SC/69A/SD & DNA/01

Sub-committees/working group name: SD & DNA

Update on genetics of Bowhead whales using autosomal SNP genotypes and mtDNA sequences

Amy B. Baird, Emily Fritsche, John Citta, John W. Bickham

Update on genetics of Bowhead whales using autosomal SNP genotypes and mtDNA sequences

Amy B. Baird[1], Emily Fritsche[1], John Citta[2], and John W. Bickham[3]

[1]Department of Natural Sciences, University of Houston – Downtown, Houston, TX 77002, USA; [2]North Slope Borough, Department of Wildlife Management, Barrow, AK 99723, USA; [3]Department of Ecology and Conservation Biology, Texas A&M University, College Station, TX 77843, USA

ABSTRACT

This report analyzes mtDNA sequence data and SNP genotypes to examine stock structure in bowhead whales, *Balaena mysticetus*, with an emphasis on points brought up during the 2018 bowhead implementation review. The focus of the study is the Bering-Chukchi-Beaufort Seas (BCB) stock, with fewer samples from the Eastern Canadian Arctic (ECWG) and Sea of Okhotsk (OKS) used for comparison. mtDNA data includes sequences from the HVR1 portion of the control region, cytochrome-b, and ND1 totaling 2494 bp. The SNP panel, described in Baird et al. (2017), contains 69 autosomal loci (analyzed here) and 7 sex chromosome markers (to be used in future studies). Population genetic analyses indicate that the BCB and ECWG populations are not highly differentiated, but statistically significantly different using the SNP loci. In contrast, the OKS population is easily distinguishable from both BCB and ECWG for both mtDNA haplotype frequencies and the SNP loci. These results are consistent with previous studies based on mtDNA control region sequences, focal microsatellites, and a smaller SNP panel.

INTRODUCTION

Bowhead whales (*Balaena mysticetus*) are divided into four recognized stocks: 1) the Bering-Chukchi-Beaufort Seas (BCB); 2) the Eastern Canadian Western Greenland (ECWG); 3) the Okhotsk Sea (OKS); and 4) East Greenland-Svalbard-Barents Sea (EGSB). These stocks are recognized based on migration patterns, geographic distribution, satellite tagging, and population genetic analyses (Baird and Bickham 2021).

The BCB aboriginal harvest is regulated by the International Whaling Commission (IWC). As part of their efforts to estimate sustainability of the harvest quota, implementation reviews (IR) are conducted approximately every 5-6 years. The previous BCB bowhead IR occurred in 2018.

As part of the 2018 bowhead IR, population genetic data were reviewed (Baird et al. 2018). These data consisted of analyses of SNP genotypes and mtDNA sequences. For mtDNA, a total of 3 loci (HVR1, cyt-b, ND1) were used, totaling 2494 bp. 383 BCB, 39 ECWG, and 7 OKS whale samples were sequenced across all 3 genes. For the SNP database, 69 autosomal were used, and 359 BCB, 27 ECWG, and 28 OKS samples were genotyped across these loci. Various analyses were conducted, including haplotype networks, Fst calculations, AMOVA, and Structure. The authors reported no evidence of sub-structuring of the BCB stock, and similar findings to previous studies indicating that both the BCB and ECWG stocks are well differentiated from the OKS stock, but that the BCB and ECWG stocks are less differentiated.

In the 2018 SC report, the following statement was made regarding these data (note that the stock abbreviations below are CAN for ECWG and OK for OKS):

> "In discussion, it was noted that the results presented in SDDNA01 [Baird et al. 2018] have implications for two aspects of bowhead whale stock structure. The primary question of interest for the IR is whether substructure exists within the B-C-B stock. While a number of SNP loci showed significant deviations from Hardy Weinberg equilibrium (HWE) within the B-C-B samples, only about half of these loci exhibited heterozygote deficiencies. This pattern is inconsistent with what would be expected if the deviations from HWE were the result of a Wahlund effect (i.e. due to population substructure). An alternative explanation for the deviations from HWE is that the loci could be under selection pressure. In response, Baird noted that some of the SNPs occur within protein-coding loci, which are more likely to be under selection than non-coding regions. The Working Group further noted that several of the comparisons previously explored using microsatellite genotypes and mtDNA sequence data (see review IWC 2008) had been re-examined using the SNP dataset, including: temporal comparisons (whales sampled in the spring vs the fall), spatial comparisons (St. Lawrence Island vs Barrow), and the potential for age structure, using length (large vs small) as a proxy for age. No significant differences were identified. Based on these results, the Working Group agreed that the results presented were consistent with a lack of substructure within the B-C-B stock. The second question of interest to the SC relates to the degree of mixing between the B-C-B stock and the eastern Canadian Arctic (CAN) stock. Comparisons between these two strata revealed only small, and in some cases statistically insignificant, levels of genetic differentiation in both the mitochondrial and the SNP data. While this pattern could be related to historical connectivity between the two stocks, it could also, or additionally, be driven by some degree of contemporary gene flow. Some evidence of recent movements between these two regions exists (harpoon recovery, reviewed in Rugh et al. 2003; satellite tagging, Quakenbush et al. 2012). To provide increased resolution on the genetic structure within and between these two stocks, the Working Group recommended that the authors: (1) analyse the data using ordination methods, such as PCA and DAPC, which can potentially discriminate between groups with low levels of differentiation; and (2) analyse additional samples from the CAN stock in order to increase the power to detect genetic differentiation and to potentially allow for the detection of whales moving between regions via genetic mark-recapture. Frasier et al. (2015) [SC/67b/ForInfo31] was reviewed as part of a joint session with the Ad hoc Working Group on Abundance Estimates, Status and International Cruises. A summary of the discussion can be found in Annex Q under Agenda Item 3.1.1. Attention: SC, C-A The Committee reviewed the results of new genetic analyses of bowhead whales within the Bering-Chukchi-Beaufort Sea (BCB) stock and between the BCB stock and the Eastern Canadian and Okhotsk Sea stocks. The Committee: (1) agrees that the results were consistent with a lack of substructure within the B-C-B stock; (2) agrees that the results suggested that some level of historic or contemporary gene flow could exist between the B-CB and the Eastern Canadian stock; and (3) although not of immediate management concern, agrees that additional genetic analyses be conducted prior to the next Implementation Review to explore potential differentiation within and connectivity between the B-C-B and the Eastern Canadian stock, as detailed in Annex I."

Here, we present an update on bowhead population genetics to address the suggestions made by the SD/DNA working group in 2018. We have included additional BCB samples from years spanning 2018-2022 and are working toward obtaining more ECWG samples from Canadian collaborators. We are including results from ordination analyses as requested by the working group, and are not focusing on areas which have previously been agreed upon, including sub-structuring of the BCB stock.

METHODS

Sampling. —BCB samples were obtained from whales harvested during aboriginal hunts, biopsies from tagged whales, and whales found dead. Canadian (ECWG) and Okhotsk (OKS) samples were obtained from biopsies.

mtDNA sequence data.—Bowhead whale DNA was extracted, amplified, and sequenced using the methods presented in LeDuc et al. (2008) and Phillips et al. (2011). Three mitochondrial loci were used: the hyper variable region-1 (HVR1) of the mtDNA control region, the complete cytochrome-b (cytb) gene and the complete ND1 gene. These resulted in sequence lengths of 397bp, 1140bp, and 957bp, respectively. Acquired sequences were compared to existing haplotypes from previous studies deposited in GenBank and a haplotype code was assigned to each individual. For any known relatives (e.g. mother/fetus pairs), one individual was removed from further analysis. For the mtDNA analyses listed below, only individuals sequenced for all 3 mtDNA loci were included.

Arlequin version 3.5 (Excoffier et al. 2005) was used to compute mtDNA haplotype frequencies, search for shared haplotypes among stocks, compute standard diversity indices, calculate Tajima's D and Fu's Fs, and estimate among-population Fst. Arlequin was also used to conduct an analysis of molecular variance (AMOVA) to examine the source of genetic variation (either among-population or within-population).

PopART (http://popart.otago.ac.nz) was used to create a mtDNA haplotype network using the TCS network method (Clement et al. 2002).

GenAlEx (version 6.5; Peakall and Smouse 2012) was used to calculate Principal Coordinates Analysis (PCoA).

Single nucleotide polymorphisms (SNPs).—Baird et al. (2017) described the SNP panel used in this study and the quality control measures employed to determine reliability of data. In total, the panel yielded 69 autosomal loci, 1 Y-chromosome locus, and 6 X-chromosome loci of high quality. For the analyses of stock structure below, only the 69 unlinked autosomal loci were used. The sex chromosome markers will be used in future studies of bowhead population dynamics. As in the mtDNA analysis, for any known pair of relatives, one individual was removed from further analysis.

Genepop version 4.7.0 (Rousset 2008) was used to calculate genotypic and genic differentiation among all pairs of stocks and Fis (inbreeding coefficient) values per locus in each population.

Arlequin version 3.5 (Excoffier et al. 2005) was used to calculate population pairwise Fst. Although a number of diversity estimators have been developed, we follow Meirmans and Hedrick (2011) who showed that Fst is the best estimator for population differentiation based upon SNP loci.

GenAlEx (version 6.5; Peakall and Smouse 2012) was used to calculate Principal Coordinates Analysis (PCoA).

RESULTS

Mitochondrial DNA results. —The mtDNA database contained sequences from 3 loci: cytochrome-b (742 samples), ND1 (666 samples) and the HVR1 portion of the control region (984 samples). 405 BCB, 39 ECWG, and 7 OKS whale samples were sequenced for all 3 loci and only these fully completed samples were used in our analyses. In combination, the 3 loci totaled 2494 base pairs. The haplotype network shown in Fig. 1 depicts the relationships and frequencies of each haplotype. Each stock contained private/unique haplotypes. BCB shared haplotypes with both ECWG and OKS, but ECWG

and OKS did not share any haplotypes. Although there are lineages of related haplotypes unique to the BCB, there was no evidence of strong phylogenetic structure in the haplotype network. OKS and ECWG haplotypes are found throughout the network. This is consistent with previous mtDNA analyses.

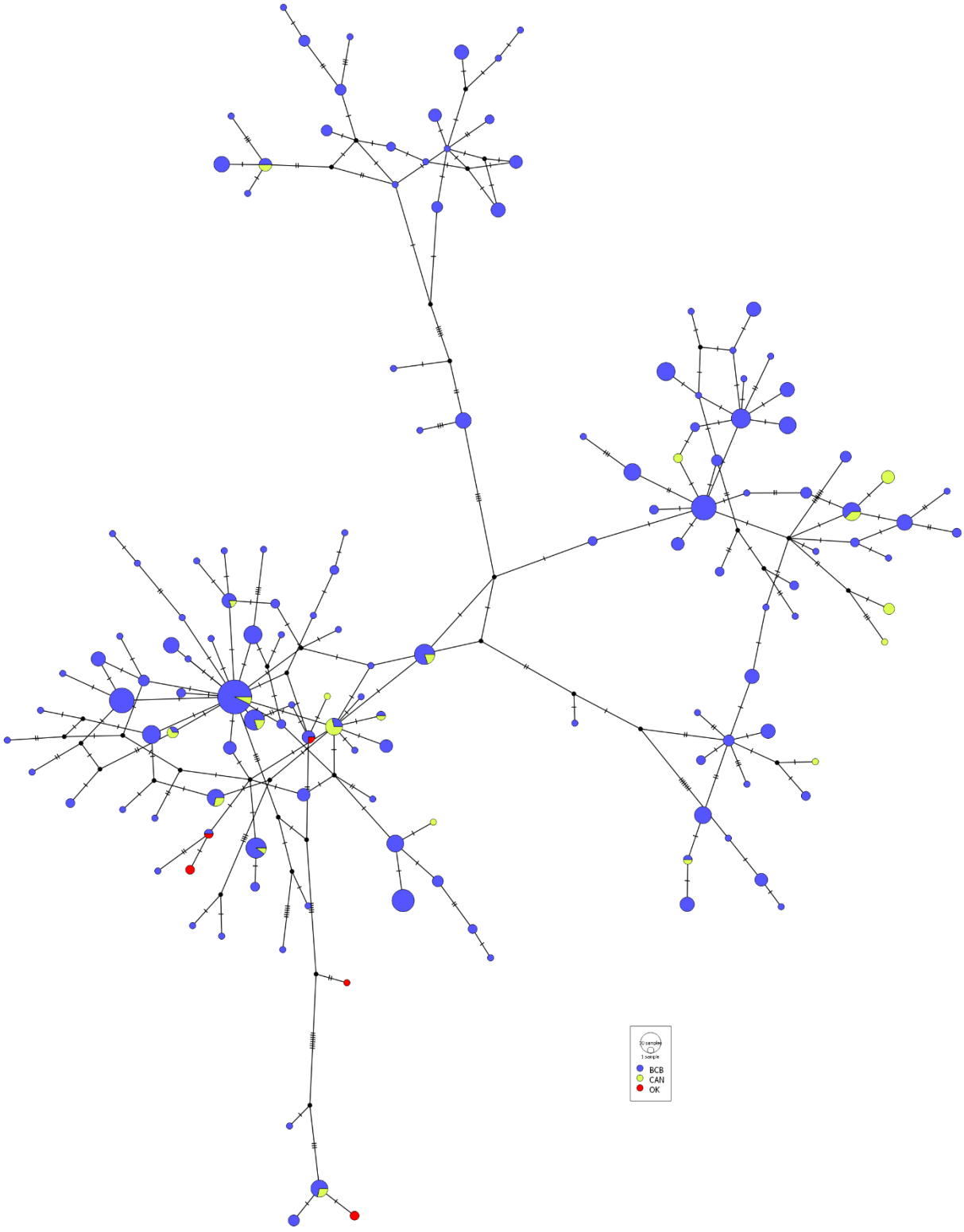

Figure 1: Haplotype network of combined Cyt-b, ND1, and HVR1 sequences. Size of circle is scaled to the observed frequency of a given haplotype. Colors indicate stocks from which the haplotype was found (see legend for key). Black circles represent unsampled haplotypes.

Within the BCB there were 120 polymorphic sites, with a nucleotide diversity (averaged over all loci) of 0.003831 +/- 0.002. Among the ECWG samples, there were 56 polymorphic sites, with a nucleotide diversity of 0.003131 +/- 0.001655. Among the OKS samples, there were 27 polymorphic sites, with a nucleotide diversity of 0.005613 +/- 0.003292.

Results of neutrality tests, including Tajima's D and Fu's FS, are shown in Table 1. Both BCB and ECWG showed significant negative values for Tajima's D and Fu's FS, while OKS was not significant.

Table 1: Results of neutrality tests for mtDNA. Data are shown by stock. P-values are in parentheses. Significant values (p<0.05) are highlighted.

	BCB (N=405)	ECWG(N=39)	OKS (N=7)
Tajima's D	-1.40149 (0.03)	-1.47608 (0.049)	1.53700 (0.969)
Fu's FS	-23.977 (0.003)	-24.93650 (0.00)	-0.99561 (0.179)

Fst was calculated using all samples and the results are shown in Table 2. Fst was significant between OKH/BCB and OKH/ECWG (p<0.05) but not significant between BCB and ECWG (p=0.06).

Table 2: Fst values for among-stock comparisons based on full mtDNA (3 loci) sequences. All BCB samples were used in this analysis. P-values are in parentheses. Significant values are highlighted (p<0.05).

	BCB (N=405)	**ECWG (N=39)**	**OKS (N=7)**
BCB	-		
ECWG	0.01039 (0.06)	-	
OKS	0.12935 (<0.001)	0.14835 (0.009)	-

AMOVA was calculated using all BCB samples, ECWG, and OKS. Those results are shown in Table 3.

Table 3: Results of AMOVA using 3 groups: all BCB samples (N=405), all ECWG samples (N=39), and all OKS samples (N=7).

Source of Variation	Degrees of freedom	% variation
Among population	2	3.26
Within population	448	96.74
p-value = 0.01466		

The results shown of the initial AMOVA using all the 3 groups had a significant p-value (p<0.05), suggesting significant structure among groups.

Principal Coordinates Analysis (PCoA) was calculated using the concatenated sequence of the 3 genes described above. The results of this analysis are shown in Figure 2. The percentage of variation explained by the first 3 axes are as follows: Axis 1 (33.04%), Axis 2 (14.98%), Axis 3 (8.01%).

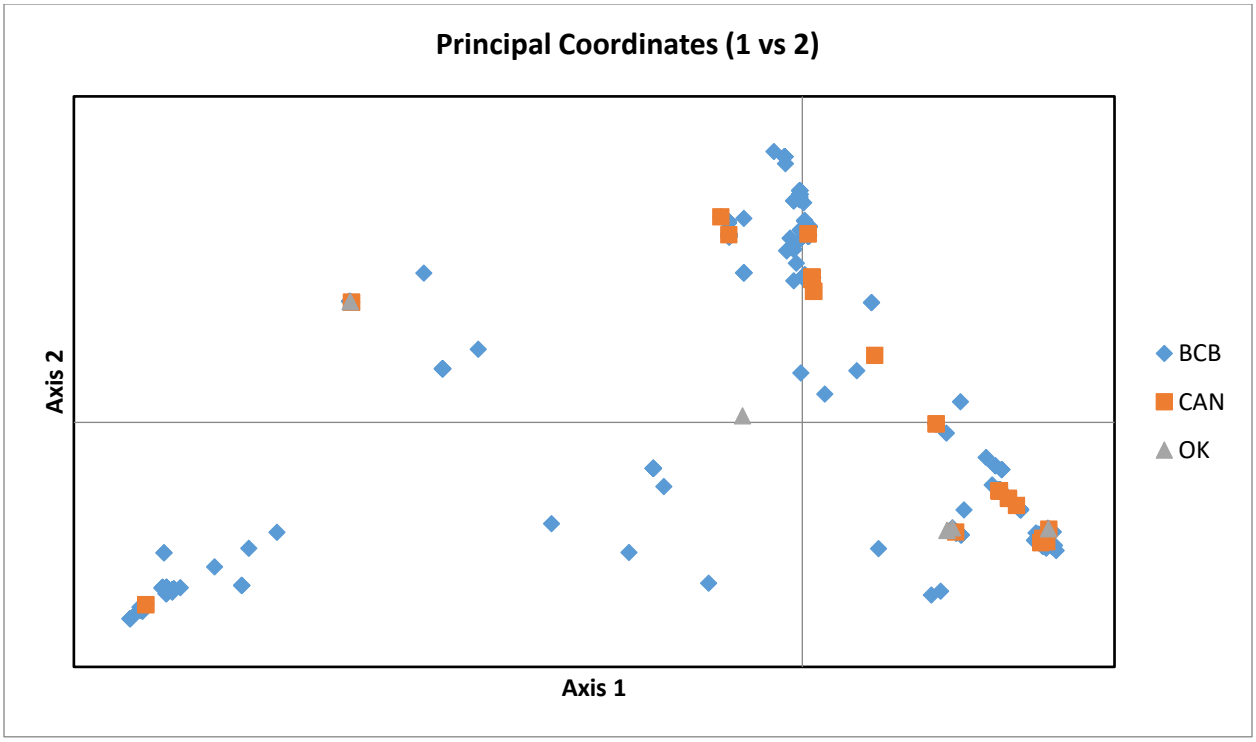

Figure 2: PCoA based on mtDNA sequences. CAN = ECWG stock; OK = OKS stock.

SNP results. —After removing duplicate samples and known close relatives, the SNP database consisted of 642 BCB, 27 ECWG, and 33 OKS samples.

In BCB, 68/69 loci were polymorphic with an average gene diversity over loci of 0.311 (+/1 0.152). For ECWG, 56/69 loci were polymorphic with an average gene diversity over loci of 0.289 (+/- 0.144). In OKS, 55/69 loci were polymorphic with an average gene diversity over loci of 0.268 (+/- 0.133).

The PCA results are shown in Figure 3. The percentage of variation explained by the first 3 axes are as follows: Axis 1 (3.12%), Axis 2 (2.98%), Axis 3 (2.89%).

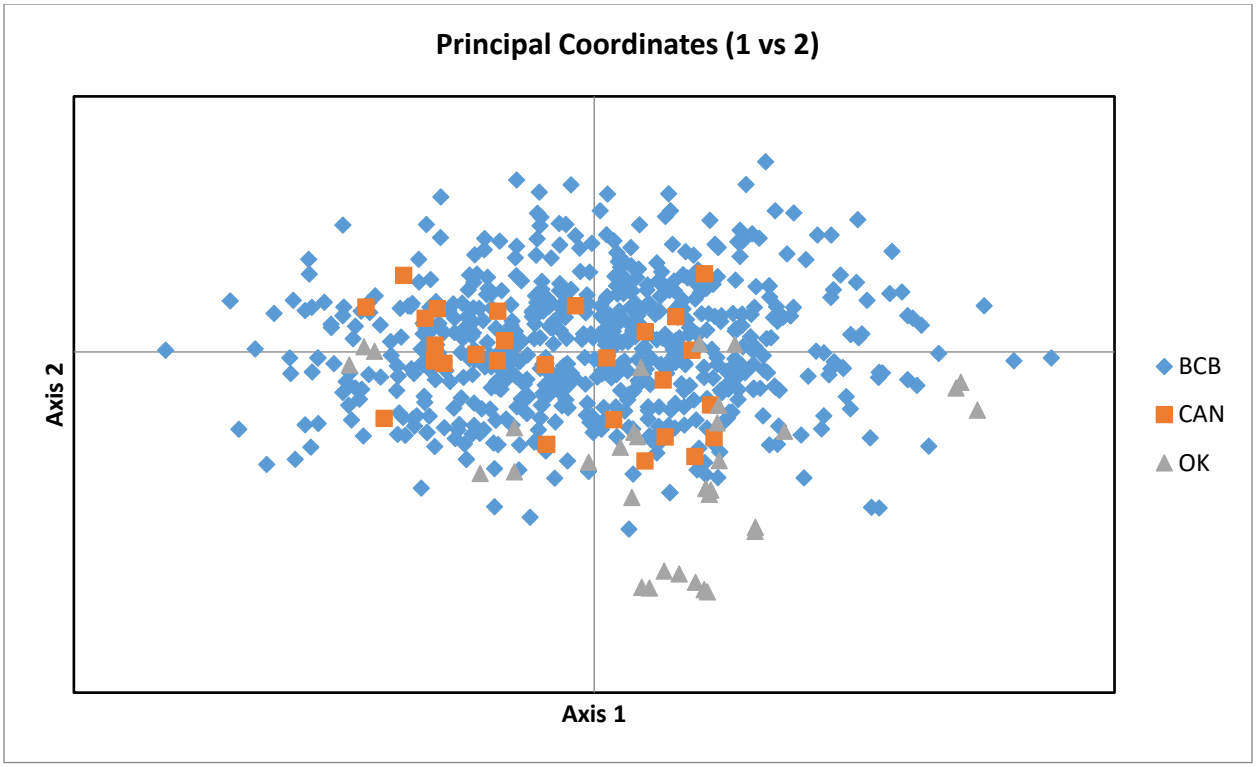

Figure 3: PCoA based on SNP data. CAN = ECWG stock; OK = OKS stock.

The population pairwise Fst values are given in Table 4. All population comparisons had significant values for Fst, including BCB/ECWG.

Table 4: Fst among pairs of stocks. All BCB samples were used in this analysis. P-values are in parentheses. Significant values are highlighted (p<0.05).

	BCB (N=642)	**ECWG (N=27)**	**OKS (N=33)**
BCB	-		
ECWG	0.00371 (0.027)	-	
OKS	0.03401(0.00)	0.0437 (0.00)	-

Analysis of genic and genotypic differentiation (using all samples) indicated that all pairs of stocks were significantly distinguishable from one another (Tables 5 and 6). The OKS stock was highly significantly different from the other 2, whereas the BCB and ECWG stocks were less so (but still significant at p<0.05).

Table 5: Genic differentiation among pairs of stocks. The analysis was done for each locus; shown are the combined p-values across all loci.

Using all BCB samples	p-value
BCB (N=642) & ECWG (N=27)	0.023
BCB & OKS (N=33)	Highly significant
ECWG & OKS	Highly significant

Table 6: Genotypic differentiation among pairs of stocks. The analysis was done for each locus; shown are the combined p-values values across all loci.

Using all BCB samples	p-value
BCB (N=642) & ECWG (N=27)	0.011
BCB & OKS (N=33)	Highly significant
ECWG & OKS	Highly significant

DISCUSSION

mtDNA. —The mtDNA haplotype network (Fig. 1) depicts 4 closely related lineages of haploytpes. As expected with the disproportionately large sampling of BCB, that stock consists of diverse haplotypes from all of those groups. The next most genetically diverse stock is ECWG, from which haplotypes in each of the lineages occur. In OKS, haplotypes are limited to one of the major lineages, although we note that sampling was extremely limited from this stock. Overall, with the possible exception of the OKS haplotypes clustered in one of the 4 lineages, there does not appear to be notable phylogeographic structuring to the haplotype network. This could be indicative of both the evolutionary history of bowhead stocks, as well as current demographics. The OKS stock is by far the smallest (Vladimirov 1994; Brownell et al. 1997), so it is not surprising that it contains fewer haplotypes. The ECWG and BCB stocks are geographically close in proximity and occasional migration could decrease the divergence between them. Clearly though, because each has private haplotypes, there is not extensive migration and exchange of haplotypes.

Results of neutrality tests, including Tajima's D and Fu's FS, demonstrate that the BCB and ECWG stocks likely have undergone recent population expansion, whereas the same cannot be said of the OKS stock. These results make biological sense given the extreme population bottleneck experienced during Yankee whaling and may also reflect recent population growth. An alternative explanation for the significantly negative values of Tajima's D and Fu's FS is that the stocks are experiencing selective sweeps. However, given what is known about the severe population bottlenecks to which these stocks were subjected, we favor the biological scenario of population expansion to explain the results.

Results from AMOVA and Fst suggest that the OKS stock is highly divergent from both BCB and ECWG.

SNPs. — Fst analyses using all samples suggest that OKS is highly distinguishable from BCB and ECWG, whereas BCB and ECWG are less so.

Genic and genotypic differentiation tests are the only analyses that clearly indicated differentiation among BCB and ECWG, though at a much less significant level than between OKS and the other stocks.

Overall conclusions and 2018 IR recommendations. — The data presented here are consistent with our previous analyses of bowhead mtDNA and SNP data, even with significantly larger numbers of BCB whales included. All recent population genetic analyses have concluded that the OKS stock is readily distinguishable from the others, but that BCB and ECWG are less distinguishable. One notable change is

that in this iteration of SNP analysis, the BCB/ECWG stocks were significantly different for Fst, though we note that the Fst value is very small.

The ordination analyses presented here have not previously been conducted on the full 3-gene mtDNA database, nor the SNP panel. Both markers show large amounts of overlap between BCB and ECWG, with OKS being slightly more differentiated. These results are not surprising, given the results from standard population genetic analyses presented here and in Baird et al. (2018). The results are also consistent with previously published mtDNA control region sequences (LeDuc et al., 2008), focal microsatellites (Givens et al., 2010), and a smaller SNP panel (Morin et al., 2012).

In 2018, it was recommended that we obtain additional ECWG samples for analysis. We agree with that recommendation and have reached out to colleagues in Canada who are amenable to that request, so work is ongoing to have samples sent to us for analysis. We anticipate that these can be added to the analyses by the next bowhead IR.

ACKNOWLEDGMENTS

We thank the Alaska Eskimo Whaling Commission (AEWC), the Barrow Whaling Captains' Association, and other village Whaling Captains' Associations for their confidence, guidance, and support of our research, especially for providing access to samples. We gratefully acknowledge funding provided by the North Slope Borough Department of Wildlife Management. We thank Andrew DeWoody for his advice on methodologies and data anlaysis. We thank Steve MacLean for the use of OKS bowhead samples.

LITERATURE CITED

Baird, A.B., G.H. Givens, J.C. George, R.S. Suydam, and J.W. Bickham. 2018. Stock structure of bowhead whales inferred from mtDNA and SNP data. Submitted to the International Whaling Commission Scientific Committee, Document SC/67B/SDDNA/01. Available from https://archive.iwc.int/pages/search.php?search=!collection29559&bc_from=themes

Baird, A.B., and J.W. Bickham. 2021. The stocks of bowheads. Pp. 19-29 in The Bowhead Whale: *Balaena mysticetus*: Biology and Human Interactions (J.G.M. Thewissen and J.C. George, eds.). London: Elsevier, xxiii + 640 pp.

Brownell, R. L., Jr., A. M. Burdin, S. A. Blokhin, and A. A. Berzin. 1997. Observations on bowhead whales (Balaena mysticetus) in the Shantar Archipelago, western Okhotsk Sea. International Marine Biological Research Institute Reports 7: 1-7.

Clement, M., Q. Snell, P. Walker, D. Posada, and K. Crandall. 2002. TCS: Estimating gene genealogies. Parallel and Distributed Processing Symposium, International Proceedings, 2, 184.

Excoffier, L., G. Laval, and S. Schneider. 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1: 47-50.

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

LeDuc RG, Martien KK, Morin PA, Hedrick N, Robertson KM, Taylor BL, Mugue NS, Borodin RG, Zelenina DA, Litovka D, George JC. 2008. Mitochondrial genetic variation in bowhead whales in the western Arctic. Journal of Cetacean Research and Management 10: 93–97.

Morin, P.A., Archer, F.I., Pease, V.L., Hancock-Hanser, B.L., Robertson, K.M., Huebinger, R.M., Martien, K.K., Bickham, J.W., George, J.C., Postma, L.D., Taylor, B.L., 2012. Empirical comparison of single nucleotide polymorphisms and microsatellites for population and demographic analyses of bowhead whales. Endangered Species Research 19, 129-147.

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

Phillips, C. D., T. S. Gelatt, J. C. Patton, and J. W. Bickham. 2011. Phylogeography of Steller sea lions: relationships among climate change, effective population size, and genetic diversity. Journal of Mammalogy 92: 1091-1104.

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

Vladmimirov, V. L. 1994. Recent distribution and abundance levels of whales in Russian far-eastern seas. Russian Journal of Marine Biology 20: 1-9.