# The genetic consequences of non-equilibrial dynamics in bowhead whales

Karen K. Martien, Eric Archer, Bonnie J. Ripley and Barbara L. Taylor
Southwest Fisheries Science Center, 8604 La Jolla Shores Dr., La Jolla, CA 92037 USA

## ABSTRACT

Archer et al. (2007) used a realistic model of bowhead whale population dynamics and genetics to determine whether the genetic heterogeneity that has been observed in bowhead whales from the Bering-Chukchi-Beaufort Seas region is consistent with the hypothesis of a single population that is out of genetic equilibrium due to the effects of commercial whaling. They found that most empirical results were consistent with a single population. However, the complexity of their model precluded them from varying the minimum population size reached during the simulation, the degree of age selectivity in the historical harvest, or the variance in reproductive success among males. We used a simpler model to examine the impact of these factors on the degree of genetic disequilibrium in the simulated populations. Our model used the same life history parameters as the one presented by Archer et al., but simulated historical commercial whaling as a single catastrophic year. None of the factors we examined had a substantial impact on genetic equilibrium, indicating that the conclusions of Archer et al. are also unlikely to be sensitive to these factors. However, our results do reveal some interesting properties of non-equilibrial populations that warrant further investigation.

## INTRODUCTION

Genetic heterogeneity has been found in bowhead whales in the Bering-Chukchi-Beaufort Seas region (hereafter referred to as BCB bowheads) (Givens et al. 2004, Givens et al. 2007, Jorde et al. 2006, Jorde et al. 2007, LeDuc et al. 2007). Such heterogeneity has been hypothesized to either represent the presence of two stocks or a single stock out of genetic equilibrium. Realistic simulations were used to mimic the historical catch and population trajectory of bowhead whales (Archer et al. 2007). These simulations found that most empirical results were consistent with a single stock scenario. However, the complexity of these simulations precluded exploration of some plausible scenarios of single stock dynamics. For example, reducing the population to the low hundreds resulted in such high chances of extinction using the realistic model that obtaining a sufficient number of replicates for analysis was infeasible. Here we use a simpler model that uses the same life history parameters, but simulates the historical catch as a single catastrophic year. Using this model allows us to explore three factors: 1) the effect of a lower minimum population size, 2) the effect of a historical catch that was not random with respect to age and left few adults, and 3) the effect of mating system. The realistic historical catch model assumes random mating, however, sperm competition is plausible in bowhead whales and would result in unequal reproductive success of males. These factors reduce the effective population size and may have consequences on the magnitude of non-equilibrial genetic patterns. This paper examines how these factors affect Hardy-Weinberg equilibrium and investigates the magnitude of expected genetic differences between cohorts.

Bowhead whales are currently not in genetic equilibrium because they experienced a large reduction in abundance from commercial whaling and have a long generation time (for a more detailed review see Taylor et al., 2007). Commercial whaling reduced abundance by approximately 90% over a short period concentrated between 1848 and 1909. Generation time is estimated to be 36 years for the current expanding population (Taylor et al. 2007) meaning that only 2.5 generations have passed since the dramatic reduction. Because bowhead whales can live over 100 years, it is certain that some whales that were born prior to the end of commercial whaling are still living today. The genes of these old whales represent frequencies of the unexploited population, while those of recent cohorts represent the smaller, yet still diverse, gene pool that survived commercial whaling.

Genetic differences between cohorts can also arise through unequal reproductive success. The effective population size for nuclear markers is a factor not only of overall abundance but also reproductive variation in males and females. Although the mating system of bowhead whales is not fully understood, the testes of males are exceptionally large and vary in size between individuals. Males with testes five times larger than average are sometimes observed (O'Hara et al. 2002). It is not known whether these 'supermales' enjoy a reproductive advantage. Nonetheless, the large and highly variable testes size in bowheads is consistent with sperm competition and may result in more variable reproductive success in males. The reduced effective population size that would result from highly variable male reproductive success is expected to increase the magnitude of genetic changes resulting from population reduction followed by rapid expansion.

We used simulations to explore the genetic consequences of commercial harvest and high variance in male reproductive success in bowhead whales. We calculated genetic differentiation between three age cohorts from our simulated populations – the *Old* cohort (born prior to 1932), the *Middle* cohort (born between 1932 and 1992), and the *Young* cohort (born after 1992). We also tested for Hardy Weinberg equilibrium in the simulated populations. For both of these analyses, we examined the impact of the severity of the depletion due to commercial whaling, the degree of variance in male reproductive success, the number of animals sampled, and the degree of age-bias in the sample.

## METHODS

We used the R-based package Rmetasim, which is a library of functions to perform individual-based population genetic simulations (Strand 2002). All individuals in our model belong to a single population. Each individual has a multilocus genotype and an mtDNA haplotype. Individuals are structured demographically with an age- or stage-based matrix population model (see '*Demography'* section below; Caswell 2001). At each time step individuals are randomly assigned births, stage transitions, and deaths according to the rates specified in the matrix model (used as distributions to incorporate demographic stochasticity). Offspring

genotypes are determined by parental genotypes assuming random mating (unless another mating system is specified), independently segregating alleles, and neutrality of markers. R and Rmetasim are available freely from www.cran.r-project.org. This work was done using R v. 2.4.1 and Rmetasim v. 1.1.008. For all parameters not explicitly defined here we used the program default values.

**Model structure**

We used a model very similar to that described by Archer et al. (2007). Both models used the same demographic parameters (see *Demography*) and the same mechanism for initialization and burn-in (see *Initialization and Burn-in*). The two models differ in the way that harvest and sampling were simulated. Archer et al. (2007) matched the historical harvest and sampling routines in their model as closely as possible to empirical data. In each simulation year, the number of animals killed in the simulated population was equal to the estimate used in the Aboriginal Whaling Management Procedure (George and Zeh, pers. comm.). Samples drawn in the Archer et al. model were also closely matched to empirical data with respect to age and sex.

For this paper, we developed a simplistic model of bowhead whale history in which all whaling mortality occurs in a single year, after which the population is allowed to recover. The model was divided into three time periods (termed "epochs" in Rmetasim) with different demographic parameters. The first epoch was the burn-in phase during which the simulated population was brought into genetic and demographic equilibrium. The second epoch represented a compression of all of the whaling mortality, which actually occurred over a 60 year-period, into one year of catastrophic mortality. This year was set at 1909—the likeliest year of lowest population abundance (Brandon and Wade, 2006). We examined two values for the size of the population at the nadir ($N_{low}$): 300 and 1873. These values bracket the 95% confidence intervals in Brandon and Wade (2006). Varying $N_{low}$ allowed us to examine what effect this parameter had on the degree of genetic disequilibrium in the final population.

Within our model, whaling mortality was biased with respect to age, with older animals suffering higher rates of mortality. Such selectivity was not considered by Archer et al. (2007). Though there is no evidence of intentional selectivity on the part of commercial whalers, the fact that bowhead whales are age and sex segregated on migration may have resulted in selectivity based on availability (Bockstoce 1986). Size selectivity in the commercial hunt, by whatever mechanism, is suggested by the fact that the average size of whales caught by whalers declined notably in the early decades of the hunt (George et al. 2007). To simulate selective harvest, we first reduced older juveniles and adults by 90%. Individuals were then killed at random (irrespective of stage membership) until the abundance equaled $N_{low}$.

During the third epoch, the population was allowed to recover for 92 years (1910-2002). During the recovery phase, the population experienced exponential growth at a rate chosen such that the final abundance at the end of the simulation was consistent with the latest abundance estimates (see *Demography)*.

**Initialization and Burn-in**

The simulated populations were initialized at carrying capacity of $K$ = 12,100 and in stable age distribution. The mtDNA haplotype and microsatellite allele frequencies of the simulated populations were initialized from distributions generated by the coalescent model SIMCOAL v2.1.2 (Laval and Excoffier 2004). Initializing from a coalescent rather than from random haplotype and allele frequencies resulted in a considerably shorter burn-in, as the frequency distributions generated by SIMCOAL are already very close to equilibrium.

SIMCOAL requires as input an estimate of effective population size ($N_e$). $N_e$ can be calculated based on the loss of heterozygosity over time in the absence of mutation according to the relationship

$$N_e = \frac{m}{2 \cdot \left(1 - \left(\frac{H_t}{H_0}\right)^{\frac{g}{t}}\right)}$$

where:

$H_0$ = initial heterozygosity
$H_t$ = heterozygosity at time $t$
$t$ = elapsed time in years
$g$ = generation time (= 37 years)
$m$ = multiplier chosen to start the simulation burn-in phase closer to equilibrium (= 1.45).

We estimated $N_e$ for our model by running 20 replicate projections, each parameterized with the same life history matrices as in the full model, but with only a single microsatellite locus and mtDNA sequence. Both loci were initialized with 1000 alleles and had mutation rates set to zero. By projecting the population forward 4,000 years and calculating heterozygosity for each marker at the beginning and end of the projection, we were able to estimate $N_e$ for each marker using equation 1. It is necessary to estimate $N_e$ separately for the two marker types because the haploid and uni-parentally inherited nature of mitochondrial DNA results in a markedly smaller effective population size for the mitochondrial genome. We calculated the average effective population size ($\overline{N}_e$) for each marker type as the harmonic mean of $N_e$ from the 20 replicate projections.

We used SIMCOAL to generate initial sequences for a single mtDNA marker and 33 microsatellite loci. The sample size generated by SIMCOAL was $\overline{N}_e$ for the mtDNA sequences and the smaller of $\overline{N}_e$ and 1000 for the microsatellite loci. The mtDNA sequence was specified to be 397 bp, with a Ts:Tv of 10:1, and a mutation rate of $9.4 \times 10^{-3}$. For the microsatellites, two groups of loci were simulated representing the 11 "original" and 22 "new" loci used in Givens et al. (2007). Average mutation rates were set at $3.0 \times 10^{-4}$ and $1.5 \times 10^{-3}$ for the original and new loci respectively. Mutation parameters were tuned by Archer et al. (2007) to produce diversity comparable to that observed.

Following initialization, simulated populations were projected forward for 4,000 years in order to ensure that they had reached demographic and genetic equilibrium. Previous examinations of the trajectories of the number of mtDNA haplotypes, microsatellite alleles, and heterozygosity in both markers indicated that 4,000 years was a sufficient amount of time to ensure that these values were relatively stable (see Archer et al. 2007 for equilibrium diagnostics). A sample of all markers was independently generated from SIMCOAL for each burn-in replicate.

**Demography**

We used the 5% and 95% confidence levels for parameters estimated in Brandon and Wade (2006) as guidelines for the minimum and maximum values for our model. In general, our demographic model categorizes whales into five categories: juvenile (1+) (*J1*), sexual immatures with same survival as adults (*J2*), adult female (*AF*), adult male (*AM*), and supermale (*SM*). In a stage-based model, individuals have a probability of moving to the next stage based on the duration of their current stage. This has the unfortunate property of allowing some individuals to remain in any stage indefinitely. Because of the long durations of stages for bowhead whales, the simplest (5 stage) model we formulated had unrealistic age distributions within stages (e.g. 50 year-old juveniles or 500 year-old whales). Therefore it was necessary to divide categories into more stages simply to make model behavior more realistic; the simplest demographic model we developed that represented bowhead life history well has juveniles split into five stages, for a total of eight stages (Fig. 1).

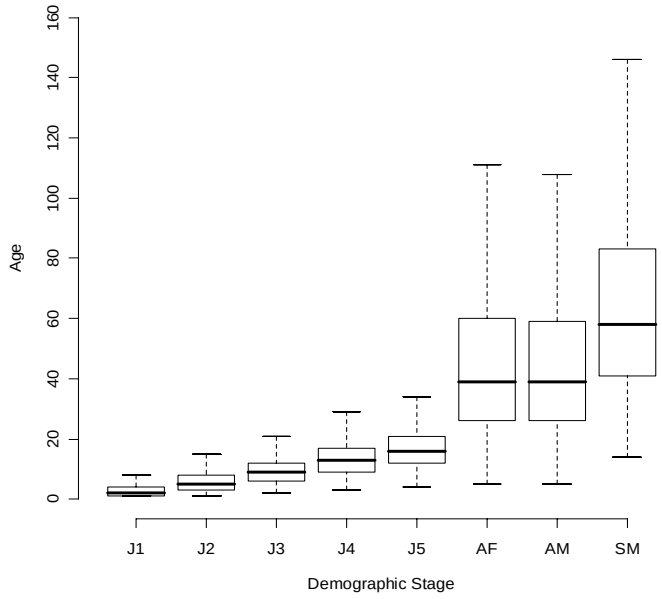


*Figure 1. Box and whisker plots of age-in-stage distributions at model year 500. Distributions are from 100 replicates of Scenario 1. The bar in the box is the median, the top and bottom of the box are the upper and lower quartiles (the distance between them is the inter-quartile range or IQR), the whiskers extend to the highest and lowest data points no further than 1.5*IQR above the top of the box or below the bottom of the box. Outliers--points beyond 1.5*IQR—are not shown.*

Projection matrix parameters were calculated using the fixed stage duration method (Caswell 2001, section 6.103). In this method, survival rates for each stage ($\sigma_i$), and durations for each stage ($T_i$) are used to calculate stage transition probabilities ($\gamma_i$):

(1)
$$\gamma_i = \frac{\left(\frac{\sigma_i}{\lambda}\right)^{T_i} - \left(\frac{\sigma_i}{\lambda}\right)^{T_i - 1}}{\left(\frac{\sigma_i}{\lambda}\right)^{T_i} - 1}$$

Then the stage transitions and survival parameters are, respectively:

(2) $G_i = \sigma_i \gamma_i$ and

(3) $P_i = \sigma_i (1-\gamma_i)$ .

The sum of the durations of the five juvenile stages was set to the estimated age at sexual maturity and durations for adult stages were based on a maximum average lifespan of about 100 years. Maturing juveniles were assumed to become male or female with equal probability. Reproduction ($F_6$) represents the proportion of females reproducing multiplied by calf (0+) survival. As the parameter with the most uncertain, it was solved for after other parameters were set, with a cap of 0.33 (all females having calves every 3 years with 100% survival to age 1). Parameters were defined for three projection matrices: one for the stable, near-*K*, pre-whaling population, one for the growing population starting with 300 whales, and one for the growing population starting with 1873 whales (Table 1). The parameters that were changed to achieve different trajectories were: age at sexual maturity, juvenile and adult survival, timing of transition from juvenile to adult survival rate, and reproductive rate. The population growth rates for the matrices are: $\lambda = 1.000$, $\lambda = 1.041$ and $\lambda = 1.026$, respectively. The reproductive parameters ($F_6$) for these matrices are 0.235 for the population at *K* and 0.295 for both growing populations.

*Table 1. Demographic parameters for each population growth rate: at K and growing from 300 and 1873 whales. For each stage (same notation as Figure 1), lower level parameters T (stage duration) and $\sigma$ (survival in stage) are used to calculate $\gamma$ with Equation 1. Matrix model parameters P(survival in stage) and G (stage transitions) are calculated from $\gamma$ by Equations 2 and 3.*

| | $\lambda = 1.000$ | | | | | $\lambda = 1.041$ | | | | | $\lambda = 1.026$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Stage** | **T** | **σ** | **γ** | **P** | **G** | **T** | **σ** | **γ** | **P** | **G** | **T** | **σ** | **γ** | **P** | **G** |
| J 1 | 4 | 0.800 | 0.173 | 0.661 | 0.139 | 2 | 0.925 | 0.470 | 0.490 | 0.435 | 3 | 0.925 | 0.299 | 0.648 | 0.277 |
| J 2 | 4 | 0.978 | 0.242 | 0.741 | 0.236 | 3 | 0.985 | 0.315 | 0.675 | 0.310 | 3 | 0.925 | 0.299 | 0.648 | 0.277 |
| J 3 | 4 | 0.978 | 0.242 | 0.741 | 0.236 | 3 | 0.985 | 0.315 | 0.675 | 0.310 | 3 | 0.925 | 0.299 | 0.648 | 0.277 |
| J 4 | 4 | 0.978 | 0.242 | 0.741 | 0.236 | 3 | 0.985 | 0.315 | 0.675 | 0.310 | 3 | 0.985 | 0.320 | 0.670 | 0.315 |
| J 5 | 4 | 0.978 | 0.242 | 0.741 | 0.236 | 3 | 0.985 | 0.315 | 0.675 | 0.310 | 3 | 0.985 | 0.320 | 0.670 | 0.315 |
| AF | 50 | 0.978 | 0.011 | 0.967 | 0.011 | 50 | 0.985 | 0.004 | 0.981 | 0.004 | 50 | 0.985 | 0.006 | 0.979 | 0.003 |
| AM | 50 | 0.978 | 0.011 | 0.967 | 0.011 | 50 | 0.985 | 0.004 | 0.981 | 0.004 | 50 | 0.985 | 0.006 | 0.979 | 0.003 |
| SM | 25 | 0.978 | 0.030 | 0.948 | 0.029 | 25 | 0.985 | 0.019 | 0.966 | 0.019 | 25 | 0.985 | 0.023 | 0.962 | 0.023 |

We examined sensitivity to variance in male reproductive success by including 'supermales' in some of our simulations. Supermales are those which have a disproportionate contribution to offspring compared to regular adult males; the mechanism for this dominance could be social or at the sperm competition level (Brownell and Ralls 1986), since some males have extremely large testes (O'Hara et al. 2002). Unequal reproductive success among females has also been suggested for great whales (Rosenbaum et al. 2002) but was not incorporated into this model because for the population to grow to the present size, high reproductive success among all females was necessary.

Rmetasim has a separate matrix that can be used to define the odds of reproductive success of males in different stages. For the models with supermales, their odds of fathering offspring were set at 10:1 compared to the regular males. Supermales were assumed to have the same mortality rate as regular adult males. We hypothesized that ca. 1% of all adult males are supermales. Therefore the probability of transition from AM to SM was set to 0.001 (in scenarios including supermales), which results in the approriate proportion supermales at simulation year 500.

In all simulations that include supermales, only one supermale survives whaling. This is an extreme assumption that maximizes the probability of observing genetic differentiation between age cohorts, as it ensures that a large fraction of the animals born in the years immediately following the harvest epoch are sired by a single male. We examined all four combinations of presence/absence of supermales and size of the nadir, resulting in four different simulation scenarios. For convenience, we refer to scenarios by number (Table 2) in the remainder of the paper.

Table 2. Definitions of the four scenarios examined in this paper. Throughout the results and discussion, scenarios are referred to by number.

| Scenario | Supermales present? | Size of nadir |
|---|---|---|
| 1 | Yes | 300 |
| 2 | No | 300 |
| 3 | Yes | 1873 |
| 4 | No | 1873 |

**Age cohort analyses**

To test for age-based genetic differences within the simulated populations, we defined three age cohorts (*Old*, *Middle*, and *Young*) between which we tested for genetic differentiation. We examined all possible breakpoints between cohorts to determine what cohort definitions would maximize the likelihood of detecting differences. To do this, we divided the simulated population into two groups by placing a breakpoint at a given birth year and calculated $F_{ST}$ between the two groups. To control for the difference in

sample size in the two groups, we also calculated $F_{ST}$ for a sample of 100 individuals from each group. We repeated the $F_{ST}$ calculations for every possible birth year, with the constraint that there had to be at least 200 individuals in each group, and looked for breakpoints at which $F_{ST}$ was maximized. We refer to this analysis as the *Breakpoint* analysis.

We measured genetic differentiation between cohorts using a permutation $\chi^2$ test with 500 permutations (Roff and Bentzen 1989). We used Fisher's method (Ryman and Jorde 2001) to summarize *p*-values across replicates.

**Hardy Weinberg equilibrium analyses**

We used Genepop v3.3 (Raymond and Rousset 1995) to test for Hardy Weinberg equilibrium (HWE) in our simulated populations. We ran the test for heterozygote deficiency for each locus using an MCMC burn-in of 30,000 iterations and a final chain length of 10,000 with batch size 100. We calculated Hardy-Weinberg disequilibrium across all loci for a given replicate using Fisher's method (Ryman and Jorde 2001). We also used Fisher's method to summarize the *p*-values across replicates. This final application of Fisher's method amounts to assuming that each locus in each replicate is an independent test of HWE for the simulated population.

**Sampling**

Samples were drawn from the simulated populations at the end of the recovery phase. We used two different sampling schemes that varied in the degree to which sampling was age-biased. The first was the *random* scheme, in which individuals were drawn at random with respect to the age distribution of the population at the end of the simulation. This resulted in no age bias in the sample. The second was the *equal* scheme, in which we attempted to equalize the number of samples in each age cohort to the extent possible. This resulted in an extremely age-biased sample. For this scheme, samples were first drawn from the *Old* cohort, as that was always the smallest. If possible, one-third ($n/3$) of the samples were taken from this cohort. If the total abundance of this cohort was less than $n/3$, then all individuals from the cohort were sampled. The remaining samples were taken equally from the *Middle* and *Young* cohorts.

For the cohort analyses, we drew a total of $n = 150$, and $n = 500$ samples from the simulated population. In a similar model, Ripley et al. (2006) used sample sizes of approximately 700 to 1100 and found that in many cases, *p*-values for comparisons between cohorts were skewed toward large values. We chose smaller sample sizes to see whether that would result in a reduction or elimination of the skew. We did not vary sample size for the HWE analyses; they were all conducted with $n = 500$.

**RESULTS**

**Characteristics of Simulated Populations**

The 8-stage model resulted in an age distribution that is reasonable for bowhead whales (Figure 2). Slightly different stage distributions occurred at the end of burn-in for the scenarios with 300 and 1,873 whales (Figure 3) due to differences in older juvenile survival. At the end of the simulation, abundances of M and O cohorts differ between scenarios with low and high $N_{low}$. Mean and standard deviations for number of individuals in M and O from the first 100 replicates of Scenario 1 are 5,786 ± 556 and 158 ± 16.7; the same statistics for Scenario 3 are 6,471 ± 49.7 and 414 ± 19.7. In scenarios for which $N_{low} = 300$ (Scenarios 1 and 2), population size distributions at the end of the simulation were normally distributed with mean ± sd of 10,801 ± 1,034. Scenarios in which $N_{low} = 1873$ (Scenarios 3 and 4) had reached carrying capacity by the end of the simulation (N = 12,096 ± 1.02).

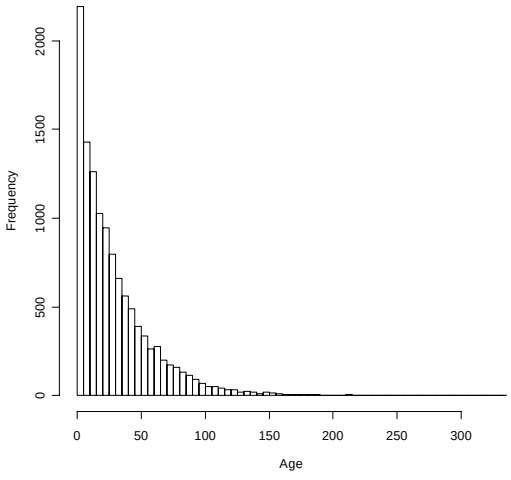


*Figure 2. Histogram of whale age just prior to commercial whaling (year 500) from a randomly selected replicate (500) of SC1.*

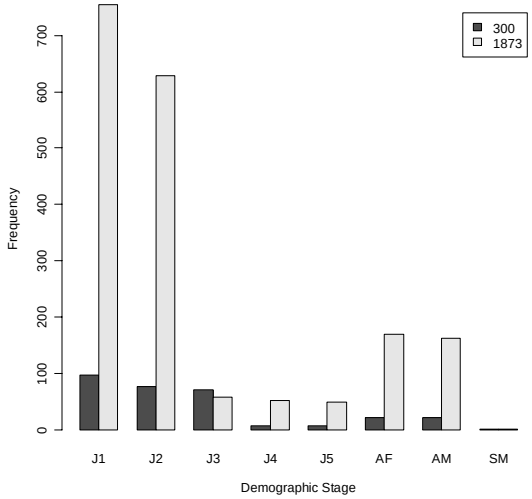


*Figure 3. Population stage structure immediately after commercial harvest, for scenarios allowing 300 or 1873 whales to survive.*

**Breakpoint analysis**

The breakpoint analysis aimed to determine breakpoints that would maximize the likelihood of detecting between cohort differences. We summarized results using two methods: 1) for each replicate we determined the breakpoint that resulted in the largest $F_{ST}$ (Figure 4), and 2) we calculated the median $F_{ST}$ across replicates for each breakpoint (Figures 5). For method 1, when all individuals on both sides of the breakpoint were included in the $F_{ST}$ calculation, there is a strong peak for the mtDNA marker around age 70 at the end of the simulation, indicating that the greatest differentiation occurs when the population is divided between animals greater than and animals less than 70 years old. The model stops in 2002, which would make this breakpoint at birth year 1932. There is also a much smaller peak at an age of approximately 10 years of age, which corresponds to a birth year of ~1992. The microsatellite markers also show peaks at around 70 years of age and 10 years of age. However, the peak at 10 years is much stronger for the microsatellites than for the mtDNA. For both marker types, peak heights are diminished when the sample size across the breakpoint is equalized, but the overall patterns are the same (not shown).

To determine whether the patterns revealed in our breakpoint analysis were strictly due to the effects of commercial harvest or would also occur in a population that is in equilibrium, we repeated the analysis on the populations at the end of the burn-in period. We found very similar patterns in the burn-in population, with highest $F_{ST}$ occurring when either the very oldest or the very youngest whales were separated from the remainder of the population. Commercial whaling simply truncated the age distribution of the population, thereby forcing the breakpoint separating the old whales to a younger age (note the different x-axis scales in figure 4).

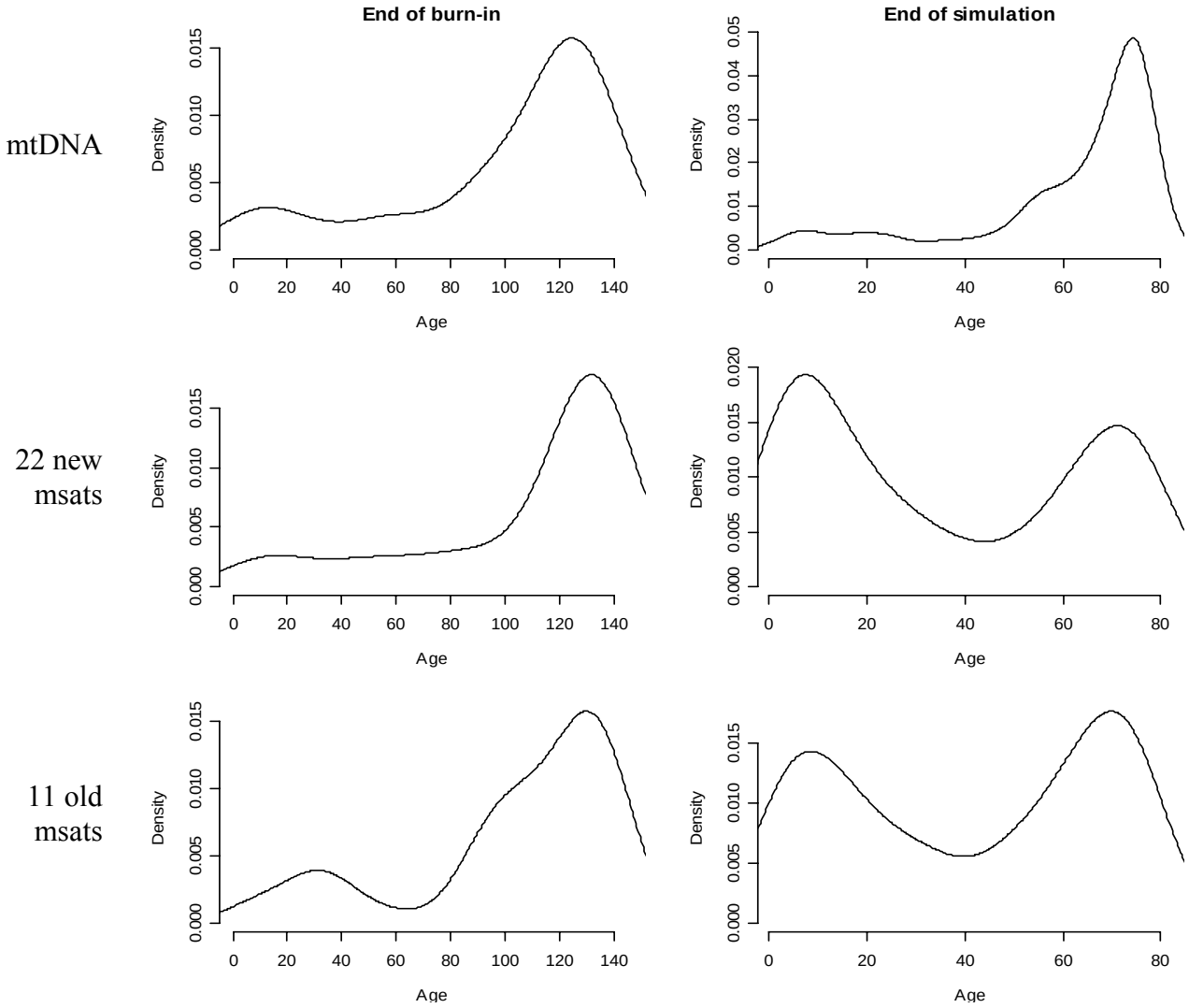


*Figure 4. The density of the distribution of maximum $F_{ST}$ as a function of the age at which samples are broken into cohorts for scenario 1 ($N_{low}$ = 300, with supermales). Results are summarized separately for the mtDNA data (top), the 22 'new' microsatellite loci (middle), and the 11 'old' microsatellite loci (bottom). Results for the population at the end of the burn-in period are on the left and at the end of the simulation on the right.*

The plots of median $F_{ST}$ (Figure 5) show a very different pattern for the two marker types. For the mtDNA, $F_{ST}$ is slightly negative until the breakpoint is about 50 years of age. As the breakpoint is moved back in time to separate older animals, $F_{ST}$ steadily increases, indicating that oldest animals are the most genetically distinct from the rest of the population. To determine whether this pattern is inherent in the life history of bowheads or is due to the effects of whaling, we repeated this analysis for the population at the end of the burn-in. The pattern in the burn-in population was quite different, with $F_{ST}$ remaining slightly negative until the breakpoint was moved to an age 110 years old or greater, at which point variance increased and $F_{ST}$ trended toward higher values. As should be the case, this pattern did not depend on the presence of supermales.

The pattern for the microsatellite markers was the opposite pattern of that seen for the mtDNA. For microsatellites, $F_{ST}$ was close to zero when the breakpoint was placed at a very young age. However, $F_{ST}$ decreased as the age at the breakpoint increased. The negative $F_{ST}$'s observed for the microsatellite markers at the end of the simulation (after the effects of commercial whaling) indicate that older animals are more genetically similar to younger ones than would be expected. Results were similar regardless of whether or not supermales were included in the simulation. We again compared these results to those observed for the population at the end of the burn-in, and found that $F_{ST}$ remains close to zero for all but the oldest cohort breakpoints, again indicating that the pattern we see in the population at the end of the simulation is due to the effects of commercial whaling. Again, the results for both marker types were consistent regardless of whether equal samples were drawn on either side of the breakpoint (shown in Figure 5) or all individuals were included in the $F_{ST}$ calculation (data not shown).

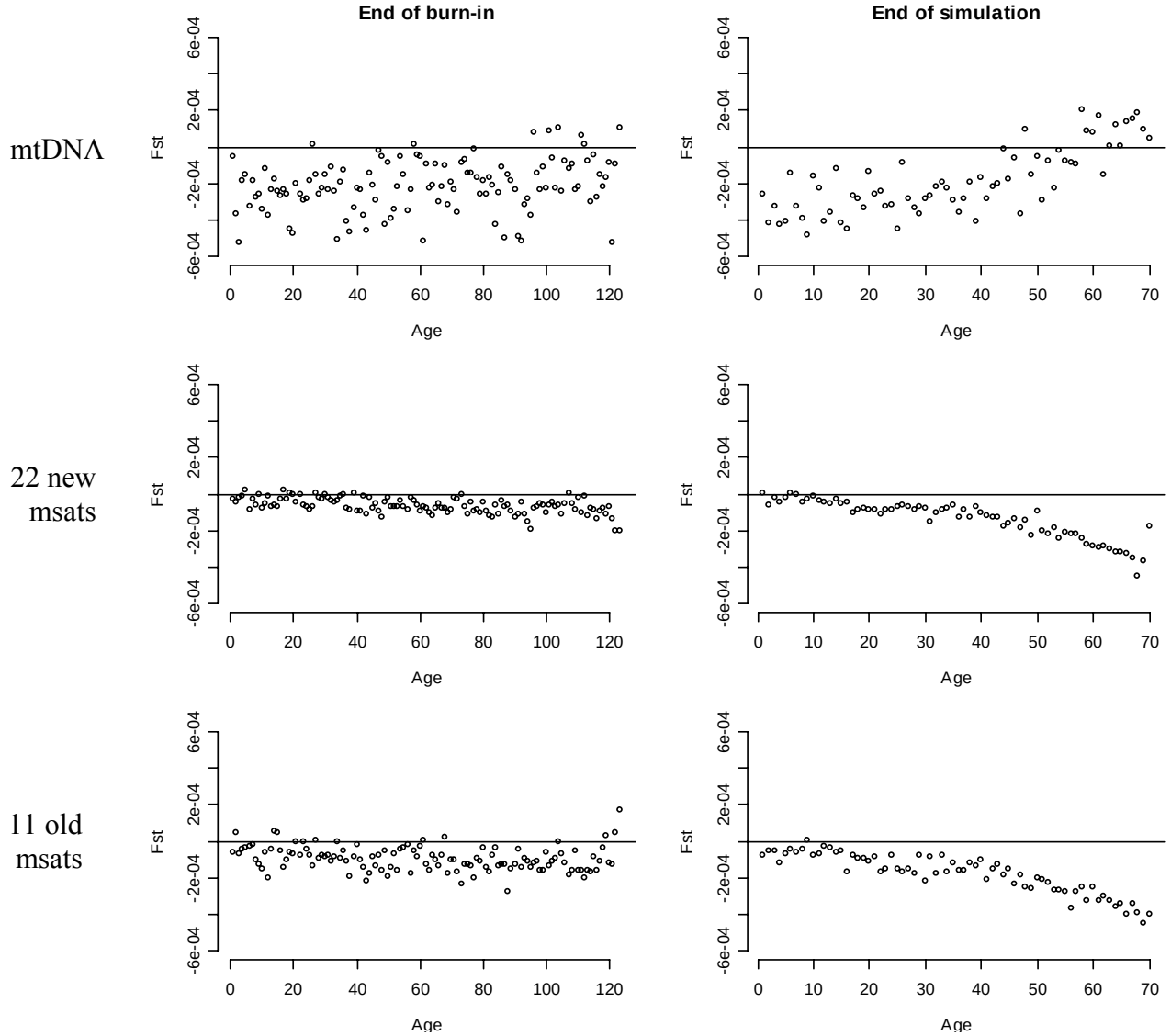


*Figure 5. The median value of $F_{ST}$ as a function of the age at which the boundary between cohorts is placed for scenario 1. The left column shows the results at the end of the burn-in epoch, while the right column shows the results at the end of the simulation. Results are summarized separately for the mitochondrial data (top row), the 22 'new' microsatellite loci (middle row) and the 11 'old' microsatellite loci (bottom row).*

**Cohort comparisons**

Based on the results of the Breakpoint analysis, we defined three cohorts: *Old* (*O*) = animals greater than 70 years old at the end of the simulation; *Middle* (*M*) = animals between 10 and 70 years old at the end of the simulation, and *Young* (*Y*) = animals less than 10 years old at the end of the simulation. Average sample sizes in the cohorts for the two different sampling schemes and sample sizes are shown in Table 3. Because the *random* scheme did not result in any animals being sampled in the *Old* cohort, cohort comparisons were only conducted for the *equal* scheme.

*Table 3. Average sample sizes in each of the age cohorts.*

| $N_{low}$ | Sampling scheme | *Young* | *Middle* | *Old* |
|---|---|---|---|---|
| 300 | *Random* | 230 | 270 | 0 |
| 300 | *Equal* | 167 | 167 | 154 |
| 1,873 | *Random* | 224 | 276 | 0 |
| 1,873 | *Equal* | 167 | 167 | 166 |

*mtDNA* – When the nadir was 300 (Scenarios 1 and 2) and sample size equaled 500, the *Old* cohort was significantly differentiated from the other two cohorts, but there was no evidence of differentiation between the *Middle* and *Young* cohorts (Table 4). We conducted one sensitivity test in which we repeated the cohort comparison for the model with a nadir of 300 and no supermales, but reduced sample size to 150. With the smaller sample size, we were unable to detect significant differentiation between any of the cohorts (Table 4).

Scenarios 3 and 4 (nadir of 1,873) produced conflicting results. Scenario 3, which included supermales, exhibited strong differentiation for all three cohort comparisons. In contrast, none of the cohorts were significantly differentiated for scenario 4, which did not include supermales. Because the mtDNA genome is maternally inherited, the difference in variance in male reproductive success between scenarios 3 and 4 was not expected to have any influence on mtDNA differentiation. Thus, results for these scenarios should be viewed as preliminary until the difference between them can be further investigated.

*Table 4. Results of $\chi^2$ comparisons of mtDNA haplotype frequency distributions between age cohorts. Significance of replicates was based on a critical value of 0.05. Fisher's method* P*-values summarize across replicates (Ryman and Jorde 2001).*

| | Sample | Proportion replicates significant | | | Fisher's method *p*-value | | |
|---|---|---|---|---|---|---|---|
| Scenario | size | *Old* v. *Middle* | *Old* v. *Young* | *Middle* v. *Young* | *Old* v. *Middle* | *Old* v. *Young* | *Middle* v. *Young* |
| 1 | 500 | 0.12 | 0.16 | 0.06 | 0.00006 | 0.00001 | 0.65 |
| 2 | 500 | 0.11 | 0.12 | 0.04 | 0.00002 | 0.00003 | 0.664 |
| 2 | 150 | 0.08 | 0.07 | 0.03 | 0.588 | 0.161 | 0.165 |
| 3 | 500 | 0.08 | 0.14 | 0.10 | 0.00068 | 0.00031 | 0.005 |
| 4 | 500 | 0.06 | 0.05 | 0.03 | 0.582 | 0.170 | 0.555 |

*Microsatellites* – There was no evidence of differentiation between cohorts at the microsatellite loci (Table 5). On the contrary, our results indicate a strong bias toward large *p*-values in the microsatellite data (Figure 6). This bias is seen in Table 5 by noting that for many comparisons, Fisher's method *p*-values are close to one. The bias is strongest in comparisons involving the *Old* cohort, for which it is present for all scenarios and sample sizes. For the comparison of *Middle* versus *Young*, the bias is less consistent. The bias does not appear to be effected by the size of the nadir. However, when we reduced the sample size to only 150, the bias toward large *p*-values was not only eliminated for the comparison of *Middle* versus *Young*, but the trend was reversed, with more replicates than expected by chance exhibiting significant differentiation between these cohorts (Figure 7). The Fisher's method *p*-value for this comparison was still not statistically significant, but was nearly so at the $\alpha = 0.10$ level (Table 5).

*Table 5. Results of $\chi^2$ comparisons of microsatellite allele frequencies between age cohorts. Significance of replicates was based on a critical value of 0.05. Fisher's method* P*-values summarize across replicates (Ryman and Jorde 2001).*

| | Sample | Proportion replicates significant | | | Fisher's method *p*-value | | |
|---|---|---|---|---|---|---|---|
| Scenario | size | *Old* v. *Middle* | *Old* v. *Young* | *Middle* v. *Young* | *Old* v. *Middle* | *Old* v. *Young* | *Middle* v. *Young* |
| 1 | 500 | 0.02 | 0.02 | 0.05 | 1.000 | 1.000 | 0.503 |
| 2 | 500 | 0.00 | 0.00 | 0.04 | 1.000 | 1.000 | 0.93 |
| 2 | 150 | 0.00 | 0.00 | 0.09 | 1.000 | 1.000 | 0.103 |
| 3 | 500 | 0.02 | 0.02 | 0.02 | 0.946 | 0.991 | 0.932 |
| 4 | 500 | 0.01 | 0.01 | 0.02 | 1.000 | 1.000 | 1.000 |

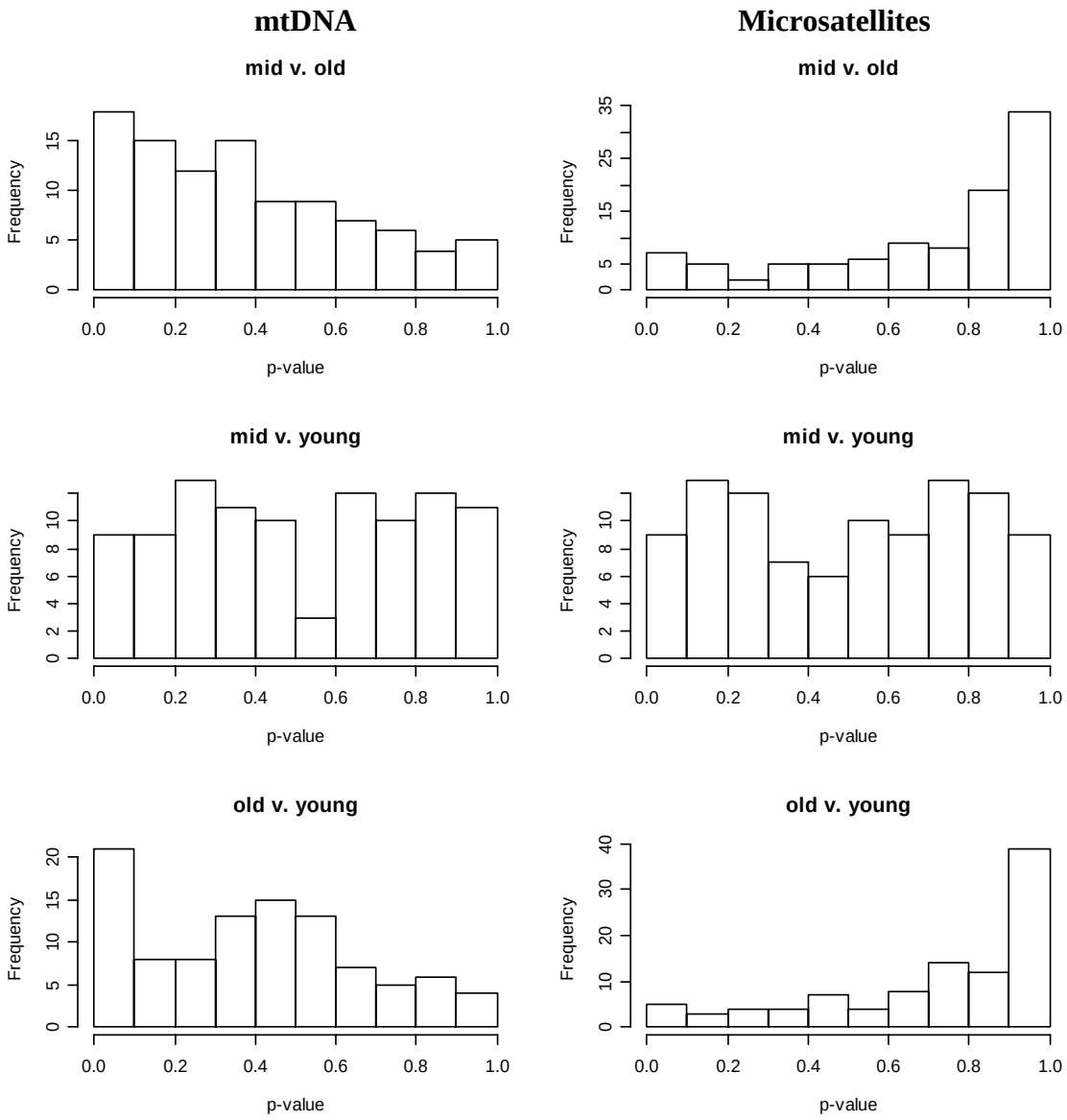


*Figure 6. Distribution of* p*-values for the three pairwise comparisons between age cohorts. Comparisons for mtDNA are shown on the left, while microsatellites are on the right. Results are for Scenario 1 with a sample size of 500.*

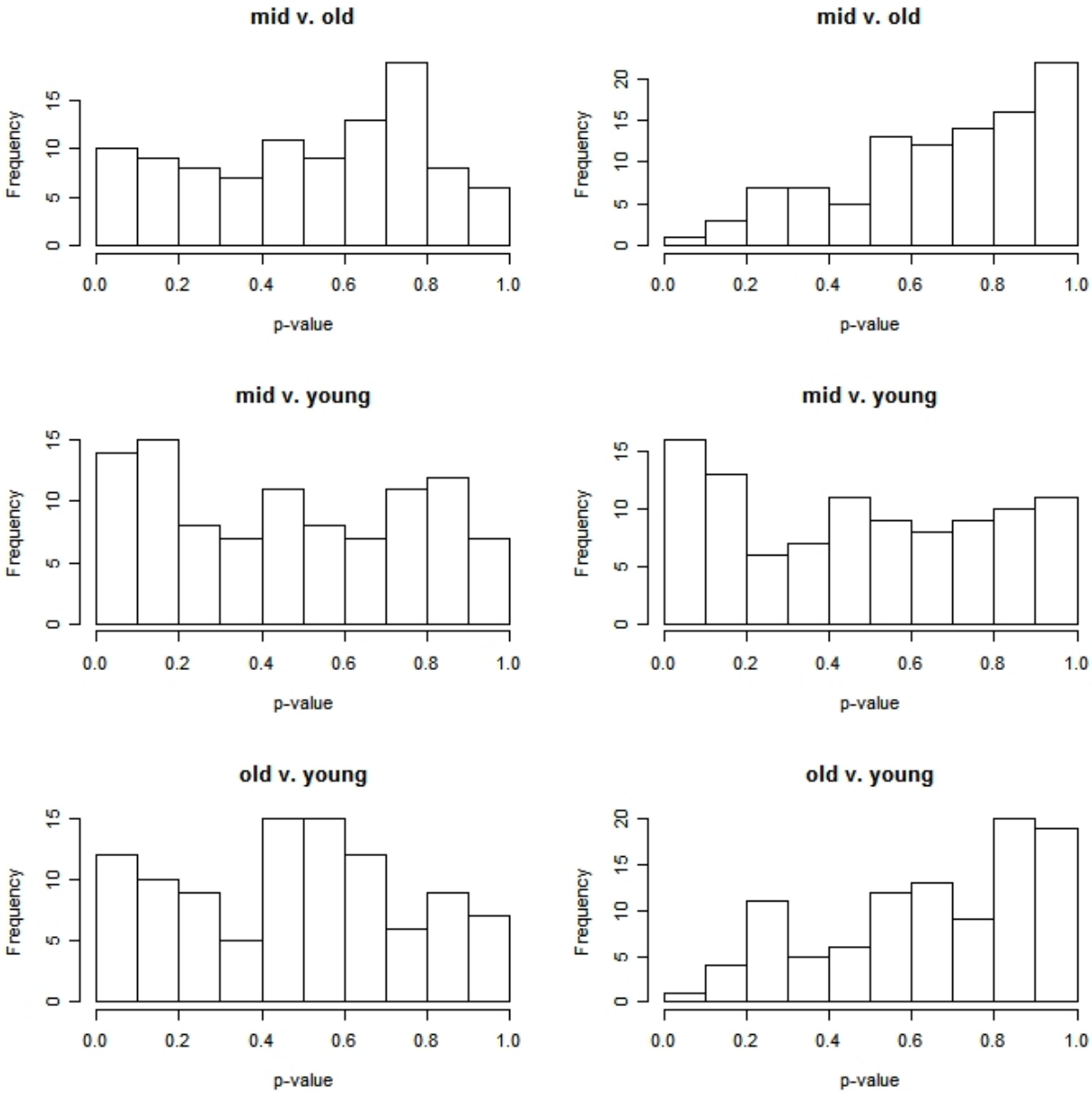


*Figure 7. Distribution of* p*-values for pairwise comparisons between age cohorts using microsatellite data. Results are for Scenario 2 with a sample size of 150.*

**Hardy Weinberg equilibrium analysis**

We tested for Hardy Weinberg equilibrium (HWE) in the simulated populations using two different sampling schemes, *random* and *equal*. For all HWE analyses, the total sample size was 500. We analyzed 100 replicates for each scenario. For scenarios with a nadir of 1,873 (scenarios 3 and 4), the results of our analyses conform to the expectations of the null hypothesis that the populations are in HWE; the proportion of loci, summarized across all replicates, for which the *p*-value was less than 0.05 was close to the expectation of 0.05 (Table 6). Similarly, the proportion of replicates for which the Fisher's method *p*-value was statistically significant was also close to 0.05.

For scenarios with a nadir of 300, however, we found that the populations were out of HWE less often than would be expected by chance when we used the *equal* sampling scheme (Table 6). This bias was absent for the *random* sampling scheme (which did not include any individuals from the *Old* cohort; see above

Table 6. *A summary of the results of the HWE analyses for the four scenarios and three sampling schemes. The 'Overall p-value' represent's Fisher's method* p-*value, and was calculated by assuming that each locus in each replicate constituted an independent test of HWE. 'Loci out of HWE' represents the proportion of loci across all replicates for which the HWE p-value calculated by Genepop was less than or equal to 0.05. 'Replicates out of HWE' shows the proportion of replicates for which the Fisher's method p-value for all loci combined was less than 0.05. If the null hypothesis that the data conformed to HWE expectations were true, we would expect the values in the last two columns to equal 0.05, and the values in the middle column ('Overall Fisher's exact p-value) to be random draws from a uniform distribution.*

| Scenario | Sampling scheme | Overall *p*-value | Loci out of HWE | Replicates out of HWE |
|---|---|---|---|---|
| 1 | Random | 0.847 | 0.047 | 0.047 |
| | Equal | 1.000 | 0.0373 | 0.01 |
| 2 | Random | 0.933 | 0.043 | 0.043 |
| | Equal | 1.000 | 0.0352 | 0.02 |
| 3 | Random | 0.78 | 0.050 | 0.050 |
| | Equal | 0.596 | 0.0503 | 0.05 |
| 4 | Random | 0.147 | 0.056 | 0.056 |
| | Equal | 0.844 | 0.0482 | 0.07 |

**DISCUSSION**

The goal of the analyses presented herein was to explore the impact on genetic equilibrium of several factors that could not be varied in the model presented by Archer et al. (2007). Specifically, we wanted to see whether a reduction in minimum population size and increase in variance in male reproductive success would cause a greater degree of genetic disequilibrium within the population, as evinced by stronger genetic differentiation between age cohorts or a higher incidence of Hardy Weinberg disequilibrium. For the most part, we did not observe any such effects, indicating that the conclusions presented by Archer et al. (2007) would not have been substantially different had they been able to simulate a lower minimum population size or higher variance reproductive success.

Though our model does not effect the conclusions of Archer et al. (2007), it did produce some interesting results that warrant further investigation. The cohort comparisons for scenarios 1 and 2 revealed significant differentiation in mtDNA haplotype frequencies between the *Old* cohort and the two younger cohorts (Table 4). Ripley et al. (2006) found a similar result with an earlier version of the model we presented here. They found that the differentiation in haplotype frequencies was due to the fact that some haplotypes in their oldest cohort were present only in males. Since males do not pass on their mtDNA when they reproduce, these represented 'ghost' haplotypes that were doomed to extinction when the males carrying them died. When Ripley et al. excluded males from their oldest cohort, they found that the genetic differentiation between it and the other cohorts was greatly diminished.

The cohort comparisons using microsatellite markers resulted in biased *p*-value distributions – large *p*-values were observed more often than expected by chance alone. This bias likely results from the fact that our strata do not represent random samples from the population, as is assumed under the null hypothesis; rather, the fact that we have stratified by age means that in any pairwise comparison, the older cohort likely contains many of the parents and grandparents of individuals in the younger cohort. This hypothesis is supported by the breakpoint analysis, which showed that median $F_{ST}$ became increasingly negative as the breakpoint between cohorts was moved to older ages. Negative $F_{ST}$ indicate that the groups being compared are more similar than would be expected by chance.

Contrary to our expectations, neither the effects of commercial whaling nor variance in male reproductive success caused significant Hardy Weinberg disequilibrium in our simulated populations. The size of the nadir did have an effect, but it was opposite to what we expected: a smaller nadir reduced the incidence of disequlibrium to below what would be expected by chance alone. In other words, the *p*-value distributions for the HWE analyses were biased in the same manner as for the cohort comparisons. We believe this is again due to the inclusion of an unusually large number of close relatives in our samples. This conclusion is supported by the fact that the skew is affected by sampling scheme; the scheme that results in the largest number of *Old* animals (the *equal* scheme) exhibits the strongest skew. The skew is absent in the *random* scheme, in which no *Old* animals were sampled.

**ACKNOWLEDGEMENTS**

Funding was provided by the US National Marine Fisheries Service. Thanks to our steering committee: Judy Zeh, Bill Koski, Paul Wade, Andre Punt, Tore Schweder. Thanks to the following for critical discussions and advice: Craig George, Rick LeDuc, Phil Morin, and Allan Strand

**REFERENCES**

Archer, E., Martien, K.K. And Taylor, B.L. 2007. Use of an individual-based simulation of BCB bowhead whale population dynamics to examine empirical genetic data. SC/59/BRG17.

Bockstoce, J.R. 1986. Whales, Ice and Men. University of Washington Press. Seattle, WA. 400 pp.

Brandon, J. R. and P. R. Wade 2006. Assessment of the Bering-Chukchi-Beaufort Seas stock of bowhead whales using Bayesian model averaging. J. Cetacean Res. Manage. 8:225-239.

Brownell, R.L. and Ralls, K. 1986. Potential for sperm competition in baleen whales. Rep. Intl. Whal. Commn. Si8:97-112.

Caswell, H. (2001) Matrix Population Models: Construction, Analysis and Interpretation. 2nded. Sinauer Associates, Sunderland, Massachusetts, USA.

Excoffier, L. , G. Laval, and S. Schneider (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online (submitted).

Falush, D., M. Stephens, and J.K. Pritchard. 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics 164:1567-1587.

George, J.C., Bockstoce, J.R., Punt, A.E. and Botkin, D.B. 2007. Preliminary estimates of bowhead whale body mass and length from Yankee commercial oil yield records. 11pp. SC/59/BRG5.

Givens, G., Bickham, J.W., Matson, C.W., Ozaksoy, I., Suydam, R.S., and George, J.C. 2004. Examination of Bering-Chukchi-Beaufort Seas bowhead whale stock structure hypotheses using microsatellite data. SC/56/BRG17.

Givens, G., Huebinger, R.M., Bickham, J.W., George, J.C. and Suydam, R.S. 2007. Patterns of genetic differentiation in bowhead whales (*Balaena mysticetus*) from the western Arctic. SC/59/BRG14.

Jorde, P.E., Schweder, T., Bickham, J.W., Givens, G.H., Suydam, R., and Stenseth, N.C. 2006. Detecting genetic structure in migrating bowhead whales off the coast of Barrow, Alaska. In Prep. for Molecular Ecology.

Jorde, P.E. And Schweder, T. 2007. Further analysis of stock structure for BCB bowhead whales using microsatellite DNA data. SC/59/BRG27.

Kitakado, T., Pastene, L.A., Goto, M., and Kanda, N. 2007. Updates of stock structure analyses of B-C-B stock of bowhead whales using microsatellites. SC/59/BRG30.

Leduc, R.G., Martien, K.K., Morin, P.A., Hedrick, N., Robertson, K.M., Taylor, B.L., Mugue, N.S., Borodin, R.G., Zelenina, D.A. and George, J.C. Mitochondrial genetic variation in bowhead whales in the western Arctic. 11pp. SC/59/BRG9.

O Hara, T. M. George, J. C. Tarpley, R. J. Burek, K. Suydam, R. S. (2002) Sexual maturation in male bowhead whales (*Balaena mysticetus*) Journal of Cetacean Research and Management 4(2): 143-148.

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

Roff, D. A. and P. Bentzen. 1989. The statistical analysis of mitochondrial DNA polymorphisms: $\chi^2$ and the problem of small samples. Mol. Biol. Evol. 6:539-545.

Rosenbaum, H. C., M. T. Weinrich, S. A. Stoleson, J. P. Gibbs, C. S. Baker, and R. DeSalle. (2002) The effect of differential reproductive success on population genetic structure: correlations of life history with matrilines in humpback whales of the Gulf of Maine. Journal of Heredity 93(6): 389-399.

Strand, A. (2002) METASIM 1.0: an individual-based environment for simulating population genetics of complex population dynamics. Molecular Ecology Notes 2(3): 373-376.

Suydam, R.S. and J.C. George. 2006. Subsistence harvest of bowhead whales (Balaina mysticetus) by Alaskan Eskimos, 1074-2003. SC/56/BRG12.

Taylor, B.L., S. J. Chivers, S. Sexton and A. E. Dizon. 2000. Estimating dispersal rates using mitochondrial DNA data and incorporating uncertainty. Conservation Biology:1287-1297.

Waples, R. 1989. Temporal variation in allele frequencies: testing the right hypothesis. Evolution 43:1236-1251.