A simulation-based approach to evaluating population structure in non-equilibrial populations

FREDERICK I. ARCHER*[1], KAREN K. MARTIEN[1], BARBARA L. TAYLOR[1], RICHARD G. LEDUC[1], BONNIE J. RIPLEY[2], GEOF H. GIVENS[3], AND JOHN C. GEORGE[4]

**Contact e-mail: eric.archer@noaa.gov*

ABSTRACT

The standard null model of panmixia used to test for population subdivision is based on a set of assumptions that can be violated given recent events likely to result in non-equilibrial genetic composition coupled with the complex life histories of many species. Bowhead whales (*Balaena mysticetus*) represent such a species. Bowhead whales also have a well-documented history of severe commercial harvest in the recent past which would be expected to leave a population out of genetic equilibrium. They also have a very long life span, overlapping generations, and age and sex-structured migrations. In addition, samples come from whales killed in a hunt known to be non-random with respect to size at different whaling villages. Sampling of such a population could lead to erroneous conclusions regarding population structure, which could have real consequences for aboriginal whaling. To better interpret the results of standard population genetic analyses, an individual-based model of bowhead whale population dynamics and genetics was created using the R package *rmetasim*. The model re-created as closely as possible all aspects of the demography, genetics, and whaling history of bowhead whales. Simulated datasets were generated by sampling from the simulated population in a way that matched the age, sex and geographic distribution of empirically collected samples. The empirical bowhead datasets were compared to null distributions generated from the simulated datasets for a variety of genetic analyses. The analysis indicates that the empirical genetic data sampled from the Bering-Chukchi-Beaufort (BCB) stock of bowhead whales are more consistent with the model of a population with the same whaling history and demographic composition as BCB whales than they are with a single, randomly-mating population in genetic equilibrium under a standard Wright-Fisher model. Additionally, it was demonstrated that by failing to account for the unique features of the population dynamics of the species, standard tests of genetic differentiation based on panmixia may produce misleading results. The approach outlined will likely prove useful for evaluating population structure in other species likely to be out of equilibrium.

KEYWORDS: WHALING – ABORIGINAL; WHALING – HISTORICAL; ARCTIC; MANAGEMENT; MODELLING; AGE DISTRIBUTION; GENETICS; BOWHEAD WHALE; BERING SEA; CHUKCHI SEA; BEAUFORT SEA; NORTHERN HEMISPHERE

INTRODUCTION

The Wright-Fisher model of population genetics (Fisher, 1930; Wright, 1931) forms the basis of the null hypotheses describing no population structure for most genetic analyses. Under this model, the hypothetical single population is assumed to be at genetic equilibrium (the rate of genetic drift equals that of mutation), is panmictic (every individual has an equal chance of mating with every other individual), has non-overlapping generations, and experiences no immigration or emigration. While not an explicit assumption of the Wright-Fisher model, most studies also assume that samples have been collected at random and thus accurately represent the genetic frequency distribution in the population at large.

In most real populations, one or more of these assumptions are often not met, potentially leading to problems in interpreting the results of standard genetic tests. Bowhead whales (*Balaena mysticetus*) represent a good example of such a population. In the late 19th and early 20th centuries, bowhead numbers were very rapidly reduced by whaling followed by a recovery in only two and a half generations (Bockstoce, 1986; Bockstoce and Burns, 1993; Brandon and Wade, 2006), guaranteeing the population or populations to be strongly out of genetic equilibrium. Sampling is also not random. Most samples are from animals killed by Alaskan native subsistence hunters, with some villages preferring to kill large (and hence older) whales, while others prefer smaller (younger) whales (Noongwook *et al.*, 2007; Suydam and George, 2004). Further, hunting primarily occurs during migration and often in short time periods, and whales are known to segregate by size and reproductive condition during migration (Angliss *et al.*, 1995). Thus, samples of bowhead whales stratified by sampling location could represent different demographic components of the population.

Results of previous analyses of bowhead whales migrating past Barrow, Alaska have been interpreted to suggest that more than one stock may exist in the Beaufort-Chukchi-Bering (BCB) Sea (Jorde *et al.*, 2007). However, these results could also reflect age stratification of migration. Bowhead whales can live for over 100 years (George *et al.*, 1999), thus it is likely that some whales that were born prior to the end of commercial whaling are still alive today. Their genes represent frequencies of the unexploited population, while those of recent cohorts represent the smaller, yet still diverse, gene pool that survived commercial whaling. These differences between genetic frequencies of cohorts resulting from non-equilibrial dynamics are called the Generational Gene Shift (GGS) hypothesis. In addition to GGS, it has been demonstrated that both non-random sampling as well as sampling from age-structured populations can lead to results

[1] *Southwest Fisheries Science Center, 8604 La Jolla Shores Blvd., La Jolla, CA 92038 USA.*
[2] *Grossmont College, 8800 Grossmont College Dr., El Cajon, CA 92020 USA.*
[3] *Dept. of Statistics, 1877 Campus Delivery, Colorado State University, Fort Collins, CO 80523 USA.*
[4] *North Slope Borough Dept. of Wildlife Management, PO Box 69, Barrow, AK 99723 USA.*

that differ significantly from those expected from a panmictic population (Waples, 1990; Waples and Teel, 1990; Waples and Yokota, 2007). It is possible that all of these factors are present in the bowhead whale data.

In this paper, an individual-based simulation is described that attempts to both capture the population dynamics that lead to GGS and non-equilibrial genetic samples, and match the non-random empirical samples as closely as possible with respect to birth year and sex. The aim is to generate an alternative to the standard null distribution that will allow testing for population structure without assuming equilibrium conditions, panmixia, or random sampling. The analyses conducted include evaluations of Hardy-Weinberg equilibrium and measures of population structure (F_{st}, χ^2, and Φ_{st}) using strata considered possible for bowhead whales. A single population was simulated and evaluated the likelihood of obtaining observed results without invoking a multiple-stock scenario. Issues with applying the approach to multiple stocks are addressed in the discussion.

METHODS

The simulation is based on the *rmetasim* package (version 1.1.008 – Strand, 2002), run in the R statistical environment (version 2.4.1 – R Development Core Team, 2006). *Rmetasim* is a library of functions that performs individual-based population genetic simulations. Each individual has a multi-locus genotype and a mitochondrial DNA (mtDNA) haplotype. Individuals are structured demographically with a stage-based matrix population model (see 'Demography' section below; Caswell, 2001). At each time step individuals are randomly assigned their 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, independently segregating alleles, and neutrality of markers. For all parameters not explicitly defined here the program default values were used.

During the simulation, a set of individuals are selected to mimic the 1,099 BCB bowhead whales in the recorded harvest between 1937 and 2006 (Braund *et al.*, 1988; Suydam and George, 2004). While the analysis of Braund *et al.* (1988) cautioned that 'these data represent minimum numbers' of landed whales and some landed whales likely went unrecorded, harvest numbers were relatively low in the 1930s-60s and in many years fewer than 10 were taken, so the overall number of missed harvests through this period should be very small. Harvests increased significantly in the 1970s, but it is unlikely that any landed whales went unrecorded, due to careful monitoring by NOAA; however a few struck but lost whales may not have been reported during this period. Potential mortalities from these events were not simulated.

During the last 25 years (and predominantly during the last decade), tissue samples were collected both from whales killed during Alaskan subsistence hunts as well as from biopsies of live whales (O'Hara *et al.*, 1998; Suydam and George, 2004). As noted above, this sampling was not random, due to hunting preferences, biopsy opportunities, and important aspects of bowhead whale migratory behaviour and distribution. A variety of life history data was also collected from many of the whales killed during the hunts, some of which were used to characterise the demographic composition of the simulated population as described below.

Demography

Rmetasim version 1.1.008 incorporates density dependent population growth, as described in Martien (2006). Density dependence is implemented by interpolating between matrices that represent survival and reproduction rates at carrying capacity and near zero population density. Although this version of *rmetasim* only allows for linear interpolation between these matrices, the program was modified to allow for non-linear density dependence. The value of a given element (the probability of transition between stages) of the life history matrix in year t is given by:

$$x_t = x_0 + (x_{\max} - x_0)\left(1 - \left(\frac{N_t}{K}\right)^z\right)$$

where:

x_t is the value of the element in year t
x_0 is the value of the element at carrying capacity
$x_{\max}$ is the maximum value of the element (near zero population size)
N_t is the size of the population at the start of year t
K is the carrying capacity of the population
z is the shape parameter.

The demographic matrices used for this study are for a stage-based model with the following 7 stages: 5 juvenile stages (J1–J5), adult females (F), and adult males (M) (Ripley *et al.*, 2006). Juvenile stages were added to assure that individuals did not remain or advance through being juveniles in a biologically unrealistic way while allowing the model to avoid having a different stage for all 120 ages (most of which have identical probabilities of birth and death). Stage transition probabilities were calculated using the fixed stage duration method (Caswell, 2001). The life history parameter estimates presented in Brandon and Wade (2006) were used to develop two matrices, one for which the intrinsic population growth rate $\lambda = 1.00$, the other for which $\lambda = 1.042$ (Table 1). These matrices were used to represent vital rates (age at sexual maturity, juvenile and adult survival, timing of transition from juvenile to adult survival rates, and

Table 1

Demographic parameters at carrying capacity (λ=1.00) and near zero population size (λ=1.042). For each stage, stage duration (T) and age-specific survival (σ) are used to calculate the matrix model parameters P (survival in stage) and G (stage transition probability) according to the fixed stage duration model (Caswell 2001; Ripley *et al.* 2006).

	λ=1.00				λ=1.042			
Stage	T	σ	P	G	T	σ	P	G
J 1	4	0.800	0.661	0.139	2	0.925	0.490	0.435
J 2	4	0.978	0.741	0.236	3	0.985	0.675	0.310
J 3	4	0.978	0.741	0.236	3	0.985	0.675	0.310
J 4	4	0.978	0.741	0.236	3	0.985	0.675	0.310
J 5	4	0.978	0.741	0.118	3	0.985	0.675	0.155
F	50	0.978	0.967	0.011	50	0.985	0.981	0.004
M	50	0.978	0.967	0.011	50	0.985	0.981	0.004

reproductive rates) at carrying capacity and near zero population size, respectively. z was set to 4, which is the posterior median from Brandon and Wade's (2006) backward projection model (referred to as '1848DD' in their paper).

Genetic initialisation and burn-in

The simulated populations were initialised using mitochondrial haplotype and microsatellite allele frequency distributions generated by the coalescent program *SIMCOAL* v2.1.2 (Laval and Excoffier, 2004). Initialising from a coalescent rather than with random allele and haplotype frequencies greatly reduced the number of generations required for the simulations to reach equilibrium (Martien *et al.*, 2009). In order to initialise *SIMCOAL*, the average effective population size (Ne) at carrying-capacity was estimated using *rmetasim*.

The relationship between a change in heterozygosity and effective population size is given by,

$$\mathrm{E}\left[\frac{H_t}{H_0}\right] = \left(1 - \frac{1}{2N_e}\right)^{\frac{t}{g}}$$

where:

H_0 = initial heterozygosity
H_t = heterozygosity at time t
t = elapsed time in years
g = generation time

After rearranging the above and including a multiplier *(m)* chosen to start the simulation burn-in phase close to equilibrium, N_e was estimated as,

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

A generation time (g) of 37 years was used, representing the average age of reproductive females (Taylor *et al.*, 2007). A value for m of 1.45 was empirically found to be satisfactory and was used in all simulations. N_e was estimated for mtDNA and microsatellites separately. For microsatellites, the above equation actually estimates $2N_e$, which is the value required by *SIMCOAL*. The average effective population size ($\tilde{N}_e$) used to initiate *SIMCOAL* was the harmonic mean of N_e from 20 *rmetasim* population projections at carrying capacity, each initialised with the same survival and reproduction matrices as in the full simulation, and lasting 4,000 years (t). The sample size generated by SIMCOAL was $\tilde{N}_e$ for the mtDNA sequences and the smaller of $\tilde{N}_e$ and 1,000 for the microsatellite loci.

Parameters for both mtDNA sequences and microsatellite loci were set to mimic the empirical data as closely as possible. The mtDNA sequence was specified to be 397bp, with a Ts:Tv of 10:1, and a mutation rate of 9.4×10^{-3} (LeDuc *et al.*, 2005; LeDuc *et al.*, 2008). For the microsatellites, two groups of loci were simulated representing 11 'original' loci which were typed from a variety of different cetaceans including bowhead whales, and 22 'new' loci which were screened from a CA-enriched library using bowhead whale samples. A detailed description of the development of these two sets of loci can be found in Givens *et al.* (2010) and Huebinger *et al.* (2006). Average mutation rates were set at 3.0×10^{-4} and 1.5×10^{-3} for the original and new loci respectively. Mutation parameters were tuned to produce diversity comparable to that observed in the empirical bowhead dataset, as has been done previously (Taylor *et al.*, 2000). The same mutation rates that were used in *rmetasim* were used in the *SIMCOAL* initialisations.

In order to ensure that the simulated populations were in equilibrium, a burn-in phase was conducted following initialisation. 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 (Fig. 1). A sample of all markers was independently generated from *SIMCOAL* for each burn-in replicate.

Simulated whaling and sampling

For each burn-in replicate, multiple replicates of simulated harvest of whales designed to mimic the historical kill were conducted. The historical kill encompasses the commercial harvest and Russian and Alaskan subsistence catches from 1848 to 2006. The harvest data used in the model are the same data being used in the IWC Aboriginal Whaling Management Plan (IWC, 2003). In each year of a whaling replication, the first whales removed from the simulated populations were those included in the empirical genetic dataset. These consisted of whales for which biological samples and measurements were collected from the Alaskan subsistence catch (available from 1974 to 2006). For each sampled whale in the empirical dataset, an individual of the same age and gender was randomly selected from the simulated population. If the gender of the sampled whale was unknown, then it was randomly selected using the ratio of known-gender whales killed in that year. A 50:50 ratio was assumed if no known-gender whales were available in a particular year (empirical sex ratio for all whales harvested from 1974–2006 for which sex was identified was 487(F):468 (M), i.e. very close to parity).

In order to match the age of simulated individuals to those of harvested whales, the age of each harvested whale was determined in a hierarchical fashion based on the quality of data available. Many of the whales included in the empirical dataset were aged and had estimates of standard errors from one of the methods given in Lubetkin and Zeh (2006). For these whales, ages were randomly sampled from a normal distribution and rounded to the nearest whole age.

For whales that were not aged, a Classification and Regression Tree (CART – Breiman *et al.*, 1984), as implemented in the R package *rpart* v3.1–34 (Therneau and Atkinson, 2006), was used to estimate the age bin to which they belonged based on morphological characteristics (gender, body length, baleen length, anterior flipper length, peduncle girth, and length of the peduncle white patch). The CART tree was created from 177 known-aged samples using ten age bins (Fig. 2), which were selected from an exploratory series of CART regression trees. Bins were selected probabilistically based on the distribution of bin membership from the training data in the node to which an unknown sample was assigned (Table 2). The age for each sample was then chosen at random from all individuals in the simulated population within the chosen age bin. If the

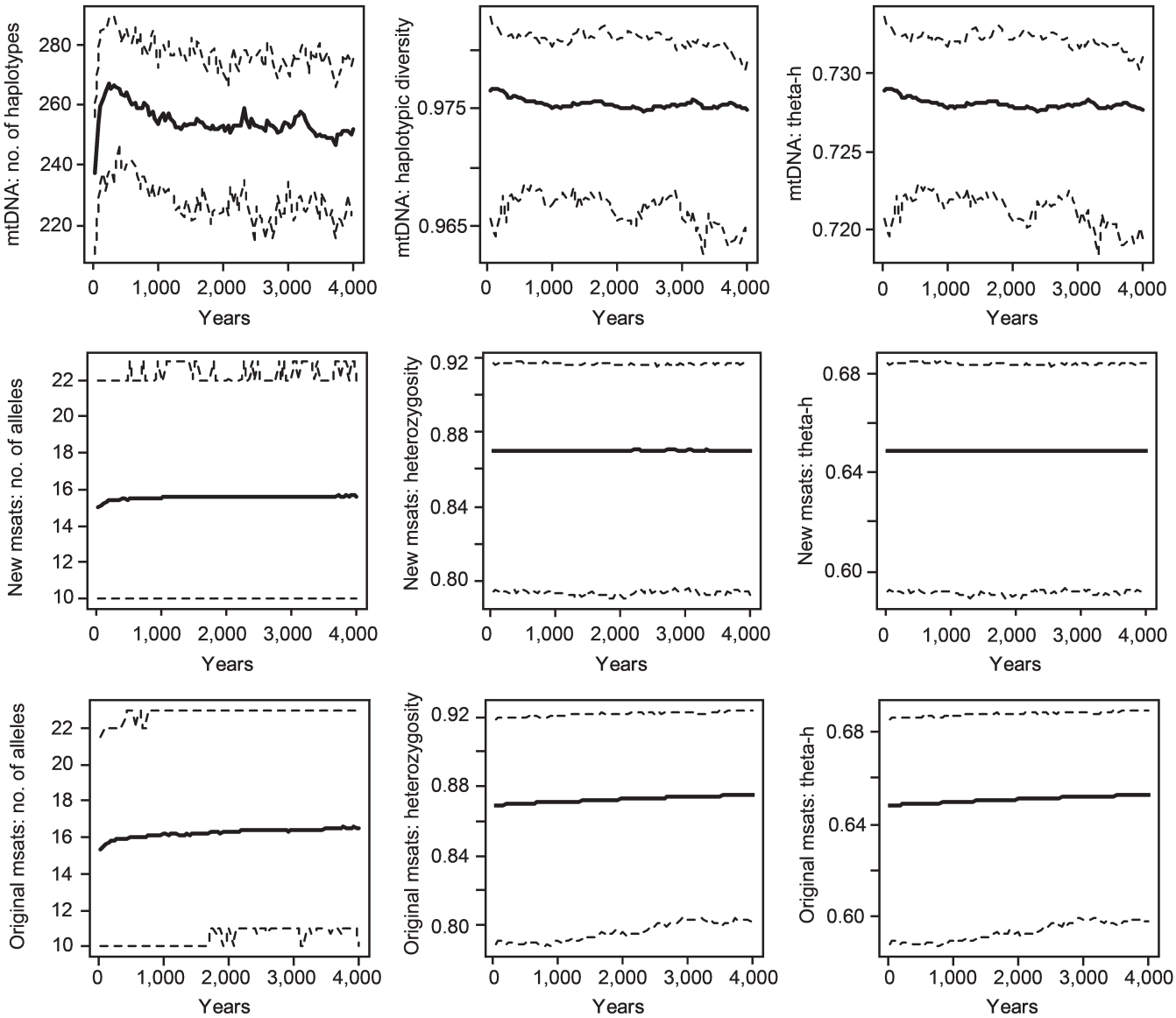

Fig. 1. Mean values (solid lines) and 95-percentiles (dashed lines) for number of haplotypes (mtDNA) and alleles (microsatellites), heterozygosity, and Theta (θ)-h during 50 burn-ins.

morphometric data necessary to classify a sample at a particular split in the CART tree was missing, surrogate variables were used if available. If there was insufficient morphometric data for the CART algorithm, then an age was chosen at random from the simulated population. In all cases, the age distribution being chosen from was that of the simulated population immediately following burn-in, which was considered a stable age distribution.

In some cases, no individuals in the simulated population were found to match the age and gender of a harvested whale exactly. When this occurred, all individuals within a progressively increasing age window around the whale under consideration were examined. In each age window, probabilities were assigned to each individual based on their gender and the size of the window. The probability of choosing an individual of the same sex as the sample under consideration ranged from one for an age window of zero (all simulated individuals were of the same age as the sampled whale) to 0.5 when all individuals in the population were considered. The probability of choosing an individual of the opposite sex was one minus this value. If no individuals in the window had probabilities greater than a randomly chosen value, then the age window was increased and the new set of individuals were reconsidered. In this manner, all simulated individuals were matched to a unique sampled whale.

Following the removal of any biologically sampled whales, the un-sampled portion of the recorded catch in that year was then removed from the simulated populations. In all cases, whaling was restricted to individuals older than one year. The genetic data of the simulated whales selected to be killed each year were saved if genetic data were available for their matched harvested counterparts. If biopsies were collected in a given year, the genetic data of an equivalent number of randomly selected simulated individuals still alive in the population were also saved. Following a simulated year of whaling and sampling, the population was projected forward one year and the whaling for the next year would occur as described above.

In order to ensure that the abundance trajectories from the simulations were similar to those of historical trend analyses (Brandon and Wade, 2006), two abundance ‘gates’ were

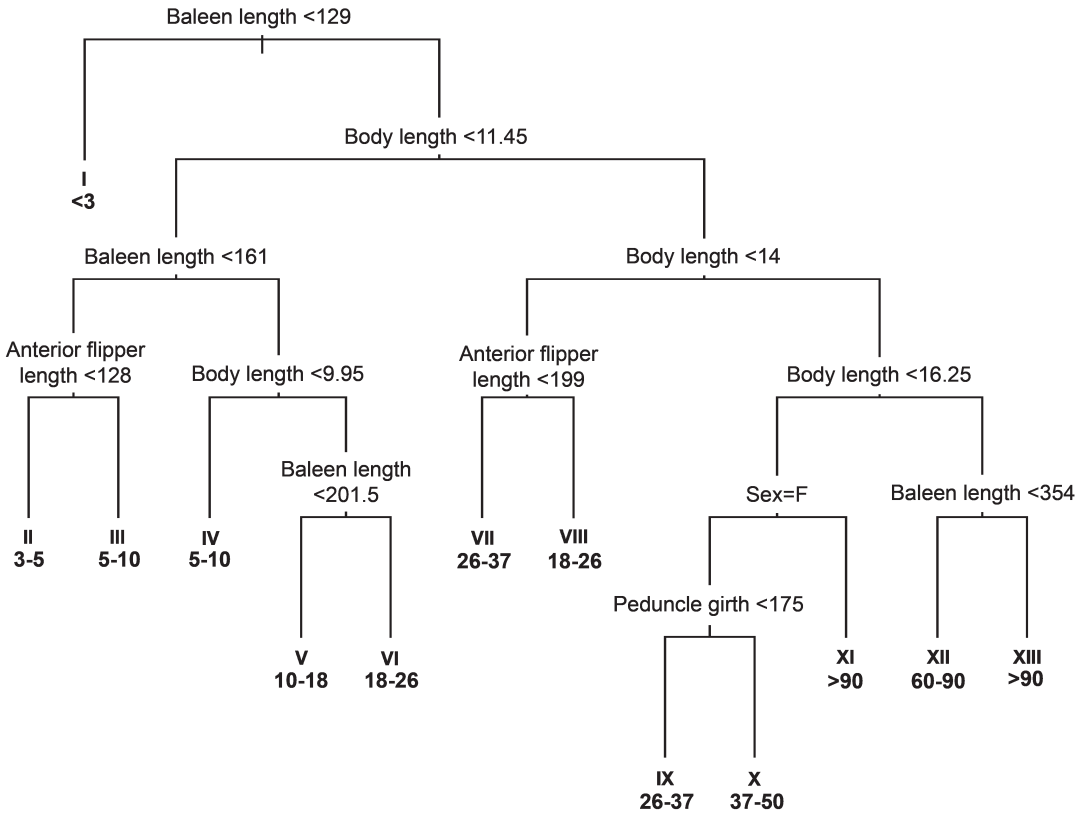

Fig. 2. CART tree with primary splits used for age estimation. Cases meeting the criteria at a node are sent down the left. Roman numerals are leaf identifiers corresponding to rows in Table 2. Values below leaf identifiers are estimated age bin of leaf.

established that replicates had to pass through. Replicates that had trajectories outside of the 99% confidence intervals of the first and last years of estimated population abundance (1978 and 2001), including those that went extinct, were discarded. For each of 50 burn-in replicates, the first ten successful whaling replicates were saved, producing a total of 500 replicates. The final output of each replicate was a simulated genetic sample representing the demographic composition of the empirical harvest sample and all individuals surviving in each of the simulated populations. Annual population abundances were saved for comparison with trajectories from historical trend analyses (Brandon and Wade, 2006).

Introduction of errors

Microsatellite datasets inevitably contain genotyping errors. Error rates reported in the literature range from 0.1% to 48% (Morin *et al.*, 2009). To examine the effect of genotyping errors on the analytical methods applied to the bowhead whale dataset, genotyping errors were introduced into the simulated dataset. By comparing genotypes for duplicate samples included in the original empirical dataset, Morin *et al.* (2009), estimated an overall error rate of 0.01 for the bowhead whale microsatellite data. Of these, 40% were apparent cases of allelic dropout, i.e. the individuals were scored as homozygotes in one genotyping attempt and as heterozygotes in another.

Table 2

Probability of assignment to age bins for leaves of the CART tree in Fig. 2. Bins are inclusive of the lower boundary.

	Age bin									
Leaf	< 3	3-5	5-10	10-18	18-26	26-37	37-50	50-60	60-90	≥ 90
I	0.84	0.11	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00
II	0.00	0.86	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00
III	0.21	0.21	0.36	0.21	0.00	0.00	0.00	0.00	0.00	0.00
IV	0.00	0.06	0.88	0.00	0.00	0.00	0.06	0.00	0.00	0.00
V	0.00	0.00	0.36	0.57	0.00	0.00	0.07	0.00	0.00	0.00
VI	0.00	0.00	0.33	0.00	0.50	0.17	0.00	0.00	0.00	0.00
VII	0.00	0.00	0.00	0.33	0.17	0.50	0.00	0.00	0.00	0.00
VIII	0.00	0.00	0.00	0.13	0.53	0.27	0.00	0.00	0.07	0.00
IX	0.00	0.00	0.00	0.00	0.00	0.53	0.29	0.18	0.00	0.00
X	0.00	0.00	0.00	0.00	0.00	0.14	0.43	0.29	0.00	0.14
XI	0.00	0.00	0.00	0.11	0.00	0.00	0.22	0.11	0.22	0.33
XII	0.00	0.00	0.00	0.00	0.00	0.11	0.11	0.11	0.67	0.00
XIII	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.17	0.17	0.50

The number of genotyping errors introduced into a simulated dataset was determined by drawing a random deviate from a binomial distribution given the overall error rate (0.01) and the number of alleles in the dataset (18,314). The alleles to which the errors were applied were chosen at random from the entire dataset. When an error occurred, it had a 0.4 probability of being an instance of allelic dropout, in which case the allele in question was set equal to the other allele the individual possessed at that locus, making the individual homozygous at that locus. Otherwise, the allele was replaced by a different allele chosen at random from the allele frequency distribution for the appropriate locus.

Genetic analyses

A suite of standard population genetic algorithms were used to analyse both the genetic samples from the simulations as well as the matching empirical genetic data. *Genepop* v3.3 (Raymond and Rousset, 1995) was used to run the Hardy-Weinberg test of heterozygote deficiency on the 213 samples from Barrow using both the 11 'original' and 22 'new' loci. For this test, an MCMC burn-in of 30,000 iterations was used, with a final chain length of 10,000 and batch size of 100. The Hardy-Weinberg disequilibrium was also calculated across all loci using Fisher's method (Ryman and Jorde, 2001).

To examine the effects of GGS, several stratification schemes were used. In the first, in order to examine the magnitude of extreme GGS, samples were stratified into three age cohorts based on the year of catch and the estimated ages of the samples (George *et al.*, 1999; Lubetkin and Zeh, 2006; Rosa *et al.*, 2004). The first stratum was composed of animals born prior to 1950, encompassing animals born during the low point in the population's history (prior to 1918). The second stratum, those born 1950–79, represented animals born during the rapid growth phase. The final stratum was those born after 1979, representing animals born recently during the period when most of the samples were taken.

Additionally, to replicate actual analyses that have been conducted previously, stratifications based on sampling site (Barrow versus St. Lawrence Island) were also examined, as well as those based on season of collection (spring or autumn) at each of those sites. Fig. 3 shows the distribution of body lengths of samples from each of these three stratifications.

For the village and season comparative tests, all empirical samples were used. For the tests comparing cohorts only whales that were actually aged were used in order to reduce the large uncertainties that are introduced when using length to approximate age.

The F_{st} for mtDNA and microsatellites was calculated following Weir and Cockerham (1984), and χ^2 for both sets of markers was calculated following Roff and Bentzen (1989). The AMOVA Φ_{st} metric was calculated for mtDNA data using the R package *ade4* (Chessel *et al.*, 2004).

For each test, the empirical test statistic was compared to null distributions generated by 500 replicates of standard permutation methods (representing a null model of panmixia) and the distribution generated by the simulation. For both comparisons, *p*-values reported in this paper refer to the proportion of replicates with test-statistics equalling or exceeding the value obtained from the empirical data using the same (matched) samples; *p*-values ≤0.05 were considered to indicate empirical results inconsistent with the model.

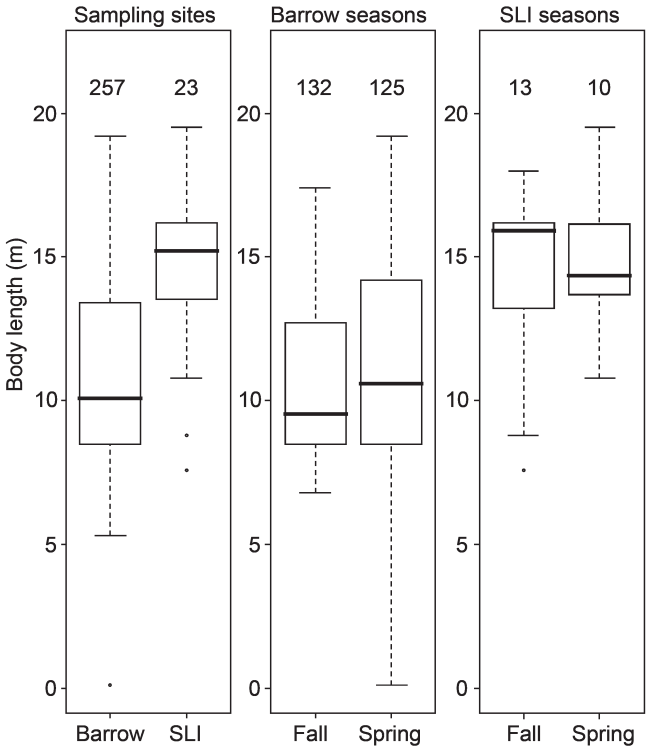

Fig. 3. Distribution of body length of samples stratified by sampling site, and season of collection within each site. Numbers above each distribution indicate sample size. 'SLI' = St. Lawrence Island.

To evaluate the relative support for the panmixia and simulation null models, the ratio between the likelihoods of each model given the empirical data for each pairwise test conducted was calculated. In order to estimate likelihoods, a Gamma distribution (chosen because of its' relative flexibility) was fitted to the distribution of test statistics from replicates in each null model. As both F_{st} and Φ_{st} can have negative values, each null distribution was rescaled to have a minimum value of zero prior to fitting the gamma. Using the estimated parameters of the gamma distribution the likelihood of the observed test statistic plus the fixed rescaling constant was then calculated. Finally, the log of likelihood(data|panmixia)/likelihood(data|simulation) was calculated. Log-likelihood ratios less than one support the simulation model, while those greater than one support the panmixia model.

RESULTS

Simulation diagnostics

The population trajectories for the 500 replicates are given in Fig. 4. At the nadir, the median abundance was 1,197 with a range of 806–1,608. Four percent of the replicates ended with an abundance greater than 12,000 in 2006. Fig. 5 shows the distribution of ages within each stage at the end of burn-in (a) and at the end of the simulation (b). The mean age of all reproductive individuals was 48 (95-percentile = 13–129) at the end of burn-in and 33 (95-percentile = 12–76) at the end of the simulation. At the end of the simulation,

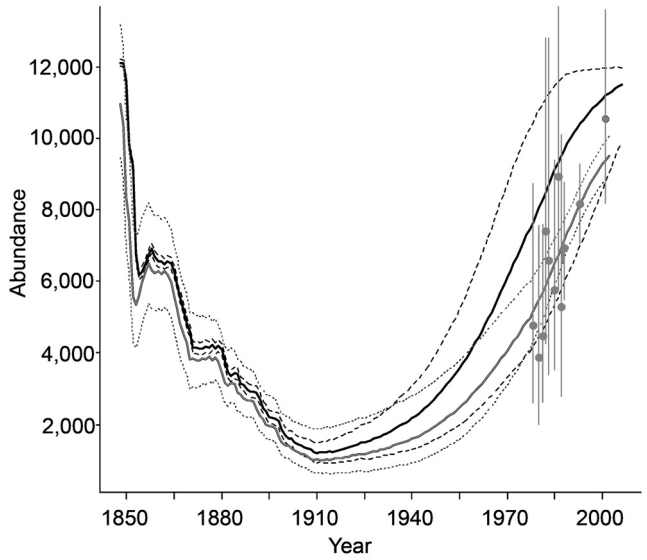

Fig. 4. Median population abundance for 500 simulation replicates from 1848 to 2006 shown in black. Dashed lines bound the 90-percentile of abundance in each year. Grey lines and points indicate trajectory and median abundance estimates from surveys with 90% CIs reproduced from Brandon and Wade (2006).

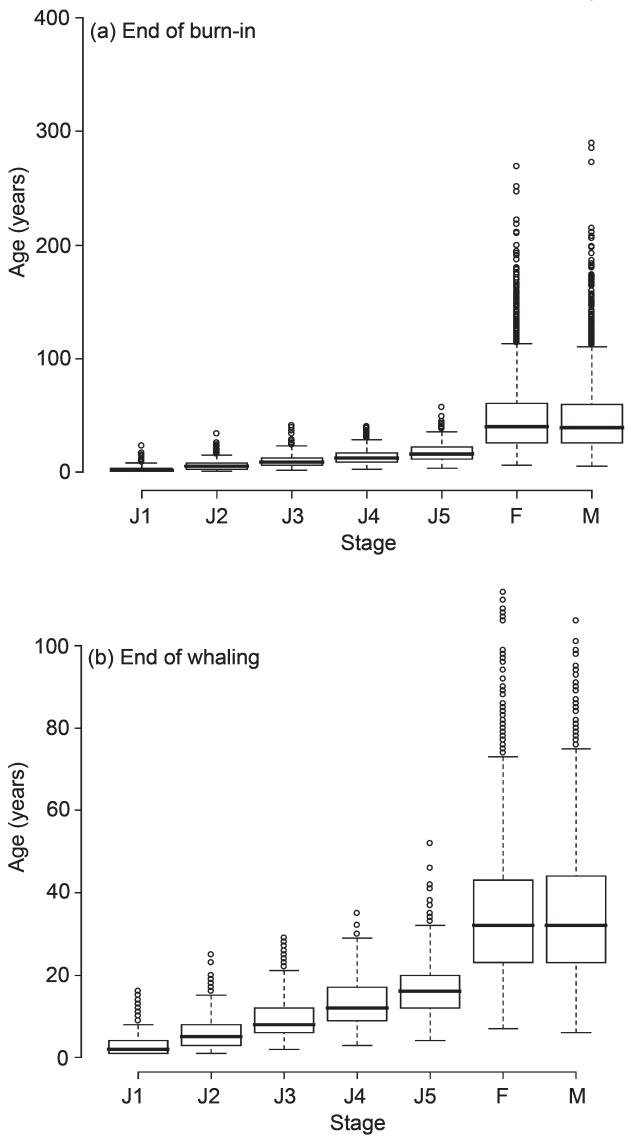

Fig. 5. Distribution of age within each demographic stage for all simulated individuals at the end of burn-in (a), and the end of whaling (b), for an example simulation replicate.

approximately 48% of the individuals were reproductive adults and the sex ratio was not significantly different from 50:50.

Genetic diversity of the empirical data, as measured by the number of alleles (haplotypes for mtDNA), heterozygosity, and Φ_H, was similar to the distributions of these metrics from the matched simulated samples (Fig. 6). Only measures of heterozygosity and Φ_H for mtDNA were outside of the simulated distributions, a result of the skewed haplotypic frequency distribution in the empirical data.

Genetic analyses

In the empirical data, nine of the 33 loci were found to be out of Hardy-Weinberg equilibrium (HWE) with a combined *p*-value using Fisher's method of 2.3×10^{-6}. In the simulation, the median number of loci out of HWE was two, with a maximum of five (Fig. 7a). When errors were added to the simulated data, the median number out of HWE increased to three with a maximum of 11 (Fig. 7b). The *p*-value for the test with errors was 0.006.

There was a relatively uniform distribution of MCMC HWE *p*-values across loci without errors included (Fig. 8a). The combined *p*-value using Fisher's method for the empirical data was 2.3×10^{-6}, which was less than the minimum value in the simulation of 0.0025. The 95-percentile of the simulated distribution was 0.022–0.989. When errors were introduced into the simulated data, the distributions of the MCMC HWE *p*-values and the Fisher's method *p*-values were highly skewed (Fig. 8b). The median of Fisher's method *p*-values was 2.5×10^{-2}, with a 95-percentile of 3.14×10^{-5}–5.1×10^{-1}. Sixty one percent of this distribution was ≤0.05. The empirical value was at the lower 1% of this distribution, making it inconsistent with the simulation.

In the analyses of the empirical data stratified by age cohorts, sampling sites, and seasons within sampling sites, only four of the 60 tests (30 pairwise tests each for panmixia and simulation null models) indicated significant genetic differentiation ($p \leq 0.05$) (Table 3). χ^2-tests of cohorts born before 1950 versus those born after 1979 for both mtDNA and microsatellites were significant using the permutation test for panmixia ($p = 0.012$ and 0.048 respectively) but not significant using the simulation generated *p*-value ($p = 0.186$ and 0.494 respectively). The likelihood for the simulation null model was 12.5 times more likely than panmixia for mtDNA and 3.4 times more likely for microsatellite data. The mtDNA χ^2-test between the 1950–79 and after-1979 cohorts, which had a small, but non-significant panmixia *p*-value (0.088) was also non-significant (0.312) under the better supported simulation null model. Conversely, the same test for microsatellites and the χ^2-tests between cohorts born before 1950 and those born 1950–79 supported the panmixia model. For both markers, F_{st}-tests supported the simulation model, while the three Φ_{st}-tests supported the panmixia model.

The other two tests that indicated significant differentiation were inconsistent with our simulation model. They were the mtDNA F_{st}-test between autumn and spring samples from St. Lawrence Island ($p = 0.008$), and the microsatellite F_{st}-test between Barrow and St. Lawrence Island ($p = 0.042$). In both of these comparisons, although

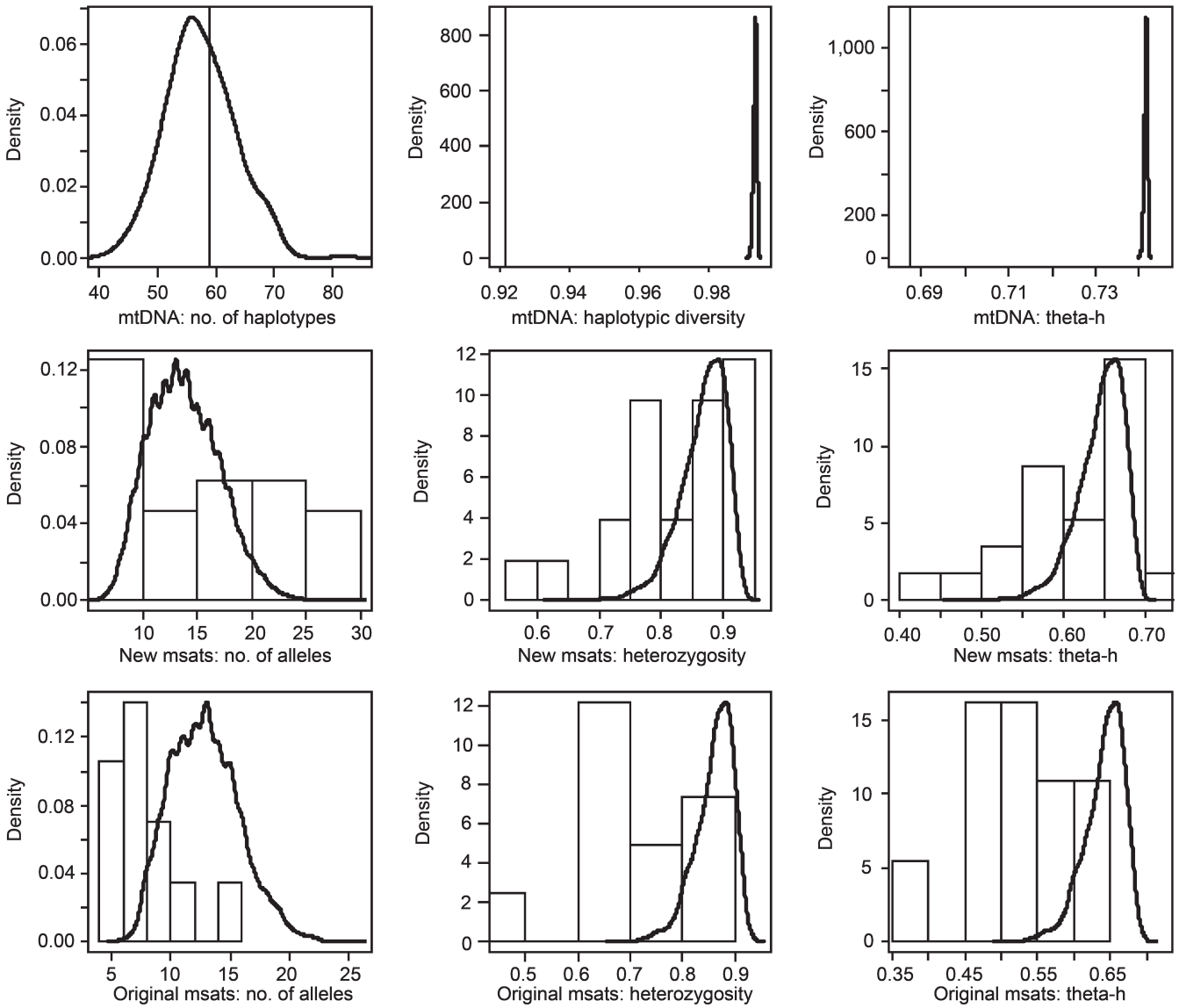

Fig. 6. Distribution of single-locus measures of genetic diversity in the empirical data (histograms over the 33 loci), and 500 replicates of the simulation (bold lines). For mtDNA, the empirical data value is given by a single line.

the *p*-value for panmixia was non-significant, the panmixia model was less than 1.4 times more likely than the simulation model. Additionally, the panmixia *p*-values for both microsatellite tests of the sampling sites were only slightly higher than the critical α ($p = 0.066$ for both).

The introduction of errors into the simulated microsatellite data did not make a substantial change in the results of any of the F_{st} or χ^2-tests. While the *p*-value of some results changed slightly, there was not a consistent pattern in the direction of the change (results not shown).

DISCUSSION

Results indicate the benefit of using simulation models to develop null distributions when assumptions of Wright-Fisher null distributions are known to have been violated. The two significant permutation tests for panmixia (χ^2 for two extreme age cohorts) were not significant using the null distributions from the simulations and were also the tests which most strongly supported the simulation null model (i.e. had the largest negative log-likelihood ratio values). In many respects, the results of the analysis indicate that the empirical genetic data sampled from BCB bowhead whales are more consistent with the model of a single, randomly-mating population with a history of whaling and subsequent recovery mimicking the true bowhead whale history than they are with standard null distributions generated under assumptions of panmixia. None of the 15 pairwise cohort tests conducted (three comparisons for five measures of genetic differentiation) exhibited a significant difference between the simulated replicates and empirical data.

However, the number of loci actually out of HWE significantly exceeded the numbers estimated to be out of HWE in the simulations, which suggests that, in some respect, our model may not fully capture some component of the process that generated these samples. Given the findings of Morin *et al.* (2009), an attempt was made to simulate some of the genotyping errors that may lead to Hardy-Weinberg disequilibrium. However, when errors were introduced in the simulated data, the difference between the two decreased dramatically but remained significant. This was the only analysis for which the introduction of genotyping errors into the simulated datasets had a

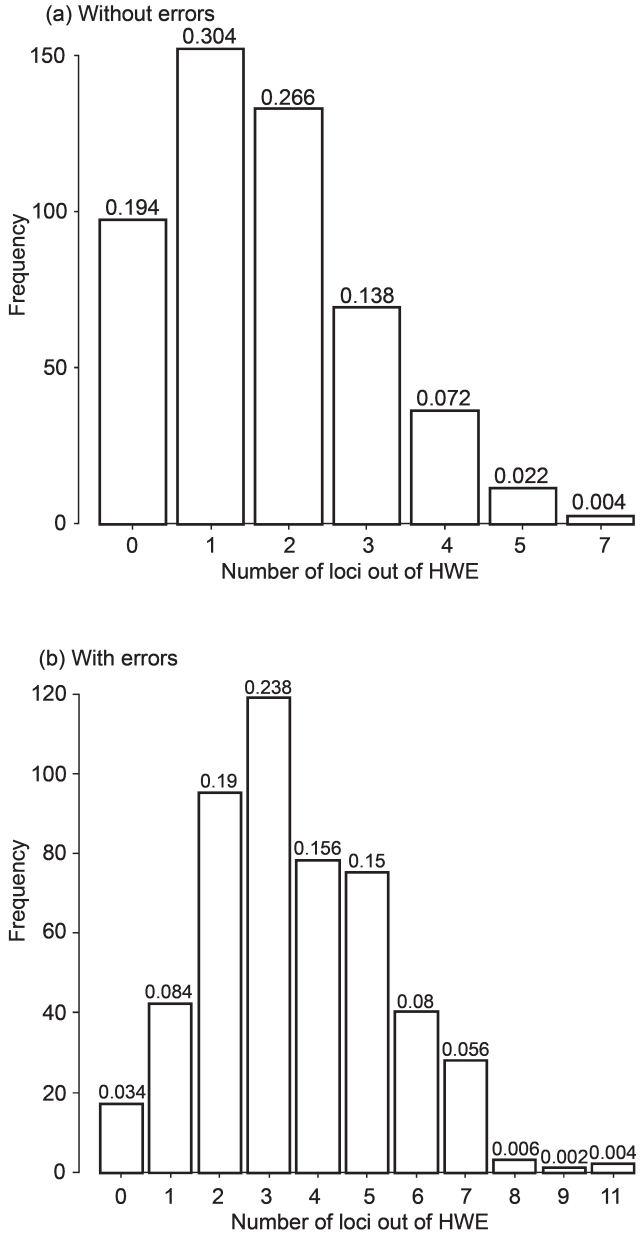

Fig. 7. Distribution of the number of loci out of Hardy-Weinberg equilibrium in the 500 replicates without (a) and with (b) errors included. Numbers above the bars are the fraction of the total number of replicates represented by that bar.

substantial impact on the results. While the comparison between the empirical and simulated results was still significant even with errors introduced, the results support those of Morin *et al.* (2009) in highlighting the sensitivity of HWE to genotyping errors. For instance, in the simulation without genotyping errors only 0.026 of the simulated datasets had five or more loci out of HWE. When genotyping errors were introduced, this frequency jumped to 0.298.

The introduction of errors showed a greater effect on the distribution of the overall Fisher's method *p*-value for HWE. Without errors, this distribution was relatively uniform as would be expected under a standard null hypothesis. When errors were included, the distribution became highly skewed towards very small *p*-values. While this skew was not large enough to make the empirical finding of overall Hardy-Weinberg disequilibrium consistent with our model, the implication is that with even the relatively low error rate identified by Morin *et al.* (2009), there is a large probability (61% of the replicates had a $p \leq 0.05$) of falsely assessing widespread disequilibrium.

A very simplistic model was used for introducing genotyping errors to the simulated datasets. Though the estimated allelic dropout rate from the empirical data was incorporated, all errors were random with respect to the loci at which they occurred and the alleles and individuals that were affected. In reality, genotyping errors are often not random (Bonin *et al.*, 2004; Gagneux *et al.*, 1997; Morin *et al.*, 2009). Some loci or samples may be more susceptible to errors than others. Allele length and frequency may also affect the likelihood of the allele being correctly scored. Stutter bands and slippage would result in the mis-scored allele being very close in length to the correct allele, rather than reflecting the overall allele frequency at the locus. Since it was not possible to quantify the various biases inherent in genotyping errors, a simplistic model was chosen that only included allelic dropout and random errors. If other realistic biases had been incorporated into the simulated genotyping errors, an even stronger impact on the expected distribution of the number of loci out of HWE (Morin *et al.*, 2009) would have been expected. The susceptibility of HWE to genotyping errors makes reliance on this metric as the sole source of data indicating population structure questionable practice.

The second analysis that indicated a lack of consistency between the empirical data and simulation model was the F_{st} test of mtDNA sequences from samples taken at St. Lawrence Island in the autumn and spring. The magnitude of the observed mtDNA F_{st} value (0.054) in this test results from the difference in the frequency of one haplotype (BH42) between spring and autumn samples (6 in autumn, 1 in spring). Given that five of the six autumn samples that possessed haplotype BH42 came from one location in St. Lawrence Island (Savoonga), and sample sizes are relatively small in both strata, it is possible that these samples do not adequately represent the haplotypic distributions of St. Lawrence Island whales in these seasons.

The final analysis that was inconsistent with the model described here was the F_{st} test between Barrow and St. Lawrence Island for the microsatellite markers. It is possible that this result is being influenced either by the unusual distribution of the autumn Savoonga samples mentioned above or by the loci that were found to be out of HWE in the Barrow samples. The effect of the latter were partially examined by running this analysis again after removing the six samples most influential on HWE, which were identified by Morin *et al.* (2009). The removal of these samples did not significantly change the lack of consistency between the empirical and simulated data either with or without errors introduced into the simulated data.

By mimicking the demographic composition and whaling history of BCB bowhead whales, the model was able to capture non-equilibrial effects such as GGS that a standard panmixia null model could not. This was most clearly shown in the extreme example of the age cohort analyses. However, it was also evident in analyses which compared whales sampled in Barrow versus those sampled from St. Lawrence Island. Although the microsatellite F_{st} test indicated significant differentiation using the simulated null

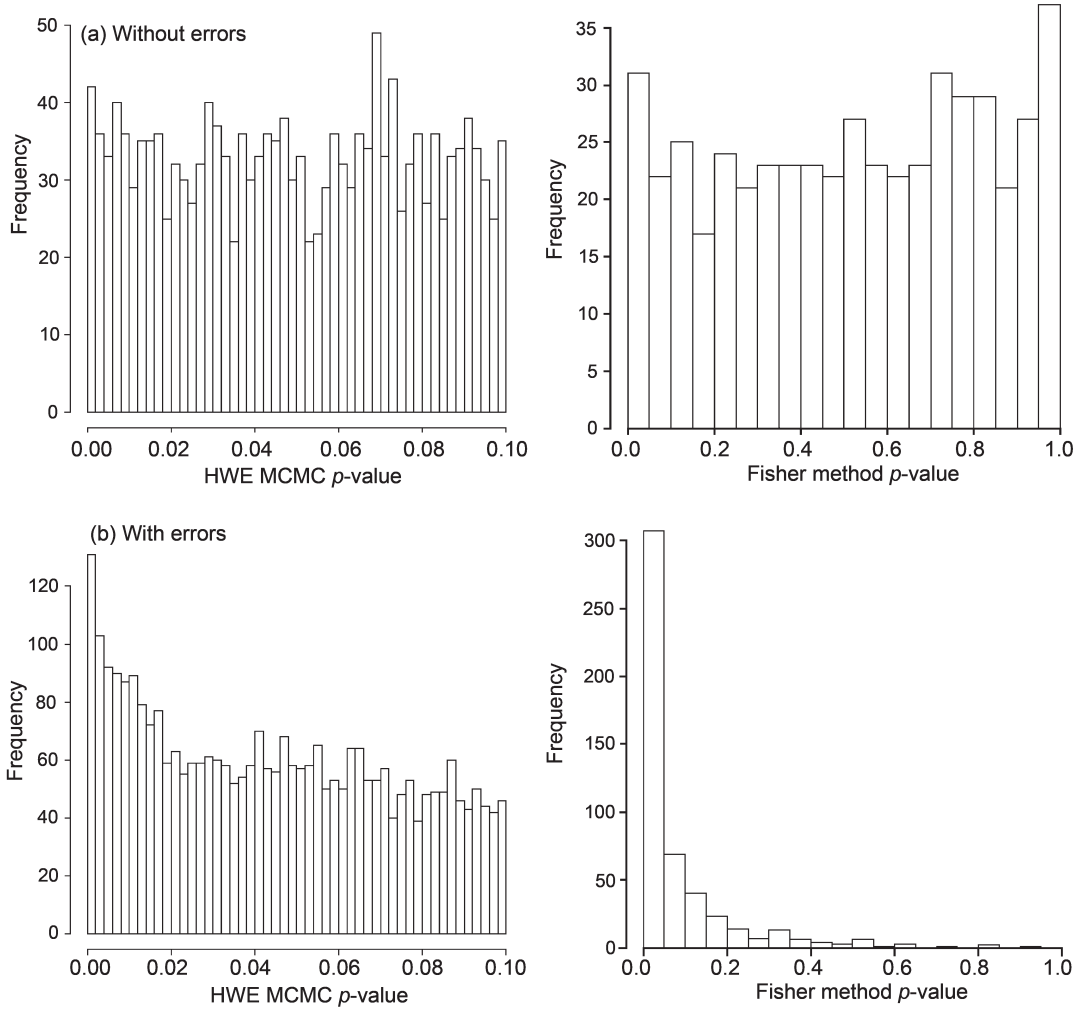

Fig. 8. Distribution of p-values for Hardy-Weinberg equilibrium (HWE) without (a) and with (b) errors included. Figures on left are distributions of locus *p*-values from *Genepop* (truncated to values ≤ 0.1). Figures on right are distributions of overall *p*-value using Fisher's method.

distribution, the likelihood ratio suggests that the panmixia model is only very slightly favoured. Given that there is a significant difference in ages (as inferred from length – Fig. 3) between these two sites, it is expected that this stratification would behave more like the cohort analyses. Additionally, the St. Lawrence Island microsatellite sample size was small (n = 25) relative to mtDNA (n = 52), so conclusions must be treated as preliminary. Genotyping of more samples from this region with Single Nucleotide Polymorphisms (SNPs) is underway which ought to provide more resolution.

The fact that both microsatellite panmixia test results are only slightly greater than the critical α could lead one to incorrectly infer the presence of population structure were there is none. Thus, the results presented here suggest that in some cases where standard permutation tests for panmixia may indicate significant genetic differentiation, if the population demographic history is taken into account, the simulated distribution will more appropriately reflect the genetic distribution of the null model being tested.

Simulation construction

The simulation described in this paper represents a null hypothesis based on a very specific model of a single population that is out of genetic equilibrium due to its population history. One of the strengths of this simulation is that by matching the age and sex characteristics of the empirical samples where possible, this null hypothesis inherently incorporates any potential demographic biases in the sampling process. The model relies on several parameters controlling the population dynamics and genetic diversity. When possible, empirical data and parameter values from independent sources were used. When these were not available, parameters were iteratively tuned to ensure that other aspects of simulation either fit published results or matched the empirical data as closely as possible. It is important to note that this process does not ensure that the parameter values are accurate with respect to a 'true' single population; the current best statistical estimates of stock structure and biological/demographic parameters are of varying precision.

An example is the procedure by which carrying capacity (K) was selected. With the value of the logistic growth shape parameter (z) set at the median posterior value from Brandon and Wade's (2006) backward projection model, a value of K was selected such that the majority of the replicates did not go extinct and passed through the abundance 'gates'. Under these constraints it can be seen that many of the population dynamics parameters, most notably K, z, and the population growth rate (r – not specified in the model, but resulting from

Table 3

Comparison of panmixia and simulation null models for tests of genetic differentiation of several stratifications of the data. Sample size of each strata given in parentheses. Panmixia and simulation *p*-values are from 500 replicates of each test. Likelihoods estimated from gamma distributions fit to replicates from each null model. Log-likelihood ratio is log(Panmixia likelihood/Simulation likelihood). Log-likelihood ratios less than 1 support the simulation model, while those greater than 1 support the panmixia model. Pairwise comparisons within each marker type are sorted by increasing Log-likelihood ratio. Values in grey indicate *p*-values ≤ 0.05. SLI=St. Lawrence Island.

Stratification	Strata	Test	Observed value	Panmixia *p*-value	Simulation *p*-value	Panmixia likelihood	Simulation likelihood	Log-likelihood ratio
(a) mtDNA								
Age cohorts	Before 1950 (21) v. after 1979 (34)	χ^2	35.582	0.012	0.186	0.004	0.048	0.08
	1950-1979 (25) v. after 1979 (34)	χ^2	34.331	0.088	0.312	0.039	0.058	0.67
	Before 1950 (21) v. after 1979 (34)	F_{st}	0.009	0.190	0.186	16.031	20.528	0.78
	1950-1979 (25) v. after 1979 (34)	F_{st}	0.007	0.261	0.216	20.523	25.555	0.80
	Before 1950 (21) v. 1950-79 (25)	F_{st}	-0.007	0.631	0.728	29.291	35.237	0.83
	Before 1950 (21) v. after 1979 (34)	Φ_{st}	0.012	0.230	0.242	11.580	9.223	1.26
	1950-1979 (25) v. after 1979 (34)	Φ_{st}	-0.015	0.778	0.658	27.910	20.842	1.34
	Before 1950 (21) v. 1950-79 (25)	Φ_{st}	0.005	0.321	0.284	14.524	10.299	1.41
	Before 1950 (21) v. 1950-79 (25)	χ^2	22.758	0.842	0.844	0.083	0.054	1.54
Sampling sites	Barrow (258) v. SLI (52)	F_{st}	-0.002	0.655	0.772	119.446	125.890	0.95
	Barrow (258) v. SLI (52)	χ^2	52.939	0.780	0.802	0.030	0.025	1.20
	Barrow (258) v. SLI (52)	Φ_{st}	-0.002	0.575	0.482	75.804	56.751	1.34
Barrow seasons	Autumn (133) v. Spring (125)	F_{st}	0.000	0.355	0.406	142.269	179.073	0.79
	Autumn (133) v. Spring (125)	Φ_{st}	0.002	0.218	0.232	57.302	48.569	1.18
	Autumn (133) v. Spring (125)	χ^2	57.078	0.573	0.586	0.053	0.040	1.33
SLI seasons	Autumn (13) v. Spring (11)	χ^2	16.519	0.172	0.672	0.126	0.131	0.96
	Autumn (13) v. Spring (11)	Φ_{st}	-0.014	0.485	0.492	8.479	6.949	1.22
	Autumn (13) v. Spring (11)	F_{st}	0.054	0.080	0.008	2.318	1.673	1.39
(b) Microsatellites								
Age cohorts	Before 1950 (14) v. after 1979 (24)	χ^2	355.637	0.048	0.494	0.004	0.014	0.29
	Before 1950 (14) v. 1950-79 (16)	F_{st}	-0.004	0.804	0.844	76.232	97.654	0.78
	Before 1950 (14) v. after 1979 (24)	F_{st}	0.002	0.335	0.272	71.310	88.219	0.81
	1950-1979 (16) v. after 1979 (24)	F_{st}	-0.003	0.790	0.858	107.732	118.951	0.91
	1950-1979 (16) v. after 1979 (24)	χ^2	324.783	0.281	0.888	0.015	0.007	2.14
	Before 1950 (14) v. 1950-79 (16)	χ^2	269.402	0.774	0.998	0.015	0.001	15.00
Sampling sites	Barrow (213) v. SLI (25)	χ^2	464.636	0.066	0.126	0.003	0.005	0.60
	Barrow (213) v. SLI (25)	F_{st}	0.002	0.066	0.042	89.679	72.476	1.24
Barrow seasons	Autumn (115) v. Spring (98)	χ^2	425.940	0.206	0.374	0.010	0.011	0.91
	Autumn (115) v. Spring (98)	F_{st}	0.001	0.128	0.098	301.650	293.580	1.03
SLI seasons	Autumn (14) v. Spring (11)	F_{st}	-0.007	0.934	0.988	43.487	17.524	2.48
	Autumn (14) v. Spring (11)	χ^2	254.095	0.978	1.000	0.003	$1.22 \cdot 10^{-4}$	24.59

the reproduction and survival matrices), will be closely correlated such that multiple combinations would work. While the goal was not to estimate these parameters, through iterative testing it was determined that, given the historical catch record, there was a small range over which they could vary and still meet the extinction and abundance constraints. In the tests, choices of K outside of the range of approximately 11,900–12,400 would not produce useable replicates. This range is well within the 95% credibility interval for K (9,112–13,610) from the Brandon and Wade (2006) assessment model most similar to the simulation used in this study, which is expected given that this study used the same historical catch and abundance data.

In this simulation, the carrying capacity of the population was assumed to be the same now as it was prior to the onset of commercial whaling. This could have been violated if there has been a substantial change in the ecosystem or the range of the population has either expanded or contracted. The 2001 abundance estimate of 10,545 suggests that the population is very close to the carrying capacity estimated in Brandon and Wade (2006), making it unlikely that there has been a decrease in carrying capacity. Moreover, estimates from George *et al*. (2004) indicate that the bowhead whale population continues to increase more than 3% annually. Among a variety of possible explanations for this finding would be an increase in carrying capacity. Whether or not there has indeed been a significant increase in carrying capacity will require future surveys.

Similarly, had bowhead whales been subjected to substantial population fluctuations prior to commercial whaling, some of the parameters used in the 4,000 year burn-in may not reflect reality as that phase simulated a population at demographic and genetic equilibrium. However, because the burn-in was tuned to produce a population with the approximate amount of diversity as seen in the current empirical data, deviations would not be expected to substantially affect the results. The difference between the actual effective population size (N_e) and what was calculated to initialise the population at equilibrium would be offset by a difference in the actual mutation rate and the one used in the study.

Another result of the model constraints is that it was not possible to directly control population abundance at the nadir or the range over which it varied. As a population reduced to a very small size will be unable to contain the entire genetic diversity of its larger progenitor, this factor is likely to greatly affect the degree of genetic disequilibrium within the population. The smaller the nadir, the stronger the signal of a GGS is expected to be, in which the genotypes of individuals born before and after the nadir will appear to have come from two different distributions. Therefore, it is important to note that the results of this study are conditional

on the nadir being approximately 1,100. This is consistent with the suggestions that the population size might have been 1,000 or lower at the end of commercial whaling (Bockstoce and Botkin, 1983; Bockstoce and Burns, 1993).

In the absence of GGS, sampling from an age-structured population as well as non-random sampling can also lead to inferences of populations that are out of genetic equilibrium (Waples, 1998; Waples and Yokota, 2007). Though there is no evidence that whalers were intentionally selective in their hunting, the fact that bowhead whales segregate by age and sex during migration may have resulted in selectivity on the basis of availability (Bockstoce, 1986). Evidence for some selectivity can be found in the fact that the average size of whales killed decreased between the beginning of the fishery and 1874, the only period for which such data are available (Bockstoce and Burns, 1993). If a similar kind of selectivity continued throughout the commercial hunt, it would add to effects of GGS as the portion of the population that survived through the nadir would tend to represent younger cohorts. As a result of the complexity of this simulation and the constraints to mirror bowhead whale population trajectories as discussed above, the sensitivity of the results to parameters such as selectivity of harvest, differences in the age-structure of the population, or variance in reproductive success were not examined. These items are being further explored with a simpler form of the simulation previously described by Ripley *et al*. (2006) and presented in Martien *et al.* (2009).

In theory, the methods described here could be extended to construct any number of alternative null hypotheses. For example, while this simulation models a single stock, a variety of two-stock hypotheses have been proposed for BCB bowhead whales (IWC, 2008) and it would be a productive exercise to use the methods presented here to explore their relative likelihoods. Genetic simulations for any of these two-stock hypotheses would require several important refinements such as defining the population dynamics of each stock, as well as the annual partitioning of the historical catch among stocks. The stocks would also have to be initialised at their appropriate pre-whaling genetic conditions, which are a result of the relative abundances and degree of gene-flow. Finally, during the 'whaling' phase of the simulation, empirical samples and their simulated equivalents would need to be assigned to stock, introducing further uncertainty.

In summary, the creation of more appropriate null distributions for common tests for population structure is a potentially important strategy when there are known reasons for the population to be in disequilibrium due to historical or demographic factors. In such cases, the simulated null distributions may provide a better basis for inference than reliance on the standard Wright-Fisher assumption of equilibrium.

ACKNOWLEDGMENTS

The authors would like to thank Allan Strand for the design of and continued support with *rmetasim* throughout this project, Ryan Huebinger and John Bickham for the microsatellite genotypes, and the SWFSC Population Identity Program's Molecular Genetics Lab for the mtDNA sequencing. Judy Zeh and June Morita were invaluable in compiling and interpreting age estimates and providing thoughtful insights on estimating age from morphometric characters. We thank the whale hunters of Alaska and Russia for their collaboration and support. We appreciate the efforts of the support staff at the Alaska Eskimo Whaling Commission, NMML, SWFSC, and NSB (including NSB Mayors George Ahmaogak and Edward Itta) who contributed to and supported the bowhead stock structure research. We acknowledge the efforts of Maggie Ahmaogak and Jessica LeFevre at the Alaska Eskimo Whaling Commission for helping to secure funding for stock structure studies. We are grateful to Alaska Senator Ted Stevens for his support of these, and Stanley Speaks with the Bureau of Indian Affairs for helping secure additional funding (through BIA) for this project. Rollie Schmitten, Bill Hogarth, Mike Tillman, Tim Smith, Doug DeMaster, Dave Rugh, and Paul Wade of the NMFS, have offered advice and support of these studies. Suggestions from our IWC colleagues in Norway, Russia, and Japan have improved the studies outlined above. The National Park Services Program, Shared Beringian Heritage Program gave funding support for the biopsy work in Chukotka (through the North Slope Borough) for the years 2003–06. Geof Givens work was supported by the North Slope Borough (Alaska) and the National Oceanic and Atmospheric Administration (through the Alaska Eskimo Whaling Commission). This paper has had the benefit of thoughtful reviews from Robin Waples and John Brandon. All work in this paper conforms to the legal requirements for conservation and animal welfare of the Department of Commerce and the United States of America.

REFERENCES

Angliss, R.P., Rugh, D.J., Withrow, D.E. and Hobbs, R.C. 1995. Evaluations of aerial photogrammetric length measurements of the Bering-Chukchi-Beaufort Seas stock of bowhead whales (*Balaena mysticetus*). *Rep. int. Whal. Commn* 45: 313–24.

Bockstoce, J.R. 1986. *Whales, Ice and Men: the History of Whaling in the Western Arctic*. University of Washington Press, Seattle. 400pp.

Bockstoce, J.R. and Botkin, D.B. 1983. The historical status and reduction of the western Arctic bowhead whale (*Balaena mysticetus*) population by the pelagic whaling industry, 1848–1914. *Rep. int. Whal. Commn* (special issue) 5: 107–41.

Bockstoce, J.R. and Burns, J.J. 1993. Commercial whaling in the North Pacific sector. pp.563–77. In: Burns, J.J., Montague, J.J. and Cowles, C.J. (eds). *The Bowhead Whale*. Society for Marine Mammalogy, Lawrence, Kansas.

Bonin, A., Bellemain, E., Bronken Eidesen, P., Pompanon, F., Brochmann, C. and Taberlet, P. 2004. How to track and assess genotyping errors in population genetics studies. *Mol. Ecol.* 13: 3261–73.

Brandon, J. and Wade, P.R. 2006. Assessment of the Bering-Chukchi-Beaufort Sea stock of bowhead whales using Bayesian model averaging. *J. Cetacean Res. Manage*. 8(3): 225–40.

Braund, S.R., Stoker, S.W. and Kruse, J.A. 1988. Quantification of subsistence and cultural need for bowhead whales by Alaska eskimos. Paper TC/40/AS2 presented to the IWC Aboriginal Whaling Sub-Committee, May 1988, Auckland, New Zealand (unpublished). 118pp. [Paper available from the Office of this Journal].

Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. 1984. *Classification and Regression Trees*. Chapman and Hall, New York.

Caswell, H. 2001. *Matrix Population Models*. 2nd ed. Sinauer Associates, Inc, Sunderland, Massachusetts, USA. i–xxii+722pp.

Chessel, D., Dufour, A.-B. and Thioulouse, J. 2004. The ade4 package-I-One-table methods. *R News* 4: 5–10.

Fisher, R.A. 1930. *The Genetical Theory of Natural Selection*. Oxford University Press (Clarendon), London/New York.

Gagneux, P., Boesch, C. and Woodruff, D.S. 1997. Mircosatellite scoring errors associated with noninvasive genotyping based on nuclear DNA amplified from shed hair. *Mol. Ecol.* 6: 861–68.

George, J.C., Bada, J., Zeh, J., Scott, L., Brown, S.E., O'Hara, T. and Suydam, R. 1999. Age and growth estimates of bowhead whales (*Balaena mysticetus*) via aspartic acid racemization. *Can. J. Zool.* 77: 571–80.

George, J.C., Zeh, J., Suydam, R. and Clark, C. 2004. Abundance and population trend (1978–2001) of western arctic bowhead whales surveyed near Barrow, Alaska. *Mar. Mammal Sci.* 20(4): 755–73.

Givens, G.H., Huebinger, R.M., Patton, J.C., Postma, L.D., Lindsay, M., Suydam, R.S., George, J.C., Matson, C.W. and Bickham, J.W. 2010. Population genetics of bowhead whales (*Balaena mysticetus*) in the western Arctic. *Arctic* 63(1): 1–12.

Huebinger, R.M., Bickham, J.W., Patton, J.C., George, J.C. and Suydam, R.S. 2006. Progress report on the development of new microsatellite markers for bowhead whales. Paper SC/57/BRG19 presented to the IWC Scientific Committee, May 2006, St Kitts and Nevis (unpublished). 3pp. [Paper available from the Office of this Journal].

International Whaling Commission. 2003. Report of the Scientific Committee. Annex E. Report of the standing working group on the development of an aboriginal subsistence whaling management procedure (AWMP). *J. Cetacean Res. Manage.* (Suppl.) 5:154–255.

International Whaling Commission. 2008. Report of the Scientific Committee. Annex E. Report of the standing working group on the development of an aboriginal subsistence management procedure. *J. Cetacean Res. Manage.* (Suppl.) 10:121–49.

Jorde, P.E., Schweder, T., Bickham, J.W., Givens, G.H., Suydam, R. and Stenseth, N.C. 2007. Detecting genetic structure in migrating bowhead whales off the coast of Barrow, Alaska. *Mol. Ecol.* 16(10): 1993–2004.

Laval, G. and Excoffier, L. 2004. SIMCOAL 2.0: a program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history. *Bioinformatics* 20: 2485–87.

LeDuc, R.G., Dizon, A.E., Burdin, A.M., Blokhin, J.C., George, J.C. and Brownell, R.L., Jr. 2005. Genetic analyses (mtDNA and microsatellites) of Okhotsk and Bering/Chukchi/Beaufort Seas populations of bowhead whales. *J. Cetacean Res. Manage.* 7(2): 107–12.

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., Litovka, D. and George, J.C. 2008. Mitochondrial genetic variation in bowhead whales in the western Arctic. *J. Cetacean Res. Manage.* 10(2): 93–98.

Lubetkin, S.C. and Zeh, J. 2006. Deriving age-length relationships for bowhead whales (*Balaena mysticetus*) using a synthesis of age estimation techniques. Paper SC/58/BRG14 presented to the IWC Scientific Committee, May 2006, St. Kitts and Nevis, West Indies (unpublished). 20pp. [Paper available from the Office of this Journal].

Martien, K. 2006. Progress on TOSSM dataset generation. Paper SC/58/SD2 presented to the IWC Scientific Committee, May 2006, St Kitts and Nevis, West Indies (unpublished). 17pp. [Paper available from the Office of this Journal].

Martien, K.K., Gregovich, D., Bravington, M.V., Punt, A.E., Strand, A.E., Tallmon, D.A. and Taylor, B.L. 2009. TOSSM: an R package for assessing performance of genetic analytical methods in a management context. *Molecular Ecology Resources* 9(6): 1456–59.

Morin, P.A., LeDuc, R.G., Archer, F.I., Martien, K.K., Huebinger, R., Bickham, J.W. and Taylor, B.L. 2009. Significant deviations from Hardy-Weinberg equilibrium caused by low levels of microsatellite genotyping errors. *Mol. Ecol. Resources* 9: 498–504.

Noongwook, G., Huntington, H.P. and George, J.C. 2007. Traditional knowledge of the bowhead whale (*Balaena mysticetus*) around Saint Lawrence Island, Alaska. *Arctic* 60(1): 47–54.

O'Hara, T.M., House, C., House, J.A., Suydam, R.S. and George, J.C. 1998. Viral serologic survey of bowhead whales in Alaska. *J. Wildl. Dis.* 34(1): 39–46.

R Development Core Team. 2006. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL: *http://www.R-project.org.*

Raymond, M. and Rousset, F. 1995. GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. *J. Hered.* 86: 248–49.

Ripley, B.J., Martien, K. and Taylor, B.L. 2006. A simulation approach to understanding non-equilibrial dynamics in a recovering long-lived species: the bowhead whale. Paper SC/58/BRG13 presented to the IWC Scientific Committee, May 2006, St. Kitts and Nevis, West Indies (unpublished). 12pp. [Paper available from the Office of this Journal].

Roff, D.A. and Bentzen, P. 1989. The statistical analysis of mtDNA polymorphisms: chi-square and the problem of small samples. *Mol. Biol. Evol.* 6(5): 539–45.

Rosa, C., George, J.C., Zeh, J., Botta, O., Zauscher, M., Bada, J. and O'Hara, T.M. 2004. Update on age estimation of bowhead whales (*Balaena mysticetus*) using aspartic acid racemization. Paper SC/56/BRG6 presented to the IWC Scientific Committee, July 2004, Sorrento, Italy (unpublished). 15pp. [Paper available from the Office of this Journal].

Ryman, N. and Jorde, P.E. 2001. Statistical power when testing for genetic differentiation. *Mol. Ecol.* 10: 2361–73.

Strand, A.E. 2002. METASIM 1.0: an individual-based environment for simulating population genetics of complex population dynamics. *Mol. Ecol. Notes* 3: 373–76.

Suydam, R.S. and George, J.C. 2004. Subsistence harvest of bowhead whales (*Balaena mysticetus*) by Alaskan Eskimos, 1974 to 2003. Paper SC/56/BRG12 presented to the IWC Scientific Committee, July 2004, Sorrento, Italy (unpublished). 12pp. [Paper available from the Office of this Journal].

Taylor, B.L., Chivers, S.J., Sexton, S. and Dizon, A.E. 2000. Evaluating dispersal estimates using mtDNA data: comparing analytical and simulation approaches. *Conserv. Biol.* 14: 1287–97.

Taylor, B.L., Chivers, S.J., Larese, J. and Perrin, W.F. 2007. Generation length and percent mature estimates for IUCN assessments of cetaceans. US NMFS Administrative Report LJ-07-01, Southwest Fisheries Science Center.

Therneau, T.M. and Atkinson, B. 2006. *rpart: Recursive Partitioning. R package version 3*. 1–34.

Waples, R.S. 1990. Conservation genetics of Pacific salmon. III. Estimating effective population size. *J. Hered.* 81(3): 277–89.

Waples, R.S. 1998. Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. *J. Hered.* 89: 438–50.

Waples, R.S. and Teel, D.J. 1990. Conservation genetics of Pacific salmon. I. Temporal changes in allele frequency. *Cons. Biol.* 4: 144–56.

Waples, R.S. and Yokota, M. 2007. Temporal estimates of effective population size in species with overlapping generations. *Genetics* 175(219–233).

Weir, B.S. and Cockerham, C.C. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38: 1358–70.

Wright, S. 1931. Evolution in Mendelian populations. *Genetics* 16: 97–159.

Date received: April 2009.
Date accepted: January 2010.