Comparison of Methods for Molecular Assessment of Sex Chromosome Polymorphisms and levels of Genetic Diversity in the Bowhead Whale

John W. Bickham[1], Gary W. Stuart[2], Heather K. Downing[3], John C. Patton[3], John C. George[4], and Robert S. Suydam[4].

[1]Battelle Memorial Institute, Houston, TX 77042; [2]Department of Biology, Indiana State University, Terre Haute, IN 47809; [3]Department of Forestry and Natural Resources, Purdue University, West Lafayette, IN 47907; [4]North Slope Borough, Department of Wildlife Management, Barrow, AK 99723, USA

Abstract

Advances in the techniques of molecular genetics provide the opportunity to expand our understanding of heretofore poorly known components of population biology. The inheritance patterns, expected neutral mutation rates, and effective population sizes of the X and Y chromosomes differ from each other, as well as from those of both the maternally inherited mitochondrial DNA and the bi-parentally inherited autosomes which are the typical elements used in population genetics. We used two analytical approaches to discover polymorphisms in order to better comprehend the population biology of the X and Y chromosomes in bowhead whales. A targeted gene approach called exon-primed intron crossing (EPIC) was used to sequence 21,750 nucleotides for USP9Y and 11,150 nucleotides for USP9X. We observed one Single Nucleotide Polymorphism (SNP) and one single-base insertion or deletion (indel) in USP9Y and 8 variable sites in USP9X; 6 SNPs and two variable microsatellite repeats. We also sequenced the transcriptome of two bowhead whales and searched for polymorphisms among RNA transcripts of a 3,800 bp region from each of USP9X and USP9Y genes. We observed one SNP in USP9Y and 7 variable sites, including 6 SNPs and one indel, in USP9X. To estimate diversity levels for the X and Y-chromosomes we sequenced a complex microsatellite region in intron 43, 7 variable sites of intron 45, and 7 variable sites of the 3'UTR of USP9X from 15 whales, and the 2 variable sites from intron 37 and one SNP from the 3'UTR of USP9Y for 19 whales. Haplotype diversity was $H = 0.935$ for USP9X, and $H = 0.11$ for USP9Y. Variation in the X chromosome is of a level comparable to that expected from theoretical mutation rates for this element. However, much less variation than expected was observed in the Y chromosome based on theoretical mutation rates and from previous studies on human Y chromosome variation. Our data suggest that bowheads have experienced a Y-chromosome selective "sweep" in the recent evolutionary past which contrasts markedly with a previously presented estimate of 1.2 million years for the time to most recent common ancestor for mtDNA. This paper describes two systems for the analysis of X and Y chromosome variation which yielded similar results. EPIC has the advantage of being applicable to species for which no genome or transcriptome data are available, but is highly labour intensive. When genome or transcriptome data are available screening for variable sites is much faster but still requires the development of PCR primers for population assessments.

Introduction

Recent interest in genetics of the bowhead whale (*Balaena mysticetus*) has stemmed from the endangered status of populations that were decimated by commercial whaling, the need to successfully manage aboriginal hunts in Alaska and Russia, and aspects of its unique biology including its longevity. It is the longest lived mammal with estimates > 200 years for some individuals (Bockstoce and Burns, 1993; George et al., 1999; Givens et al., 2010). The understanding and preservation of the genetic diversity of species and populations are key metrics in conservation biology. Since the decline of genetic diversity is well known to be associated with increased probability of extinction, accurate measures of genetic diversity and

50 their temporal trends are useful data for population managers (Morin et al., 2012; Phillips et
51 al., 2012).
52
53 One of the most commonly used genetic markers for population genetic studies of cetaceans
54 is the mitochondrial DNA (mtDNA) control region. Since mtDNA is maternally inherited,
55 data from this marker is informative only for the female component of population genetics.
56 Roman and Palumbi (2003) used this marker to estimate long-term female effective
57 population sizes (N_{ef}) and from N_{ef} they calculated total population sizes for pre-whaling
58 populations of North Atlantic fin, humpback, and minke whales. Compared to present day
59 population estimates, historical population sizes for fin and humpback whales was 6 and 20
60 times current population estimates, respectively. If the historic population estimates based on
61 mtDNA are accurate, then the North Atlantic populations of the three whale species studied
62 by Roman and Palumbi (2003) have not recovered to near pre-whaling levels and might not
63 qualify for harvest. However, caution must be used in the interpretation of population genetic
64 estimators such as effective population size because many assumptions that go into the
65 calculations are in themselves highly uncertain. Therefore, estimates should be based not
66 upon one locus but on multiple loci including especially those linked to the different
67 mammalian inheritance patterns; mtDNA (strict maternal inheritance), bi-parentally inherited
68 autosomal loci, X-linked loci (two copies in females and one in males), and Y-linked loci
69 which are not recombining and show strict paternal inheritance. For bowhead whales
70 population genetics data exist for mtDNA (LeDuc et al., 2008; Bickham et al., 2012),
71 autosomal microsatellites (Givens et al., 2010), and autosomal Single Nucleotide
72 Polymorphisms (SNPs) (Morin et al., 2012). This paper presents the results of sequence
73 analyses of paralogous X- and Y-chromosomal loci useful in determining levels of sequence
74 diversity characteristic of these two distinct inheritance systems each of which could provide
75 a new and useful perspective on the evolutionary history and population genetics of
76 bowheads and related great whales.
77
78 Two methods of sequence analysis were used to search for variation in X- and Y-linked
79 genes including exon-primed, intron crossing (EPIC) amplification with PCR and traditional
80 DNA sequencing methods (Palumbi and Baker, 1994) and sequencing of RNA using next
81 generation transcriptome sequencing. For both analyses we selected paralogous genes
82 located in the X added region (XAR) and the Y added region (YAR) of mammals. These two
83 regions were translocated from an autosome to the X and Y chromosomes prior to the
84 diversification of Eutherian mammals approximately 105 MYA (Ross et al., 2005). The
85 genes contained in these regions do not undergo recombination and have retained similar
86 gene structure and related functions. However, they show different levels of variation and
87 different mutation rates stemming from their respective patterns of inheritance and effective
88 population sizes (Ross et al., 2005). The Ubiquitin Specific Peptidase-9 gene (USP9) was
89 selected for analysis by EPIC because of the large number of introns of a size range amenable
90 to this method. "Islands" of clustered exons are found throughout the gene, and are flanked
91 on each side by relatively small introns which were the targets of the sequencing efforts. The
92 exons provided conserved sequence for the primers and allowed for the successful
93 amplification of intervening introns by intron spanning. Clusters of these exon islands are
94 flanked by very large introns, which are too long to allow intron spanning. Regions of the
95 gene that are not accessible by intron spanning can be investigated by genomics methods. In
96 this case, we studied the 3' untranslated region (3'UTR) of USP9 through transcriptome
97 analysis.
98

USP9 functions to bind and modify ubiquitin which is a regulatory protein found in all eukaryotic cells and involved in cellular control (Hochstrasser, 2009). This protein is involved in a number of physiological processes, including conjugating with other proteins to mark them for degradation by the ubiquitin-proteasome system. USP9 codes for an enzyme which removes ubiquitin from attached proteins by hydrolyzing the isopeptide bonds. USP9X is found on the X chromosome, and USP9Y is found on the Y chromosome.

There is obvious value in direct comparisons of population genetic or evolutionary studies between maternally inherited mtDNA and paternally inherited Y-chromosomal genes. Nevertheless, there are no population genetic studies of Y-linked genes for any cetacean species and in part this is due to the difficulty of discovering polymorphic markers for the Y chromosome. This problem has only been adequately addressed in humans where the phylogeography of males has been well studied but based on an extensive effort to sequence multiple genes. In particular, Shen et al. (2000) surveyed approximately 41,000 nucleotides per individual from three genes in a sample of 70 human Y-chromosomes. They observed 51 polymorphic sites in their study, or around 1 variant per 900 base pairs. This is an indication of how many base pairs need to be surveyed to find an equivalent amount of variation, assuming that variation in bowheads is comparable to that in humans. There is evidence that baleen whales differ among species in levels of Y-chromosome polymorphism. Hatch et al. (2006) sequenced 1,049 bp of Y-chromosome DNA from two anonymous loci to investigate phylogenetic relationships among baleen whales. They examined 43 whales including 28 fin whales (*Balaenoptera physalus*), and two individuals each of minke whales (*B. borealis*), blue whales (*B. musculus*), sei whales (*Balaenoptera* borealis), Bryde's whales (*B. edeni*), humpback whales (*Megaptera novaeangliae*), gray whales (*Eschrictius robustus*) and bowhead whales. A single sperm whale (*Physeter macrocephalus*) was examined as an outgroup. The authors reported 7 Y-chromosome haplotypes for fin whales and two haplotypes each in humpback, minke and Bryde's whales. Bowheads, gray whales, blue whales and sei whales were monomorphic at these loci. Nonetheless, given the levels of variability in humans and some great whales, it is clear that we would expect to observe multiple haplotypes in bowheads with sufficient Y-chromosome sequence data. The problem is, how much sequence is enough and what is the best way to obtain it.

The overview of male evolutionary history as determined by population genetics studies of the human Y chromosome and studies of cetacean Y-chromosome phylogenies has set the framework for this study. Shen et al. (2000) have shown that polymorphic genes on the Y-chromosome are evolutionarily neutral, with few recurrent substitutions, and represent a useful system for the reconstruction of evolutionary history. And Nishida et al. (2003) and Hatch et al. (2006) used Y-chromosome linked loci to investigate the phylogenetic relationships among cetacean species. Both studies showed the utility of Y-linked markers to reconstruct accurately the phylogenetic history of cetaceans.

Likewise, the X-chromosome is potentially useful as a marker in population and evolutionary genetics. The use of paralogous genes found on the X and Y offers a convenient system of analysis because both chromosomes are present as only a single copy in males and thus haplotype identification is straight-forward even for highly variable regions. Haplotype identification is problematic when highly variable diploid loci are used, and requires the development of allele-specific sequencing primers (Baird et al., 2006), or cloning. The objectives of this study are 1) to compare two methods of sequence analysis of X- and Y-linked genes in bowhead whales, 2) identify a sufficiently large number variable sites for meaningful analysis of the paternal and X-chromosomal evolutionary histories of the

bowhead whale, and 3) to estimate relative levels of variability of bowhead X and Y chromosomes.

Methods

Search for variable sites using Exon-primed, intron crossing sequencing.--Spleen samples were obtained from 26 bowhead whales taken during subsistence hunts from 2008-2010 at Barrow, Alaska (24) and Kaktovik, Alaska (2). To extract DNA, tissue samples were incubated at 55°C in Longmire's solution with proteinase K for 24 hrs. Digested tissue samples were then extracted with phenol:chloroform and precipitated as described in Sambrook et al. (2001).

After extraction, samples were amplified for different regions along the USP9 gene using primers from adjacent exons for intron spanning (Tables 1 and 2, and Appendx 1). The PCR protocol was essentially that of Cronin et al. (1996), as modified by proper annealing temperatures of individual primer pairs. Appendix 1 lists the primers and temperatures used for amplification across the spanned introns. Amplification mixture consisted of 5 µl of each primer, 10X dNTPs, 10X BSA, and10X salt solution, respectively, 1.5 units Taq polymerase, and 25 µl Sigma water, for a total reaction volume of 50 µl per sample. PCR reactions were carried out on a BioRad MyCycler or an ABI GeneAmp 2700 thermocycler. The cycle-sequence protocol was: 4 min at 94°C, followed by 36 cycles of a 30 sec denaturation step at 94°C, a 30 sec annealing step of 50-65°C (depending on specific needs of the primer pair), and a 1.5-3.5 min extension step at 70°C, depending on the expected size of the products. For a few primer pairs, a touchdown procedure was used, where after three cycles the initial annealing temperature was reduced by three degrees for the subsequent three cycles, and then returned to the previous annealing temperature for the duration of the amplification reaction. PCR primers were mainly developed from GenBank data using the cow sequence. Amplified DNA products were separated by gel electrophoresis in a 0.8% agarose medium and stained with ethidium bromide; bands were visualized on a BioRad GelDox XR imaging system.

PCR products were either excised from the gel and purified with a Qiagen™ Band Excision Clean-up Kit (Fig. 1) or directly purified via Qiagen™ PCR Clean-up kit, depending on the complexity of the generated bands. An appropriate amount of purified DNA was then dried down and an ABI BigDye v 3.1 sequencing reaction was performed using the recommended protocol. Sequencing was performed on an ABI 3730 automated sequencer at the Purdue Genomics Core Facility. The sequences obtained were then blasted against known sequences in GenBank to confirm that they were indeed USP9 and determine their X or Y-chromosome specificity. X and Y specific primers used in further amplifications or sequencing were developed from these preliminary sequences as a perfect match to the *Balaena mysticetus* sequence; X- and Y-chromosome primer sequences are reported in Tables 2 and 3, respectively. Sequences of amplification products as well as those taken from GenBank were aligned and assembled using Sequencher™ 4.7.

Search for variable sites using RNA sequencing of X- and Y-chromosome genes.—The entire USP9X (approximately 11,500 bp) and USP9Y (approximately 8,500 bp) genes were sequenced from two male bowheads using transcriptome sequencing methods. The assembly was constructed from 51 contigs and represents both genes with the exception of a few gaps. We searched for variable sites within an alignment of approximately 3,800 bp which included the 3'Untranslated Region (3'UTR) of the USP9X and USP9Y from the two individuals.

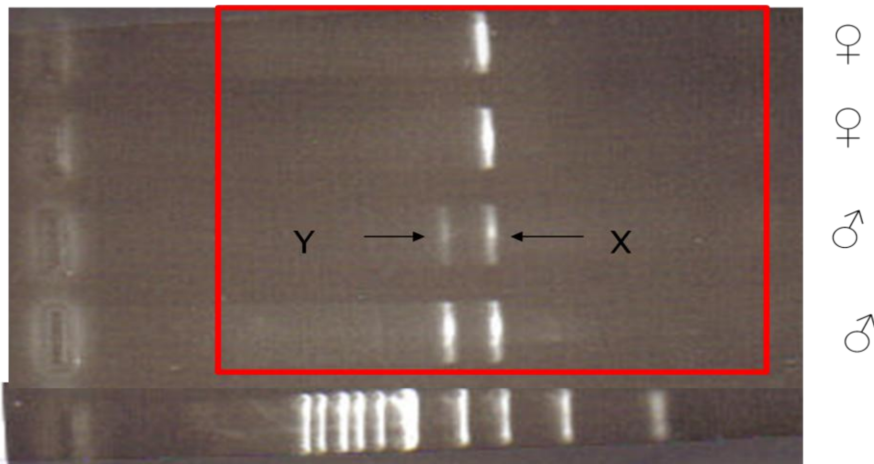

Figure 1.—The results of amplification of a fragment of the USP9Y and USP9X genes in the region between exons 14 and 15. Different banding patterns can be observed in males and females, with the common band being from USP9X and the band exclusively seen in male samples being from the USP9Y gene.

Assessment of X and Y Haplotype Diversity in Bowhead Whales.—Haploytpe diversity was assessed for variable sites in 3 regions of the X-chromosome (USP X intron 43, intron 45, and 3'UTR). Variable sites within the 3'UTR were initially identified from transcriptome sequences. PCR primers were subsequently designed to amplify the region of each variable site to allow efficient survey. Haplotype diversity was calculated using all three regions combined from a subset of individuals. Haplotype diversity was also calculated for the Y-chromosome using variable sites found in intron 37 (2 sites) and the 3'UTR (1 site).

Results

The EPIC method (Palumbi and Baker, 1994) provided successful amplifications of nine regions of USP9Y and eight regions of the USP9X. Figure 1 shows the products of amplification of the fragment between exons 14 and 15 run out on a 0.8% agarose gel. The difference in banding patterns on the gel of amplified USP9X and USP9Y regions for males and females are due to differing sizes of the introns for the two paralogous genes. The common band between the four individuals is from USP9X. The extra band of the two male samples is a region from USP9Y. Band excision and gel DNA extraction was used to separate and sequence the X and Y-chromosome fragments of the amplifications.

Using EPIC derived data we sequenced more than 32,000 nucleotides for USP9Y and USP9X for varying numbers of individuals. As can be seen in Table 1, this includes approximately 21,750 nucleotides sequenced for nine regions of USP9Y. For USP9Y we sequenced a total of 138,350 bases which yields 6.36 X coverage of the 9 regions studied. Two variable positions were found in intron 37; an A/T transversion was found at position 295 and a single base indel (deletion or insertion) was found at position 544. Table 2 reports the results of EPIC sequence analysis of 11,150 base pairs for USP9X. For USP9X we sequenced a total of 94,150 bases which yields 8.44 X coverage of the 8 regions studied. Eight variable sites were found; two transitions in intron 18 (A/G,C/T), one transition in intron 25 (C/T), one transition in intron 37 (A/G), a highly complex microsatellite region in intron 43, and in intron 45 there were three transitions (A/G, A/G, C/T) and a variable tetrameric microsatellite repeat (TATG). For the complex microsatellite region in intron 43 (Figure 2), we sequenced

23 males from which 11 haplotypes were identified (Table 3). Because males have only one X chromosome, these sequences each represent distinct haplotypes, thus allowing unambiguous identification of the variants on the X-chromosome of each male. This is advantageous, as interpreting the complex patterns in a female heterozygote could be difficult.

```
GCAGGTGAGGGTCTCTCTCCCTCTCTCTCTCTTGCTCTCTCTTGCTCTCTCTTGCTCTCTCTTGCTCTCTCTTGCTCTCTCTTGCTCTC
GCAGGTGAGGGTCTCTCTCCCTCTCTCTCTCTTGCTCTCTCTTGCTCTCTCTTGCTCTCTCTTGCTCTCTCTTGCTCTCTCTTGCTCTC
GCAGGTGAGGGTCTCTCTCCCTCTCTCTCTCTTGCTCTCTCTTGCTCTCTCTTGCTCTCTCTTGCTCTCTCTTGCTCTCTCTTGCTCTC
GCAGGTGAGGGTCTCTCTCCCTCTCTCTCTCTTGCTCTCTCTTGCTCTCTCTTGCTCTCTCTTGC::::::::::::::::::::::::
GCAGGTGAGGGTCTCTCTCCCTCTCTCTCTCTTGCTCTCTCTTGCTCTCTCTTGCTCTCTCTTGC::::::::::::::::::::::::
GCAGGTGAGGGTCTCTCTCCCTCTCTCTCTCTTGCTCTCTCTTGCTCTCTCTTGCTCTCTCTTGC::::::::::::::::::::::::

TCTTGC::TCTCTCTCTCTCGCGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTC:::::TATATATATATATGTGTATATATATAC
TCTTGC::TCTCTCTCTCTCGCGCGCGCGC:::TCTCTCTCTCTCTCTCTCTCTC::TATATATATATATATATGTGTATATATATAC
TCTTGC::TCTCTCTCTCTCGCGCGCGCGC:::::TCTCTCTCTCTCTCTCTCTCTATATATATATATATATATGTGTATATATATAC
::::::TCTCTCTCTCTCT::CGCGCGCGC:::::TCTCTCTCTCTCTCTCTCTC::TATATATATATATATATGTGTATATATATAC
::::::TCTCTCTCTCTCT::CGCGCGCGC:::TCTCTCTCTCTCTCTCTCTCTC:::::TATATATATATATGTGTATATATATAC
::::::TCTCTCTCTCTCT::CGCGCGCGC:::::::::TCTCTCTCTCTCTCTCTATATATATATATATATATGTGTATATATATAC

ACACACACATATAATTTTGTAGAAATCT
ACACACACATATAATTTTGTAGAAATCT
ACACACACATATAATTTTGTAGAAATCT
ACACACACATATAATTTTGTAGAAATCT
ACACACACATATAATTTTGTAGAAATCT
ACACACACATATAATTTTGTAGAAATCT
```

Figure 2.—Nucleotide sequence of a segment of intron 43 of USP9X of 6 bowhead whales showing allelic variation in a complex repeat. Six variable microsatellites were identified and scored in this region, of which 5 are illustrated here by different colors. A sixth variable site was identified in one whale (not shown).

Table 1.—The number of sequenced base pairs for 2-12 bowhead whales in 9 regions of the USP9Y gene using the targeted gene method. "EX" denotes an exon.

Region	**Fragment size (bp)**	**Variable sites**
EX 8- EX 11	**2,400**	
EX 14- EX 15	**1,800**	
EX 17- EX 20	**3,600**	
EX 23- EX 24	**550**	
EX 25- EX 26	**950**	
EX 28- EX 31	**2,800**	

269

EX 34- EX 39	**7,200**	**A/T transversion, single base indel**
EX 40- EX 42	**950**	
EX 43- EX 46	**1,500**	
TOTAL:	**21,750 nucleotides**	**2 variable sites**

270
271 Table 2.—The number of sequenced base pairs for 2-12 individuals in the USP9X gene using
272 the targeted gene method. "EX" denotes an exon, "in" denotes an intron.

Region	Fragment Size (bp)	Variable Sites
EX 14- EX 15	**1,300**	
EX 18- EX 19	**1,000**	**A/G, C/T transitions**
EX 25- EX 26	**1,050**	**C/T transition**
EX 28 -EX 31	**2,100**	
EX 36- EX 38	**2,150**	**A/G transition**
EX 40-in41	**1,000**	
EX 43- EX 45	**1,250**	**Complex repeat with six variable microsatellites: (GC(TC)3TT)N, (CT)N, (CG)N, (TC)N, (TA)N, (TG)N**
EX 45- EX 46	**1,300**	**G/A, G/A, G/A, G/T, C/T SNPs; 3-base indel, 76-base indel, 1 variable tetrameric microsatellite**
TOTAL:	**11,150 nucleotides**	**18 variable sites**

273
274 Using transcriptome sequencing we searched for variable sites within a 3,800 bp segment that
275 included the 3'UTR of USP9X and USP9Y for 2 individuals. We found a single variable
276 site, a G/A transition, in USP9Y, and 8 variable sites in USP9X including 6 SNPs (T/A, T/C,
277 T/C, G/C, A/C, T/G) and 2 indels (one single base and one 2-base indel).
278
279 **Discussion**
280 One objective of this study was to develop a system for sequence analysis of X- and Y-linked
281 genes to include a sufficiently large number of nucleotides for meaningful documentation of
282 the paternal and X-chromosomal evolutionary histories of the bowhead whale. We employed
283 two methods of SNP discovery, namely the targeted gene approach using EPIC and the
284 transcriptome approach. Using the targeted gene approach, which is a method that can be
285 applied to organisms for which genome sequence is not available, we report sequence
286 variation for approximately 21,000 bp of the bowhead Y and 11,000 bp of the bowhead X.
287 These 32,000 nucleotides are located in several 'island' regions of exons, with spans of

288 introns of generally less than 3 kilobase pairs (kbp) in length, allowing the use of a standard
289 PCR protocol to amplify them.
290
291 Table 3.—Haplotypes identified from 23 whales sequenced for a complex repeat containing 6
292 variable microsatellite loci in intron 43 of USP9X. The numbers in the columns refer to the
293 repeat size and the colors of the columns refer to the colors identifying the loci in Figure 2.
294 (TG)N is not figured.

SAMPLE	$(GC(TC)_3TT)_N$	$(CT)_N$	$(CG)_N$	$(TC)_N$	$(TA)_N$	$(TG)_N$	in43 X hap
BA MYS 10KK01	3	5	5	10	8	2	AABCCA
BA MYS 10KK02	3	6	4	8	9	2	ABAADA
BA MYS 08B07	6	5	5	10	9	2	BABCDA
BA MYS 09B06	6	5	5	13	6	2	BABFAA
BA MYS 10B04	3	6	4	11	6	2	ABADAA
BA MYS 10B07	3	6	4	8	9	2	ABAADA
BA MYS 10B08	6	5	5	11	8	2	BABDCA
BA MYS 10B16	3	6	4	10	8	2	ABACCA
BA MYS 10B17	6	5	5	11	8	2	BABDCA
BA MYS 11B01	6	5	5	11	8	2	BABDCA
BA MYS 11B02	3	6	4	10	7	2	ABACBA
BA MYS 11B10	6	5	5	11	8	2	BABDCA
BA MYS 11B11	6	5	5	12	7	2	BABEBA
BA MYS 11B12	6	5	5	13	6	2	BABFAA
BA MYS 11B13	3	6	4	10	8	2	ABACCA
BA MYS 11B14	6	5	5	10	8	2	BABCCA
BA MYS 11B15	6	5	5	11	8	2	BABDCA
BA MYS 11B16	6	5	5	11	8	2	BABDCA
BA MYS 12B03	6	5	5	12	7	2	BABEBA
BA MYS 12B06	6	5	5	10	8	2	BABCCA
BA MYS 12B11	6	5	5	11	8	2	BABDCA
BA MYS 12B13	3	6	4	10	8	2	ABACCA
BA MYS 12B14	5	6	5	10	8	3	CBBCCB

295
296 On the Y chromosome, 21,750 base pairs of the USP9 gene were sequenced with 6.36 X
297 coverage (i.e., 6.36 individuals were sequenced on average). From these sequences, we
298 observed two variable sites in intron 37. Position 295 was a T/A SNP in which 3 whales had
299 T and 26 whales had A. Position 455 was a one-base indel in which one whale had an
300 apparent insertion of an A and 20 individuals lacked the insertion (Table 1). This compares
301 with our survey of 11,150 nucleotides on the USP9X gene with 8.44 X coverage where a total
302 of 9 SNPs, two indels, one variable microsatellite, and one complex microsatellite region
303 were discovered. For the complex microsatellite in intron 43 we sequenced 23 males and
304 observed 11 different haplotypes (Table 3). Because the X in males is haploid it is possible
305 to obtain unambiguous sequence data for each variant of the complex repeat (Figure 2). For
306 intron 45 of the USP9X we analyzed 7 variable sites, including 5 SNPs and 2 indels from16
307 whales. A total of 9 haplotypes were identified (Table 4).
308

309 The transcriptome analysis yielded sequence of the entire 11,500 bp mRNA of the USP9X
310 gene and the entire 8,500 bp of the USP9Y mRNA. We examined 3,800 nucleotides of RNA
311 transcripts from the 3' end of the USP9X and USP9Y genes for variable sites from 2 whales.
312
313 Table 4.—Haplotpes identified from 16 male bowhead whales sequenced for intron 45 of
314 USP9X. There are 7 variable sites including 5 SNPs and 2 indels. The sequence position for
315 each variable site is given on the top row.

Whale	133	308	474	740	743 to 755	1012	1376 to 1542	Haplotype
BA MYS 10KK01	G	G	A	G	-	C	+	H
BA MYS 10KK02	G	G	A	G	+	C	-	F
BA MYS 10B16	G	G	A	G	-	C	+	H
BA MYS 10B17	G	G	A	G	+	T	+	A
BA MYS 11B01	G	G	A	G	+	T	+	A
BA MYS 11B02	G	G	G	G	+	C	+	G
BA MYS 11B10	G	G	A	G	+	T	+	A
BA MYS 11B11	G	A	A	G	+	C	+	C
BA MYS 11B12	G	A	A	G	+	C	+	C
BA MYS 11B13	G	G	A	G	+	C	+	E
BA MYS 11B14	G	G	A	G	+	C	+	B
BA MYS 11B15	G	G	A	T	+	T	+	K
BA MYS 12B03	G	A	A	G	+	T	+	D
BA MYS 12B11	G	G	A	G	+	T	+	A
BA MYS 12B13	A	G	A	G	-	C	+	J
BA MYS 12B14	G	G	A	G	+	C	+	E

316
317 From the Y chromosome we identified a G/A SNP at position 366. Twenty whales were
318 eventually sequenced for this marker; 3 had A and 17 had G. On the X-chromosome we
319 identified 7 variable sites including six SNPs (A/T, T/C, T/C, G/C, A/C, and G/T) and one
320 single-base pair indel. These 7 variable sites defined 4 X-chromosome haplotypes from 16
321 sequenced whales.
322
323 The neutral mutation rate of the Y-chromosome is expected to be greater than the autosomes
324 which in turn are expected to be greater than the X-chromosome. This is because 100% of
325 the Y-chromosomes pass through the male, 50% of the autosomes pass through the male, and
326 33.3% of the X-chromosomes pass through the male. Higher mutation rates are predicted for
327 the Y-linked genes because of the greater number of germ cell divisions in spermatogenesis
328 than in oogenesis (Miyata et al., 1987). Thus, the higher the percentage of chromosomes in a
329 population that have passed through the male, the higher the relative mutation rate. Because
330 spermatogenesis produces so many more gametes than oogenesis, mutations are much more
331 likely to occur in the male lineage. This relationship of Y>autosomes>X has been confirmed
332 empirically (Miyata et al., 1987). One of the most thorough studies to have compared X and
333 Y-chromosome sequence evolution (Slattery and O'Brien, 1988) showed Y-chromosomal
334 sequences evolved 2 times faster than X-chromosomal sequences in a paralagous gene pair,
335 Zfx and Zfy. But in population genetics, variation within a species is due to both mutation
336 rate and effective population size N_e. Because N_e for males is much smaller than females,
337

338 Table 5.—Haplotpes identified from 16 male bowhead whales sequenced for the 3'UTR of
339 USP9X. There are 7 variable sites including 6 SNPs and a single base indel. The sequence
340 position for each variable site is given on the top row.

Whale	60	161	178	222	237	974	1342	Haplotype
BA MYS 10KK01	T	C	-	C	G	A	T	B
BA MYS 10KK02	A	T	+	T	C	C	T	A
BA MYS 10B16	T	C	-	C	G	A	T	B
BA MYS 10B17	A	T	+	T	C	C	T	A
BA MYS 10B20	A	T	+	T	C	C	T	A
BA MYS 11B01	A	T	+	T	C	C	T	A
BA MYS 11B02	A	T	+	T	C	C	G	C
BA MYS 11B10	A	T	+	T	C	C	T	A
BA MYS 11B11	A	T	+	T	C	C	T	A
BA MYS 11B12	A	T	+	T	C	C	T	A
BA MYS 11B13	A	C	+	T	C	C	T	D
BA MYS 11B14	A	T	+	T	C	C	T	A
BA MYS 11B15	A	T	+	T	C	C	T	A
BA MYS 12B11	A	T	+	T	C	C	T	A
BA MYS 12B13	A	T	+	T	C	C	T	A
BA MYS 12B14	A	T	+	T	C	C	T	A

341
342 due to the high variance in reproductive success for males, new Y-chromosome variants
343 come to fixation very quickly. Thus, the number of variants within a population will be
344 reflected in the time since the last Y-chromosome sweep or the Time to Most Recent
345 Common Ancestor (TMRCA; Walsh, 2001).
346
347 The two methods (transcriptome and targeted gene method) used to identify variable sites on
348 the X and Y- chromosomes of bowheads yielded similar results. For equivalent rough order-
349 of-magnitude sequencing efforts we observed about 7-8 times as many variable sites on the
350 X-chromosome as on the Y-chromosome. To estimate diversity levels for the X -
351 chromosome we sequenced a complex microsatellite region in intron 43, 7 variable sites of
352 intron 45, and 7 variable sites of the 3'UTR of USP9X from varying numbers of whales but
353 including 15 whales for all sites (Tables 3-5). To estimate diversity for the Y-chromosome
354 we sequenced the 2 variable sites from intron 37 and one variable site from the 3'UTR of
355 USP9Y for 19 whales (Table 6). These data revealed 4 distinct USP9Y haplotypes as
356 follows: haplotype A, 14 whales; haplotype B, 1 whale; haplotype C, 2 whales; haplotype D,
357 2 whales. From the fifteen bowheads sequenced for all variable sites in intron 43, intron 45
358 and the 3'UTR, there were 11 haplotypes as follows: haplotype 1, 4 whales; haplotype 2, 2
359 whales; haplotypes 3-11, one whale each. For comparison, we calculated by hand the
360 haplotype diversity estimates for the X and Y chromosomes of bowheads using the combined
361 datasets using the following equation:
362

$$H = \frac{N}{N-1}(1 - \sum_i x_i^2)$$

363
364
365 For the X chromosome $H = 0.935$, and for the Y chromosome $H = 0.11$.
366

367 Table 6.—Haplotypes determined from 3 variable sites in the USP9Y gene for 19 bowhead
368 whales. The variable sites included 2 SNPs and one indel.

	intron 37		3'UTR	**Haplotype**
SAMPLE	295	544	366	
BA MYS 10KK01	A	-	G	A
BA MYS 10KK02	A	-	G	A
BA MYS 08B07	A	-	A	C
BA MYS 10B04	A	-	G	A
BA MYS 10B07	A	-	G	A
BA MYS 10B16	A	-	G	A
BA MYS 10B17	T	-	G	D
BA MYS 11B01	A	-	G	A
BA MYS 11B02	A	-	G	A
BA MYS 11B10	A	-	G	A
BA MYS 11B11	A	+	G	B
BA MYS 11B12	T	-	G	D
BA MYS 11B13	A	-	G	A
BA MYS 11B14	A	-	G	A
BA MYS 11B15	A	-	G	A
BA MYS 12B03	A	-	G	A
BA MYS 12B11	A	-	G	A
BA MYS 12B13	A	-	G	A
BA MYS 12B14	A	-	A	C

369
370
371 In comparison to the study by Shen et al. (2000), it is clear that bowhead whales have less
372 variation in the Y-chromosome than is seen in humans. Shen et al. (2000) observed
373 approximately 1 variant per 900 base pairs in the SMCY gene on the human Y chromosome.
374 In bowheads we observed only 2 variants per 21,750 base pairs using the targeted gene
375 method, and one variable site in 7,600 base pairs of the transcriptome. These studies are not
376 comparable on the basis of sample size. The sample size for the human study is 1,000
377 individuals. Our sample size from the Y-chromosome averages 6.36 individuals over the
378 21,750 base pairs studied and was only 2 individuals for the transcriptome method.
379 Nonetheless,
380
381 **Conclusions**
382
383 This study describes methods to analyze X and Y-chromosomal genetic loci for population
384 studies of bowhead whales. The conserved PCR primers will allow the development of
385 equivalent analyses in other cetacean species. In addition, the data show that X-chromosome
386 haplotype diversity is high in bowheads but Y-chromosomal diversity is very low. This is
387 most likely due to a recent selective sweep and indicates that male reproductive success likely
388 is highly variable. This would be expected if sperm competition plays a role in bowhead
389 reproductive strategy as has been suggested and it is consistent with the observation of "super
390 males" with extremely large testes.
391
392
393

Acknowledgments
We thank the Alaska Eskimo Whaling Commission (AEWC) and the Barrow Whaling Captains' Association for their confidence, guidance and support of our research. We gratefully acknowledge funding provided by the North Slope Borough Department of Wildlife Management and National Oceanic and Atmospheric Administration (through the AEWC).

Literature Cited

Baird, AB, Hillis DM, Patton JC Bickham JW (2009). Speciation by monobrachial centric fusions: A test of the model using nuclear DNA sequences from the bat genus *Rhogeessa*. *Molecular Phylogenetics and Evolution,* 50, 256-267.

Bickham JW, Huebinger RM, Philips CD, Patton JC, Postma LD, George JC, Suydam RS (2012) Assessing molecular substitution patterns in the mitochondrial control region compared to protein coding genes in bowhead whales: update of SC/63/BRG13. Paper SC/64/AWMP9 submitted to the International Whaling Commission Scientific Committee. 12pp.

Bockstoce, JR, Burns, JJ (1993) Commercial whaling in the North Pacific sector. In J.J.Burns, J.J. Montague, and C.J. Cowles, eds. *The bowhead whale*. Allen Press, Lawrence. Pp. 563-576.

Cronin, MA, Bodkin J, Ballachey B, Estes J, Patton JC (1996) Mitochondrial DNA variation among subspecies and populations of sea otters (*Enhydra lutris*). *Journal of Mammalogy*, 77, 546-577.

George, JC, Bada, J, Zeh, J, Scott, L, Brown, SE, O'Hara, T, Suydam, R (1999) Age and growth estimates of bowhead whales (Balaena mysticetus) via aspartic acid racemization. *Canadian Journal of Zoology*, 77, 571-580.

Givens GH, Huebinger RM, Patton JC, Postma LD, Lindsay M, Suydam RS, George JC, Matson CW, Bickham JW (2010) Population genetics of bowhead whales (*Balaena mysticetus*) from the western Arctic. *Arctic* 63, 1-12.

Hatch LT, Dopman EB, Harrison RG (2006) Phylogenetic relationships among the baleen whales based on maternally and paternally inherited characters. *Molecular Phylogenetics and Evolution* 41, 12-27.

Hochstrasser, M (2009). Origin and function of ubiquitin-like proteins. *Nature* 458, 422–429.

Miyata T, Hayashida H, Kumar K, Mitsuyasu K, Yasunaga T (1987) Male-driven molecular evolution: A model and nucleo tide sequence analysis. Cold Spring Harbor Symp. Quant. Biol. 52, 863–867.

Morin PA, Archer FI, Pease VL, Hancock-Hanser B, Robertson KM, Huebinger RM, Martien KK, Bickham JW, George JC, Postma LD, Taylor BL (2012) An empirical comparison of SNPs and microsatellites for population structure, assignment, and demographic analyses of bowhead whale populations. Paper SC/64/AWMP3 submitted to the International Whaling Commission Scientific Committee. 27 pp.

444 Nishida S, Pastene LA, Goto M, Hiroko K (2003) SRY gene structure and phylogeny in the
445 cetacean species. *Mammal Study* 28, 57–66.
446
447 Palumbi SR, Baker SC (1994) Contrasting population structure from nuclear intron
448 sequences and mtDNA of humpback whales. *Molecular Biology and Evolution* 11, 426-435.
449
450 Phillips CD, Hoffman JI, George JC, Suydam RS, Huebinger RM, Patton JC, and Bickham
451 JW (2012) Molecular insights to historical demography in Bowhead whales. Paper
452 SC/64/AWMP1 submitted to the International Whaling Commission Scientific Committee.
453 20 pp.
454
455 Ross MT et al. (2005) The DNA sequence of the human X chromosome. *Nature,* 434, 325-
456 337.
457
458 Sambrook J, MacCallum P, Russel DW (2001) Molecular cloning: A Laboratory Manual
459 Coldspring Harbor Press, 3rd addition.
460
461 Semino et al. (2010) The Genetic Legacy of Paleolithic *Homo sapiens sapiens* in Extant
462 Europeans: A Y chromosome Perpsective. *Science*, 290, 1155-1159.
463
464 Shen, et al. (2000) Population genetic implications from sequene variation in four Y
465 chromosome genes. *PNAS*, 97, 7354-7359.
466
467 Slattery JP, O'Brien SJ (1998) Patterns of *Y* and *IX* chromosome DNA sequence divergence
468 during the Felidae radiation. *Genetics* 148**,** 1245–1255.
469
470 Thompson R, Pritchard JK, She P, Oefner PJ, Feldman MW (2000) Recent common ancestry
471 of human Y chromosomes: Evidence from DNA sequence data. *PNAS*, 97, 7360-7365.
472
473 Underhill PA, Passarino G, Lin AA, Shen P, Mirazon Lahr M, Foley RA, Oefner, PJ, Cavalli-
474 Sforza LL (2000) The phylogeography of Y chromosome binary haplotypes and the origins
475 of modern human populations. *Annals of Human Genetics*, 65, 43-62.
476
477 Walsh, B. 2001. Estimating the Time to the Most Recent Common Ancestor for the Y
478 chromosome or Mitochondrial DNA for a Pair of Individuals. *Genetics,* 158, 897–912.

479 Appendix 1.—List of intron-spanning PCR primer sequences, their annealing temperatures, and the
480 chromosome or chromosomes from which they amplify a product. The primers are placed in exons as listed,
481 and they span the intervening intron the size of which is listed for the human genome.
482

Span	Human Intron Size	Primer 1	Sequence	Primer 2	Sequence	Temp	Whale Amp
Exon 3-4	11,044						
Exon 4-5	1368	USPX4Y F	TCACCTGATTCTTCCA ATGAGA	USPX5 R	CAAAACAGGAACCACCCAT C	60	No
		USPX4 F	CTTCATCGCCTGATTC ATCC			60	No
Exon 5-6	2923						
Exon 6-7	1350	USPX6 F	ATGGATGAGGCTGTG AGTGG	USPX7 R	GATTCACATGGACGTGTAC CA	60	No
Exon 7-8	8819						
Exon 8-9	312	**USPX8** F	TGGCACAYTAAATGG GTTCC				
		USPX8 F2	GATCGTTTTATTAATG GATCAGC			57	
Exon 9-10	100			USPX10 R	GCTCAGCTGTCAGCCATTCT	60	XY
		USPX9 F	CCATTTGGACAATGCT ATGAG				
Exon 10-11	2417			USPX11 R	CCTGTGCTGCCCAGATATTA	58	Y
Exon 11-12	1121	USPX11 F	GGAATGGATACAGCA AAAYAATATC	USPX12 R	GGAGAAAAATYCCAAGCCA AY	57	No
Exon 12-13	17,736						
Exon 13-14	1105	USPX13 F	CAAATGCAAGTAAAA AGCAACG	USPX14 R	GCWGGAATTACCCAYTTGT CA	60	No
Exon 14-15	1835	USPX14 F	GAYCGAGAYACACAA AAGATCCAGTGG	USPX15 R	ATGTAAGTTGCAAGGTTTTC TGCTACC	60	XY
		USPX14 F2	TTCGCACAAATGACAA GTGG	USPX15 R2	TTCTGCTACCAAAGTAACTA AAGCA	59	No
Exon 15-16	10,453						
Exon 16-17	2426	USPX16 F	YCCACAAACAGTGAG GCTWG	USPX17 R	TCCWATTAATTCCAAATCAT CCA	59	No
		New USP9X16F2 Y	CCACAAACTGTGAGG CTTGGA	New USP9X17R2 Y	TTTCCCTTCTCGACAGTTAA CWGC	63	
Exon 17-18	1543	USPX17 F	GGAAGTGCTTAGCAG AAAATGC	USPX18 R	CCTGTAGCCTCGGACCAA		X
Exon 18-19	1078	USPX18 F	AGATTGCCAGCAGAG CTATAG	USPX19 R	TTGACATAGGTAGAATCAT YCTTTC		No
		New USPX18 F	ATGATGAYATTGCCAR CAGAGCTA	New USP9X19 R	CCCTTAAAACAGTYAATACT CGAACCA	65	
Exon 19-20	1157	USPX19 F (old)	GATACATTGTGTGTTT TGGATGGTGA	USPX20 R (old)	GGCTACATTTGCTTTTATAC GACTG		X
		New USPX19 F	TYCAGTCTTGCTTTGA TCGTTTGA	New USPX20 R	TSAGAATACATCGTCGTACT GAACCA	63	
Exon 20-21	347	USPX20 F (old)	GGCAGACAGGTTGAT GACTTGGA				
		New USPX20 F	CAAACCAKGGCAGAC AGGTTGAT				

Exon 21-22	767			USPX22 R (old)	AGGTCTGCAACTTGCCAAA GGAA		XY
				New USPX22 R	TRARTCAGCAACTTGCCARA GGAA	64	No
Exon 22-23	6554						
Exon 23-24	184	New USPX23 F1Y	GCAAAACTTGGAGAA AAYAACCTTGG	New USPX24 RY	CAACTGGCTGACAAGCTTC TGCTA	65	Y
		New USPX23 F2	CTTCTGCCTCCCGAGT SCTGT				Y
		New USPX23 F3	CCTTCTGCCTCCCRAG TGCTAT				Y
Exon 24-25	3604						
Exon 25-26	964	New USPX25 F	TCATGATCAAGCAGT GGTGCTACA	New USPX26 R	ATCCTGATGCCCAGGTRATT TTYT	65	XY
Exon 26-27	1407	New USPX26 F1	TACARAAAATCAYCTG GGCATCAG	New USPX27 R	TGCCAGGCTTTTTCTTTAYT AAGTGYAT	65	
		New USPX26 F2	AAATCAYCTGGGCATC AGGATGTG				
Exon 27-28	17,040						
Exon 28-29	384	New USPX28 F	AGAAYTGTTCGTCAGT TGGCACAG	New USPX29 RB	GCATAATTAAGAAGGTGCC GWAGAA	63	XY
				New USPX29 RC	GCRTARTTAAGAAGATGCC GYAGAA	63	Y
Exon 29-30	815	New USPX29 F	ACAGCAATATTAATGT RCCCAATGCTGA	New USPX30 R	AGGTGRCCTTCCAAGATAG GCTCT	64	
Exon 30-31	1048	New USPX30 F	SAGGTGAAACAGGTR TTGAAGAGCCWA	New USPX31 R	CCAGCAYTGATRGTAGCTG GTGAAC	64	XY
Exon 31-32	3072						
Exon 32-33	2074	New USPX32 F	TGAAGCACTTACKGM GTGGGAATA	New USPX33 R	TGTAGTYGRGAAGCAGCTA AATGACC	64	XY
Exon 33-34	15,066						
Exon 34-35	4045	New USPX34 F	AAAGCTTTAGGACAT CCAGCTATGC	New USPX35 R	TCACAACGATAWGCRTTTG CACCT	63	XY
	2600 (cow)						
Exon 35-36	1808	New USPX35 F	CAAAGGMGATTTATT GGAAGGTGCAA	New USPX36 R	AATGTCCACCACTTGCTTGA CCAC	65	Y
Exon 36	780 Exon	New USPX36 F	AATTTCCWCGAGAGC TGGAYATGG	New USPX36 RY	AATGGCTGGTGACATAATC TGATGG	65	XY
				New USPX36 RX	AATYTGATGGGGTCTTGYG GTGAT	65	X
Exon 36-37	400	New USPX36 FX	CCYCATCARATTATTA TGCCATCAGC	New USPX37 R1	CCTGTRGTAAARAGGAACC TAGCAGCA	63	X
		New USPX36 FY	AARACCCCATCARATT ATGTCACCA	New USPX37 R2	GTGGTAAAGAGGAARCTAG CWGCAA	63	X
Exon 37-38	1403	New USPX37 F	TTCAGCTTGCWGCTA GRTTCCTCT	New USPX38 R	AWGRTGAAGRACAAGGCC CATCT	63	XY
Exon 38-39	594	New USPX38 F	GCATTGTGCATTCTAC TCCGTCAC	New USPX39 R1	TAAATGACGRCCATGCTCW GAMAC	65	Y
				New USPX39 R2	MACTTCCCTTCTSAGGAGA TTTARYACTG	62	Y
Exon 39-40	3045						

Exon 40-41	410	New USPX40 F	AGGTRTGGCAGAAAA RACACAGCTT	New USPX41 R	GAGAACTGAGGATTCTCCC AGCAG	64	X
Exon 41-42	82	New USPX41 F	TCRCAGCCTATAATGC CAATTCAGC	New USPX42 R	RAAGTARATCCAARTATGG CCGAAG	63	Y
Exon 42-43	9011						
Exon 43-44	134	New USPX43 F	CACTTAAAGGAATCCC AGATGATCG	New USPX44 R	CACAAGCTTTTGCMAGTGT CATCC	64	XY
Exon 44-45	718	New USPX44 F	GGGYAATGGAGATCT TAAAAGAAAGTGG	New USPX45 R	TGAYAGTGAGATCCAGATG RATGAGG	63	XY
Exon 45-46	1610	New USPX45 F	ATGACCAAGAWGCCC CAGATGAG	New USPX46 R	ATGTGCTGCTGGTCCTGTAT ATGG	64	XY
		BMY USPX45 F2	AGGAGCCAGATGACC AAGATGC	BMY USPX46 R2	CAAGTGATGTGCTGCTGGT CCT		

483
484
485
486

487 Appendix 2.—List of X-chromosome PCR and sequencing primers for bowheads. Primers
488 are located in exons and span the intervening intron. Forward (primer 1) and reverse (primer
489 2) primers and their sequences are indicated along with their annealing temperature (Temp).

Location	Primer 1 Name	Primer 1 Sequence	Primer 2 Name	Primer 2 Sequence	Temp
Exon 14 Exon15	BOW USP9X X14 F	GACAAATGGGTAATTCCTGCA	BOW USP9X X15 R	GGGACTTCGCTGAGTTTG ACTG	58
	BMY USPin14XF Sq	GGATTCCTTTATCATATCCACCTC			59
	BMY USPin14 630F X	TTGCAAAGTAAACCAATTACGG			59
			BMY USPin14XR Sq	TCCCATACTACAGTAGTAC CAAAGC	57
			BMY USPin14 800R X	CTTATTTCCATCAAGTATG CCAGA	59
Exon 17 Exon 18	BMY USPX17F X	TGCTTCAGCTTGATCCTTCC	Use with USPX18 R		58
Exon 18 Exon 19	Use with USPX18 F		BMY USPX19R X	TCATAGGAGGCTTTCAAA CGA	59
			BMY USPin18X dnR	GAGTTTGGGATTAGCACA TGC	59
			BMY USPX19X UpR	CACCTACAAATAATGGGC CAAA	60
Exon 25 Exon 26	BMY USPin25X 300F	CCCTTAACATGAATTTTACCTAACG	BMY USPin25X 800R	CAGAGGTGATCAAACCAT TCC	59
Exon 28 Exon 31	BMY USPX28X F	CTCTGCTCTTCACCGTTTTGG	BMY USPX31X R	AAACATTGGATGCAGGGA AGAT	62
	BMY USPin29X Fsq	TAAGTAGTTTTGTCAAATGCCTTGCT	BMY USPin29X Rsq	GCTAATCTGAAGGCTAAA AGCTACC	60
	BMY USPin30X Fsq	GGTTAATTTATGGTATCATGTCTTTGG			60
Exon 36 Intron 37	New USPX36 FX	CCYCATCARATTATTATGCCATCAGC	BMY USPin37X 750R	GAAAGAACCAGCAATGGT AGC	59
			USP9 in37X 650R	ACAAATCCATGACCAAAC TCTTCC	62
Exon 37 Exon 38			BMY USP9X38 dnR	AAGCTTTTGATTACCTGAC TAGAAGG	59
Exon 43 Exon 45	New USPX43X F	GCACTTAAAGGAATYCCAGATGACC	New USPX45 RX	TGATACTGAGATCCAGGT GAATGGG	64
Exon 45 Exon 46	BMY USPX45 F2	AGGAGCCAGATGACCAAGATGC			64
			BMY USPX46 R2	CAAGTGATGTGCTGCTGG TCCT	62
Intron 45	BMY USPin45X 380F	AAGACCACATTTCCTGTTAGGC			59
Intron 45	BMY USP9in45X 385F	GCACTAAGATATGGCAGTGAAGG			60
Intron 45	BMY USP9in45X 1000F	CCTCAGTACATTTCTAACGAGTTCC			59
Intron 45			BMY USPin45X 600R	TGGGGTTTAGAAAGAGAC GCT	60
Intron 45			BMY USPin45X 1240R	TCTGTACGTTCCCCACCAG T	60
3'UTR	BMY USP3' X F1	AAGCTTTGGGTATCTTATTTGCAG	BMY USP3' X R1	AGCCCTGCTAGGAAACCA GT	60
3'UTR	BMY USP3' X F1 Seq	CACCAGATAGATCCTGTTTTTGC			60
3'UTR	BMY USP3' X F2 Seq	GGGCTTTGCCCTATAGGATCT			60

490 Table 3.—List of Y-chromosome specific PCR and sequencing primers for bowhead whales.
491 The primers are located in exons or introns as indicated, and span the intervening intron.
492 Forward (primer 1) and reverse (primer 2) primers and their sequences are indicated along
493 with their annealing temperature (Temp).
494

Location	Primer 1 Name	Primer 1 Sequence	Primer 2 Name	Primer 2 Sequence	Temp
Exon 8 Exon 10	BMY USPin8YF Sq	ACTGGATATCGCAGAAGTATTGC			58
Exon 10 Exon 11	BMY USPX10F Y2	ACTCATCGACATGGTCATCCTGA	BMY USPX11upR	GAAGACTGTCTCGCAAGACTATGGA	63
	BMY USPX10Y F Sq	TTAATGGAAAGATGAATGCACTG			58
	BMY USPin10F Y	ATGCATATTCTGTGCCAGCA			60
			BMY USPin10R Y Sq	CAAATGGGAGACAGGAAAGCT	60
Intron14 Exon15			BMYS USP9in14Y 1420 R	AGWTTTGCCYATGGATACCACCAGA	66
	BMYS USP9in14Y 190 F	AGKGGGAAAACAGKAAGCTTAATGA			65
	BOW USP9Y X14 F1	AGAACTTCGCACAAATGACAAG	BOW USP9Yin14dnR	TGCCTCAGAACGTATTTATTCC	58
	BOW USP9Y X14 F2	AATTCCTGCCCTGAAACAAATA	BOW USP9Y X15 R	GGGACTTCGCTGAGTTTGACTA	58
Exon 17 Intron 18	Use with New USP9X17 F		BMY USP9in18Y R	CACTGAGTTGGCATTTGGTAAGG	63
			BMY USP9X18UpRY Sq	TTCAATCCTCACCTGATTGACC	60
Intron 18 Exon 20	BMy USP9in18Y F	TTGTCCCTTACCAAATGCCAAC	Use with New USP9X20 R		63
	BMY USP9X19dnFY Sq	GCCCAATATTTATAGGTGGTTATCC			60
Intron 19 Exon 22	BMY USPin19F1 Y	TTCTGCATTCATCGAGGCAAAG	Use with New USPX22 R		62
	BMY USPin19F2 Y	CCCAGCTTTATGTCTTGGCCATT			62
Exon 23 Exon 24	BMY USPX23dnF Y	TCCTTCTGCCTCCCGAGT			60
Exon 25 Exon 26	BMY USPin25upF Y	AAGTTCTGGGATCTTGGTTGG	With NEW USPX26 R		60
Exon 28 Exon 30	Use with New USP9X28 F		BaMys USP9in29Y R	ACAATGTTGCCAAACCTTTTCATTC	64
	BMY USPin29dnF Y	TGGCCTATAATATTTTTCCTTCG			58
	USPin28F Y Sq	TCCTTAAGGGAAAAGTATACTGC			55
Intron 29 Exon 31	Use with NEW USPX29 F		BMY USPX30R	GCACCTCCTTTCTCACAACC	59
			BMY USPX30R Y	AAGATCGGCTCTTCAACACCT	60
	BMY USP9in29F Y	GAATGAAAAGGTTTGGCAACATTG			
	BMY USP9in29F Y2 Sq	TTTAAAATAAGGATCCGATAGGTT			56
			BMY USPin29R Y	ACAATGTTGCCAAACCTTTTCATTC	63
			BMY USPin29R Y2	CCAAAATATTAATGGTGCCAAGG	61
Exon 34 Intron 35	New USPX34 F	AAAGCTTTAGGACATCCAGCTATGC	BaMys USP9in35Y R	TGTTTCAGGGCAGAAATGGAAA	64
	BMy USP9X34Y Fsq	CTTTGCKGATCAGAAGATTTGC	BMy USP9in34Y R3	TCCACAGAGGAGAAAAAGGTC (Sq)	58

	BMY USP9in34Y F2	GCACATCAGATCATTAGCAAGAGC		Can use with BaMys USP9in35Y R	62
	BMy USP9in34Y F3	GGAGTCATTTATTCACCTAGCTTG (Sq)			
			BMY USPIn34Y R2	GTAGTAATTCATGGCAAGGTAAAGG	58
			BMY USPin34Y 1800R	AGACACTTCCATTTGGCAGAGC	62
			BMy USP9in35Y upR	CTAAAATTAACCACCCAATTAAAAG G (seq)	60
	BMY USPin34Y 2480F	TGTGCGTTAGGCTAGAACTTGG			61
Intron 35 X37	USP9in35Y F	TTAGCAGTGGGAGTGTGGAACC		Use with New USPX37 R2	63
	BMY USPin35Y 550F	CATCTCGAAGTTATTACCTGAACTGC			59
			BMY USPin35Y 700R	CATGCTGCATTTAATCCATACTACC	58
			BaMys USP9in35Y R	TGTTTCAGGGCAGAAATGGAAA	62
	BaMys USP9X36Y Fsq	GCAAGTGGTGGACATTATTATTCC			60
			BaMys USP9X36Y Rsq	CGATTTCTCTCACCATCTTTACC	59
	Use with USP9X36 F		BMY USPin37 upR	CCACAGTTAGGGGAAAATAAAGC	60
Intron 36 Exon 38	BaMys USP9in36Y F	GGATTGGTTAAGTACTGAGGTAAGTG C	BMY USP9X38 dnR	AAGCTTTTGATTACCTGACTAGAAG G	59
			OR Use with USP9X38 R		59
	BMy USP9in36Y F2	GGCTAAGAATTTTTTTAAACGTGGA			61
	BMy USP9in37Y F2	GCTTTATTTTCCCCTAACTGTGG (seq)			60
			BMY USP9in37Y Rsq	CACATGGCTTGTGGGATCTTA	60
Intron 37 Exon 39	BaMys USP9in37Y F	GAGTTCGTATGGTTCATTATTATGTTGC		use with USP9X39 R2	62
			BaMys USP9in38Y R	TGGGAAATGAAATATATAGCTTGG (seq)	58
			BMY USP9X38 dnR	AAGCTTTTGATTACCTGACTAGAAG G (seq)	60
			BMY USP9in38Y Rsq	AAACCACTTGGAGTTTTATGCT	57
Exon 40 Intron 41	New USPX40 F	AGGTRTGGCAGAAAARACACAGCTT	BaMys USP9in41Y R	ATGAATACGCAACCTAAAGACAGG	61
Exon 43 Exon 45	Bmy USP9X43 dnF Y	GCAGTTGTCCAGTTGCTTACCAAA	BMy USP9X45 R Y	GGATATAATGGGGCATCTTCTGG	63
			BMY USPX45Y R2	GGGCATCTTCTGGTGGAGA	61
	BMY USPin43Y upF	GCTTACCAAATATTACAGGTGAGAA			57
Exon45 Exon46	BMY USPX45 F2	AGGAGCCAGATGACCAAGATGC			64
			BMY USPX46 R2	CAAGTGATGTGCTGCTGGTCCT	62
	BaMys USPin45Y F	GGAGTGGGGTCACGTAATAGC (seq)			60
3'UTR	BMY USP3' Y F1	CCAGCAGCACATCACTTGAA	BMY USP3' Y R1	TGCATTTGGCAACCAGTATT	60
3'UTR	BMY USP3' Y Seq	GGACCCTGATGATGTCCACT			59

495