & Meuwissen, T.H.E. 2001: Marker-based estimates of between and within population kinships for the conservation of genetic diversity.- Journal of Animal Breeding and Genetics 118:141-159.

Epperson, B.K, & Li, T. 1996: Measurement of genetic structure within populations using Moran's spatial autocorrelation statistics. – Proceedings of the National Academy of Sciences 93:10528-10532.

Evanno, G., Regnaut, S., & Goudet J. 2005: Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. – Molecular Ecology 14:2611-2620.

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

Falush, D., Stephens, M., & Pritchard, J.K. 2003: Inference of population structure: extensions to linked loci and correlated allele frequencies. – Genetics 164:1567-1587.

Falush, D., Stephens, M., & Pritchard, J.K. 2007: Inference of population structure using multilocus genotype data: dominant markers and null alleles. – Molecular Ecology Notes 7:574-578.

Goudet, J. 2002: FSTAT: a program to estimate and test gene diversities and fixation indices (version 2.9.3.2) Available at http://www.unil.ch/popgen/softwares/fstat.htm (last accessed on 23/5/2011).

Goudet, J., Perrin, N., & Waser, P. 2002: Tests for sex-biased dispersal using bi-parentally inherited genetic markers. – Molecular Ecology 11:1103-1114.

Guo, S.W., & Thompson, E.A. 1992: Performing the exact test for Hardy-Weinberg proportion for multiple alleles. – Biometrics 48:361-371.

Hartl, D.L., & Clark, A.G. 2007: Principles of population genetics 4th ed. – Sinauer Associates Inc., Sunderland, 652 pp.

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

Jones, K.C., Levine, K.F., & Banks, J.D. 2000: DNA-based genetic markers in black-tailed and mule deer for forensic applications. – California Fish and Game 86:115–126.

Kalinowski, S.T. 2005: HP-Rare: a computer program for performing rarefaction on measures of allelic diversity. – Molecular Ecology Notes 5:187-189.

Latch, E.K., Heffelfinger, J.R., Fike, J.A., & Rhodes O.E. 2009: Species-wide phylogeography of North American mule deer (*Odocoileus hemionus*): cryptic glacial refugia and postglacial recolonization. – Molecular Ecology 18:1730-1745.

Latch, E.K., Amann, R.A., Jacobson, J.P., & Rhodes, O.E. 2008: Competing hypotheses for the etiology of cryptorchidism in Sitka black-tailed deer: an evaluation of evolutionary alternatives. – Animal Conservation 11:234-246.

Mackie, R.J., Kie, J.G., Pac, D.F., & Hamlin, K.L. 2003: Mule deer. In: *Wild Mammals of North America* (eds Feldhamer, G.A., Thompson, B.C., Chapman, J.A.) The Johns Hopkins University Press, Baltimore, 889-905 pp.

Mackie, R.J., Pac, D.F., Hamlin, K.L., & Dusek, G.L. 1998: Ecology and management of mule deer and white-tailed deer in Montana. Report to the Montana Department of Fish, Wildlife & Parks.

Manly, B.F.J. 2007: Randomization, bootstrap and Monte Carlo methods in biology, Third Edition. – Chapman & Hall/CRC, Boca Raton, 455 pp.

Mantel, N. 1967: The detection of disease clustering and a generalized regression approach. – Cancer Research 27:209-220.

Miller, M.W., Hobbs, N.T., & Tavener, S.J. 2006: Dynamics of prion disease transmission in mule deer. – Ecological Applications 16:2208-2214.

Miller, M.W., & Conner, M.M. 2005: Epidemiology of chronic wasting disease in free ranging mule deer: spatial, temporal, and demographic influences on observed prevalence patterns. – Journal of Wildlife Diseases 41:275-290.

Peakall, R., & Smouse, P.E. 2006: GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. – Molecular Ecology Notes 6:288-295.

Pease, K.M., Freedman, A.H., Pollinger, J.P., McCormack, J.C., Buermann, W., Rodzen, J., Banks, J., Meredith, E., Bleich, V.C., Schaefer, R.J., Jones, K., & Wayne, R.K. 2009: Landscape genetics of California mule deer (*Odocoileus hemionus*): the roles of ecological and historical factors in generating differentiation. – Molecular Ecology 18:1848-1862.

Picton, H.D., & Lonner, T.N. 2008: Montana's wildlife legacy – decimation to restoration. – Media Works Publications, Bozeman, 304 pp.

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

Prugnolle, F., & de Meeus, T. 2002: Inferring sex-biased dispersal from population genetic tools: a review. – Heredity 88:161-165.

R Development Core Team. 2009: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, ISBN 3-900051-07-0, Available at: http://www.R-project.org (last accessed 23/12/2010).

Raymond, M., Rousset, F. 1995: GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. – Journal of Heredity 86:248-249.

Rosenberg, N.A. 2004: Distruct: a program for the graphical display of population structure. - Molecular Ecology Notes 4:137–138.

Rousset, F. 2008: Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. – Molecular Ecology Resources 8:103-106.

Rousset, F. 1997: Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. – Genetics 145:1219-1228.

Seidel, B., Thomzig, A., Buschmann, A., Groschup, M.H., Peters, R., Beekes, M., & Terytze, K. 2007: Scrapie agent (strain 263K) can transmit disease via the oral route after persistence in soil over years. – PloS ONE 2:e435.

Smouse, P.E., & Peakall, R. 1999: Spatial autocorrelation analysis of multiallele and multi-locus genetic micro-structure. – Heredity 82:561-573.

Tamura, K. 1992: Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C content biases. – Molecular Biology and Evolution 9:678-687.

Weir, B.S., & Cockerham, C.C. 1984: Estimating *F*-statisitcs for the analysis of population structure. – Evolution 38:1358-1370.

Wilson, G.A., Strobeck, C., Wu, C.L., & Coffin, J.W. 1997: Characterization of microsatellite loci in caribou *Rangifer tarandus*, and their use in other artiodactyls. – Molecular Ecology 6:697–699.

Williams, E.S., Miller, M.W., Kreeger, T.J., Kahn, R.H., & Thorne, E.T. 2002: Chronic wasting disease of deer and elk: a review with recommendations for management. – Journal of Wildlife Management 66:551-563.

Wright, S. 1946: Isolation by distance under diverse systems of mating. – Genetics 31:39-59.

Wright, S. 1943: Isolation by distance. – Genetics 28:114-138.

Zink, R.M., & Barrowclough, G.F. 2008: Mitochondrial DNA under siege in avain phylogeography. – Molecular Ecology 17:2107-2121.

CHAPTER FOUR

HYBRIDIZATION DID NOT APPEAR TO INCREASE DISPERSAL IN A WESTSLOPE CUTTHROAT TROUT *ONCORHYNCHUS CLARKII LEWISI* METAPOPULATION

Contribution of Authors and Co-Authors

Manuscript in Chapter 4

Author: John H. Powell

Contributions: Developed study questions, developed appropriate analysis methods, analyzed the data, and wrote the manuscript.

Co-Author: Steven T. Kalinowski

Contributions: Assisted with developing study questions, discussed analysis methods and results and edited the manuscript.

Co-Author: Megan D. Higgs

Contributions: Helped develop appropriate analysis methods, discussed results and edited the manuscript.

Co-Author: Clint C. Muhlfeld

Contributions: Conceived the study and obtained funding, developed sampling design, discussed results.

Manuscript Information Page

John H. Powell, Steven T. Kalinowski, Megan D. Higgs, Clint C. Muhlfeld
Journal Name: Canadian Journal of Fisheries and Aquatic Sciences
Status of Manuscript:
__X_ Prepared for submission to a peer-reviewed journal
____ Officially submitted to a peer-review journal
____ Accepted by a peer-reviewed journal
____ Published in a peer-reviewed journal

Publisher: NRC Research Press

Abstract

Hybridization between native westslope cutthroat trout (*Oncorhynchus clarkii lewisi*) and introduced rainbow trout (*O. mykiss*) is frequent throughout the range of westslope cutthroat trout and is spreading even though hybrids appear to be less fit. We investigated patterns of dispersal of hybridized individuals using genetic samples collected from 17 flow-connected sites within the Middle Fork Flathead River drainage, Montana to test whether differences in dispersal rate between hybridized and non-hybridized individuals could explain this apparent paradox. We genotyped individuals at 39 diagnostic single nucleotide polymorphism (SNP) loci and 31 SNP loci that were polymorphic within westslope cutthroat trout. We tested whether hybridization is associated with dispersal rates by comparing patterns of isolation by distance between hybridized and non-hybridized samples. Highly hybridized individuals were dispersing in this landscape, but we were unable to determine if this dispersal was occurring at an elevated rate. We did not detect differences in isolation by distance between hybridized and non-hybridized sites. Therefore, hybridization did not appear to increase dispersal rates among our sampling locations in the Middle Fork Flathead River.

Introduction

Hybridization is a natural evolutionary process that can play an important role in the development of new taxa (Stebbins 1959, Reiseberg 1994, Barton 2001). For example, the Hawaiian silversword alliance is thought to have originated through a whole genome duplication event caused by hybridization (Baldwin 2003). Although hybridization is a natural process, many taxa have gone extinct due to increased rates of hybridization resulting from anthropogenic changes to the environment (Allendorf et al. 2001). Therefore, to reduce the risk of extinction for taxa threatened by anthropogenic hybridization it is important to develop management policies that oppose this process (Allendorf et al. 2001). Such management actions will become more crucial with the expected increase in the rate at which these anthropogenic hybridization events occur due to increasing rates of species introduction, and habitat fragmentation and alteration (Rhymer and Simberloff 1996, Allendorf et al. 2001).

Many of the challenges of managing anthropogenic hybridization are exemplified by westslope cutthroat trout (*Oncorhynchus clarkii lewisi*). Westslope cutthroat trout had the greatest historical geographic range of any subspecies of cutthroat trout (Behnke 2002), but hybridization with introduced *Oncorhynchus* species has reduced the amount of lotic habitat occupied by westslope cutthroat trout populations that are genetically unaltered to no more than approximately 22.5% of their estimated historic range (Shepard et al. 2005). Moreover, hybridization has been identified as the primary threat to the conservation of cutthroat trout (Allendorf and Leary 1988), and appears to be spreading

even though hybrid individuals appear to be less fit (Allendorf et al. 2004, Muhlfeld et al. 2009).

One potential explanation for the observed spread of hybridization in the face of reduced fitness is that "The admixture from [rainbow trout] appears to be increasing the rate of dispersal and thereby causing the spread of introgression" (Allendorf et al. 2004, 1210). In investigating the role of dispersal in the spread of hybridization across the landscape, previous studies detected the influence of both the long distance dispersal of highly hybridized individuals from source populations, as well as dispersal among neighboring populations (Hitt et al. 2003, Boyer et al. 2008). The objective of this study is specifically to investigate the second of these processes. If low levels of hybridization increase the rate of dispersal between neighboring populations, introgression will spread across this landscape quicker than what would occur with long distance migration alone. Therefore, we test whether low levels of hybridization increase the dispersal rate among neighboring populations in the Middle Fork Flathead River.

The test we developed for assessing this 'increased dispersal' hypothesis is based on the observation that population genetic structure of westslope cutthroat trout in the North Fork Flathead River drainage follows a pattern of isolation by distance (Boyer et al. 2008), which is a positive association between the genetic divergence between populations and the geographic distance separating populations (Wright 1943). The 'increased dispersal' hypothesis predicts that if hybridization increases the dispersal rate among hybridized sites, we expect to observe a weaker relationship between genetic

divergence and geographic distance among hybrid populations in the Middle Fork Flathead River.

We present an extension of the *Structure* algorithm (Pritchard et al. 2000) that uses information from diagnostic loci to estimate westslope cutthroat trout allele frequencies at polymorphic loci in hybridized populations. This analysis method was necessary to avoid confounding estimates of dispersal rate with the stocking history on the landscape. This confounding occurs because hybridization will reduce the genetic distance between hybridized samples as a function of the proportion of rainbow trout genes in a sample (Allendorf and Leary 1988, Waples 1991). Therefore, calculating genetic distance between populations based only on the observed allele frequencies will result in biased estimates of dispersal. To avoid such biases, we used the estimated allele frequencies from the modified *Structure* algorithm to estimate isolation by distance among hybridized and non-hybridized samples collected from flow-connected (e.g. sites that a fish could both swim to and from) tributary streams and lakes to the Middle Fork Flathead River. By comparing patterns of isolation by distance between hybridized and non-hybridized sample sites we statistically test, for the first time to our knowledge, the hypothesis that low levels of hybridization with rainbow trout may increase the dispersal rate of westslope cutthroat trout.

Methods

Study Area

We used three criteria to select sites for this study. First, we only included samples from flow connected sites. This criterion was established because if a fish cannot disperse into a site, we can gain no knowledge about the effects of hybridization on dispersal at that site. Second, sites were only included from watersheds that contained both hybridized and non-hybridized samples. The inclusion of non-hybridized samples allowed us to estimate the ancestral pattern of connectivity across the landscape, which served as a reference level for comparing the dispersal rates observed among hybridized samples. Finally, because we were testing the effects hybridization with rainbow trout, and both Yellowstone cutthroat trout (*O. c. bouvieri*) and rainbow trout were introduced into our study area (Hitt et al. 2003), only those sites with westslope cutthroat trout and rainbow trout genetic contributions were included in analysis.

Based on these three requirements, we selected 17 flow-connected sites in the Middle Fork Flathead River (Figure 4.1) that had a mixed hybridization history for this study (Table 4.1). These samples are part of a larger study designed to investigate the extent of hybridization between westslope cutthroat trout, rainbow trout and Yellowstone cutthroat trout across Glacier National Park (C. Muhlfeld *unpublished data*). Between two and 20 individuals were collected at each site either by angling or electrofishing (Table 4.1). The Middle Fork Flathead River comprises a majority of the southern boundary of Glacier National Park.

Molecular Methods

Individual fish were genotyped at 39 nuclear single nucleotide polymorphism (SNP) loci with putative fixed allelic differences between westslope cutthroat trout and rainbow trout and 34 SNP loci polymorphic within westslope cutthroat trout populations on Fluidigm 96.96 microfluidic PCR chips. The 39 diagnostic loci were previously described by Amish et al. (2012) and Campbell et al. (2012), and the 34 polymorphic loci were previously described by Campbell et al. (2012). Individual loci names are listed in Table 4.2.

Preliminary Genetic Data Analysis

We removed samples from analysis that failed to amplify more than 5% of all genotyped loci. In addition, a random 5% of our samples were re-run as a quality control measure. If there was a consistent pattern of allelic dropout between the two samples the heterozygous genotype was selected at discordant loci. Where no consistent pattern of mismatched genotypes existed, loci were coded as missing data.

After screening samples for genotyping errors, we tested whether loci that were supposed to be diagnostic had unwanted polymorphisms that precluded their use in this study (see Appendix A for a description of how this was done). We also tested whether a hybridized sample was collected from a hybrid swarm (*sensu* Leary et al. 1995), which would indicate that non-native alleles are randomly distributed throughout the population. This test is necessary because hybridization between genetically divergent populations creates linkage disequilibrium, which is the non-random association of alleles across loci, throughout the genome that is dependent on the hybridization history of the population

(Pfaff et al. 2001). If this linkage disequilibrium due to hybridization is present in a sample, frequently used statistical methods assuming a binomial distribution of non-native alleles are no longer valid because the assumption that each binomial count consists of independent and identically distributed Bernoulli trials is violated. We tested for the presence of linkage disequilibrium by calculating the mixing degree statistic (Kalinowski and Powell *In Review*) in each sample, and testing whether this value was statistically different from zero at $\alpha = 0.05$.

We estimated the proportion of rainbow trout alleles in each individual, and at each site, as the sum of all rainbow trout alleles divided by the total number of amplified alleles. Because linkage disequlibrium was present in many of the samples, we were unable to construct confidence intervals for these estimates using methods based on the binomial distribution. Therefore, we constructed 95% one-sided confidence intervals for these estimates using the percentile bootstrap method of Efron (1979) as presented in Manly (2007). The upper bound of the confidence interval was the 95% quantile of the bootstrap distribution of the proportion of rainbow trout alleles present at the site, calculated by resampling across individuals and loci (Tang et al. 2005) 10,000 times. Unless otherwise stated, all analyses were performed in the statistical computing package R version 2.15.3 (R Development Core Team 2012).

Finally, we tested whether the observed genotype frequencies at each polymorphic locus conformed to Hardy-Weinberg expectations using the exact test of Guo and Thompson (1992). This test was run using the Markov chain exact test in GENEPOP on the web (Rousset & Raymond 1995, Rousset 2008). Decisions about the

nature of the departures from Hardy-Weinberg equilibrium were made following a false discovery rate correction of Benjamini and Yekutieli (2001), as suggested by Narum (2006). Loci that exhibited a consistent departure from Hardy-Weinberg equilibrium (either an excess of homozygotes or an excess of heterozygotes) in multiple samples were removed from subsequent analyses.

Statistical Analysis of Dispersal Patterns

We tested for an increase in the dispersal rate between hybridized samples by comparing the patterns of isolation by distance among these sites to that observed among pairs of sites in which at least one of the samples was non-hybridized. Isolation by distance was tested by associating the pairwise stream distance between two sites to the ratio of $\frac{F_{ST}}{1-F_{ST}}$ using a simple linear regression (Rousset 1997). F_{ST} values were estimated with Weir and Cockerham's θ (Weit and Cockerham 1984), and pairwise stream distance was calculated with the network analyst tools in ArcMap version 10.0 (Environmental Systems Research Institute, Inc., Redlands, CA).

We initially tested for statistical evidence of a relationship between genetic distance and geographic distance for the hybrid-hybrid sites using a Mantel test (Mantel 1967). We then tested for a difference between the patterns of isolation by distance among hybridized sites and among pairs of sites in which at least one of the samples was non-hybridized in a similar manner to constructing a Mantel based confidence interval as described by Manly (2007). This test is based on the logic that there should be no relationship between the stream distance matrix and the residual genetic distance matrix,

which is calculated as the difference between the estimated $\frac{F_{ST}}{1-F_{ST}}$ values (calculated using the true slope of the isolation by distance line) and the observed $\frac{F_{ST}}{1-F_{ST}}$ values. Therefore, to test whether hybrid sites have a weaker pattern of isolation by distance we estimated the relationship between the stream distance matrix and the residual genetic distance matrix where the estimated $\frac{F_{ST}}{1-F_{ST}}$ values were calculated using the non-hybridized pattern of isolation by distance. If isolation by distance was stronger among the non-hybridized sites, then the residual genetic distance matrix should have a strong negative relationship with the stream distance matrix (i.e. all estimated values would be too high). A Monte Carlo approach was used to incorporate uncertainty in the estimated slope of the isolation by distance line between pairs of sites where at least one was non-hybridized. Therefore, 10,000 slope coefficients were sampled from a normal distribution whose mean equaled the estimated slope parameter in the non-hybrid analysis and whose standard deviation equaled the estimated standard error of that parameter. Equivalence was assessed by calculating the proportion of the 10,000 samples that produced a residual matrix in which fewer than 5% of randomizations gave a lower correlation with the stream distance matrix.

The observed genetic differentiation of the Upper Park Creek and Autumn Creek sites was larger than other non-hybridized sites in the North Fork Flathead River and the Middle Fork Flathead River (see Appendix B). We hypothesize that fish at these two sites only express a resident life history (see Appendix B), and as such the processes affecting the genetic divergence at these sites are not the same as those operating at other flow-

connected sites. Therefore, we classified these sites as potential outliers and analysis was run both including and excluding them.

F_{ST} Estimation in Hybridized Samples

Pritchard et al. (2000) suggested that by calculating genetic distance based on estimated ancestral population allele frequencies reported from their program *Structure*, genetic distance between ancestral populations could be estimated. This program uses a Bayesian mixture model to simultaneously estimate the ancestry of alleles in an individual, the admixture proportions in an individual, and the allele frequencies in the ancestral populations from the same loci (Pritchard et al. 2000). We present an extension of the Gibbs sampling algorithm (A3) developed by Pritchard et al. (2000) that estimates mixture proportions at polymorphic loci from a separate panel of genotyped diagnostic loci. Therefore, in this model we are only simultaneously estimating the ancestry of alleles in an individual and the allele frequencies in the ancestral populations from the same panel of loci. Pritchard et al. (2000) suggested this extension for situations where loci with fixed allelic differences are available (similar to the current study).

To highlight the distinction between the marker types we use the superscript D or P to indicate equations whose likelihood functions are based on information from either diagnostic or polymorphic loci respectively. The major difference between our presentation and the original algorithm (A3) of Pritchard et al. (2000) is that we did not select the parameters of the prior distribution for the mixing proportions using a Metropolis-Hastings step, instead we selected a $Beta\left(\frac{1}{2},\frac{1}{2}\right)$ prior. This change was made because we are using diagnostic loci with fixed allelic differences and wanted to have a

non-informative prior on the mixing proportions. In addition, a non-informative $Beta\left(\frac{1}{2},\frac{1}{2}\right)$ prior distribution added little bias with 78 amplified alleles (approximate bias in the mean of the posterior distribution = 0.0063). In addition, our notation differs from that of Pritchard et al. (2000) in that we specify Beta (a special case of the Dirichlet) distributions for admixture proportions and allele frequencies because we were only estimating two mixing populations (westslope cutthroat trout or rainbow trout) and the frequency of two alleles at each locus (due to the use of diallelic SNP loci).

Modified Algorithm

First, we modeled $q_i|X,Z \sim Beta\left({}^{(D)}m_i + \frac{1}{2}, {}^{(D)}n_i - {}^{(D)}m_i + \frac{1}{2}\right)$, where ${}^{(D)}m_i$ is the total number of amplified alleles diagnostic for rainbow trout, ${}^{(D)}n_i$ is the total number of amplified diagnostic alleles in the ith individual, and the ½ is from the Beta prior. Consistent with the *Structure* algorithm (Pritchard et al. 2000), X is the observed genotype of the ith individual, Z is the proportion of the ith individual's genome ancestral to one of the two species, and P are the unknown allele frequencies in the ancestral populations. We next updated the allele frequencies for westslope cutthroat trout $p_{Wl\cdot}|X,Z \sim Beta\left({}^{(P)}n_{Wl1} + \frac{1}{2}, {}^{(P)}n_{Wl2} + \frac{1}{2}\right)$, where ${}^{(P)}n_{Wl1}$ and ${}^{(P)}n_{Wl2}$ are the number of amplified westslope cutthroat trout alleles of both states, and the ½ is from the Beta prior. In the same manner we updated allele frequencies for the rainbow trout population $p_{Rl\cdot}|X,Z \sim Beta\left({}^{(P)}n_{Rl1} + \frac{1}{2}, {}^{(P)}n_{Rl2} + \frac{1}{2}\right)$, where ${}^{(P)}n_{Rl1}$ and ${}^{(P)}n_{Rl2}$ are the number of amplified rainbow trout alleles of both states, and the ½ is from the Beta prior.

The number of rainbow trout alleles at a locus was modeled based on an identical distribution to that reported by Pritchard and others (2000). If $z_l^{(i,a)}$ is defined as the population of origin of allele *a* at the lth locus, then $Pr\left(z_l^{(i,a)} = wct|X,P\right) = \frac{(1-q_i)*\Pr\left(x_l^{(i,a)}|P,z_l^{(i,a)}=wct\right)}{q_i*\Pr\left(x_l^{(i,a)}|P,z_l^{(i,a)}=rbt\right)+(1-q_i)*\Pr\left(x_l^{(i,a)}|P,z_l^{(i,a)}=wct\right)}$ was the probability that allele *a* at the lth locus descended from a westslope cutthroat trout ancestor (*wct*) and $Pr\left(z_l^{(i,a)} = rbt|X,P\right) = \frac{q_i*\Pr\left(x_l^{(i,a)}|P,z_l^{(i,a)}=rbt\right)}{q_i*\Pr\left(x_l^{(i,a)}|P,z_l^{(i,a)}=rbt\right)+(1-q_i)*\Pr\left(x_l^{(i,a)}|P,z_l^{(i,a)}=wct\right)}$ was the probability that it descended from a rainbow trout ancestor. In each of these equations the probability that the observed allele was of a given state, $\Pr\left(x_l^{(i,a)}|P,z_l^{(i,a)} = wct\right)$or $\Pr\left(x_l^{(i,a)}|P,z_l^{(i,a)} = rbt\right)$, was simply the allele frequency, $p_{Wl\cdot}$ or $p_{Rl\cdot}$, in either the westslope cutthroat trout population or the rainbow trout population.

Finally, missing genotypes at a locus were modeled based on the average allele frequency weighted by the estimated proportion of an individual's genome with a given ancestry. This simple missing data model was selected based on discussion of modeling strategies in Lunn et al. (2012), and the assumption that missing alleles occurred completely at random.

We used the R2WinBUGS package in R (Sturtz et al. 2005) to estimate the number of westslope cutthroat trout alleles ($\sum_i z_l^{(i,a)} = wct$) and the westslope cutthroat trout allele frequencies ($\boldsymbol{p_W}$) for each sample from a site. Estimates were made from three independent chains run for 30,000 total iterations (saving every 15th iteration after discarding the initial 15,000 iterations) in WinBUGS version 1.4.3 (Lunn et al. 2000).

Run length was selected to achieve approximate convergence as indicated by visual inspection of trace plots and checking that the upper bound of the confidence interval for the $\hat{R}$ statistic (Gelman et al. 2004) was less than 1.1 for all sample size, westslope cutthroat trout allele frequency, mixing proportion and missing data parameters as calculated with the CODA package in R (Plummer et al. 2006). In addition we wanted to make sure the number of effective draws was greater than 500 as calculated with R2WinBUGS. Gelman et al. (2004) argued that for most applications 100 effective draws is enough for inference about the center of the posterior distribution, but increasing the number of samples increases the stability of the numeric summaries. We selected 500 effective draws as a compromise between increasing stability and run-time, and because we were estimating the median of the posterior distributions. Following estimation of westslope cutthroat trout allele frequencies and sample sizes, we used the random gametes formulation of Weir and Cockerham (1984) to calculate their F_{ST} estimator θ using the posterior median allele frequencies and sample sizes. The random gametes formulation was used due to our inability to estimate the number of individuals heterozygous at a locus after removing rainbow trout alleles.

Mixture Model Performance

We investigated the accuracy and precision of estimated θ values following the removal of rainbow trout alleles using individual-based simulations. Each simulated individual had 1040 loci with fixed differences between westslope cutthroat trout and rainbow trout, and 1040 loci that were polymorphic within westslope cutthroat trout. Rainbow trout were assumed fixed for a randomly selected allele at each of the

polymorphic westslope cutthroat trout loci. These loci were alternated across 52 linkage groups, with a recombination rate of a single crossover per linkage group per generation. The number of linkage groups was selected to represent the number found in westslope cutthroat trout (Allendorf and Thorgaard 1984), and the recombination rate was based on an estimate developed for rainbow trout (Danzmann et al. 2005). A single population of 2000 individuals, with equal number of males and females, was simulated and then split into two populations of 1000 individuals, with equal number of males and females. A Pólya urn scheme was employed to reduce the effective size of each population to approximately 230 individuals (Anderson 2001). This ratio of census to effective population size was selected to be the midpoint of estimated ratios from a Paiute cutthroat trout (*O. c. seleniris*) population (Finger et al. 2011). Populations were allowed to evolve in isolation for 0 or 50 generations, at the end of which 50 or 200 non-hybridized rainbow trout, with equal numbers of males and females, were introduced into each population. Random mating was allowed to progress for 15 generations, assuming equal fitness for all individuals, and an effective population size of 230 individuals.

One hundred samples of 20 individuals were drawn from each population and genotyped at 39 randomly selected diagnostic loci and 31 randomly selected polymorphic loci. These fish were also genotyped at an additional 57 randomly selected diagnostic loci and 65 randomly selected polymorphic loci. Therefore, the number of loci tested was a subset mimicking the number of markers available for this study, as well as a set assuming 96 diagnostic, and 96 polymorphic, loci were genotyped. The Bayesian mixture model was used to estimate the number of westslope cutthroat trout alleles, and westslope

cutthroat trout allele frequencies in each of the samples drawn from each population. Weir and Cockerham's θ (Weir and Cockerham 1984) was calculated using the random gametes formulation for the naïve sample, the median allele frequencies and median sample sizes estimated with the Bayesian mixture model, and the true sample after removing all rainbow trout alleles. We used the posterior median allele frequencies and the posterior median number of westslope cutthroat trout alleles to calculate θ. We calculated the bias, relative to the parametric value calculated from the complete populations, and variance of the estimators.

Results

Preliminary Genetic Data Analysis

We did not detect any evidence of polymorphisms at loci that have putatively fixed allelic differences between westslope cutthroat trout and rainbow trout in samples from sites in the Middle Fork Flathead River (smallest P-value = 0.055). Six sample sites included in this study did not show any evidence of hybridization with rainbow trout, and the remaining 11 sites had estimated proportions of rainbow trout genes ranging from 0.004 in the Harrison Lake sample to 0.223 in the Lower Lincoln Creek sample (Table 4.1). Of these 11 sites only one sample, from Lower Ole Creek, appeared to have been collected from a hybrid swarm (Table 4.1).

Departures from Hardy-Weinberg equilibrium consistently occurred at two loci: one departure was consistent with patterns expected due to the amplification of paralogous (duplicated) loci and the other departure was characteristic of null alleles (no-

amplification due to mutations within the primer site) at a locus. A total of 363 tests for Hardy-Weinberg equilibrium were run from the 34 polymorphic loci genotyped in samples from 17 sites. This number of tests led to an adjusted $\alpha = 0.0077$ (Benjamini-Yekutieli 2001). There were 15 individual tests, or approximately 4.1% of all comparisons, with observed departures from Hardy-Weinberg equilibrium as extreme or more extreme than that expected by random chance. Of these 15 departures, 6 occurred at the *OclVAR_RAD-69806_GL* locus and 4 occurred at the *Ocl_var_mx1_129_NC* locus. There was an excess of heterozygotes in all samples with a statistical departure from Hardy-Weinberg equilibrium at this first locus as indicated by a negative value of Weir and Cockerham's (1984) estimate of F_{IS} (range from -1.00 to -0.89). This locus was fixed for the heterozygous genotype at all six of these sites, and an additional three others in the dataset. Excess heterozygotes suggest that this primer is amplifying paralogous loci (Hohenlohe et al. 2013). All sites with evidence of a departure from Hardy-Weinberg equilibrium at the second of these loci contained an excess of homozygotes as indicated by their estimated F_{IS} value (range from 0.74 to 1.00). An excess of homozygotes in multiple samples suggest that null alleles are present at this locus. These loci were removed from further analysis. After removing the loci, we found five additional departures from Hard-Weinberg equilibrium, or approximately 1.5% of all comparisons, each of which occurred at a different locus (*Ocl_var_hsc71pro_71_00*, *OclVAR_RAD_16087_GL*, *Oclvar_fkbp2_62NC*, *Ocl_var_impa1_189_NC*, and *OclVar96899L_Garza*) in a single population. Because there was no apparent pattern to these departures, no additional loci were removed from analysis.

In addition to removing the two loci due to consistent departures from Hardy-Weinberg equilibrium, one locus (*Ocl_var_npc2_145_NC*) failed to amplify in most, or all, of the individuals in the samples from Middle Coal Creek and Lower Lincoln Creek. Because the Bayesian mixture model estimated missing alleles, and there was little to no information available from which to base estimates in these samples, this locus was removed from analysis. Therefore, estimates of population structure and dispersal rate were made using a total of 31 polymorphic loci.

Statistical Analysis of Dispersal Patterns

We did not detect a difference in the relationship between genetic divergence and geographic distance among hybridized sites and non-hybridized sites in the Middle Fork Flathead River drainage (as we expected if low levels of hybridization increased the dispersal rate between populations). We found some evidence that the genetic distances between pairs of hybridized sites were associated with the stream distance separating them (Mantel r = 0.539, P-value = 0.005). However, we did not detect evidence that this association was lower than the estimated association between pairs of sites where at least one site contained only non-hybridized individuals (mean Mantel r = 0.008, only 2.44% of all Monte Carlo samples had a P-value ≥ 0.95, Figure 4.2). This result indicates similar dispersal patterns are present among the hybridized and non-hybridized sites we sampled in this watershed. However, this finding depended on whether or not the samples from Upper Park Creek and Autumn Creek were included in the analysis (Figures 4.2 and 4.3). If the samples from Upper Park Creek and Autumn Creek are included there is evidence that the connectivity between hybridized sites may be higher than between pairs of sites

where at least one sample contained only non-hybridized individuals (mean Mantel r = -0.989, 99.93% of all Monte Carlo samples had a P-value ≥ 0.95, Figure 4.3).

When the proportion of non-native genes in simulated populations was low, the Bayesian mixture model had very little bias and had a similar variance to the estimated genetic divergence based on the true allele frequencies and sample sizes (Tables 4.3 and 4.4). However, when simulated populations were isolated after a high proportion of non-native genes were introduced, the mixture model had a relatively large bias of between 0.018 and 0.02 in later generation samples (Table 4.3). Although the naïve estimator appears to be unbiased at small F_{ST} values, it has a large negative bias (approximately -0.04) at moderate F_{ST} values and high proportions of non-native genes (Table 4.3). In general the Bayesian estimator appears to be more accurate and precise in early generations following admixture (Tables 4.3 and 4.4), and precision increases with an increase in the number of genotyped loci (Table 4.4).

Discussion

Using samples from 17 tributary streams and lakes to the Middle Fork Flathead River we investigated the patterns of dispersal among hybridized and non-hybridized westslope cutthroat trout populations. We also tested the hypothesis that low levels of hybridization will increase the dispersal rate among populations. We initially screened samples for the presence of hybridization, and compared patterns of genetic structure between hybridized and non-hybridized sites using estimates of isolation by distance among each class of sites. To complete this analysis we modified the *Structure* algorithm

based on suggestions in Pritchard et al. (2000) and used a small simulation study to investigate the performance of this Bayesian mixture model in estimating genetic distance between samples from hybridized populations.

Hybridization appears to be ongoing in the Middle Fork Flathead River, and we documented its presence in the sample from lower Ole Creek, which only contained non-hybridized fish in 1998 (Hitt et al. 2003). Previous work within this drainage has uncovered a pattern of hybridization wherein samples generally appear to be non-hybridized, but also include highly hybridized individuals (R. Leary *personal communication*). This pattern is apparent in our results as well. We only detected a single sample from the Middle Fork Flathead River that appeared to come from a hybrid swarm (Table 4.1). Also, many of the sites contained individuals for which the estimated proportion of the genomes that came from rainbow trout was higher than the overall site estimate (Figure 4.4). Furthermore, individuals with proportions of rainbow trout genes approximately consistent with levels expected for recent hybridization events (first and second generation) were detected in the lower and upper sites from Lincoln Creek, the lower site in Harrison Creek, as well as the sites in Nyack Creek, Peril Creek and Bear Creek (Figure 4.4). Based on these observations, there is strong support for the hypothesis that highly hybridized individuals are contributing to the spread of hybridization within this landscape.

Previous studies point to Abbott Creek (Hitt et al. 2003, Boyer et al. 2008), Mill Creek and the Whitefish River (Hitt et al. 2003) as likely sources for these highly hybridized individuals in this watershed. We only detected a single individual, in the

sample from lower Harrison Creek, that had a similar proportion of rainbow trout genes in its genome (0.987) to levels observed in Abbott Creek (0.975 Hitt et al. 2003, 0.916 Boyer et al. 2008), Mill creek (0.894 Hitt et al. 2003) or the Whitefish River (0.982 Hitt et al. 2003). However, under an island model of migration (Wright 1940) we would expect that historically each population exchanged approximately 2.82 migrants per generation based on the observed level of genetic divergence among sites in this watershed (average pairwise $F_{ST} = 0.08$ after removing the samples from Upper Park and Autumn Creek). Therefore, because of the low levels of genetic divergence observed in this landscape (Figure 4.2), it is unclear if the dispersal rate of highly hybridized individuals is greater than the historic levels of connectivity in this landscape.

With regard to dispersal patterns among our sampled sites, we found no difference in the patterns of isolation by distance between hybridized sites and pairs of sites where at least one sample was non-hybridized (Figure 4.2). It is important to note that this result is dependent on whether the samples from Upper Park Creek and Autumn Creek were included in the analysis (Figure 4.3). However, we believe that the allele frequencies at these sites are being affected by different ecological processes than are affecting the other flow connected sites in this watershed and, therefore, they should not be included in comparisons of patterns of isolation by distance. This contention is supported by the distribution of pairwise θ values for these sites (see Appendix B).

Although the simulated populations indicated the *Structure* algorithm is capable of estimating ancestral F_{ST} values between hybrid populations as was suggested by Pritchard et al. (2000), the success of this algorithm appears to be greatest in early

generations and in samples with moderate population structure (Tables 4.4 and 4.5). In addition we observed positive bias in the estimates of the F_{ST} between non-isolated sites at high proportions non-native genes. We do not believe either of these issues is a problem for this study for the following reasons. First, the fact that only a single sample was collected from a hybrid swarm indicates that we are estimating allele frequencies at sites with recent hybridization events. Second, while there was bias in the estimates of F_{ST} at low values and high proportions of non-native genes, we only had a single comparison of sites with this level of non-native genes (Lower Lincoln Creek and Lower Harrison Creek).

Here we have shown evidence that hybridization is spreading in this landscape due to a combination of long distance dispersal of highly hybridized individuals and exchange among neighboring populations, a finding consistent with previous studies in this landscape (Hitt et al. 2003, Boyer et al. 2008). Although we were unable to detect a difference in the dispersal rate among hybridized and non-hybridized sites in this landscape (Figure 4.2), there was an apparent negative relationship between the average proportion of rainbow trout genes and the pairwise genetic distance between sites in this watershed (Figure 4.5). We hypothesize that in this landscape the observed relationship between hybridization and genetic distance is likely due to a combination of introduction history and historic connectivity in this watershed. This hypothesis is supported by the combination of the fact that the likely sources of highly hybridized individuals are low in the Flathead River drainage (Hitt et al. 2003, Boyer et al. 2008), and the detected pattern of isolation by distance that is consistent with Boyer et al.'s (2008) findings in the North

Fork Flathead River. Additional studies comparing dispersal rates in other watersheds should help inform if this is a viable hypothesis for the spread of hybridization between westslope cutthroat trout and rainbow trout.

Tables and Figures

Table 4.1. Site information for the 17 sample locations from the Middle Fork Flathead River, with figure codes corresponding to Figure 4.1, and 95% confidence intervals for the estimated proportion of rainbow trout genes (P(RBT)) at a site and the mixing degree statistic (m_d) reported in brackets.

Figure code	Drainage	Site	n	P(RBT)	m_d
1	Middle Fork	Lower Lincoln Creek	17	0.22 [0.00, 0.35]	0.50 [0.39, 0.67]
2	Middle Fork	Upper Lincoln Creek	17	0.06 [0.00, 0.10]	0.87 [0.83, 0.91]
3	Middle Fork	Lincoln Lake	2		
4	Middle Fork	Lower Harrison Creek	20	0.22 [0.00, 0.34]	0.44 [0.25, 0.68]
5	Middle Fork	Harrison Lake	14	0.00 [0.00, 0.01]	0.99 [0.99, 1.00]
6	Middle Fork	Upper Harrison Creek	6	0.02 [0.00, 0.04]	0.92 [0.92, 1.00]
7	Middle Fork	Middle Nyack creek	18	0.06 [0.00, 0.13]	0.59 [0.42, 0.99]
8	Middle Fork	Pinchot Creek	12		
9	Middle Fork	Peril Creek	7	0.04 [0.00, 0.10]	0.76 [0.74, 1.00]
10	Middle Fork	Middle Coal Creek	16	0.01 [0.00, 0.04]	0.85 [0.83, 1.00]
11	Middle Fork	Upper Coal Creek	20		
12	Middle Fork	Lower Muir Creek	20	0.01 [0.00, 0.01]	0.96 [0.94, 1.00]
13	Middle Fork	Upper Muir Creek	19		
14	Middle Fork	Upper Park Creek	19		
15	Middle Fork	Lower Ole Creek	19	0.00 [0.00, 0.01]	1.00 [0.99, 1.00] *
16	Middle Fork	Middle Bear Creek	4	0.05 [0.00, 0.13]	0.81 [0.81, 1.00]
17	Middle Fork	Autumn Creek	14		

* indicates that the sample from Lower Ole Creek was statistically mixed

Table 4.2. Locus names for diagnostic and polymorphic loci used in this analysis.

Locus category	Locus name
Diagnostic	*OmyRD_RAD_29252_Hoh*, *OmyRD_RAD_77157_Hoh*, *OmyRD_RAD_30378_Hoh*, *OclRD_P53T7R1_Har*, *OmyRD_RAD_30423_Hoh*, *OmyRD_RAD_59515_Hoh*, *OclRD_Thymo_320Kal*, *OmyRD_RAD_48301_Hoh*, *OmyRD_RAD_49759_Hoh*, *OclRD_P53T7R2_Har*, *OmyRD_URO_302May*, *OmyRD_RAD_20663_Hoh*, *OmyRD_RAD_51740_Hoh*, *OmyRD_RAD_22111_Hoh*, *OmyRD_RAD_55820_Hoh*, *OmyRD_RAD_5666_Hoh*, *OmyRD_F5_136May*, *OmyRD_RAD_42014_Hoh*, *OmyRD_RAD_54584_Hoh*, *OclRD_CLK3W5_Har*, *OclWD_CLK3W1_Har*, *OclWD101119_Garza*, *OmyWD_RAD_76689_Hoh*, *OclWD_114315L _Garza*, *OclWD_Tnsf_387Kal*, *OmyWD_RAD_55391_Hoh*, *OclWD_P53_307Kal*, *OclWD111312_Garza*, *OclWD_107031L_Garza*, *OclWD_PrLcW1_Har*, *OmyWD_RAD_54516_Hoh*, *OclWD_105075L_Garza*, *OmyWD_RAD_52968_Hoh*, *OclWD114336_Garza*, *OclWD103713_Garza*, *OclWD107074_Garza*, *OclWD109651_Garza*, *OclWD_129170L_Garza*, *OclWD_ppie_32NC*
Polymorphic	*OclVar_115987L _Garza*, *OclVAR_RAD_51955_GL*, *Ocl_var_hsc71pro_71_NC*, *OclVar105105_Garza*, *Oclvar_arp_117NC*, *OclVAR_RAD_16087_GL*, *Oclvar_arp_314NC*, *OclVar113772_Garza*, *OclVAR_RAD_16730_GL*, *Oclvar_fKbp2_62NC*, *Oclvar_vatf_277NC*, *Oclvar_Carpa1_45NC*, *OclVAR_RAD_69806_GL*, *Ocl_var_mtap_224_NC*, *OclVar_bnip3_263NC*, *Ocl_var_npc2_145_NC*, *Ocl_var_mx1_129_NC*, *Oclvar_mkpro_69NC*, *Oclvar_ca050_39NC*, *Oclvar_pigH_42NC*, *OclVar102483_Garza*, *Oclvar_pop5_50NC*, *Ocl_var_impa1_189_NC*, *Ocl_var_bcAKala_259_NC*, *Oclvar_pnpo_56NC*, *Oclvar_srp09_172NC*, *OclVar96899L_Garza*, *OclVAR_RAD_43407_GL*, *Oclvar_ada10_100NC*, *OclVAR_cin_90NC*, *Oclvar_parp3_19NC*, *Oclvar_metB_106NC*, *OclVar128996_Garza*, *Omg_LDHB2i6_Garza*

Table 4.3. Observed bias in estimates of Weir and Cockerham’s θ calculated from the observed sample (Naïve), the posterior medians from the Bayesian mixture model (Bayesian), and the sample after removing all rainbow trout alleles (True).

		P(RBT) = 0.05				P(RBT) = 0.20			
		3 generations		15 generations		3 generations		15 generations	
		31,39	96,96	31,39	96,96	31,39	96,96	31,39	96,96
	Naïve	-0.00084	-0.00078	-0.00226	-0.00081	0.00018	-0.00135	-0.01052	-0.01015
θ = 0.0	Bayesian	0.00029	0.00044	0.00069	0.00230	0.00855	0.00839	0.01809	0.02063
	True	-0.00040	-0.00033	-0.00243	-0.00091	0.00098	0.00065	0.00034	-0.00058
	Naïve	-0.01417	-0.01225	-0.01724	-0.01508	-0.04113	-0.04058	-0.04357	-0.04753
θ = 0.1	Bayesian	-0.00264	-0.00125	-0.00680	-0.00390	0.00577	0.00447	0.00524	0.00399
	True	-0.00101	0.00099	-0.00471	-0.00097	0.00128	0.00060	0.00148	-0.00307

Table 4.4. Observed variance in estimates of Weir and Cockerham’s θ calculated from the observed sample (Naïve), the posterior medians from the Bayesian mixture model (Bayesian), and the sample after removing all rainbow trout alleles (True).

		P(RBT) = 0.05				P(RBT) = 0.20			
		3 generations		15 generations		3 generations		15 generations	
		31,39	96,96	31,39	96,96	31,39	96,96	31,39	96,96
	Naïve	0.000065	0.000029	0.000129	0.000048	0.000096	0.000041	0.000279	0.000071
θ = 0	Bayesian	0.000080	0.000033	0.000165	0.000056	0.000202	0.000064	0.000787	0.000206
	True	0.000079	0.000033	0.000165	0.000057	0.000125	0.000032	0.000405	0.000117
	Naïve	0.000893	0.000258	0.001038	0.000299	0.000684	0.000246	0.000706	0.000216
θ = 0.1	Bayesian	0.001127	0.000320	0.001210	0.000364	0.001034	0.000371	0.001566	0.000499
	True	0.001100	0.000325	0.001316	0.000370	0.001174	0.000326	0.001500	0.000445

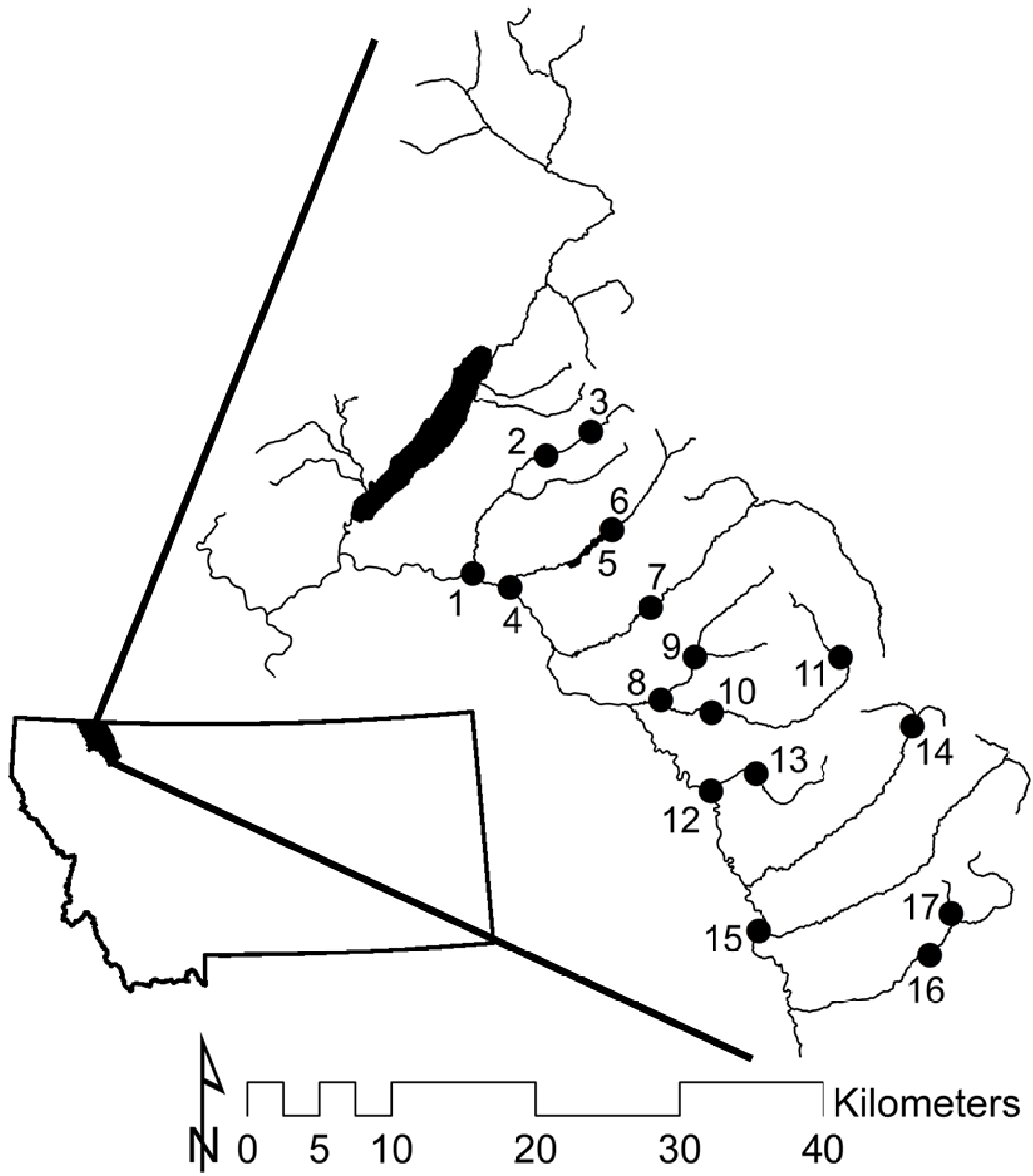

Figure 4.1. Map of the sample locations in the Middle Fork Flathead River drainage, with site labels corresponding to Table 4.1.

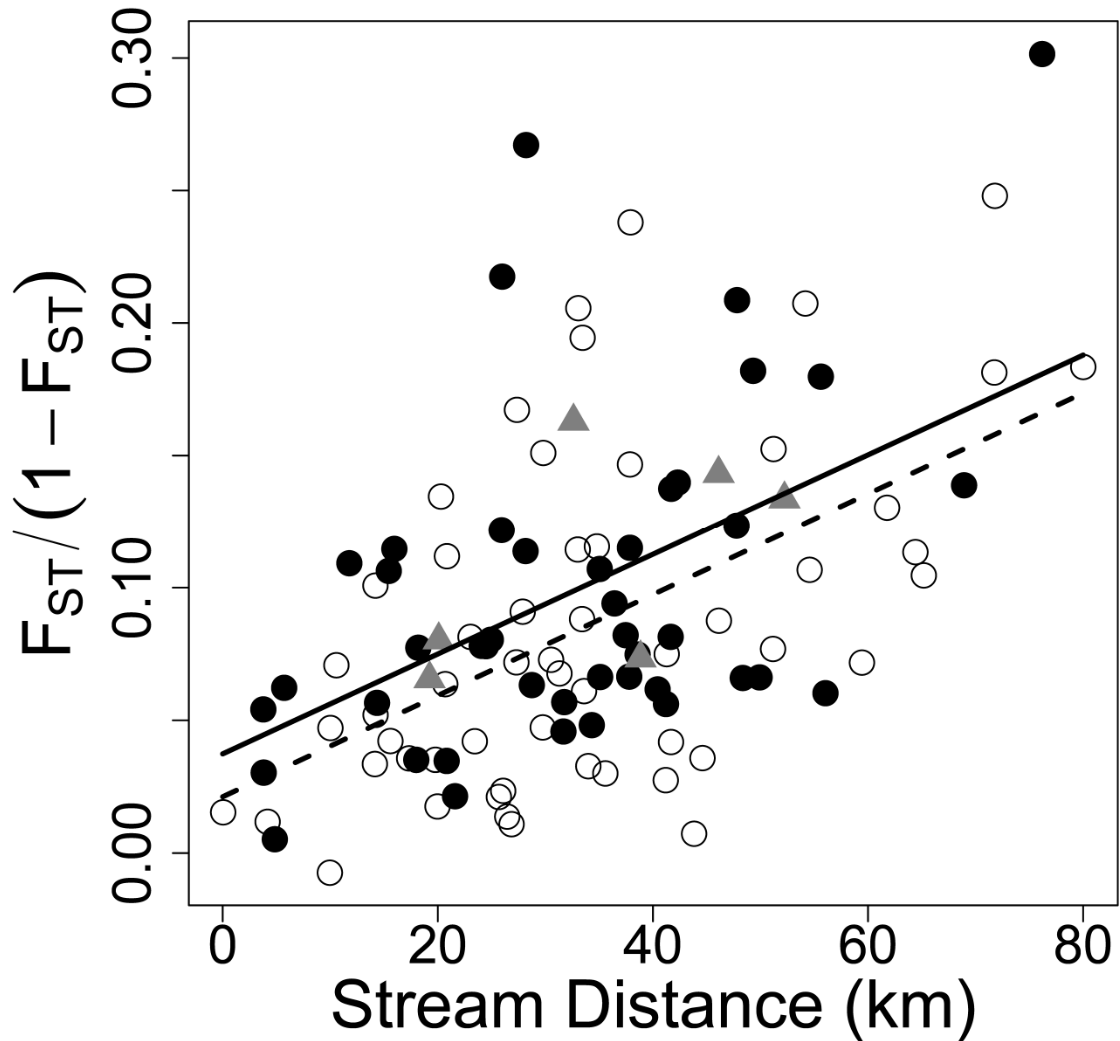

Figure 4.2. Isolation by distance for all putatively flow connected sites in the Middle Fork Flathead River drainage after removing the samples from Upper Park Creek and Autumn Creek. The dashed line plots the estimated association between genetic distance and geographic distance among hybridized sites, represented by open circles, whereas the solid line plots the estimated association among sites where at least one sample did not contain hybridized individuals. This relationship was estimated using both the pairs where a single site did not contain hybridized individuals, represented by closed circles, and pairs where neither site contained hybridized individuals, represented by gray triangles.

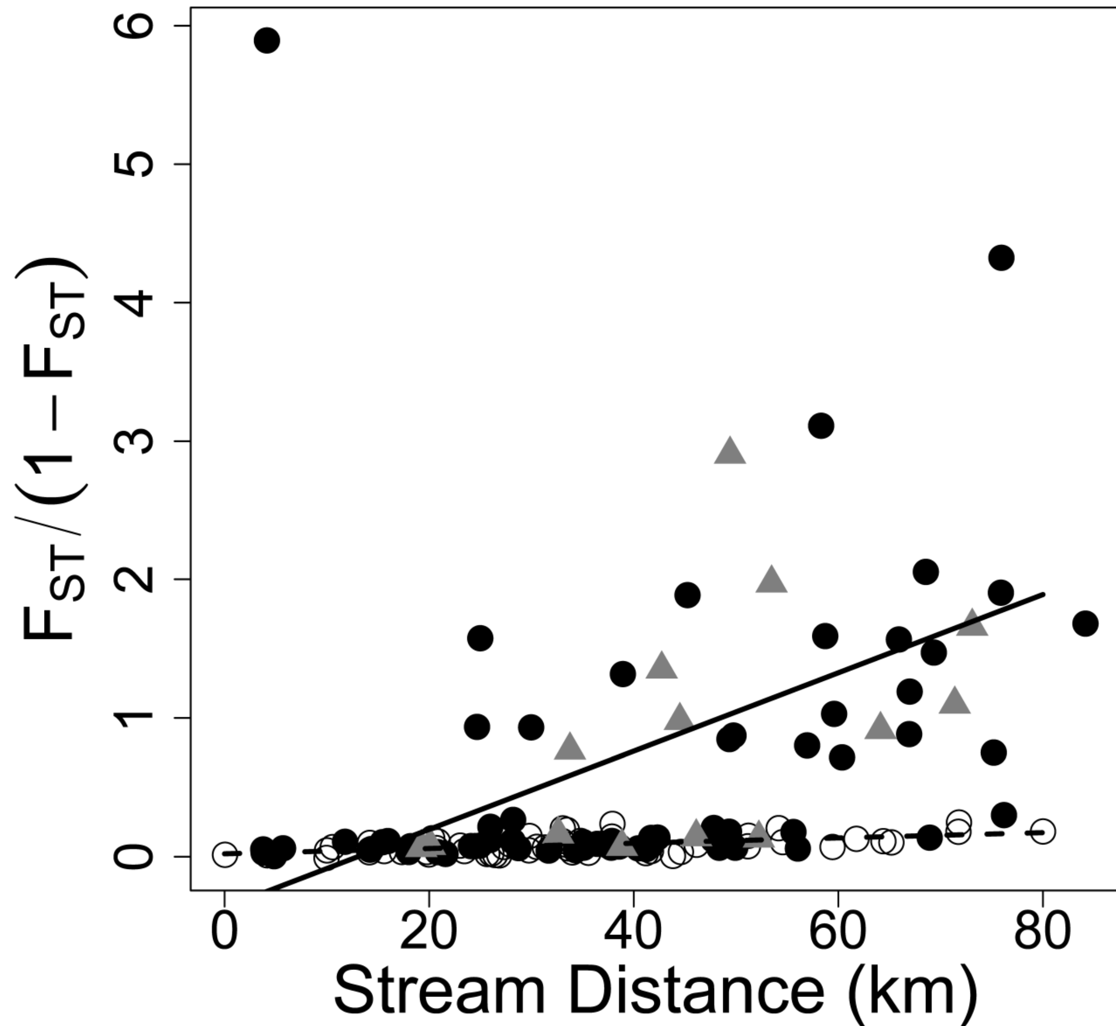

Figure 4.3. Isolation by distance for all putatively flow connected sites in the Middle Fork Flathead River drainage. The dashed line plots the estimated association between genetic distance and geographic distance among hybridized sites, represented by open circles, whereas the solid line plots the estimated association among sites where at least one sample did not contain hybridized individuals. This relationship was estimated using both the pairs where a single site did not contain hybridized individuals, represented by closed circles, and pairs where neither site contained hybridized individuals, represented by gray triangles.

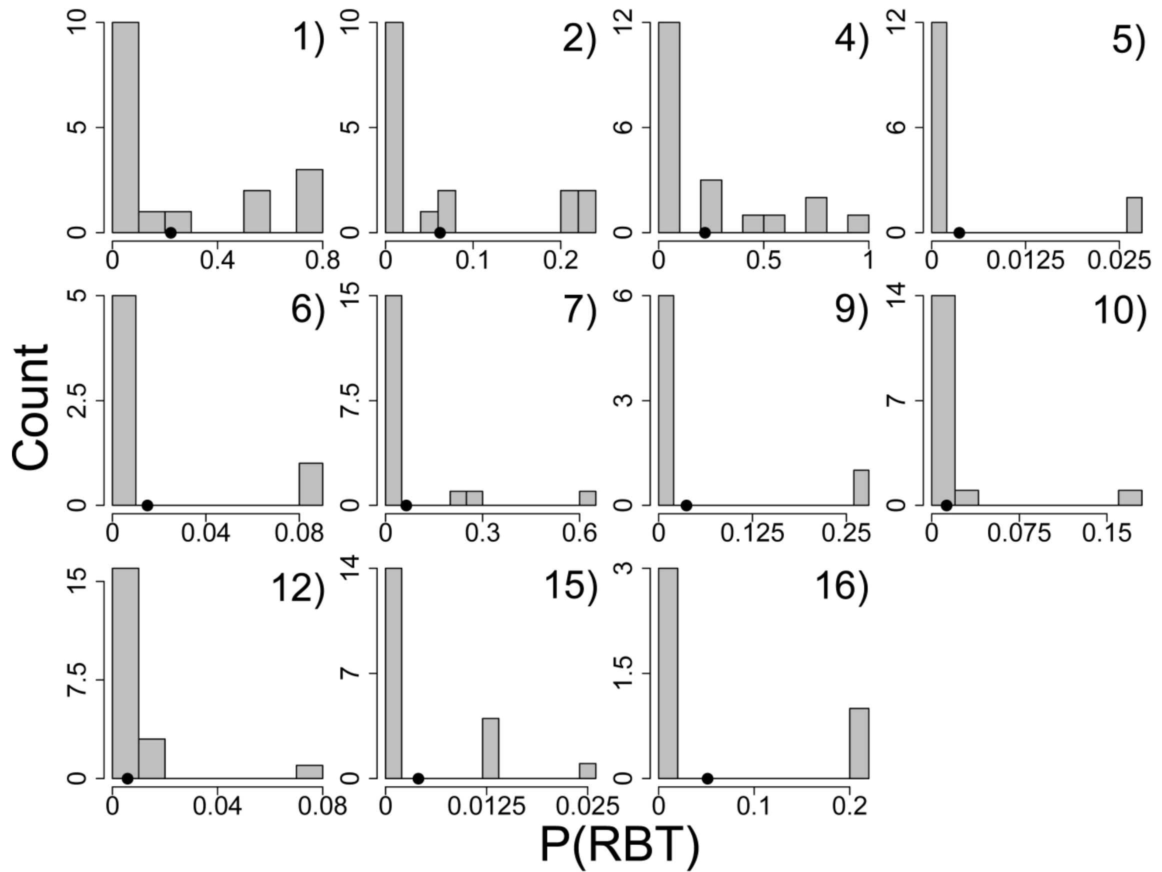

Figure 4.4. Histograms of the number of individuals in a sample with an estimated proportion of rainbow trout genes (P(RBT)) in their genome. Individual sites are plotted in each sample, labeled with the site number that corresponds to Figure 4.1 and Table 4.1. The estimated proportion of rainbow trout genes at the site is plotted as a point along the x-axis for reference.

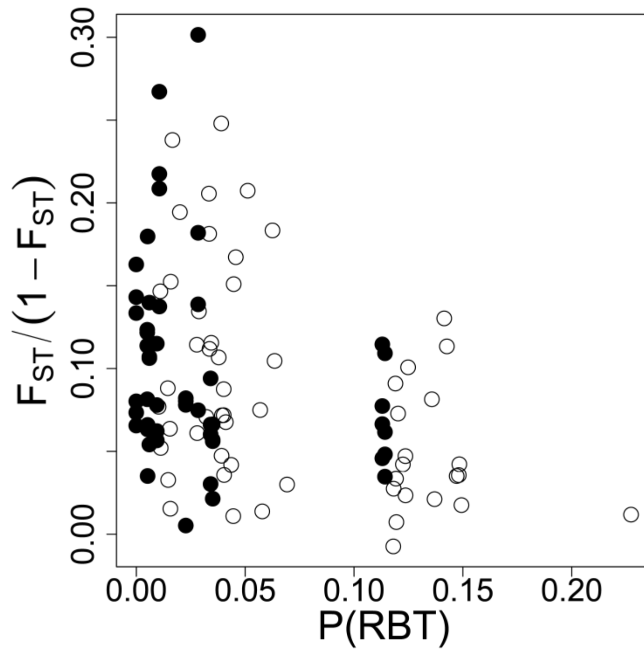

Figure 4.5. Relationship between the average proportion of rainbow trout genes (P(RBT)) and the ratio of ——— between the pairs of sites after removing samples from Upper Park Creek and Autumn Creek from analysis. The filled circles represent comparisons between samples where at least one site did not contain hybrid individuals, whereas the open circles represent comparisons between hybridized sites.

Literature Cited

Allendorf, F.W., and Leary, R.F. 1988. Conservation and distribution of genetic variation in a polytypic species, the cutthroat trout. Conserv. Biol. **2**: 170-184.

Allendorf F.W., and Thorgaard, G.H. 1984. Tetraploidy and the evolution of salmonid fishes. *In* Evolutionary genetics of fishes. *Edited by* B.J. Turner, Plenum, New York, N.Y. pp. 1-53.

Allendorf F.W., Leary, R.F., Hitt, N.P., Knudsen, K.L., Lundquist, L.L., and Spruell, P. 2004. Intercrosses and the U.S. Endangered Species Act: should hybridized populations be included as westslope cutthroat trout? Conserv. Biol. **18**: 1203-1213.

Allendorf F.W., Leary, R.F., Spruell, P., and Wenburg, J.K. 2001. The problem with hybrids: setting conservation guidelines. Trends Ecol. Evol. **16**:613-622.

Amish, S.J., Hohenlohe, P.A., Painter, S., Leary, R.F., Muhlfeld, C.C., Allendorf, F.W., and Luikart, G. 2012. RAD sequencing yields a high success rate for westslope cutthroat trout species-diagnostic SNP assays. Mol. Ecol. Resour. **12**: 653-660.

Anderson, E.C. 2001. Monte Carlo methods for inference in population genetic models. Ph.D. dissertation, Interdisciplinary Program in Quantitative Ecology and resource Management, University of Washington, Seattle, W.A.

Baldwin, B. G. 2003. A phylogenetic perspective on the origin and evolution of Madiinae. *In* Tarweeds and silverswords: evolution of the Madiinae *Edited by* S. Carlquist, B. G. Baldwin, and G. D. Carr, Missouri Botanical Garden Press, St. Louis, M.O. pp. 193-228.

Barton, N.H. 2001. The role of hybridization in evolution. Mol. Ecol. **10**: 551-568.

Behnke, R.J. 2002. Trout and salmon of North America. The Free Press, New York, N.Y.

Benjamini, Y., and Yekutieli, D. 2001. The control of the false discovery rate in multiple testing under dependence. Ann. Stat. **29**: 1165-1188.

Boyer M. C., Muhlfeld, C. C., and F. W. Allendorf. 2008. Rainbow trout (*Oncorhynchus mykiss*) invasion and the spread of hybridization with native westslope cutthroat trout (*Oncorhynchus clarkii lewisi*). Can. J. Fish. Aquat. Sci. **65**: 658-669.

Campbell, N.R., Amish, S.J., Pritchard, V.L., McKelvey, K.S., Young, M.K., Schwartz, M.K., Garza, J.C., Luikart, G., and Narum, S.R. 2012. Development and

evaluation of 200 novel SNP assays for population genetic studies of westslope cutthroat trout and genetic identification of related taxa. Mol. Ecol. Resour. **12**: 942-949.

Danzmann, R.G., Cairney, M., Davidson, W.S., Ferguson, M.M., Gharbi, K., Guyomard, R., Holm, L.-E., Leder, E., Okamoto, N., Ozaki, A., Rexroad III, C.E., Sakamoto, T., Taggart, J.B., and Woram, R.A. 2005. A comparative analysis of the rainbow trout genome with 2 other species of fish (Arctic charr and Atlantic salmon) within the tetraploid derivative Salmonidae family (subfamily: Salmoninae). Genome **48**: 1037-1051.

Efron, B. 1979. Bootstrap methods: another look at the jackknife. Ann. Stat. **7**: 1-26.

Finger, A.J., Anderson, E.C., Stephens, M.R., and May, B.P. 2011. Application of a method for estimating effective population size and admixture using diagnostic single nucleotide polymorphisms (SNPs): implications for conservation of threatened Paiute cutthroat trout (*Oncorhynchus clarkii seleniris*) in Silver King Creek, California. Can. J. Fish. Aquat. Sci. **68**: 1369-1386.

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. 2004. Bayesian data analysis, Second edition. Chapman & Hall/CRC, Boca Raton, F.L.

Guo, S.W., and Thompson, E.A. 1992. Performing the exact test for Hardy-Weinberg proportion for multiple alleles. Biometrics **48**: 361-371.

Hitt N.P., Frissell, C.A., Muhlfeld, C.C., and Allendorf, F.W. 2003. Spread of hybridization between native westslope cutthroat trout, *Oncorhynchus clarki lewisi*, and nonnative rainbow trout, *Oncorhynchus mykiss*. Can. J. Fish. Aquat. Sci. **60**: 1440-1451.

Hohenlohe, P.A., Day, M.D., Amish, S.J., Miller, M.R., Kamps-Hughes, N., Boyer, M.C., Muhlfeld, C.C., Allendorf, F.W., Johnson, E.A., and Luikart, G. 2013. Genomic patterns of introgression in rainbow and westslope cutthroat trout illuminated by overlapping paired-end RAD sequencing. Mol. Ecol. **22**: 3002-3013.

Kalinowski, S.T., and Powell, J.H. *In Review*. A Parameter to quantify the degree of genetic mixing among individuals in hybrid populations. Heredity

Leary, R.F., Allendorf F.W., Sage, G.K. 1995. Hybridization and introgression between introduced and native fish. Am. Fish. Soc. Symp. **15**: 91-101.

Lunn, D., Jackson, C., Best, N., Thomas, A., and Spiegelhalter, D. 2012. The BUGS book – a practical introduction to Bayesian analysis. CRC Press/Chapman & Hall, Boca Raton, F.L.

Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. 2000. WinBUGS – a Bayesian modeling framework: concepts, structure, and extensibility. Stat. Comput. 10:325-337.

Manly, B.F.J. 2007. Randomization, bootstrap and Monte Carlo methods in biology, Third Edition. Chapman & Hall/CRC, Boca Raton, F.L.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Res. **27**: 209-220.

Muhlfeld, C.C., Kalinowski, S.T., McMahon, T.E., Painter, S., Leary, R.F., Taper, M.L., and Allendorf, F.W. 2009. Hybridization rapidly reduces fitness of a native trout in the wild. Biol. Lett. **5**:328-331.

Narum, S.R. 2006. Beyond Bonferroni: less conservative analyses for conservation genetics. Conserv. Genet. **7**:783-787.

Pfaff, C.L., Parra, E.J., Bonilla, C., Hiester, K., McKeigue, P.M., Kamboh, M.I., Hutchinson, R.G., Ferrell, R.E., Boerwinkle, E., and Shriver, M.D. 2001. Population structure in admixed populations: effects of admixture dynamics on the pattern of linkage disequilibrium. Am. J. Hum. Genet. **68**: 198-207.

Plummer, M., N. Best, K. Cowles, and K. Vines. 2006. CODA: convergence diagnostics and output analysis for MCMC. R News **6**: 7-11.

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

R Development Core Team. 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. Available from http://www.R-project.org (last accessed 5/2012).

Raymond, M., Rousset, F. 1995. GENEPOP. version 1.2. Population genetics software for exact tests and ecumenicism. J. Hered. **86**: 248-249.

Reiseberg, L.H. 1994. The role of hybridization in evolution: old wine in new skins. Am. J. Bot. **82**: 944-953.

Rhymer, J. M., and D. Simberloff. 1996. Extinction by hybridization and introgression. Annu. Rev. Ecol. Syst. **27**: 83-109.

Rousset, F. 1997. Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. Genetics **145**: 1219-1228.

Rousset, F. 2008. Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. Mol. Ecol. Resour. **8**: 103-106.

Shepard, B. B., May, B. E., and W. Urie. 2005. Status and conservation of westslope cutthroat trout within the western United States. N. Am. J. Fish. Manag. **25**: 1426-1440.

Stebbins, G.L. 1959. The role of hybridization in evolution. Proc. Am. Philos. Soc. **103**: 231-251.

Sturtz, S., Ligges, U., and Gelman, A. 2005. R2WinBUGS: A package for running WinBUGS from R. J. Stat. Soft. **12**: 1-16.

Tang, H., Peng, J., Wang, P., and Risch, N.J. 2005. Estimation of individual admixture: analytical and study design considerations. Genet. Epidemiol. **28**: 289-301.

Waples, R.S. 1991. Genetic interactions between hatchery and wild salmonids: lessons from the Pacific Northwest. Can. J. Fish. Aquat. Sci. **48**(Suppl. 1): 124-133.

Weir, B.S., and Cockerham, C.C. 1984. Estimating *F*-statisitcs for the analysis of population structure. Evolution **38**: 1358-1370.

Wright, S. 1940. Breeding structure of populations in relation to speciation. Am. Nat. **74**: 232-248.

Wright, S. 1943. Isolation by distance. Genetics **28**: 114-138.

CHAPTER 5

GENERAL CONCLUSIONS

Dispersal is an ecological process affecting the allele frequencies of populations (Hartl and Clark 2007). Being able to quantify the effects of this process is essential to understanding the factors that shape population genetic structure. However, to correctly identify this process in the wild it is critical to quantify uncertainty in our estimates of the effects of dispersal. Here I have presented three studies that utilize statistical methods that allow for an incorporation of uncertainty due to the data generation process when investigating the effects of dispersal at different levels of biological organization.

In chapter two I presented a study investigating the effects of restricted dispersal on the lifetime reproductive success of individual female Weddell seals from a population in Erebus Bay, Antarctica. Because a pedigree was not available for this population, we had to estimate individual inbreeding coefficients based on the genotypes of individuals at 29 microsatellite loci. Because multilocus heterozygosity is only weakly correlated with inbreeding coefficients (Balloux et al. 2004, Slate et al. 2004), it is important to incorporate uncertainty associated with estimating individual inbreeding coefficients into estimates of inbreeding depression in a wild population. We used a two-step maximum likelihood analysis, fit with the data cloning algorithm (Lele et al. 2007), as a means of incorporating uncertainty in estimating individual inbreeding coefficients into the estimates of inbreeding depression in this population. We found that the model most supported by the data was one that included inbreeding depression, but that the

confidence interval around this parameter included zero. However, our two-stage analysis allowed us to estimate a range of plausible values of inbreeding depression for the lifetime reproductive success of mature female seals in this population that accounted for our uncertainty in individual inbreeding coefficients. These estimates indicated that inbreeding depression for lifetime reproductive success was at most modest in this population. In addition, while the estimated values of life history parameters from our samples matched previous work we were unable to detect an association between the colony of birth and any of the life history parameters, despite previous studies identifying these associations in this population (Hastings and Testa 1998, Hadley et al. 2008). However, because the method we present is capable of incorporating uncertainty in estimating individual inbreeding coefficients into tests for the presence of inbreeding depression in wild populations, it should be widely applicable in future studies when pedigree information is not available.

Chapter three focused on the next level in the biological hierarchy, the level of populations, in investigating the association between landscape features and dispersal in mule deer across Montana. Graves (2012) found that individual based isolation by distance analyses violate the assumption of linearity necessary for traditional analysis methods, and presented an alternate method for accounting for this non-linearity. In this chapter we took a different approach that used a logistic regression model and Mantel test confidence intervals to estimate the association between genetic distance and geographic distance. We found low levels of isolation by distance, and a lack of geographic

clustering, among individuals across this landscape. Therefore, there appeared to be few barriers to the migration of mule deer across Montana.

Chapter four addressed the role of dispersal in the progression of hybridization between westslope cutthroat trout and introduced rainbow trout in the Middle Fork Flathead River, Montana. Because hybridization appears to be ongoing in this landscape, we used a modified version of the *Structure* algorithm (Pritchard et al. 2000) to estimate the ancestral westslope cutthroat trout allele frequencies in hybridized samples. This approach allowed us to directly compare the dispersal rates between hybridized and non-hybridized samples in this landscape. We failed to detect a difference in the dispersal rate between hybridized and non-hybridized sites in this watershed. We hypothesized, therefore, that the observed patterns of hybridization in this landscape could be the result of the introduction history of rainbow trout overlaid on patterns of historically high levels of connectivity among sites in this watershed.

In conclusion, dispersal is an important ecological process whose presence, or absence, affects population allele frequencies. When investigating the role of dispersal in populations, or across a landscape, it is important to quantify, to the best of our ability, the uncertainty from multiple sources (such as other ecological and genetic processes) that is present in the data. In this dissertation I have presented three studies that highlight the role of dispersal in wild populations. By investigating the association between inbreeding and lifetime reproductive success, one study was an investigation of the role of restricted dispersal in a population of Weddell seals. In examining the association between genetic dissimilarity and features of the landscape, another study investigated

the role of dispersal within a mule deer population. The final study investigated dispersal among populations by estimating the association between genetic ancestry and dispersal rates within a watershed.

Literature Cited

Balloux, F., W. Amos, and T. Coulson. 2004. Does heterozygosity estimate inbreeding in real populations? Mol. Ecol. 13:3021-3031.

Graves, T. A. 2012. Spatial ecology of grizzly bears in northwestern Montana and estimating resistance to gene flow. Ph.D. Dissertation, Northern Arizona University, Flagstaff, A.Z.

Hadley, G. L., J. J. Rotella, and R. A. Garrott. 2008. Spatial variation in age-specific probabilities of first reproduction for Weddell seals. Oikos 117:1165-1174.

Hartl, D. L., and A. G. Clark. 2007. Principles of population genetics 4th ed. Sinauer Associates Inc., Sunderland, M.A.

Hastings K. K., and J. W. Testa. 1998. Maternal and birth colony effects on survival of Weddell seal offspring from McMurdo Sound, Antarctica. J. Anim. Ecol. 67:722-740.

Lele, S. R, B. Dennis, and F. Lutscher. 2007. Data cloning: easy maximum likelihood estimation for complex ecological models using Bayesian Markov chain Monte Carlo methods. Ecol. Lett. 10:551-563.

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

Slate, J., P. David, K. G. Dodds, B. A. Veenvliet, B. C. Glass, T. E. Broad, and J. C. McEwan. 2004. Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: theoretical expectations and empirical data. Heredity 93:255-265.

APPENDICES

APPENDIX A

A TEST FOR GENE TREE DISCORDANCE AT DIAGNOSTIC LOCI

At a single locus, gene trees and species trees can be discordant due to incomplete lineage sorting (failures to coalesce due to the presence of ancestral polymorphisms), horizontal gene transfer, gene duplication and loss, or hybridization (Degnan and Rosenberg 2009). An implicit assumption when using diagnostic loci to detect hybridization events is that hybridization is the only process leading to gene tree-species tree discordance. However, two other processes could be responsible for gene tree-species tree discordance in salmonids. Recent work has shown that genetic structure in ancestral human (*Homo sapiens*) populations can produce similar levels of apparent admixture to what has been observed between humans and Neanderthals (*Homo neandertalensis*) due to incomplete lineage sorting (Eriksson and Manica 2012). This is problematic since westslope cutthroat trout have high levels of genetic divergence across their range (Drinan et al. 2011). In addition, paralogous loci might be present in populations due to the recent genome duplication event in the common ancestor to all salmonid fishes (Allendorf and Thorgaard 1984). Regardless of whether gene tree-species tree discordance arose as an ancestral polymorphism or due to the whole genome duplication event, the presence of these polymorphic loci will bias estimates of admixture in hybrid populations. We present a statistical test for identifying these polymorphisms at putatively diagnostic loci.

This test is based on the heterogeneity test of Cavalli-Sforza and Bodmer (1971) which assumes independence across loci. This test was based on the assumption that allele frequencies in species A (P_A) and species B (P_B) were fixed for alternate alleles at diagnostic loci (P_A = 1, P_B = 0), reducing equation 8.54 ($M = \frac{P_M - P_B}{P_A - P_B}$, Cavalli-Sforza and

Bodmer 1971) to $M = P_M$. At a polymorphic locus the allele frequencies in the native population will not be fixed ($P_B > 0$) and therefore equation 8.54 becomes $M = \frac{P_M - P_B}{1 - P_B}$. Testing for the heterogeneity of ancestral gene frequencies (M) across loci should detect these differences.

Although Cavalli-Sforza and Bodmer (1971) provide a set of formulas to calculate a test statistic that is approximately χ^2-distributed, we used a permutation based test similar to that used in the multinomial regression statistic of Gompert and Buerkle (2009). Therefore, this test was performed by initially calculating a χ^2-statistic for homogenous allele frequencies across loci based on the observed number of non-native alleles, and the expected number assuming a constant proportion admixture (M) in the sample. The null distribution of this test was constructed by randomly permuting diagnostic alleles across loci 10,000 times, and calculating the permutation χ^2-statistic. The p-value for this test is the proportion of 10,000 random permutations that resulted in allele counts that were as, or more, heterogeneous than the observed data. When this test indicated allele frequency heterogeneity we removed the locus with the largest departure from the mean proportion of rainbow trout alleles and re-ran the test. The test was iterated in this manner until allele frequencies were determined to be homogenous across all remaining loci. For subsequent iterations of the test, permutation p-values were Bonferroni adjusted to control for the family-wise error rate. If loci were identified as polymorphic in multiple samples from the same watershed, these loci were removed from all samples collected within that watershed.

Because this test does not account for an increase in the heterogeneity of allele frequencies due to genetic drift (Long 1991), Type I error rates could increase with time since the admixture event (Fitzpatrick et al. 2009). This increased error rate is a problem when testing for genes under selection, which is often the focus of these tests (Long 1991, Fitzpatrick et al. 2009, Gompert and Buerkle 2009, Hohenlohe et al. 2013). In contrast we are interested in estimating the proportion admixture in the population, indicating our estimates will be less sensitive to Type I errors, particularly when a large number of loci are genotyped. Also, we did not incorporate genetic drift into our test because the equations that incorporate drift also account for error in parental population allele frequencies (Long 1991), exactly the phenomenon we are trying to identify.

Care should be used in applying this test in situations where enough time has passed since the admixture event for the linkage disequilibrium caused by admixture to decay and genetic drift to affect individual allele frequencies. Genetic drift will randomly increase, or decrease, the frequency of non-native alleles across the genome within a population, whereas polymorphic loci should be consistent across populations that share an evolutionary history. Therefore, if this test identifies a large number of loci, of which there is little overlap between samples within the watershed, this pattern could be a signal that genetic drift has occurred and the removal of loci is not appropriate.

Literature Cited

Allendorf F. W., and G. H. Thorgaard. 1984. Tetraploidy and the evolution of salmonid fishes. *In* Evolutionary genetics of fishes. *Edited by* B.J. Turner, Plenum, New York, N.Y. pp. 1-53.

Cavalli-Sforza, L.L., and W. F. Bodmer. 1971. The genetics of human populations. W.H. Freeman and Company, San Francisco, C.A.

Degnan, J. H., and N. A. Rosenberg. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol. Evol. 24:332-340.

Drinan, D. P., S. T. Kalinowski, N. V. Vu, B. B. Shepard, C. C. Muhlfeld, and M. R. Campbell. 2011. Genetic variation in westslope cutthroat trout *Oncorhyncus clarkii lewisi*: implications for conservation. Conserv. Genet. 12:1513-1523.

Eriksson, A., and A. Manica. 2012. Effect of ancient population structure on the degree of polymorphism shared between modern human populations and ancient hominins. P. Nat. Acad. Sci. USA 109:13956-13960.

Fitzpatrick, B. M., J. R. Johnson, D. K. Kump, H. B. Shaffer, J. J. Smith, and S. R. Voss. 2009. Rapid fixation of –non-native alleles revealed by genome-wide SNP analysis of hybrid tiger salamanders. BMC Evol. Biol. 9:176.

Gompert, Z., and C. A. Buerkle. 2009. A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. Mol. Ecol. 18:1207-1224.

Hohenlohe, P. A., M. D. Day, S. J. Amish, M. R. Miller, N. Kamps-Hughes, M. C. Boyer, C. C. Muhlfeld, F. W. Allendorf, E. A. Johnson, and G. Luikart. 2013. Genomic patterns of introgression in rainbow and westslope cutthroat trout illuminated by overlapping paired-end RAD sequencing. Mol. Ecol. 22:3002-3013.

Long J. C. 1991. The genetic structure of admixed populations. Genetics 127:417-428.

APPENDIX B

LANDSCAPE GENETIC ANALYSIS OF WESTSLOPE CUTTHROAT TROUT *ONCORHYNCHUS CLARKII LEWISI* POPULATIONS IN GLACIER NATIONAL PARK

Introduction

Management of population units presupposes an understanding of population structure that may or may not be present throughout the species range. Directly tracking known individuals using either radio telemetry or GPS collars is a common method to assess connectivity, but is logistically and financially prohibitive at broad spatial scales or in cases where rare but long-distance dispersals may be important (Cullingham et al. 2008). Indirect estimates of dispersal using molecular markers provides an alternative method for understanding broad-scale patterns of connectivity when sample sizes are too small to directly estimate dispersal with assignment or parentage tests using this data (Cullingham et al. 2008). Also, by relating indirect estimates of genetic divergence to landscape features, the main goal of the field of landscape genetics, managers and biologists can determine what constitutes a population (Manel et al. 2003).

Westslope cutthroat trout (*Oncorhynchus clarkii lewisi*) are a species of conservation concern within their native range in the northern Rocky Mountains. Westslope cutthroat trout have the largest native range of the 14 identified subspecies of cutthroat trout (Behnke 2002). However, their persistence is threatened by widespread hybridization with introduced *Oncorhynchus* species (Allendorf and Leary 1988). Currently, only about 23% of the historically occupied lotic habitat is thought to contain non-hybridized populations of westslope cutthroat trout (Shepard et al. 2005).

These non-hybridized populations of westslope cutthroat trout exhibit complex patterns of population structure. Substantial genetic divergence has been observed among populations of westslope cutthroat trout at both broad (Leary et al. 1988, Drinan et al.

2011) and fine geographic scales (Young et al. 2004). However, within the North Fork Flathead River the estimated average pairwise F_{ST} was much lower (Boyer et al. 2008) than in other similarly sized watersheds to the north and west (Taylor et al. 2003, Young et al. 2004). In addition to lower genetic divergence, Boyer et al. (2008) identified an association between the genetic distance and stream distance separating two sites in the North Fork Flathead River. Meeuwig et al. (2008) determined that the structure of bull trout populations in this same landscape was best explained by the presence of one- and two-way migration barriers between sites, indicating these barriers to migration are likely important in the structuring of westslope cutthroat trout populations as well.

In this study we set out to investigate the population structure of westslope cutthroat trout across a similar geographic region to Meeuwig et al. (2008). We present an indirect assessment of dispersal of non-hybridized westslope cutthroat trout within Glacier National Park, Montana. We undertook a landscape based approach to study the possible effects of isolation on the population structure of westslope cutthroat trout in tributary streams and lakes to the North Fork Flathead River and Middle Fork Flathead River. Finally, we used the observed patterns of genetic divergence between connected and isolated sites as a baseline to examine patterns of dispersal among connected sites.

Methods

Samples from 16 sites located in tributary streams and lakes to the North Fork Flathead River and the Middle Fork Flathead River (Figure B.1) were selected for this study (Table B.1). These samples were the only non-hybridized samples collected as part

of a larger study investigating the extent of hybridization between native westslope cutthroat trout and introduced rainbow trout (*O. mykiss*) and Yellowstone cutthroat trout (*O. c. bouvieri*) within Glacier National Park (C. Muhlfeld *unpublished data*). At each site, individuals were collected either by angling or electrofishing.

Individual fish were genotyped at 34 single nucleotide polymorphism (SNP) loci, previously described by Campbell et al. (2012, Table B.2), that are polymorphic within westslope cutthroat trout populations. Genotyping was done on Fluidigm 96.96 microfluidic PCR chips.

We used the exact test of Guo and Thompson (1992), as implemented with the Markov chain exact test in GENEPOP on the web (Rousset & Raymond 1995, Rousset 2008), to test whether the observed genotype frequencies conformed to Hardy-Weinberg expectations. We made a decision on whether to remove loci from analysis based on whether they consistently departed, following a false discovery rate correction (Benjamini and Yekutieli 2001, Narum 2006), from Hardy Weinberg expectations in multiple samples.

We calculated pairwise values of Weir and Cockerham's θ (Weir and Cockerham 1984) using the program *TreeFit* (Kalinowski 2009). We explored the effects of migration barriers on the structure of populations within this landscape by mapping pairwise θ values to individual segments of the watershed using the program *StreamTree* (Kalinowski et al. 2008). Results of this analysis were plotted using *Adobe Illustrator* version 13.0.2 (Adobe Systems Incorporated). Pairwise θ values assigned to individual segments were compared to the presence of known dispersal barriers within this

watershed using ArcMap version 10.0 (Environmental Systems Research Institute, Inc., Redlands, CA). Finally, the distribution of pairwise θ among connected sites was compared to the distribution of pairwise θ between barrier isolated sites and connected sites to investigate apparent patterns of connectivity. This analysis was performed in the statistical computing program R version 2.15.3 (R Development Core Team 2012).

Results and Discussion

In investigating the genetic structure of populations across this landscape we initially screened loci for departures from Hardy-Weinberg equilibrium. We calculated pairwise θ between samples collected at 16 sites located within Glacier National Park in the North Fork Flathead River and Middle Fork Flathead River drainages. We assigned θ values to individual stream segments and compared the presence of potential dispersal barriers to observed values of θ. Finally, we compared the distribution of pairwise θ for sample locations that did not conform to expectations based on known barriers to dispersal to patterns observed at the contrasting class of sites.

We detected multiple departures from Hardy-Weinberg equilibrium at two loci following a Benjamini-Yekutieli (2001) false discovery rate correction, based on 212 simultaneous comparisons, of $\alpha = 0.0084$. Of the 13 total departures from Hardy-Weinberg equilibrium 6 occurred at the *OclVAR_RAD-69806_GL* locus and 3 occurred at the *Ocl_var_mx1_129_NC* locus. Each departure at the *OclVAR_RAD-69806_GL* locus was due to an excess of heterozygotes based on Weir and Cockerham's (1984) estimate of $F_{IS} \in [-1.00, -0.89]$. This observed pattern suggests that the primers for this locus were

amplifying paralogous loci (Hohenlohe et al. 2013). Each departure from Hardy-Weinberg equilibrium at the *Ocl_var_mx1_129_NC* locus was due to an excess of homozygotes ($F_{IS} \in [0.83, 1.00]$). This excess of homozygotes in multiple samples could be due to the amplification of null alleles at this locus. These two loci were removed from further analysis. There were four additional departures from Hard-Weinberg equilibrium, each of which occurred at a different locus (*Ocl_var_hsc71pro_71_00*, *OclVar_115987L_Garza, OclVAR_RAD_16087_GL*, and *Oclvar_metB_106NC*). We did not remove any of these loci because there was no apparent pattern to these departures from Hard-Weinberg expectations.

Among samples collected in connected sites from the North Fork Flathead River and the Middle Fork Flathead River pairwise θ ($\bar{\hat{\theta}}$ = 0.067, range from 0.009 to 0.127, Table B.3 and Figure B.2) was consistent with estimates across a similar scale in the North Fork Flathead River ($\bar{\hat{\theta}}$ = 0.079 Boyer et al. 2008). The sample collected in Upper Fern Creek was included as a connected site, despite the presence of two culvert barriers between it and Lake McDonald, due to the fact that all other barriers between sample sites have been present on a geologic timescale (e.g. waterfalls). Pairwise θ was much higher at sites isolated by a geologic barrier ($\bar{\hat{\theta}}$ = 0.630, range from 0.000 to 0.952, Table B.3 and Figure B.2) in this landscape.

The observed patterns of genetic divergence conformed well to the presence of dispersal barriers (Figure B.2), as was expected based on previous work with bull trout populations in this landscape (Meeuwig et al. 2008). However, there were three samples that did not fit this general pattern. Upper Muir Creek (L in Figure B.1 and Table B.1) is

located above a waterfall barrier but has pairwise θ values (Table B.3) that are similar to those observed in the non-barrier isolated samples (Figure B.3). If the waterfall in this creek was restricting the dispersal of individuals into this creek, we would expect to see a pattern of genetic divergence similar to other isolated sites. However, the similarity between the observed levels of genetic divergence at this sample site and other non-barrier isolated sites indicates similar dispersal patterns occur at each site. Therefore, the observed genetic divergence at the Upper Muir Creek site indicates that the waterfall is not a complete, or even strong, barrier to the dispersal of individuals into this creek.

In contrast, Upper Park Creek and Autumn Creek (M and P in Figure B.1 and Table B.1 respectively) have pairwise θ values (Table B.3) falling within the range observed for the geologically isolated samples (Figures B.2 and B.3), despite a lack of known barriers to dispersal. The similarity between the observed levels of genetic divergence at these sites and other barrier isolated sites indicates there is likely no meaningful dispersal into these sites. This lack of dispersal may be explained in the Upper Park Creek site by the presence of a potential dispersal barrier (V. D'Angelo *personal communication*).

Westslope cutthroat trout populations can contain individuals expressing a range of life histories strategies (McIntyre and Reiman 1995). Neville et al. (2006) found that genetic divergence was lower in Lahontan cutthroat trout (*O. c. henshawi*) populations with a greater proportion of migratory individuals. Because migratory individuals spawn in the lower reaches of streams (McIntyre and Reiman 1995), and these lower elevation reaches are associated with increased proportion of rainbow trout genes in hybridized

populations (Weigel et al. 2003, Muhlfeld et al. 2009, Rasmussen et al. 2010, Yau and Taylor 2013), the high levels of genetic divergence observed among extant non-hybridized westslope cutthroat trout populations may be due to the loss of non-hybridized populations expressing the migratory life history.

The continued expression of the migratory life history of westslope cutthroat trout in the Flathead River drainage (Muhlfeld et al. 2005), therefore, could explain the lower average pairwise θ in the North Fork Flathead River (0.076, Boyer et al. 2008) compared to other watersheds (0.18, Taylor et al. 2003; 0.123, Young et al. 2004). We hypothesize that the increase in the genetic divergence observed at the Upper Park Creek and Autumn Creek sample locations may be due to the fact that these samples were collected from populations predominantly expressing resident life histories. The persistence of a migratory life history in the other sampled populations could also account for the low levels of observed population structure among other connected sites in the Flathead River drainage.

Tables and Figures

Table B.1. Information for the 16 sample locations from the North Fork Flathead River and Middle Fork Flathead River, with figure codes corresponding to Figure B.1.

Figure code	Drainage	Site	n
A	North Fork	Spruce Creek	19
B	North Fork	Starvation Creek	17
C	Middle Fork	Upper Fern Creek	5
D	Middle Fork	Snyder Lakes	25
E	Middle Fork	Lower McDonald Creek	5
F	Middle Fork	Avalanche Lake	4
G	Middle Fork	Mineral Creek	18
H	Middle Fork	Upper McDonald Creek	20
I	Middle Fork	Lincoln Lake	2
J	Middle Fork	Pinchot Creek	12
K	Middle Fork	Upper Coal Creek	20
L	Middle Fork	Upper Muir Creek	19
M	Middle Fork	Upper Park Creek	19
N	Middle Fork	Lake Isabel	10
O	Middle Fork	Upper Lake Isabel	10
P	Middle Fork	Autumn Creek	14

Table B.2. Locus names for the SNP loci used in this analysis.

Locus name
OclVar_115987L _Garza, *OclVAR_RAD_51955_GL*, *Ocl_var_hsc71pro_71_NC*, *OclVar105105_Garza*, *Oclvar_arp_117NC*, *OclVAR_RAD_16087_GL*, *Oclvar_arp_314NC*, *OclVar113772_Garza*, *OclVAR_RAD_16730_GL*, *Oclvar_fKbp2_62NC*, *Oclvar_vatf_277NC*, *Oclvar_Carpa1_45NC*, *OclVAR_RAD_69806_GL*, *Ocl_var_mtap_224_NC*, *OclVar_bnip3_263NC*, *Ocl_var_npc2_145_NC*, *Ocl_var_mx1_129_NC*, *Oclvar_mkpro_69NC*, *Oclvar_ca050_39NC*, *Oclvar_pigH_42NC*, *OclVar102483_Garza*, *Oclvar_pop5_50NC*, *Ocl_var_impa1_189_NC*, *Ocl_var_bcAKala_259_NC*, *Oclvar_pnpo_56NC*, *Oclvar_srp09_172NC*, *OclVar96899L_Garza*, *OclVAR_RAD_43407_GL*, *Oclvar_ada10_100NC*, *OclVAR_cin_90NC*, *Oclvar_parp3_19NC*, *Oclvar_metB_106NC*, *OclVar128996_Garza*, *Omg_LDHB2i6_Garza*

Table B.3. Pairwise θ (Weir and Cockerham 1984) with site labels corresponding to Figure B.1 and Table B.1.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
A															
B	0.009														
C	0.028	0.019													
D	0.528	0.574	0.738												
E	0.406	0.397	0.512	0.828											
F	0.303	0.304	0.426	0.805	0.535										
G	0.354	0.399	0.479	0.706	0.549	0.036									
H	0.353	0.375	0.482	0.737	0.549	0.019	0.017								
I	0.080	0.114	0.125	0.788	0.662	0.415	0.438	0.470							
J	0.026	0.058	0.033	0.611	0.404	0.326	0.383	0.385	0.127						
K	0.061	0.067	0.072	0.553	0.454	0.362	0.423	0.397	0.117	0.066					
L	0.031	0.072	0.067	0.512	0.392	0.323	0.336	0.355	0.118	0.056	0.065				
M	0.407	0.463	0.528	0.745	0.670	0.549	0.530	0.550	0.489	0.446	0.440	0.374			
N	0.506	0.553	0.678	0.853	0.775	0.709	0.665	0.668	0.772	0.518	0.519	0.468	0.367		
O	0.510	0.558	0.687	0.857	0.792	0.720	0.668	0.672	0.785	0.524	0.526	0.469	0.346	0.000	
P	0.571	0.629	0.821	0.857	0.952	0.857	0.777	0.767	0.908	0.631	0.614	0.536	0.715	0.923	0.926

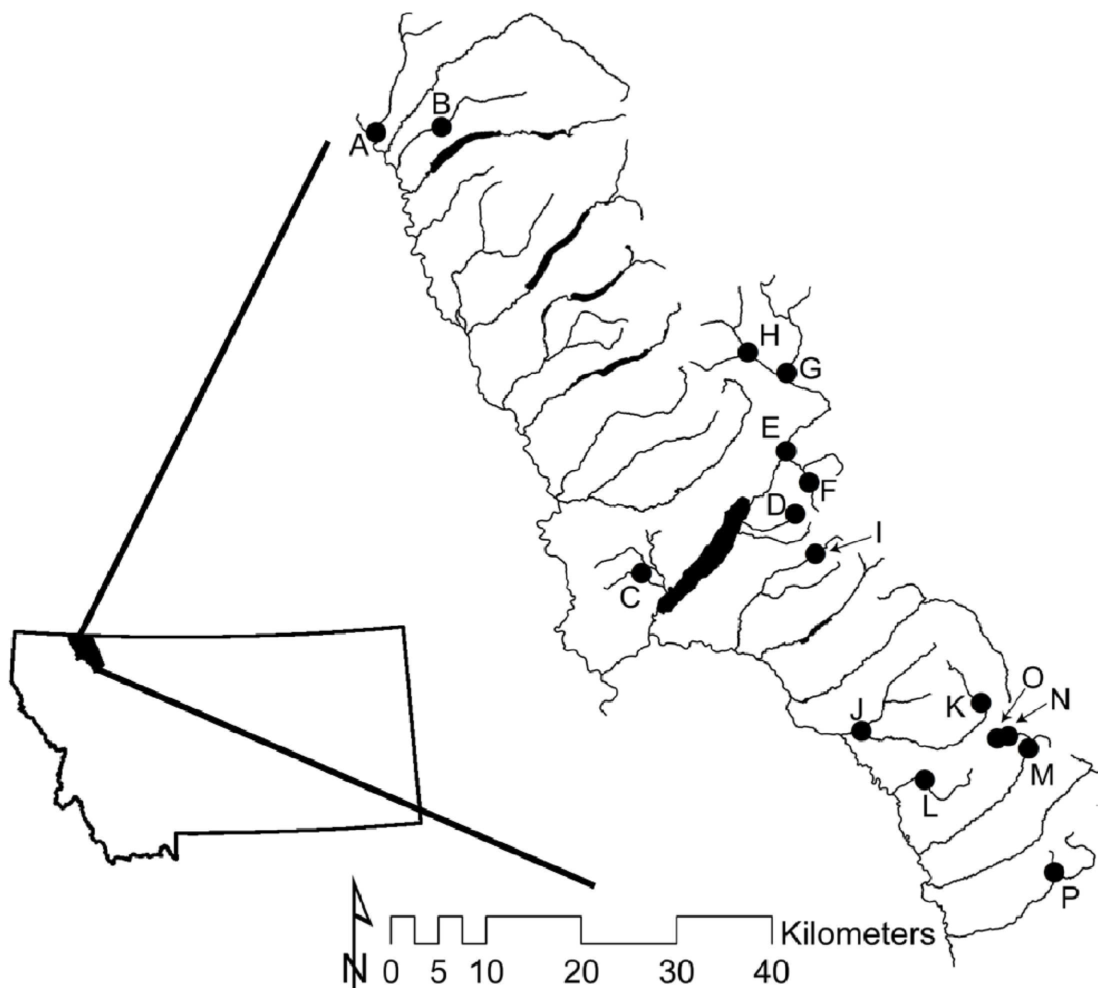

Figure B.1. Map of the sample locations in the North Fork Flathead River and Middle Fork Flathead River drainages, with site labels corresponding to Table B.1.

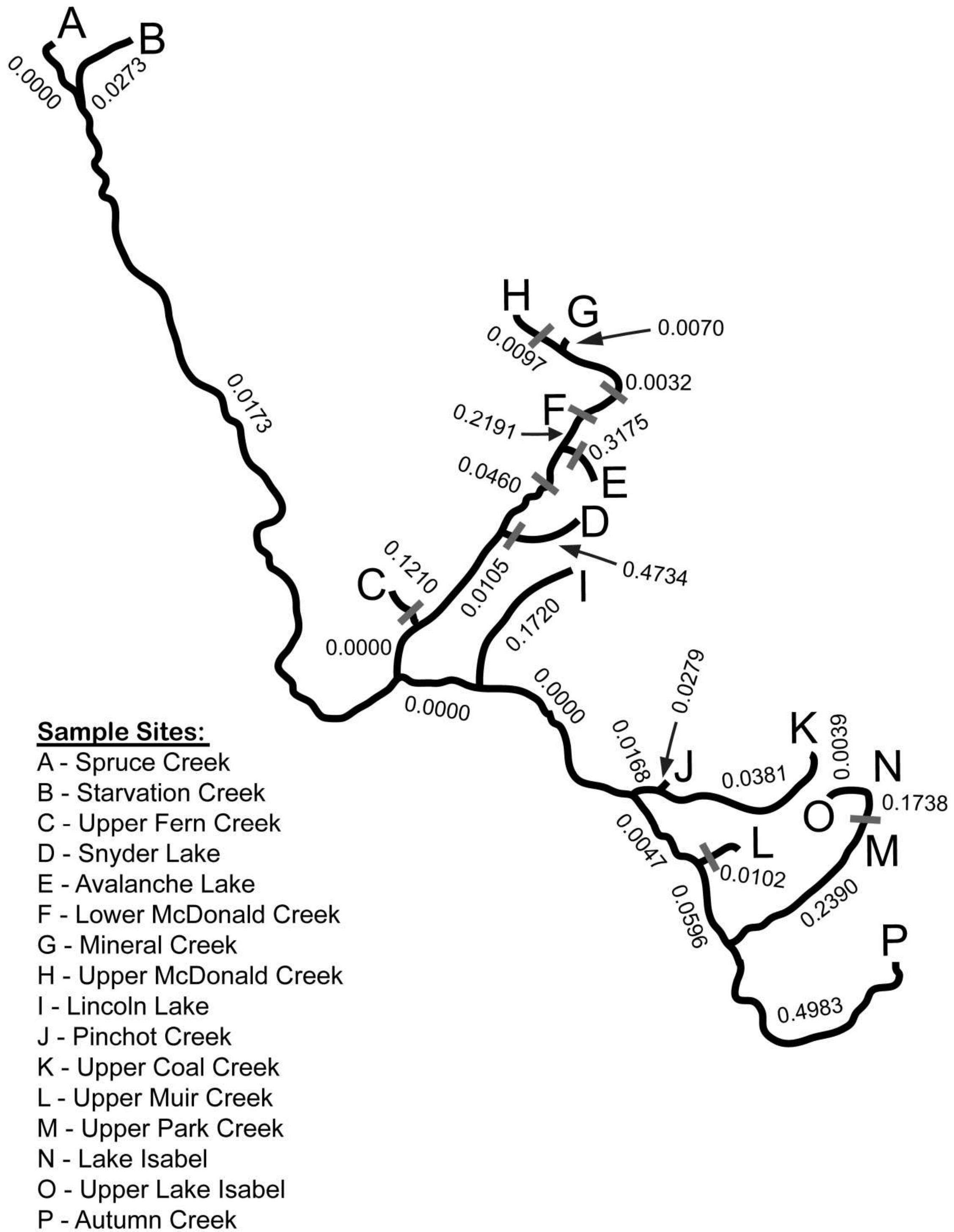

Figure B.2. Stream tree based on pairwise θ values, with site labels corresponding to Figure B.1 and Table B.1. The approximate location of barriers is indicated by gray dashes across the channel. Multiple barriers within the same segment were condensed to a single marker to increase readability.

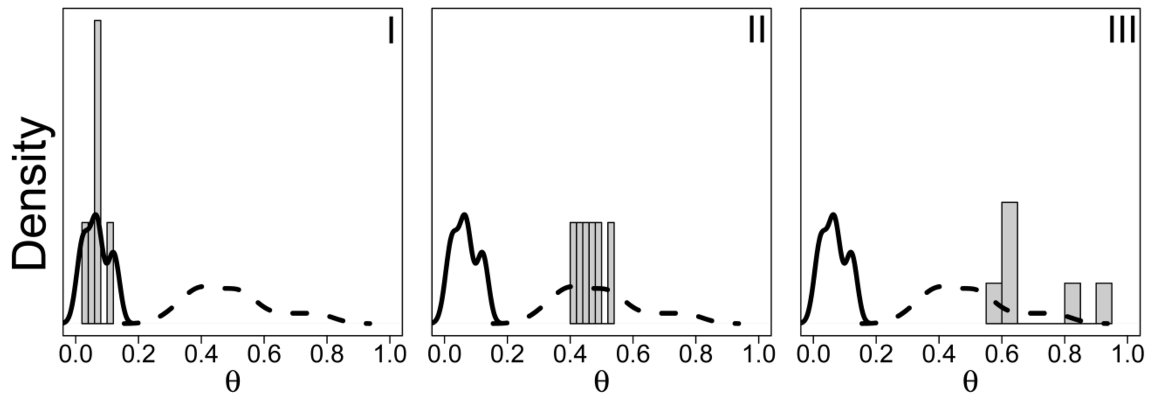

Figure B.3. Distribution of pairwise θ between the samples collected in I) Upper Muir Creek, II) Upper Park Creek, and III) Autumn Creek and other connected sites (gray bars). The solid line plots the kernel density of pairwise θ values among connected sites, and the dashed line plots the kernel density of pairwise θ values between geologic barrier isolated sites and connected sites.

Literature Cited

Allendorf, F. W., and R. F. Leary. 1988. Conservation and distribution of genetic variation in a polytypic species, the cutthroat trout. Conserv. Biol. 2:170-184.

Behnke, R. J. 2002. Trout and salmon of North America. The Free Press, New York, N.Y.

Benjamini, Y., and D. Yekutieli. 2001. The control of the false discovery rate in multiple testing under dependence. Ann. Stat. 29:1165-1188.

Boyer M. C., C. C.Muhlfeld, and F. W. Allendorf. 2008. Rainbow trout (*Oncorhynchus mykiss*) invasion and the spread of hybridization with native westslope cutthroat trout (*Oncorhynchus clarkii lewisi*). Can. J. Fish. Aquat. Sci. 65:658-669.

Campbell, N. R., S. J. Amish, V. L. Pritchard, K. S. McKelvey, M. K. Young, M. K. Schwartz, J. C. Garza, G. Luikart, and S. R. Narum. 2012. Development and evaluation of 200 novel SNP assays for population genetic studies of westslope cutthroat trout and genetic identification of related taxa. Mol. Ecol. Resour. 12:942-949.

Cullingham, C.I., B. A. Pond, C. J. Kyle, E. E. Rees, R. C. Rosatte, and B. N. White. 2008. Combining direct and indirect genetic methods to estimate dispersal for informing wildlife disease management decisions. Mol. Ecol. 17:4874-4886.

Drinan, D. P., S. T. Kalinowski, N. V. Vu, B. B. Shepard, C. C. Muhlfeld, and M. R. Campbell. 2011. Genetic variation in westslope cutthroat trout *Oncorhyncus clarkii lewisi*: implications for conservation. Conserv. Genet. 12:1513-1523.

Guo, S. W., and E. A. Thompson. 1992. Performing the exact test for Hardy-Weinberg proportion for multiple alleles. Biometrics 48:361-371.

Hohenlohe, P. A., M. D. Day, S. J. Amish, M. R. Miller, N. Kamps-Hughes, M. C. Boyer, C. C. Muhlfeld, F. W. Allendorf, E. A. Johnson, and G. Luikart. 2013. Genomic patterns of introgression in rainbow and westslope cutthroat trout illuminated by overlapping paired-end RAD sequencing. Mol. Ecol. 22:3002-3013.

Kalinowski, S. T. 2009. How well do evolutionary trees describe genetic relationships between populations? Heredity 102:506-513.

Kalinowski, S.T., M. H. Meeuwig, S. R. Narum and M. L. Taper. 2008. Stream trees: a statistical method for mapping genetic differences between populations of

freshwater organisms to the sections of streams that connect them. Can. J. Fish. Aquat. Sci. 65:2752-2760.

Leary, R. F., F. W. Allendorf, S. R. Phelps, and K. L. Knudsen. 1988. Population genetic structure of westslope cutthroat trout: genetic variation within and among populations. Proc. Mont. Acad. Sci. 48:56-70.

Manel, S., M. K. Schwartz, G. Luikart, P. Taberlet. 2003. Landscape genetics: combining landscape ecology and population genetics. Trends Ecol. Evol. 18:189-197.

McIntyre, J. D., and B. E. Reiman. 1995. Westslope cutthroat trout. *In* Conservation assessment for inland cutthroat trout. *Edited by* M. K. Young. U.S. Forest Service General Technical Report RM-256. pp. 1-15.

Meeuwig, M. H., C. S. Guy, S. T. Kalinowski, and W. A. Fredenberg. 2010. Landscape influences on genetic differentiation among bull trout populations in a stream network. Mol. Ecol. 19:3620-3633.

Muhlfeld, C. C., B. Marotz, S. R. Thorrold, and J. L. FitzGerald. 2005. Geochemical signatures in scales record stream origin in westslope cutthroat trout. Trans. Am. Fish. Soc. 134:945-959.

Muhlfeld, C.C., T.E. McMahon, M.B. Boyer, and R. E. Gresswell. 2009. Local habitat, watershed, and biotic factors influencing the spread of hybridization between native westslope cutthroat trout and introduced rainbow trout. Trans. Am. Fish. Soc. 138: 1036-1051.

Narum, S. R. 2006. Beyond Bonferroni: less conservative analyses for conservation genetics. Conserv. Genet. 7:783-787.

Neville, H., J. B. Dunham, and M. Peacock. 2006. Landscape attributes and life history variability shape genetic structure of trout populations in a stream network. Landscape Ecol. 21:901-916.

R Development Core Team. 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. Available from http://www.R-project.org (last accessed 5/2012).

Rasmussen J. B., M. D. Robinson, and D. D. Heath. 2010. Ecological consequences of hybridization between native westslope cutthroat (*Oncorhynchus clarkii lewisi*) and introduced rainbow (*Oncorhynchus mykiss*) trout: effects on life history and habitat use. Can. J. Fish. Aquat. Sci. 67:357-370.

Raymond, M., F. Rousset. 1995. GENEPOP. version 1.2. Population genetics software

for exact tests and ecumenicism. J. Hered. 86:248-249.

Rousset, F. 2008. Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. Mol. Ecol. Resour. **8**: 103-106.

Shepard, B. B., May, B. E., and W. Urie. 2005. Status and conservation of westslope cutthroat trout within the western United States. N. Am. J. Fish. Manag. **25**: 1426-1440.

Taylor, E. B., M. D. Stamford, and J. S. Baxter. 2003. Population subdivision in westslope cutthroat trout (*Oncorhynchus clarkii lewisi*) at the northern periphery of its range: evolutionary inferences and conservation implications. Mol. Ecol. 12:2609-2622.

Weigel, D. E., J. T. Peterson, and P. Spruell. 2003. Introgressive hybridization between native cutthroat trout and introduced rainbow trout. Ecol. Appl. 13:38-50.

Weir, B. S., and C. C. Cockerham. 1984. Estimating *F*-statistics for the analysis of population structure. Evolution 38:1358-1370.

Yau, M. M., and E. B. Taylor. 2013. Environmental and anthropogenic correlates of hybridization between westslope cutthroat trout (*Oncorhynchus clarkii lewisi*) and introduced rainbow trout (*O. mykiss*). Conserv. Genet. 14:885-900.

Young, S. F., J. G. McLellan, and J. B. Shaklee. 2004. Genetic integrity and microgeographic population structure of westslope cutthroat trout, *Oncorhynchus clarkii lewisi*, in the Pend Oreille basin in Washington. Environ. Biol. Fish. 69:127-142.

LITERATURE CITED

Abdelkrim, J., B. C. Robertson, J.-A. L. Stanton, and N. J.Gimmel. 2009. Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. Biotechniques 46:185-191.

Allendorf, F. W., and R. F. Leary. 1988. Conservation and distribution of genetic variation in a polytypic species, the cutthroat trout. Conserv. Biol. 2:170-184.

Allendorf F. W., and G. H. Thorgaard. 1984. Tetraploidy and the evolution of salmonid fishes. *In* Evolutionary genetics of fishes. *Edited by* B.J. Turner, Plenum, New York, N.Y. pp. 1-53.

Allendorf F. W., R. F. Leary, N. P. Hitt, K. L. Knudsen, L. L. Lundquist, and P. Spruell. 2004. Intercrosses and the U.S. Endangered Species Act: should hybridized populations be included as westslope cutthroat trout? Conserv. Biol. 18:1203-1213.

Allendorf F. W., R. F. Leary, P. Spruell, and J. K. Wenburg. 2001. The problem with hybrids: setting conservation guidelines. Trends Ecol. Evol. 16:613-622.

Amish, S. J., P. A. Hohenlohe, S. Painter, R. F. Leary, C. C. Muhlfeld, F. W. Allendorf, and G. Luikart. 2012. RAD sequencing yields a high success rate for westslope cutthroat trout species-diagnostic SNP assays. Mol. Ecol. Resour. 12:653-660.

Anderson, A., O. Wallmo. 1984. *Odocoileus hemionus*. Mammalian Species 219:1-9.

Anderson, E. C. 2001. Monte Carlo methods for inference in population genetic models. Ph.D. dissertation, Interdisciplinary Program in Quantitative Ecology and resource Management, University of Washington, Seattle, W.A.

Baldwin, B. G. 2003. A phylogenetic perspective on the origin and evolution of Madiinae. *In* Tarweeds and silverswords: evolution of the Madiinae *Edited by* S. Carlquist, B. G. Baldwin, and G. D. Carr, Missouri Botanical Garden Press, St. Louis, M.O. pp. 193-228.

Balloux, F., W. Amos, and T. Coulson. 2004. Does heterozygosity estimate inbreeding in real populations? Mol. Ecol. 13:3021-3031.

Barton, N. H. 2001. The role of hybridization in evolution. Mol. Ecol. **10**: 551-568.

Behnke, R. J. 2002. Trout and salmon of North America. The Free Press, New York, N.Y.

Bengtsson, T. and J. E. Cavanaugh. 2006. An improved Akaike information criterion for state-space model selection. Comput. Stat. Data An. 50:2635-2654.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Stat. Soc. B 57:289-300.

Benjamini, Y., and D. Yekutieli. 2001. The control of the false discovery rate in multiple testing under dependence. Ann. Stat. 29:1165-1188.

Bishop, M. D., S. M. Kappes, J. W. Keele, R. T. Stone, S. L. F. Sunden, and G. A. Hawkins. 1994. A genetic linkage map for cattle. Genetics 136:619–639.

Brown, P., and D. C. Gajdusek. 1991. Survival of scrapie virus after 3 years' interment. The Lancet 337:269-270.

Boyer M. C., C. C.Muhlfeld, and F. W. Allendorf. 2008. Rainbow trout (*Oncorhynchus mykiss*) invasion and the spread of hybridization with native westslope cutthroat trout (*Oncorhynchus clarkii lewisi*). Can. J. Fish. Aquat. Sci. 65:658-669.

Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference, 2nd ed. Springer, New York, N.Y.

Cameron, M. F., and D. B. Siniff. 2004. Age-specific survival, abundance, and immigration rates of a Weddell seal (*Leptonychotes weddellii*) population in McMurdo Sound, Antarctica. Can. J. Zoolog. 82:601-615.

Campbell, N. R., S. J. Amish, V. L. Pritchard, K. S. McKelvey, M. K. Young, M. K. Schwartz, J. C. Garza, G. Luikart, and S. R. Narum. 2012. Development and evaluation of 200 novel SNP assays for population genetic studies of westslope cutthroat trout and genetic identification of related taxa. Mol. Ecol. Resour. 12:942-949.

Cavalli-Sforza, L.L., and W. F. Bodmer. 1971. The genetics of human populations. W.H. Freeman and Company, San Francisco, C.A.

Charlesworth, B., and D. Charlesworth. 1999. The genetic basis of inbreeding depression. Genet. Res. 74:329-340.

Clopper, C. J., and E. S. Pearson. 1934. The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika 26:404-413.

Crnokrak, P., and D. A. Roff. 1999. Inbreeding depression in the wild. Heredity 83:260-270.

Cross, P. C., J. O. Lloyd-Smith, P.L.F. Johnson, and W. M. Getz. 2005. Dueling timescales of host movement and disease recovery determine invasion of disease in structured populations. Ecol. Lett. 8:587-595.

Cullingham, C. I., S. M. Nakada, E. H. Merrill, T. K. Bollinger, M. J. Pybus, and D. W. Coltman. 2011. Multiscale population genetic analysis of mule deer (*Odocoileus hemionus hemionus*) in western Canada sheds new light on the spread of chronic wasting disease. Can. J. Zoolog. 89:134-147.

Cullingham, C.I., B. A. Pond, C. J. Kyle, E. E. Rees, R. C. Rosatte, and B. N. White. 2008. Combining direct and indirect genetic methods to estimate dispersal for informing wildlife disease management decisions. Mol. Ecol. 17:4874-4886.

Cushman, S. A., K. S. McKelvey, J. Hayden, and M. K. Schwartz. 2006. Gene flow in complex landscapes: testing multiple hypotheses with causal modeling. Am. Nat. 168:486-499.

Danzmann, R. G., M. Cairney, W. S. Davidson, M. M. Ferguson, K. Gharbi, R. Guyomard, L.-E. Holm, E. Leder, N. Okamoto, A. Ozaki, C. E. Rexroad III, T. Sakamoto, J. B. Taggart, and R. A. Woram. 2005. A comparative analysis of the rainbow trout genome with 2 other species of fish (Arctic charr and Atlantic salmon) within the tetraploid derivative Salmonidae family (subfamily: Salmoninae). Genome 48:1037-1051.

Darwin, C. R. 1868. Variation of animals and plants under domestication. John Murry, London, U.K.

Davis, C. S., T. S. Gelatt, D. Siniff, and A. Strobeck. 2002. Dinucleotide microsatellite markers from the Antarctic seals and their use in other Pinnipeds. Mol. Ecol. Notes 2:203-208.

Degnan, J. H., and N. A. Rosenberg. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol. Evol. 24:332-340.

DeWoody J. A., R. L. Honeycutt, and L. C. Skow. 1995. Microsatellite markers in white-tailed deer. J. Hered. 86:317-319.

Drinan, D. P., S. T. Kalinowski, N. V. Vu, B. B. Shepard, C. C. Muhlfeld, and M. R. Campbell. 2011. Genetic variation in westslope cutthroat trout *Oncorhyncus clarkii lewisi*: implications for conservation. Conserv. Genet. 12:1513-1523.

Dulberger, J., N. T. Hobbs, H. M. Swanson, C. J. Bishop, and M. W. Miller. 2010. Estimating chronic wasting disease effects on mule deer recruitment and population growth. J. Wildlife Dis. 46:1086-1095.

Eding, H., T. H. E. Meuwissen. 2001. Marker-based estimates of between and within population kinships for the conservation of genetic diversity. J. Anim. Breed. Genet. 118:141-159.

Efron, B. 1979. Bootstrap methods: another look at the jackknife. Ann. Stat. 7:1-26.

Epperson, B. K, and T. Li. 1996. Measurement of genetic structure within populations using Moran's spatial autocorrelation statistics. P. Nat. Acad. Sci. USA 93:10528-10532.

Eriksson, A., and A. Manica. 2012. Effect of ancient population structure on the degree of polymorphism shared between modern human populations and ancient hominins. P. Nat. Acad. Sci. USA 109:13956-13960.

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

Excoffier, L., and H. E. L. Lischer. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10:564-567.

Falush, D., M. Stephens, and J. K. Pritchard. 2003. Inference of population structure: extensions to linked loci and correlated allele frequencies. Genetics 164:1567-1587.

Falush, D., M. Stephens, and J. K. Pritchard. 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. Mol. Ecol. Notes 7:574-578.

Finger, A. J., E. C. Anderson, M. R. Stephens, and B. P. May. 2011. Application of a method for estimating effective population size and admixture using diagnostic single nucleotide polymorphisms (SNPs): implications for conservation of threatened Paiute cutthroat trout (*Oncorhynchus clarkii seleniris*) in Silver King Creek, California. Can. J. Fish. Aquat. Sci. 68:1369-1386.

Fitzpatrick, B. M., J. R. Johnson, D. K. Kump, H. B. Shaffer, J. J. Smith, and S. R. Voss. 2009. Rapid fixation of –non-native alleles revealed by genome-wide SNP analysis of hybrid tiger salamanders. BMC Evol. Biol. 9:176.

Frankel, O. H., and M. E. Soulé. 1981. Conservation and evolution. Cambridge University Press, Cambridge, U.K.

Garner, A., J. L. Rachlow, and J. F. Hicks. 2005. Patterns of genetic diversity and its loss in mammalian populations. Conserv. Biol. 19:1215-1221.

Garrott, R. A., J. J. Rotella, D. B. Siniff, C. L. Parkinson, and G. E. Stauffer. 2011. Environmental variation and cohort effects in an Antarctic predator. Okios 120:1027-1040.

Gelatt, T.S., C. S. Davis, I. Stirling, D. B. Siniff, C. Strobeck, and I. Delisle. 2010. History and fate of a small isolated population of Weddell seals at White Island, Antarctica. Conserv. Genet. 11:721-735.

Gelatt, T. S., C. S. Davis, D. B. Siniff, and C. Strobeck. 2001. Molecular evidence for twinning in Weddell seals (*Leptonychotes weddellii*). J. Mammal. 82:491-499.

Gelman, A. 2006. Prior distributions for variance parameters in hierarchical models. Bayesian Anal. 1:515-533.

Gelman, A., and J. Hill. 2007. Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, New York, N.Y.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 2004. Bayesian data analysis, Second edition. Chapman & Hall/CRC, Boca Raton, F.L.

Gilpin, M. E., and M. E. Soulé. 1986. Minimum viable populations: the process of species extinctions. In: *Conservation Biology the Science of Scarcity and Diversity* (ed Soul ME), pp. 13-34, Sinauer Associates, Sunderland.

Gompert, Z., and C. A. Buerkle. 2009. A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. Mol. Ecol. 18:1207-1224.

Goudet, J. 2002: FSTAT: a program to estimate and test gene diversities and fixation indices (version 2.9.3.2) Available at http://www.unil.ch/popgen/softwares/fstat.htm (last accessed on 23/5/2011).

Goudet, J., N. Perrin, and P. Waser. 2002. Tests for sex-biased dispersal using bi-parentally inherited genetic markers. Mol. Ecol. 11:1103-1114.

Graves, T. A. 2012. Spatial ecology of grizzly bears in northwestern Montana and estimating resistance to gene flow. Ph.D. Dissertation, Northern Arizona University, Flagstaff, A.Z.

Guo, S. W., and E. A. Thompson. 1992. Performing the exact test for Hardy-Weinberg proportion for multiple alleles. Biometrics 48:361-371.

Hadley, G. L., J. J. Rotella, and R. A. Garrott. 2008. Spatial variation in age-specific probabilities of first reproduction for Weddell seals. Oikos 117:1165-1174.

Hadley, G. L., J. J. Rotella, and R. A. Garrott. 2007. Evaluation of reproductive costs for Weddell seals in Erebus Bay, Antarctica. J. Anim. Ecol. 76:448-458.

Hadley, G. L., J. J. Rotella, R. A. Garrott, and J. D. Nichols. 2006. Variation in probability of first reproduction of Weddell seals. J. Anim. Ecol. 75:1058-1070.

Hartl, D. L., and A. G. Clark. 2007. Principles of population genetics 4th ed. Sinauer Associates Inc., Sunderland, M.A.

Harris, R. B. 1986. Reliability of trend lines obtained from variable counts. J. Wildlife Manage. 50:165-171.

Hastings K. K., and J. W. Testa. 1998. Maternal and birth colony effects on survival of Weddell seal offspring from McMurdo Sound, Antarctica. J. Anim. Ecol. 67:722-740.

Hedrick, P. W., and S. T. Kalinowski 2000. Inbreeding depression in conservation biology. Annu. Rev. Ecol. Syst. 31:139-162.

Hitt N. P., C. A. Frissell, C. C. Muhlfeld, and F. W. Allendorf. 2003. Spread of hybridization between native westslope cutthroat trout, *Oncorhynchus clarki lewisi*, and nonnative rainbow trout, *Oncorhynchus mykiss*. Can. J. Fish. Aquat. Sci. 60:1440-1451.

Hohenlohe, P. A., M. D. Day, S. J. Amish, M. R. Miller, N. Kamps-Hughes, M. C. Boyer, C. C. Muhlfeld, F. W. Allendorf, E. A. Johnson, and G. Luikart. 2013. Genomic patterns of introgression in rainbow and westslope cutthroat trout illuminated by overlapping paired-end RAD sequencing. Mol. Ecol. 22:3002-3013.

Hone, J. 1999. On the rate of increase (r): patterns of variation in Australian mammals and the implications for wildlife management. J. Appl. Ecol. 36:709-718.

Ives, A. R., and M. C. Whitlock. 2002. Inbreeding and metapopulations. Science 295:454-455.

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

Jones, K.C., K. F. Levine, and J. D. Banks. 2000. DNA-based genetic markers in black-tailed and mule deer for forensic applications. – Calif. Fish Game 86:115–126.

Kalinowski, S. T. 2005. HP-Rare: a computer program for performing rarefaction on measures of allelic diversity. Mol. Ecol. Notes 5:187-189.

Kalinowski, S. T. 2009. How well do evolutionary trees describe genetic relationships between populations? Heredity 102:506-513.

Kalinowski, S. T., and J. H. Powell. *In Review*. A Parameter to quantify the degree of genetic mixing among individuals in hybrid populations. Heredity

Kalinowski, S.T., M. H. Meeuwig, S. R. Narum and M. L. Taper. 2008. Stream trees: a statistical method for mapping genetic differences between populations of freshwater organisms to the sections of streams that connect them. Can. J. Fish. Aquat. Sci. 65:2752-2760.

Kardos, M., F. W. Allendorf, and G. Luikart. 2013. Evaluating the role of inbreeding depression in heterozygosity-fitness correlations: how useful are tests for identity disequilibrium? Mol. Ecol. Resour. doi: 10.1111/1755-0998.12193

Keller L. F., and D. M. Waller. 2002. Inbreeding effects in wild populations. Trends Ecol. Evol. 17:230-241.

Kondrashov, A. 1988. Deleterious mutations and the evolution of sexual reproduction. Nature 336:435-440.

Lande, R., and D. W. Schemske. 1985. The evolution of self fertilization and inbreeding depression in plants. I. Genetic models. Evolution 39:24-40.

Latch, E. K., J. R. Heffelfinger, J. A. Fike, and O. E. Rhodes. 2009. Species-wide phylogeography of North American mule deer (*Odocoileus hemionus*): cryptic glacial refugia and postglacial recolonization. Mol. Ecol. 18:1730-1745.

Latch, E. K., R. A. Amann, J. P. Jacobson, and O. E. Rhodes. 2008. Competing hypotheses for the etiology of cryptorchidism in Sitka black-tailed deer: an evaluation of evolutionary alternatives. Anim. Conserv. 11:234-246.

Leary, R. F., F. W. Allendorf, S. R. Phelps, and K. L. Knudsen. 1988. Population genetic structure of westslope cutthroat trout: genetic variation within and among populations. Proc. Mont. Acad. Sci. 48:56-70.

Leary, R.F., F. W. Allendorf, G. K. Sage. 1995. Hybridization and introgression between introduced and native fish. Am. Fish. Soc. Symp. 15:91-101.

Lele, S. R., K. Nadeem, and B. Schmuland. 2010. Estimability and likelihood inference for generalized linear mixed models using data cloning. J. Am. Stat. Assoc. 105:1617-1625.

Lele, S. R, B. Dennis, and F. Lutscher. 2007. Data cloning: easy maximum likelihood estimation for complex ecological models using Bayesian Markov chain Monte Carlo methods. Ecol. Lett. 10:551-563.

Lesaffre, E., and A. Lawson. 2012. Bayesian biostatistics. John Wiles & Sons, West Sussex, U.K.

Long J. C. 1991. The genetic structure of admixed populations. Genetics 127:417-428.

Lunn, D. 2003. WBDevDJLTruncatedNormal documentation. Available at http://www.winbugs-development.org.uk/. Accessed on December 29, 2013.

Lunn, D., C. Jackson, N. Best, A. Thomas, and D. Spiegelhalter. 2012. The BUGS book – a practical introduction to Bayesian analysis. CRC Press/Chapman & Hall, Boca Raton, F.L.

Lunn, D. J., A. Thomas, N. Best, and D. Spiegelhalter. 2000. WinBUGS – a Bayesian modeling framework: concepts, structure, and extensibility. Stat. Comput. 10:325-337.

Mackie, R. J., J. G. Kie, D. F. Pac, and K. L. Hamlin. 2003. Mule deer. *In* Wild Mammals of North America. *Edited by* G. A. Feldhamer, B. C. Thompson, and J. A. Chapman. The Johns Hopkins University Press, Baltimore, M.D. pp. 889-905.

Mackie, R. J., D. F. Pac, K. L. Hamlin, and G. L. Dusek. 1998. Ecology and management of mule deer and white-tailed deer in Montana. Report to the Montana Department of Fish, Wildlife & Parks.

Manel, S., M. K. Schwartz, G. Luikart, P. Taberlet. 2003. Landscape genetics: combining landscape ecology and population genetics. Trends Ecol. Evol. 18:189-197.

Manly, B. F. J. 2007. Randomization, bootstrap and Monte Carlo methods in biology, Third Edition. Chapman & Hall/CRC, Boca Raton, F.L.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Res. 27:209-220.

McIntyre, J. D., and B. E. Reiman. 1995. Westslope cutthroat trout. *In* Conservation assessment for inland cutthroat trout. *Edited by* M. K. Young. U.S. Forest Service

General Technical Report RM-256. pp. 1-15.

Meeuwig, M. H., C. S. Guy, S. T. Kalinowski, and W. A. Fredenberg. 2010. Landscape influences on genetic differentiation among bull trout populations in a stream network. Mol. Ecol. 19:3620-3633.

Miller, M. W., and M. M. Conner. 2005. Epidemiology of chronic wasting disease in free ranging mule deer: spatial, temporal, and demographic influences on observed prevalence patterns. J. Wildlife Dis. 41:275-290.

Miller, M. W., N. T. Hobbs, and S. J. Tavener. 2006. Dynamics of prion disease transmission in mule deer. Ecol. Appl. 16:2208-2214.

Morris, W. F., and D. F. Doak. 2002. Quantitative conservation biology. Sinauer Associates, Inc., Sunderland, M.A.

Morton, N. E., J. F. Crow, and H. J. Muller. 1956. An estimate of the mutational damage in man from data on consanguineous marriages. Proc. Nat. Acad. Sci. USA 42:855-863.

Muhlfeld, C. C., S. T. Kalinowski, T. E. McMahon, S. Painter, R. F. Leary, M. L. Taper, and F. W. Allendorf. 2009. Hybridization rapidly reduces fitness of a native trout in the wild. Biol. Lett. 5:328-331.

Muhlfeld, C. C., B. Marotz, S. R. Thorrold, and J. L. FitzGerald. 2005. Geochemical signatures in scales record stream origin in westslope cutthroat trout. Trans. Am. Fish. Soc. 134:945-959.

Muhlfeld, C.C., T.E. McMahon, M.B. Boyer, and R. E. Gresswell. 2009. Local habitat, watershed, and biotic factors influencing the spread of hybridization between native westslope cutthroat trout and introduced rainbow trout. Trans. Am. Fish. Soc. 138: 1036-1051.

Narum, S. R. 2006. Beyond Bonferroni: less conservative analyses for conservation genetics. Conserv. Genet. 7:783-787.

Neville, H., J. B. Dunham, and M. Peacock. 2006. Landscape attributes and life history variability shape genetic structure of trout populations in a stream network. Landscape Ecol. 21:901-916.

Paetkau, D., and C. Strobeck. 1995. The molecular basis and evolutionary history of a microsatellite null allele in bears. Mol. Ecol. 4:519-520.

Peakall, R., and P. E. Smouse. 2006. GENALEX 6: genetic analysis in Excel. Population

genetic software for teaching and research. Mol. Ecol. Notes 6:288-295.

Peakall, R., and P. E. Smouse. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. Bioinformatics 28:2537-2539.

Pease, K. M., A. H. Freedman, J. P. Pollinger, J. C. McCormack, W. Buermann, J. Rodzen, J. Banks, E. Meredith, V. C. Bleich, R. J. Schaefer, K. Jones, and R. K. Wayne. 2009. Landscape genetics of California mule deer (*Odocoileus hemionus*): the roles of ecological and historical factors in generating differentiation. Mol. Ecol. 18:1848-1862.

Pemberton, J. 2004. Measuring inbreeding depression in the wild: the old ways are the best. Trends Ecol. Evol. 19:613-615.

Pfaff, C. L., E. J. Parra, C. Bonilla, K. Hiester, P. M. McKeigue, M. I. Kamboh, R. G. Hutchinson, R. E. Ferrell, E. Boerwinkle, and M. D. Shriver. 2001. Population structure in admixed populations: effects of admixture dynamics on the pattern of linkage disequilibrium. Am. J. Hum. Genet. 68:198-207.

Picton, H. D., and T. N. Lonner. 2008. Montana's wildlife legacy – decimation to restoration. Media Works Publications, Bozeman, M.T.

Plummer, M., N. Best, K. Cowles, and K. Vines. 2006. CODA: convergence diagnostics and output analysis for MCMC. R News 6:7-11.

Ponciano, J. M., M. L. Taper, B. Dennis, and S. R. Lele. 2009. Hierarchical models in ecology: confidence intervals, hypothesis testing, and model selection using data cloning. Ecology 90:356-362.

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

Proffitt, K. M., R. A. Garrott, and J. J. Rotella. 2008. Variation in offspring sex ratio among individual Weddell seal (*Leptonychotes weddellii*) females of different quality. Behav. Ecol. Sociobiol. 62:1679-1687.

Proffitt, K. M., R. A. Garrott, J. J. Rotella, and K. E. Wheatley. 2007. Environmental and senescent related variations in Weddell seal body mass: implications for age-specific reproductive performance. Oikos 116:1683-1690.

Prugnolle, F., and T. de Meeus. 2002. Inferring sex-biased dispersal from population genetic tools: a review. Heredity 88:161-165.

R Development Core Team. 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. Available from http://www.R-project.org (last accessed 5/2012).

R Development Core Team. 2009. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. Available from http://www.R-project.org (last accessed 23/12/2010).

Ralls, K., J. D. Ballou, and A. Templeton. 1988. Estimates of lethal equivalents and the cost of inbreeding in mammals. Conserv. Boil. 2:185-193.

Ralls, K., K. Brugger, and J. Ballou. 1987. Inbreeding and juvenile mortality in small populations. Science 206:1101-1103.

Ramsey, F. L., and D. W. Schafer. 2002. The statistical sleuth a course in methods of data analysis, Second Edition. Duxbury, Pacific Grove, CA.

Rasmussen J. B., M. D. Robinson, and D. D. Heath. 2010. Ecological consequences of hybridization between native westslope cutthroat (*Oncorhynchus clarkii lewisi*) and introduced rainbow (*Oncorhynchus mykiss*) trout: effects on life history and habitat use. Can. J. Fish. Aquat. Sci. 67:357-370.

Raymond, M., F. Rousset. 1995. GENEPOP. version 1.2. Population genetics software for exact tests and ecumenicism. J. Hered. 86:248-249.

Reiseberg, L. H. 1994. The role of hybridization in evolution: old wine in new skins. Am. J. Bot. 82:944-953.

Rhymer, J. M., and D. Simberloff. 1996. Extinction by hybridization and introgression. Annu. Rev. Ecol. Syst. **27**: 83-109.

Rosenberg, N. A. 2004. Distruct: a program for the graphical display of population structure. Mol. Ecol. Notes 4:137–138.

Rotella, J. J., W. A. Link, J. D. Nichols, G. L. Hadley, R. A. Garrott, and K. M. Proffitt. 2009. An evaluation of density-dependent and density-independent influences on population growth rates in Weddell seals. Ecology 90:975-984.

Rousset, F. 2008. Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. Mol. Ecol. Resour. **8**: 103-106.

Rousset, F. 1997. Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. Genetics 145:1219-1228.

Schuelke, M. 2000. An economic method for the fluorescent labeling of PCR fragments. Nat. Biotechnol. 18:233-234.

Seidel, B., A. Thomzig, A. Buschmann, M. H. Groschup, R. Peters, M. Beekes, and K. Terytze. 2007. Scrapie agent (strain 263K) can transmit disease via the oral route after persistence in soil over years. PloS ONE 2:e435.

Shepard, B. B., May, B. E., and W. Urie. 2005. Status and conservation of westslope cutthroat trout within the western United States. N. Am. J. Fish. Manag. **25**: 1426-1440.

Slate, J., and J. M. Pemberton. 2002. Comparing molecular measures for detecting inbreeding depression. J. Evol. Biol. 15:20-31.

Slate, J., P. David, K. G. Dodds, B. A. Veenvliet, B. C. Glass, T. E. Broad, and J. C. McEwan. 2004. Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: theoretical expectations and empirical data. Heredity 93:255-265.

Smouse, P.E., and R. Peakall. 1999. Spatial autocorrelation analysis of multiallele and multi-locus genetic micro-structure. Heredity 82:561-573.

Sólymos, P. 2010. dclone: data cloning in R. The R Journal 2:29-37.

Spiegelhalter, D., A. Thomas, N. Best, and D. Lunn. 2003. WinBUGS user manual, version 1.4, January 2003. Available at http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/manual14.pdf. Accessed December 29, 2013.

Stebbins, G. L. 1959. The role of hybridization in evolution. Proc. Am. Philos. Soc. 103:231-251.

Sturtz, S., U. Ligges, and A. Gelman. 2005. R2WinBUGS: A package for running WinBUGS from R. J. Stat. Soft. 12:1-16.

Szulkin M., N. Bierne, and P. David. 2010. Heterozygosity-fitness correlations: a time for reappraisal. Evolution 64:1202-1217.

Tamura, K. 1992. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C content biases. Mol. Biol. Evol. 9:678-687.

Tang, H., J. Peng, P. Wang, and N. J. Risch. 2005. Estimation of individual admixture: analytical and study design considerations. Genet. Epidemiol. 28:289-301.

Taylor, E. B., M. D. Stamford, and J. S. Baxter. 2003. Population subdivision in westslope cutthroat trout (*Oncorhynchus clarkii lewisi*) at the northern periphery of its range: evolutionary inferences and conservation implications. Mol. Ecol. 12:2609-2622.

Tuyl, F., R. Gerlach, and K. Mengersen. 2008. A comparison of Bayes-Laplace, Jeffreys, and other priors: the case of zero events. Am. Stat. 62:40-44.

Vogl, C., A. Karhu, G. Moran, and O. Savolainen. 2002. High resolution analysis of mating systems: inbreeding in natural populations of *Pinus radiata*. J. Evol. Biol. 15:433-439.

Waples, R. S. 1991. Genetic interactions between hatchery and wild salmonids: lessons from the Pacific Northwest. Can. J. Fish. Aquat. Sci. 48(Suppl. 1):124-133.

Weigel, D. E., J. T. Peterson, and P. Spruell. 2003. Introgressive hybridization between native cutthroat trout and introduced rainbow trout. Ecol. Appl. 13:38-50.

Weir, B. S., and C. C. Cockerham. 1984. Estimating *F*-statistics for the analysis of population structure. Evolution 38:1358-1370.

Wilson, G. A., C. Strobeck, C. L. Wu, and J. W. Coffin. 1997. Characterization of microsatellite loci in caribou *Rangifer tarandus*, and their use in other artiodactyls. Mol. Ecol. 6:697–699.

Williams, E. S., M. W. Miller, T. J. Kreeger, R. H. Kahn, and E. T. Thorne. 2002. Chronic wasting disease of deer and elk: a review with recommendations for management. J. Wildlife Manage. 66:551-563.

Wright, S. 1940. Breeding structure and populations in relation to speciation. Am. Nat. 74:232-248.

Wright, S. 1943. Isolation by distance. Genetics 28:114-138.

Wright, S. 1946. Isolation by distance under diverse systems of mating. Genetics 31:39-59.

Yau, M. M., and E. B. Taylor. 2013. Environmental and anthropogenic correlates of hybridization between westslope cutthroat trout (*Oncorhynchus clarkii lewisi*) and introduced rainbow trout (*O. mykiss*). Conserv. Genet. 14:885-900.

Young, S. F., J. G. McLellan, and J. B. Shaklee. 2004. Genetic integrity and microgeographic population structure of westslope cutthroat trout, *Oncorhynchus*

clarkii lewisi, in the Pend Oreille basin in Washington. Environ. Biol. Fish. 69:127-142.

Zink, R. M., and G. F. Barrowclough. 2008. Mitochondrial DNA under siege in avian phylogeography. Mol. Ecol. 17:2107-2121.