GENETIC ANALYSES OF COMMON MURRE (*Uria aalge*) POPULATIONS FROM BRITISH COLUMBIA TO CALIFORNIA, WITH EMPHASIS ON RESTORATION PLANNING IN WASHINGTON

Final Report submitted to:
Tenyo Maru Trustee Committee

July 1999

Kenneth I. Warheit
Washington Department of Fish and Wildlife
Wildlife Management
600 Capitol Way N.
Olympia, WA 98501-1091
(360) 902-2595
warhekiw@dfw.wa.gov

Participating Laboratories:

- Dr. Vicki Friesen Laboratory, Queen's University, Kingston Ontario, Canada
- Washington Department of Fish and Wildlife Genetics Laboratory

Subcontractor:

- Dr. Vicki Friesen, Queen's University

ACKNOWLEDGEMENTS

I thank the *Tenyo Maru* Trustee Committee and the Washington Department of Fish and Wildlife for funding this research. I received immeasurable laboratory assistance from personnel in Dr. Vicki Friesen's laboratory at Queen's University, Kingston, Ontario. In particular, I thank Tim Birt, Brad Congdon, Christine Crossman, Deb Harrison, Gabriela Ibarguch, Monica Kidd, and Hollie Walsh. I also received assistance at the WDFW Genetic Laboratory from Jim Shaklee and Sewall Young. Christine Crossman, Gabriela Ibarguch, and Tim Birt designed the microsatellite protocols used in these analyses. I obtained Common Murre blood or tissue directly or indirectly from David Ainley, Jack Broughton, Ted Buerger, Pat Gearin, Roy Lowe, Julia Parrish, Bill Sydeman, and Chris Thompson. I also thank the Burke Museum, University of Washington (Sharon Birks, Scott Edwards, and Sievert Rohwer) for assistance. Finally, I thank Vicki Friesen for her assistance throughout the entire project. Dr. Friesen was a subcontractor on this project and is co-author on all products generated from this research.

INTRODUCTION

The purpose of oil spill-related seabird restoration programs is to restore, enhance, or protect seabird *populations* affected by spills. Although individual projects are frequently aimed at particular colonies, nest sites, or some ecological component associated with seabird community interactions (e.g., food-web interactions), the justification for focusing at these spatial scales or ecological levels is that restoration here benefits the whole population (see *Apex Houston*, *Nestucca*, and *Exxon Valdez* restoration plans). Unfortunately, restoration plans frequently are designed using assumptions about how a particular population was affected by a spill (injury assessment), and how might that population recover following the implementation of the restoration plan. In making these assumptions, trustee agencies presuppose parameters (e.g., colony numbers or attendance, reproductive success) and the geographic extent of the effected population. It would be difficult for trustee agencies to design restoration plans or set restoration goals without good estimates or assumptions about colony attendance or reproductive output. Likewise, without estimates or good assumptions about the geographic extent of an injured population (note: this is distinct from the geographic extent of the injury), trustee agencies may not know where to focus its restoration activities, what restoration activities are most promising in restoring injured populations, or whether a population would recover naturally without human intervention.

In general, trustee agencies are aware of the prerequisite requirements of data on colony attendance or reproductive success and frequently collect such data before restoration plans are designed. Unfortunately, little or no data are collected to help delineate the geographic extent of an affected population, despite the fact that the need to outline population boundaries is even more basic than the need to estimate colony attendance or reproductive success (i.e., defining the geographic extent of populations determine which colonies should be counted and where reproductive success data should be collected). For example, the trustee agencies involved with designing the *Apex Houston* Restoration Plan restricted restoration options for Common Murres to one or several colonies in the southern part of central California. Their justification was that the central California population of Common Murres is geographically and genetically isolated from other populations, and the viability of this population would be compromised without restoring these more southerly colonies. Although the trustee agencies did not possess data on the genetic structure of this population, colony attendance patterns suggested that the population may be isolated from other populations. In this example, the *basis to the restoration plan was a presumed geographic structure of a population of Common*

Murres, and the health of that population. Although the attendance patterns do suggest anecdotally that this population is isolated, the trustees' argument would have been more convincing if they had explicit data on murre dispersal or genetic data on the geographic structure of the population.

Geographic Structure of Populations – A population can be defined as a nexus of individuals that interbreed and are part of a common gene pool. Within this population individuals breed randomly. That is, within the geographic limits of the population there is a theoretically equal probability that an individual of one sex will breed with an individual of the other sex (i.e., breeding is panmictic). This implies that for seabird species such as Common Murre, there is sufficient immigration among colonies to maintain a common gene pool. However, once immigration among colonies becomes geographically structured (i.e., there is a higher probability that there is gene flow between two colonies than between these two colonies, as a whole, and a third colony), the population may become subdivided into subpopulations, with each subpopulation independently accumulating new alleles through mutation and losing alleles through genetic drift. As such, each subpopulation may ultimately exhibit diagnostic genetic profiles. If two Common Murre colonies, for example, exchange no individuals (i.e., there is no dispersal or immigration between the two) the two colonies may be parts of different populations. Therefore, identifying the degree to which colonies exchange immigrants (i.e., measuring gene flow among colonies) is the key to delimiting population boundaries.

There are two basic ways to estimate gene flow among colonies. The first method is a direct procedure and is based on mark-recapture techniques. This requires that individuals be marked (i.e., banded) and recaptured or re-sighted. The idea here is that individuals can be "followed" and movement between colonies can be directly observed. This method requires that a sufficient number of individuals be banded, and a large subset of these individuals be re-sighted. For Common Murres, there are perhaps over 700,000 nesting individuals in Oregon (Manuwal et al. *in prep.*), and roughly 10,000-15,000 nesting in Washington State (Wilson 1995, 1996; Parrish 1998; Warheit *in press*). To my knowledge not a single individual Common Murre in either state is banded, and implementing a banding program sufficient to estimate gene flow among colonies is not only extremely expensive and long-term, but logistically impossible.

The second method in estimating gene flow among colonies, is an indirect procedure that uses the products of gene flow as data. That is, genes, or more precisely DNA sequences (i.e., the sequence of

nucleotide basepairs along coding or non-coding regions of mitochondrial or nuclear DNA), are examined to estimate the degree of gene flow among colonies, subpopulations, or populations. This procedure requires that DNA from individuals be obtained through either destructive sampling (i.e., killing) or non-destructive sampling (e.g., bleeding, or the use of existing tissue or blood).

Project Justification – The recovery or recovery rate of a Common Murre population following a natural or anthropogenic disturbance will depend on the geographic structure of that population. Small genetically isolated colonies (or subpopulations) should expect a longer period of recovery than a colony that is part of a larger, widespread population. If there is little or no dispersal among isolated colonies or subpopulations, recovery following disturbance must be through local recruitment (i.e., birds that are hatched at the colony return to that colony to breed, rather than that colony being re-population through immigration). If there is a high degree of dispersal among colonies that are part of a larger population, natural recovery following a disturbance may be relatively quick due to the influx of immigrants (see Warheit et al. 1997 for brief discussion). Knowledge of the genetic structure of these colonies/populations would allow for a better assessment of whether active hands-on restoration is needed, and if so, what types of restoration are best prescribed (e.g., restoration projects aimed at encouraging immigration, versus projects that are designed to increase recruitment). Natural recovery through immigration would indicate that hands-on restoration is not needed. Furthermore genetic analyses of populations may point to a colony or geographic region to concentrate restoration efforts or to help set restoration goals. In the following report, I present data and analyses on the population structure of Common Murres from British Columbia to California. I use these data to describe patterns of gene flow among these populations and conclude, based on five microsatellite loci, that there is gene flow among all localities, but in particular, between Washington and British Columbia. I place these data in the context of Common Murre restoration and make recommendations concerning the recovery of Common Murres following the 1991 *Tenyo Maru* oil spill.

METHODS

Samples – No Common Murres were killed as a direct result of this study. I obtained all DNA material from existing specimens already accessioned in museums or from other researchers who obtained carcasses or blood in conjunction with their own work. For the data I will be summarizing in this report, I used DNA samples from four localities, one each from British Columbia, Washington,

Oregon, and California (Table 1; Fig. 1). I made considerable effort to obtain samples from a second locality in Washington (e.g., Point Grenville). However, no such specimens existed *a priori* in museums, and both permitting issues and inclement weather during the past three breeding seasons prevented new material from being collected. All samples were collected either directly on a breeding colony (e.g., blood samples from SE Farallon Island, California) or near a breeding colony during the breeding season. The breeding status of each bird was confirmed by the presence of an incubation patch (Swartz 1966). Therefore, I assume that each bird from each locality was a confirmed breeding or juvenile bird from that locality and represented the population of which that colony is a part. These four localities span the entire range of one Common Murre subspecies (*U. a. californica*) and the southern part of another subspecies (*U. a. inornata*), with the colonies in Washington and Oregon generally considered to be the area of overlap (Fig. 1; see Storer 1952, Tuck 1961). Because these colonies are linearly arranged along the coast, one initial hypothesis for the occurrence of gene flow among these colonies is that localities would be isolated by distance. That is, the further two colonies are apart, the fewer number of individuals would be migrating between the two sites. Therefore, my initial hypothesis was that if the Washington locality was not isolated from the other colonies, it would receive the most numbers of migrant individuals from British Columbia and the second most from Oregon (~330 and ~ 400 km from Cape Flattery, respectively).

The sample sizes used in the analyses are shown in Table 1. Because most of the Washington samples have not yet been sequenced for three of the four sequence loci (Table 1), the analyses in this report will be limited to microsatellite loci. In addition, I have more samples from both British Columbia and Oregon, which were obtained only recently, and would nearly double the amount of material from each locality. The DNA from these samples have not yet been extracted. I will provide the results from these additional samples and all the results from the sequenced loci in an addendum to this Final Report.

Laboratory Procedures – DNA was extracted from each sample using either phenol-chloroform (Sambrook et al. 1989) or ammonium acetate. The ammonium acetate procedure used here is based on the protocols in the Gentra Systems™ (Minneapolis, MN) Puregene DNA isolation kit. For both these procedures, a small amount of blood or tissue are used, cells are lysed, protein is removed, and DNA is precipitated using either isopropanol or ethanol. The alcohol is poured off and the precipitated DNA is dried. The pure DNA is then re-hydrated with either TE buffer (10mM Tris, 0.1

mM EDTA, pH 7.4) or sterile deionized water and stored at 4°C or -20°C. I will refer to the pure DNA isolated from each sample as template DNA.

Three different classes of genetic markers were used during this study: nuclear introns, mitochondrial DNA, and microsatellites. Both the intron and mitochondrial DNA markers are sequenced, while microsatellite loci are normally analyzed by comparing the molecular weights (i.e., size of DNA fragments) of the different alleles for each locus. The sizes of these fragments are determined by electrophoretic mobility. Introns are sections of DNA that are not translated into protein, but intervene the coding portion of a gene (exon). We sequenced only one nuclear intron for this project (an α-enolase intron). Mitochondrial DNA is non-nuclear circular DNA that is maternally inherited (haploid). The gene arrangement in the mitochondria of birds differs from that of other vertebrates (Fig. 2; Quinn 1997). We sequenced a 307 base pair (bp) portion of the cytochrome *b* locus and two portions of the control region, one each at the locus' 5′ (392 bp) and 3′ (361 bp) ends.

Finally, microsatellite DNA are non-coding repetitive DNA with short repeat units. For example, dinucleotide and tetranucleotide microsatellites have repeat units of two and four nucleotides, respectively (Fig. 3). We used five nuclear dinucleotide microsatellite loci in this study with the repeat motifs described in Figure 3. I will revisit these repeat motifs later in this report. These intron, microsatellite, and mitochondrial loci were chosen because they present different, but relatively rapid mutation rates, and as such are very useful in population-level studies. Microsatellites are considered to be hypervariable and have become the DNA marker most used in studies attempting to quantify the geographic structure of populations. For this reason, we focused primarily on these microsatellite markers (see Table 1).

Each locus was isolated in the template DNA using polymerase chain reaction (PCR) (Palumbi 1996). PCR is a process whereby millions to billions of copies of target loci are generated for each template DNA. In other words, the PCR process allows us to isolate a specific area along a chromosome (i.e., a gene or locus) and generate a huge number of copies of this specific area. The PCR process requires that each locus be targeted using a pair of oligonucleotide primers. These primers are short pieces of single-stranded DNA usually around 20 or so basepairs long, that anneal to specific sites on the template DNA at a specific temperature. Each primer is then extended along the template DNA by the addition of complementary nucleotides. A polymerase enzyme catalyzes the synthesis of new

strands of DNA and is thermally stable at high temperatures. *Taq* polymerase is the enzyme used in most PCR reactions and is derived from a thermal-vent bacterium, *Thermus aquaticus*. The PCR process as whole requires a series of boiling or denaturing (94°C), annealing (usually 45°C to 65°C), and extension (72°C) steps; the length of time and temperature for each step is experimentally determined. Dr. Vicki Friesen's laboratory at Queen's University, Kingston, Ontario designed the primer pairs, except that for cytochrome *b*, and developed the initial PCR protocols for each of the loci used in this study occurred (Table 2). These protocols were refined and microsatellite multiplexes (more than one microsatellite locus amplified in a single PCR reaction) were designed at the Washington Department of Fish and Wildlife genetics laboratory (Table 2).

The PCR products were visualized and sequenced (intron, mitochondrial DNA, microsatellites) or sized (microsatellites) using either a manual electrophoretic apparatus or semi-automated sequencer. All electrophoretic procedures at the Queen's University laboratory were conducted on manual apparatus using ^{33}P isotope and audoradiographic film. Sequences or fragment sizes were read directly from the developed and dried film. The procedures in the WDFW Genetics Laboratory differed from those at Queen's University. At WDFW the microsatellite or isolated and cleaned sequencing product was loaded on a 4.4% or 4.0% polyacrylamide gel, respectively, in an ABI 377 semi-automated sequencer (Perkin-Elmer Biosystems). A laser within the sequencer excites a fluorescent dye that is incorporated within the forward primer (microsatellite) or a dideoxynucleotide terminator (sequences) (see Table 2). Each of these dyes emits light along a specific portion of the spectrum, and that light is captured by a CCD camera. The information is encoded digitally and transferred to a Macintosh computer using ABI-Collection software (Perkin-Elmer Biosystems), where it is stored for analysis.

Statistical Analysis – I sized microsatellite fragments using GeneScan 2.1 and Genotyper 2.0 software (Perkin-Elmer Biosystems). For sequence data I used Sequence Analysis 3.3 and Factura 2.2 (Perkin-Elmer Biosystems) to call all bases and to identify ambiguous points along the sequence resulting from the presence of heteroplasmy or a heterozygote, or from a poor gel image. DNA sequences were aligned using Sequence Navigator (Perkin-Elmer Biosystems). I used a combination of statistical packages to calculate (1) descriptive statistics such as allele frequencies, expected and observed heterozygosities, and deviations from Hardy-Weinberg equilibrium (*Genepop* [Raymond and Rousett 1995], *Fstat* [Goudet 1995], *Arlequin* [Schneider et al. 1997]); (2) F-statistics for

population subdivision and inbreeding coefficients (*Fstat*, *Arlequin*); and (3) molecular distances, molecular variances, phenograms, and gene flow (*Arlequin*, *Microsat* [Minch et al. 1995], *Migrate* [Beerli 1997], *Phylip* [Felsenstein 1993]).

RESULTS AND DISCUSSION

For the remainder of this report I will describe and discuss the microsatellite results only. There is sufficient data here using only the results from the microsatellite analyses to reach a *tentative* conclusion concerning the structure of Common Murre populations from British Columbia to California. *However, there are potential problems with the homology of the microsatellite size categories (see below), and I cannot confirm these tentative conclusions until each of the microsatellite loci are sequenced.* Therefore, I will addend this final report when I have completed the sequencing of the microsatellites. The addendum to the final report will also include all the sequencing results from both the intron and mitochondrial DNA analyses.

Genetic Diversity – Genetic diversity indices provide information on the amount of genetic variability that exists within and among populations and can be summarized for each locus or across all loci. Genetic diversity can be described simply as the actual number (Table 3) or relative frequency of alleles (Table 4) within a population. More importantly, genetic diversity can be represented as the frequency of heterozygotes in a population (Table 3). The degree of heterozygosity is important because each heterozygote carries two different alleles and *prima facie* represents the existence of genetic variation (Weir 1996). Furthermore, a statistical comparison between the observed and expected heterozygosity may provide information on the degree to which a population has lost genetic variability (see Avise 1994 for short discussion on the deleterious effects to a population of reduced heterozygosity and loss of genetic variation). In Table 3 we present the number of alleles; the observed and expected heterozygosity; and an estimate of the frequency of null alleles (r), based on Brookfield (1996) for each microsatellite locus and population. Null alleles are alleles that do not amplify during the PCR process and are therefore unknown. We also present data for all loci combined and include the F_{IS} statistic (inbreeding coefficient), which measures the extent of nonrandom mating in a population. High F_{IS} values, such as those shown in Table 3, can result from inbreeding, the existence of null alleles, or unknown genetic structure within each population.

The genetic diversity for each population appears similar, although Washington has either the same number or more alleles for each locus. However, the greater number of alleles from the Washington population is probably the result of a much greater sample size, compared with the other three localities (see Table 1), than an actual increase in genetic variability. For example, I randomly selected a total of 30 individuals from the Washington sample of 104 individuals, and counted the number of alleles for each locus in this new, reduced data set. The mean number of alleles per locus decreased from 9 to 7.4, with this new mean being comparable with the mean number of alleles per locus for the other three populations (see Table 3). All populations also show an overall deficit in heterozygotes, suggesting that each population may have experienced a loss of genetic variability. A loss in genetic variability may be a reasonable hypothesis for populations with relatively few individuals (e.g., Washington and British Columbia); however, the Common Murre population in Oregon and California are or have been relatively large (770,000 and 350,000, respectively [Manuwal et al. *in prep.*) and do not appear to be populations that have experienced a genetic bottleneck. More likely, the high r-values for each locus, except uapi-23 and uaa5-8, suggest that null alleles may be present in each population, especially for the 12a12 locus, and may better explain both the high F_{IS} values and the fact that there appears to be an overall heterozygote deficiency for each population. In other words, the level of genetic diversity within each population appears similar, and the lower than expected heterozygosity may be due to the presence of null alleles rather than a high incidence of inbreeding.

Another way of looking at genetic diversity is to pool data across all populations and compare within-population genetic diversity with among-population genetic diversity. This procedure will provide the first hint of the degree to which these populations are geographically structure. F_{ST} or *theta* is a measure of geographic structuring and is calculated as the proportion of total genetic variance that can be accounted for by differences *among* populations (Table 5). F_{ST} values that are not significantly different from zero indicate that the variance associated with the distribution of alleles lies entirely within populations rather than among populations. Except for 12a12, the F_{ST} values calculated for these populations are all equal to or less than zero (Table 5) suggesting that there is very little geographic structure among the populations. Two of the five loci show F_{IS} values that are significantly greater than zero, indicating a significantly high inbreeding coefficient. However, as explained above, the high F_{IS} values for these loci may be the result of null alleles. The null allele hypothesis can be tested by redesigning the primers used for these microsatellite loci and re-running all homozygote individuals. If a point mutation occurred in the sequence to which the primer anneals,

a redesigned set of primers may permit amplification of the null allele. I will attempt this experiment in the next few months.

Genetic Differentiation among Populations – The genetic diversity data provide us with some information about the geographic structure of Common Murres from British Columbia to California. However, genetic diversity data do not provide information concerning what populations may be similar or divergent, and the data do not provide a statistical test for differences in allele frequencies among the populations. The relative allele frequencies for each microsatellite locus is shown in Table 4. Population-specific alleles (private alleles) occur in only two of the five loci (uapi-23 and 14b29), at a frequency of 30% of the total number of alleles for both loci combined. That means that 30% of all alleles for these two loci occur in only one population; however, the total number of genotypes possessing these private alleles is only 15. In other words, of the 979 total microsatellite genotypes identified (Table 1), only 1.5% are restricted to a single population, and of those 15 genotypes with private alleles, 80% (12 genotypes) reside in the Washington population (Table 4). As I discussed above, the relatively large number of alleles for the Washington population may be due to a large sample size rather than an increase in the overall genetic diversity.

Differences in allele frequencies among populations can be tested using a chi-square Markov-chain procedure (e.g., Raymond and Rousset 1995). However, this procedure assumes a binomial distribution of the alleles and requires that the populations are in Hardy-Weinberg equilibrium (Weir 1996). Populations are in Hardy-Weinberg equilibrium when observed and expected heterozygosities are equal. I have already shown in Table 3 that there is a significant heterozygote deficiency in all populations for two loci and in one population for two other loci, indicating that for these loci, these populations are not in Hardy-Weinberg equilibrium. Weir (1996) suggested that when Hardy-Weinberg equilibrium and a binomial distribution cannot be assumed, a numerical resample technique can be used to test for population differentiation. Therefore, I tested for differences in allele frequencies using a bootstrap procedure. For each bootstrap run, individuals from each population were selected at random, *with replacement*, while preserving the original sample sizes for the populations. Allele frequencies for each locus and population were calculated for each bootstrap and their means and standard deviations were calculated for a total of 1,000 bootstrap runs. We considered populations with non-overlapping 95% confidence intervals to be significantly different in their allele frequencies for that particular allele.

Three of the five loci show significantly different allele frequencies, with each population being significantly different from another in at least one locus (Table 4). Washington appears most divergent, having significantly different allele frequencies than the other three populations in both the uapi-23 and 14b29 loci. Furthermore, these significant differences cannot be attributed entirely to the relatively high proportion of private alleles in the Washington population. For uapi-23, Washington was significantly different than the other populations in the frequency of two alleles that are not private and involves a total of 16 genotypes from Washington, or 8% of the total number of genotypes from this population. Washington was significantly different in the relative frequencies for four alleles from 14b29 (Table 4), and although one of these alleles was a private allele for Washington, five genotypes from Washington possessed this allele (rather than the normal one or two genotypes for the other private alleles). While Washington appeared to be the most divergent in terms of relative allele frequencies, British Columbia and California, and Oregon and California differed in their allele frequencies for only one locus (12a12 and 14b29, respectively). The fact that there are significant differences in allele frequencies suggests that there is some degree of geographic structure among these populations.

Genetic Distances among Populations – A genetic distance is a statistic that allows us to compare the genetic structure of two populations and to determine how similar these populations are to each other, compared with the genetic structure of other populations. The genetic distance can be thought of as the time since two populations diverged from a common ancestral population (Weir 1996), with large distances indicating a longer period of divergence than a small distance. There are a plethora of genetic distance measures for use with microsatellite data, each incorporating assumptions about the type and rate of mutations, the effects of fluctuations in population size, and if the measure maintains linearity with increasing divergence among populations (see Goldstein et al. 1995, Slatkin 1995, Weir 1996, [Goldstein and Pollack also provide a very good discussion of genetic distances at http://lotka.stanford.edu/microsat/microdist.html]). There are two types of mutation models that are used in these distance measures: stepwise mutation (Ohta and Kimura 1973) and infinite alleles (Kimura and Crow 1964). The stepwise mutation model assumes that changes in allelic size (i.e., microsatellite length) is the result of a stepwise increase in the number of repeat units. That is, microsatellites mutate by adding one or a few number of repeat units producing alleles of consecutive lengths (e.g., 12a12, see Table 4). The infinite-alleles model assumes that all alleles are equally probable and there is no requirement that microsatellites increase in size by adding only one or a few repeats units at a time. Although most recent work on microsatellites suggest that a stepwise

mutation model is more appropriate (e.g., Primmer et al. 1998), some of the loci used in this analysis consist of alleles that are not sequential in length (12a12 and uapi-23). Therefore, we present two different distance measures, one assuming a stepwise mutation model (delta mu squared [$(\delta\mu)^2$]) and the other assuming an infinite-alleles model (coancestry coefficient, *Theta* [θ], or F_{ST} [= -ln(1- θ)]) (Figure 4). For these analyses, we used only the three loci with significant differences in allelic frequencies among the populations (see Table 4). The unrooted phenograms that accompany each distance matrix are simply graphical representations of the matrix and are not hypotheses concerning phylogeographic relationships of the populations. The $(\delta\mu)^2$ distance measure indicates that Washington and Oregon are 10 to 16 times more similar to each other than either is to California or British Columbia. In addition, based on these three loci and the $(\delta\mu)^2$ distance measure, California and British Columbia appear to be more similar to each other than to the other populations, and Oregon and California are the most divergent of all pairwise comparisons of populations (Figure 4). The F_{ST} measure paints an extremely different picture than the $(\delta\mu)^2$ measure. With the F_{ST} distances, Oregon and California show a genetic distance that is essentially zero, while Washington and British Columbia are three to 12 times more similar to each other than either is to California or Oregon (Figure 4). One point to consider in interpreting the conflicting results between these two distance measures is the fact that all three loci show non-sequential allelic patterns (see Table 4), which would favor the infinite-allele mutation model and the F_{ST} distance.

Analysis of Molecular Variance – I tested several hypotheses for the geographic structure of these Common Murre populations using analysis of molecular variance (AMOVA). AMOVA uses an ANOVA framework to partition variation among and within an *a priori*-defined hierarchy. In this analysis, I first considered all four populations, without any geographic structure and included all five loci (Table 6). The relative amount of among population variation increased when we limited the analysis to only those loci with significant among-population allelic variation. I then considered two alternative hypotheses concerning the geographic structure of these populations, suggested by the genetic distances shown in Figure 4. It is clear from these results that a WA-BC and OR-CA grouping provided additional information on the geographic structure of these populations, and suggests, based on these loci, that a northern (WA and BC) and southern (OR-CA) grouping of these birds is the best explanation for the among-group genetic variation. However, since the AMOVA calculation uses a form of *theta*, my use of the AMOVA does not resolve the conflicting results presented in Figure 4. Furthermore, despite the fact that the amount of genetic variation between the

WA-BC and OR-CA groups is nearly 13 times that found within each of these groups (Table 6), most of the overall genetic variation still resides *within* each population. Therefore, I conclude that although there does appear to be some geographic structure, it is minor compared with the amount of variation within each population.

Gene Flow – Gene flow is the exchange of genetic material between two populations as a result of migration of individuals (Avise 1994) and has been defined as the effective population size times migration rate (N_em). High gene flow between two populations indicates that there is a large number of individuals moving between these populations and the populations may be functioning as a single population. Just as there are several different methods of calculating the genetic distance between populations, there are also several ways of estimating gene flow. I provide two diagrams for the pattern of gene flow among the four populations (Fig. 5). Figure 5a is base on F_{ST} values and assumes that population sizes are equal and mutation rates are negligible compared with migration rates. Clearly these assumptions are violated: population sizes in Oregon and California are orders of magnitude larger than that in Washington and British Columbia (e.g., 10-15,000 in WA and 700,000 in Oregon). Furthermore, with the high mutation rate in microsatellites it is not clear if mutation rates are negligible compared with migration rates. Nevertheless, the pattern shown in Figure 5a indicates considerable gene flow between California and Oregon, slightly less gene flow between Washington and British Columbia, and minor gene flow among the other pairwise comparisons. Because this estimate of gene flow is based on F_{ST} values, it corresponds precisely with the genetic distances shown in Figure 4. Figure 5b is based on an analysis using the program *Migrate* (Beerli, 1997) and uses a maximum likelihood procedure and coalescence theory to estimate gene flow. Because the program also estimates the effective population size for each population it makes no assumption about equal population sizes and provides estimates for gene flow in both directions. Based on this analysis most of the gene flow appears as emigration out of Washington, with minor gene flow from Oregon and British Columbia into Washington and from California and Oregon into Oregon and British Columbia, respectively. All other gene flow estimates are negligible. The results presented in Figure 5b are consistent with a third method of calculating gene flow based on the private alleles (Slatkin 1985), in that both methods suggest that Washington is a nexus of gene flow among these four localities. However both methods may be biased by the large sample size and therefore large number of private alleles in the Washington population (see above). The two features that are consistent between Figures 5a and 5b are the fact that both methods indicate considerable gene flow

between Washington and British Columbia and less, but still relatively high gene flow between Washington and Oregon.

Microsatellite Sequences – As discussed above, microsatellite alleles are actually lengths of DNA, measured in base pairs, and defined by their relative molecular weights and mobility in an electrophoretic field. Different alleles are formed by the addition (and presumably the deletion) of repeat units. In at least three of the five microsatellites used in this analysis the repeat motif is complex (Figure 3). This creates a problem in interpreting the data in that there is an uncertainty as to the homology of alleles with the same size class. Alleles that are homologous are the same by decent; that is, they share a common ancestral allele. Alleles that are homoplasious are the same by convergence; that is, they do not share a common ancestor and are a product of parallel mutations. All calculations of populations parameters (e.g., genetic distance or gene flow) should be based on *homologous* alleles. At a minimum, homologous alleles should be the same size and possess the same repeat motif. However, the repeat motifs shown in Figure 3 leave open the possibility that two individuals may converge on the same allele size through a different allelic pattern For example, the Common Murre individual that was used to develop the uapi-23 microsatellite primers had an allelic pattern for the microsatellite portion of its sequence of $(CT)_8(TT)(CT)_{11}$ for a total of 40 bp. However, another individual can also produce a microsatellite allele of 40 bp for uapi-23 with an allelic pattern of $(CT)_7(TT)(CT)_{12}$. These two alleles would not be homologous and should be treated as distinct.

By sequencing a small subset of the individuals for each locus I tested the hypothesis that for each locus all alleles exhibiting the same size class are homologous. I sequenced four individuals exhibiting a total of six size classes for 14b29 (Figure 6). The four individuals produce two general allelic patterns. with the size class of 133 bp exhibiting both patterns (Figure 6). This means that there are at least two different 133 bp alleles in the data and I should treat them as distinct. Furthermore, because the second allelic pattern in Figure 6 has an interrupting sequence (TA), two sequences of the same size can be generated many different ways even if they maintain the same allelic patterns (e.g., $(CA)_9(TA)(CA)_4$; $(CA)_{10}(TA)(CA)_3$; $(CA)_{12}(TA)(CA)$ would each produce the same length fragment, but would not be homologous). I also sequenced three individuals who were homozygous for the 171 bp allele in uapi-23. This means that I sequenced a total of six single stands of DNA each a total of 171 bp long and all assumed to be homologous for the 171 bp allele. These sequences resulted in three or four allelic patterns and a maximum total of four different alleles for

these three individuals (Figure 7). In other words, while I assumed that each of these individuals shared two copies of the same allele, they actually possessed four different alleles for the 171 bp size class, and two of these individuals were actually heterozygous for this locus. These results suggest that a simple fragment analysis of these microsatellite loci, as I conducted here, is not sufficient and all alleles need to be confirmed by a separate analysis using sequences. However, if the various repeat motifs for each size class are randomly distributed with respect to locality, the fact that each size class is composed of non-homologous alleles should not affect my overall interpretation of the results. This issue becomes important only if the various repeat motifs segregate according to locality. In this case, my pooling of non-homologous alleles will mask geographic structure. An example using the size classes and not the repeat motifs may be instructive: If one population is fixed for a size class of 171 bp for some locus, and another population is fixed for a size class of 173 bp, a test for differences in allele frequencies would show that these populations are significantly different. However, if I pooled the 171 and 173 bp size classes as a single class, the populations would show no significant differences in allele frequencies. The only way to determine if the repeat motifs for these microsatellites are indeed geographically structured would be to sequence each individual for each locus.

For the remainder of this report, I will assume that the repeat motifs for these microsatellites are *not* structured geographically. However, I will revisit this assumption in the addendum to this report.

COMMON MURRE RESTORATION IN WASHINGTON

There are a variety of techniques available for restoring seabird populations affected by oil spills (Warheit et al. 1997). The efficacy of each of these techniques will depend on the ecological and demographic factors that are most limiting for the affected population. For example, following an decline associated with an oil spill, a seabird population has failed to recover naturally due to a high incidence of predation on chicks by non-indigenous mammalian predators. The demographic factor that is limiting population growth here would be survival of chicks. A restoration plan designed to decrease adult mortality in gill nets or a plan intended to increase immigration into this population would not address the ecological or demographic factor that is most important to this population, and may fail to achieve its stated goal. Depending on demographic factors, natural populations have an ability to be restored naturally, without human intervention. That is, if recruitment into a population through reproduction and/or immigration, is greater than the mortality and emigration rate for that

population, the population should grow. If a population is not growing following the effects from an oil spill, human intervention may help augment the natural process by enhancing one or several demographic factors. Immigration has played an important role in the restoration (natural or otherwise) of many seabird species (Nur and Ainley 1992, Warheit et al. 1997). In fact, the *Apex Houston* Restoration Plan for Common Murres relies nearly entirely on increasing the rate of immigration to particular colonies by using social attraction devices. The success of a restoration plan that depends on immigration will be directly related to the probability that a particular colony will receive immigrants, and the probability that a colony will receive immigrants depends on the geographic structure of that population.

I have shown with these data that there is little geographic structure to the Common Murre populations from British Columbia to California. Although there may be some structure along a north-south gradient, there appears to be sufficient gene flow in and out of Washington to support a restoration plan that requires an increase in immigration. A high incidence of gene flow may also permit, with sufficient time, a natural recovery of Common Murres in Washington, without the need to implement a hands-on restoration program. However, the recovery of Common Murres in Washington may also be affected by factors other than immigration (e.g., high adult mortality associated with their bycatch in gill nets), and all factors need to be considered before deciding what is the most appropriate restoration option.

The Point Grenville social attraction restoration project for Common Murres proposed by the *Tenyo Maru* Trustee Committee assumes that immigration into colonies around Point Grenville is compromised by the lack of social attraction. I will not address the issue concerning social attraction here. However, since this project's success depends on our ability to increase the rate of immigration to these colonies, we require some indication that immigration itself would be a likely event (regardless of whether it is facilitated through a social attraction technique). First, we know that the rapid increase in the number of Common Murres breeding on Tatoosh Island during the late 1980s and early 1990s was the result of individuals moving into the colony rather than recruitment from chicks hatched on island (Paine et al. 1990; Warheit, in press). This was our first hint that immigration may play an important role in the status of a particular Common Murre colony in Washington. Second, from the data presented here, we can assume that there is an exchange of individuals from at least Oregon and British Columbia into and out of Washington. Although it is difficult to calculate the number of immigrants expected to arrive at Washington colonies from

colonies outside the state, we can assume for the purpose of designing restoration plans that immigration is not an unlikely event.

A few caveats are in order. First, as discussed above, I am assuming that the non-homologous nature of some (or all?) the size classes for these microsatellite loci are not affecting the analysis. That is, there is no geographic structure to the different repeat motifs. Second, I am also assuming that these five microsatellite loci are sufficient to capture geographic structure among these populations, if such structure exists. It is possible that these particular microsatellites have an unusually slow rate of mutation, which would make them inappropriate for this type of population analysis. Furthermore, these microsatellite loci could have an extremely rapid rate of mutation, thereby dramatically increasing the rate of convergence in allele sizes among these localities. For this reason, additional loci with relatively rapid mutation rates, such as the two mitochondrial control region loci I have been working with (see above), are needed to help confirm the conclusions based strictly on these microsatellites. If I can show that there is no geographic structure to the different microsatellite repeat motifs, and the sequence data from the other loci I have been working with (Table 1) confirm the results based on the microsatellites, the conclusions I reached in this report are well supported.

LITERATURE CITED

Avise, J.C. 1994. Molecular Markers, Natural History, and Evolution. Chapman and Hall, New York.

Beerli, P. 1997. *MIGRATE: documentation and program, part of LAMARC.* Distributed over the Internet, http://evolution.genetics.washington.edu/lamarc.html.

Brookfield, J.F.Y. 1996. A simple new method for estimating null allele frequency from heterozygote frequency. *Molecular Ecology* 5:453-455.

Felsenstein, J. 1993. PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle.

Goldstein, D.B., A. Ruiz Linares, L.L. Cavalli-Sforza, and M.V. Feldman. 1995. An evaluation of genetic distances for use with microsatellite loci. *Genetics.* 139:463-471.

Goudet, J. 1995. FSTAT (version 1.2): a computer program to calculate F-statistics. *Journal of Heredity* 86:485-486

Kimura, M., and J.F. Crow. 1964. The number of alleles that can be maintained in a finite population. *Genetics* 49:725-738.

Kocher, T.D., W.K. Thomas, A. Meyer, S.V. Edwards, S. Paablo, F.X. Villablanca, and A.C. Wilson. 1989. Dynamics of mitchondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proc. Natl. Acad. Sci.* USA 86:6196-6200.

Manuwal, D. A., H. R. Carter, R. W. Lowe, U. W. Wilson, J. E. Takekawa, and W. J. Sydeman. *in prep.* Assessment of the biology, population and conservation of the common murre (*Uria aalge californica*) in California, Oregon, Washington, and British Columbia, Canada. USFWS.

Minch, E., A. Ruiz Linares, D. Goldstein, M. Feldman, L.L. Cavalli-Sforza. 1995. MICROSAT (version 1.5): a computer program for calculating statistics on microsatellites. Distributed over the Internet, http://lotka.stanford.edu/microsat/microsat.html.

Nur, N., and D.G. Ainley. 1992. Comprehensive review and critical synthesis of the literature on recovery of marine bird populations from environmental perturbations. *Final Report to Exxon Valdez Restoration & Planning Work Group.* Anchorage, AK.

Ohta, T., and M. Kimura. 1973. The model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a genetic population. *Genet. Res.* 22:201-204.

Paine, R., J.T. Wooton, and P.D. Boersma. 1990. Direct and indirect effects of Peregrine Falcon predation on seabird abundance. *Auk* 107:1-9.

Palumbi, S.R. 1996. Nucleic acids II: the polymerase chain reaction. Pages 205-247 *in* D.M. Hillis, C. Moritz, and B.K. Mable (eds). *Molecular Systematics, Second Edition.* Sinauer Associates, Inc., Sunderland, MA

Parrish, J.K. 1998. *Attendance and reproductive success of Tatoosh Island common murres. Final Report 1997. Tenyo Maru* Trustee Committee.

Primmer, C.R., H. Ellegren. 1998. Patterns of molecular evolution in avian microsatellites. *Molecular Biology and Evolution* 15:997-1008.

Quinn T.W. 1997. Molecular evolution of the mitochondrial genome. Pages 3-28 *in* Mindell, D.P. (ed.). *Avian Molecular Evolution and Systematics.* Academic Press, San Diego.

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

Sambrook, J., E.F. Fritsch, T. Maniatis. 1989. *Molecular Cloning. A Laboratory Manual.* 2nd Edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Schneider, S., J-M. Kueffer, D. Roessli, and L. Excoffier. 1997. *Arlequin ver. 1.1: A software for population genetic data analysis.* Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Slatkin, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139:457-462

_____. 1985. Rare alleles as indicators of gene flow. Evolution 39:53-65.

Storer, R.W. 1952. A comparison of variation, behavior, and evolution in the seabird genera *Uria* and *Cepphus*. *Univ. Calif. Publ. Zool.* 52:121-222.

Swartz, L.G. 1966. Sea-cliff birds. Pp. 611-678 *in* Wilimovsky, N.J. and J.N. Wolfe (eds.). *Environment of the Cape Thompson region, Alaska*. U.S. At. Energy Comm., Oak Ridge, TN.

Tuck, L.M. 1961. *The Murres*. Monogr. No. 1. Can. Wildl. Serv.

Warheit, K.I. *In press*. Common Murre (*Uria aalge*). *In*: Larsen, E.M. (ed.). Management recommendations for Washington's Priority Species, Volume IV: Birds. Wash. Dept. Fish and Wildl., Olympia.

_____, C.S. Harrison, and G.J. Divoky (eds). 1997. *Exxon Valdez* Oil Spill Seabird Restoration Workshop. *Exxon Valdez Oil Spill Restoration Project Final Report, Project 95038*. Technical Publication Number 1. Pacific Seabird Group, Seattle. 171 + x pp.

Weir, B.S. 1996. *Genetic Data Analysis II: Methods for Discrete Population Genetic Data*. Sinauer Assoc., Inc., Sunderland, MA.

Wilson, U.W. 1995. Washington common murre colony surveys 1995. Final Report. *Tenyo Maru* Trustee Committee.

_____. 1996. Washington common murre colony surveys 1996. Final Report. *Tenyo Maru* Trustee Committee.