

Relatedness among individuals in BCB bowhead microsatellite samples

AUTHORS: HANS J. SKAUG¹ AND GEOFF H. GIVENS²

¹*Department of Mathematics, University of Bergen, Norway*

²*Department of Statistics, Colorado State University, 1877 Campus Delivery, Fort Collins CO 80523.*

Contact e-mail: skaug@math.uib.no

ABSTRACT

We analyze 33 microsatellite loci for 282 bowhead samples to identify potentially related individuals. Possible related pairs found include many cases where one individual was caught in Barrow and the other individual was caught at St. Lawrence Island, Chukotka, or whaling villages along the Alaskan coastline.

KEYWORDS: BOWHEAD, RELATEDNESS, LOD SCORES, FDR.

INTRODUCTION

The International Whaling Commission (IWC) is currently conducting a review of its management procedure for aboriginal hunting of bowhead whales that inhabit the Bering-Chukchi-Beaufort Seas region of the north Pacific. An important component of this work is analysis of potential stock structure including study of microsatellite data from available samples. For example, Givens *et al.* (2007) summarize many key findings from microsatellite analysis. Yet Morin *et al.* (2007a) illustrate that some microsatellite analyses are subject to disturbingly high influence by the genotype for a single whale or several whales. Although Morin *et al.* (2007a) focus on rare and potentially mis-scored alleles, the implication is clearly that inclusions of closely related individuals in the bowhead microsatellite dataset might also skew results to a surprising degree. It is therefore important to carefully examine the available data for evidence of related individuals. Moreover, the frequency and patterns of relatedness might provide some information about stock structure or bowhead behaviour. Givens *et al.* (2007) eliminated known relatives (mother/fetus pairs); here we adopt a systematic quantitative method to estimate relatedness among individuals.

There have been several studies in which DNA-profiles have been used to detect instances of paternity in whale populations (Clapham, Palsboll, 1997; Garrigue *et al.*, 2004; Nielsen *et al.*, 2001; Skaug, Øien, 2005). More general relationships can be detected, provided that the number of loci in the DNA-profile is high enough. Pairs of related individuals have been detected in North Atlantic minke whales based on 25 microsatellite loci (Skaug *et al.*, 2005). For BCB bowheads more than 30 microsatellite loci are available, and the population size is smaller than for North Atlantic minke whales, so it is natural to expect that pairs of related individuals can be detected.

MATERIAL AND METHODS

Origin and nature of samples

We analyzed 282 individuals scored on at least 30 of 33 suitable loci, as detailed by Givens *et al.* (2007). These whales were sampled during the annual aboriginal hunt by villagers in Alaska and the Chukotka Peninsula, Russia. A few samples were taken by non-lethal biopsy. No animals from Canada or the Sea of Okhotsk were included here.

Classification rule

A commonly used measure of similarity of two DNA-profiles is the LOD score (Skaug *et al.*, 2005). Denote by D_1 and D_2 the DNA-profiles of two individuals. The LOD score is defined as

$$LOD = \log \left(\frac{\Pr(D_1, D_2 \mid \text{half-siblings})}{\Pr(D_1, D_2 \mid \text{unrelated})} \right),$$

where $\Pr(D_1, D_2 \mid \text{half-siblings})$ and $\Pr(D_1, D_2 \mid \text{unrelated})$ are the probabilities of D_1 and D_2 under two different hypotheses of relationship. When there are missing values at a locus (i.e. alleles that cannot be determined) in either D_1 or D_2 , the locus is discarded when calculating the LOD score for that particular pair.

From a sample of n individuals one can calculate $n(n-1)/2$ different pairwise LOD scores. A natural criterion is to declare all pairs with a LOD score above a certain threshold as being related. To avoid problems caused by multiple testing we use threshold based on the false discovery rate (FDR, Skaug *et al.*, 2007). The FDR procedure of Benjamini and Hochberg (1995) controls the proportion of ‘false positives’, i.e. the proportion of pairs that we claim are related but truly are unrelated.

The null hypothesis in each of the pairwise tests is that the two individuals are unrelated. The alternative hypothesis is not specified, although the definition of the LOD score seems to indicate that the alternative is a half-sib relationship. This particular definition of the LOD score was also used in Skaug *et al.*, (2005). When two individuals have been judged to be related, one may try to infer the relationship from auxiliary data such as mtDNA, sex, age and body length. The microsatellite profiles themselves are useful for distinguishing between parent-offspring relationships and other types of relationship, although there are problems with this related to the fact that the pairs are identified based on the microsatellite profiles.

RESULTS AND DISCUSSION

Table 1 shows the list of the 32 pairs with the highest values of LOD . The number 32 was chosen because it gives approximately a false discovery rate of about 1% in the current situation. The interpretation of this choice is that the expected number of false pairs on the list is less than one.

Despite our use of the FDR to control for Type I errors, it is likely that the number of related pairs listed here is too high, for many reasons including the fact that the bowhead population is known not to be in Hardy-Weinberg equilibrium currently. Therefore we suggest that this list be used only as a list of pairs of potential interest for further study, rather than as a list of certain relatives.

There is also one instance where analysis of SNPs data has identified a likely duplicate sample (i.e., two samples with different labels believed to have originated from the same whale), namely 01B16 and 01B12 (Morin *et al.*, 2007b). Our analysis failed to detect this pair, as the two microsatellite profiles are rather different (Table 3).

Most of the identified pairs involve two Barrow animals. Because the vast majority of the samples are from Barrow, this is expected. Apparent relatedness is also observed along the portion of the migratory corridor along the Alaskan coastline from Point Hope through Barrow to Kaktovik.

More notable is the number of pairings of a Barrow individual with an individual from either Chukotka or Savoonga. If true, such pairings would constitute strong evidence against any multi-stock hypothesis of spatial segregation that wholly separated whales in such regions. We are currently investigating ways to evaluate the numbers of pairs of various types to assess just how unusual the apparent predominance of Chukotka and Savoonga inclusions in pairings might be. A first approach is to compare the proportion of observed Barrow-Savoonga pairs to what one would expect under the hypothesis that the true related pairs that are present in the sample are randomly mixed between the areas. The proportion of observed Barrow-Savoonga pairs is $4/32 = 0.13$. There are 206 Barrow animals and 15 Savoonga animals in the sample, and hence the corresponding expected proportion is $206 \cdot 15 / (282 \cdot 281/2) = 0.08$, notably lower than the observed fraction. It is not trivial to assign a measure of significance to this apparent excess of Barrow-Savoonga pairs.

REFERENCES

- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate - a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B - Methodological* **57**, 289-300.
- Clapham PJ, Palsboll PJ (1997) Molecular analysis of paternity shows promiscuous mating in female humpback whales (*Megaptera novaengliae*, Borowski). *Proceedings of the Royal Society of London Series B-Biological Sciences* **264**, 95-98.
- Garrigue C, Dodemont W, Steel D, Baker CS (2004) Organismal and 'gametic' capture-recapture using microsatellite genotyping confirm low abundance and reproductive autonomy of humpback whales on the wintering grounds of New Caledonia. *Marine Ecology-Progress Series* **274**, 251-262.
- Givens GH, Huebinger RM, Bickham JW, George JC, Suydam R (2007) Patterns of genetic differentiation in bowhead whales (*Balaena mysticetus*) from the western Arctic. Paper SC/59/BRG14 to be presented to the Scientific Committee of the International Whaling Commission, May, 2007.
- Morin, PA, Hancock, B.L. and George, J.C. (2007b) Development and application of single nucleotide polymorphisms (SNPS) for bowhead whale population structure analysis. Paper SC/59/BRG8 to be presented to the Scientific Committee of the International Whaling Commission, May, 2007.
- Morin PA, LeDuc RG, Archer E, *et al.* (2007a) Estimated genotype error rates from bowhead whale microsatellite data. Paper SC/59/BRG15 to be presented to the Scientific Committee of the International Whaling Commission, May, 2007.
- Nielsen R, Mattila DK, Clapham PJ, Palsboll PJ (2001) Statistical approaches to paternity analysis in natural populations and applications to the North Atlantic humpback whale. *Genetics* **157**, 1673-1682.
- Skaug HJ, Palsboll PJ, Berube M (2007) Screening large collections of DNA-profiles for pairs of related individuals.
- Skaug HJ, Palsboll PJ, Berube M, Rew MB (2005) Application of high resolution DNA profiling to management of Northeastern Atlantic minke whales, p. Paper SC/57/SD52 presented in Ulsan.
- Skaug HJ, Øien N (2005) Genetic tagging of male North Atlantic minke whales through comparison of maternal and foetal DNA-profiles. *Journal of Cetacean Research and Management* **7**, 113-117.

TABLES AND FIGURES

	ID	Year	Village	S	Age (years)	Length (m)	mtDNA	# matches by locus
1	97B30	1997	Barrow	F		8	BH4	111111 11 1-11-111-11111111 10
	03B2	2003	Barrow	M		13.8	BH42	
2	96B21	1996	Barrow	M		13.4	BH42	1 10111-100-11111-101--111 11 0
	92B3	1992	Barrow	F	59	14.6	BH42	
3	97B18	1997	Barrow	M		10.8	BH39	11111111 -111-111 11-111 10 11 -0
	97B12	1997	Barrow	M		15.3	BH4	
4	05S1	2005	Savoonga	F		8.8	BH47	1 11111-1 1110 011-1111 11 11
	02B17	2002	Barrow	F	6.75	9.3	BH34	
5	BWCH14*	2005	Chukotka				BH24	111111-1-111-111 1 11 111111 111
	02B3	2002	Barrow	F	110	19.2	BH24	
6	97B11	1997	Barrow	M		13.6	BH3	11111111-111- 111 -1111101 1 1 1
	96B6	1996	Barrow	F	30.6	12.7	BH58	
7	96B5	1996	Barrow	F		14.9	BH61	11111111111111-111- 11-11 11 1-1
	93B11	1993	Barrow	M		10	BH61	
8	97B22	1997	Barrow	F		9.4	BH10	11111111 -111- 11-11-111 0 1 1 111
	03B2	2003	Barrow	M		13.8	BH42	
9	96S2	1996	Savoonga	F		17.9	BH34	11-11 1 -101111011011 111011 1111
	02B17	2002	Barrow	F	6.75	9.3	BH34	
10	93B9	1993	Barrow	F		9.2	BH42	11110 1 -000-0 0111-11 11-1- 1-1
	02B14	2002	Barrow	F		8.5	BH42	
11	03S2	2003	Savoonga	F		16.2	BH42	11111111 111111-11 -111101-0 1--1
	00B5	2000	Barrow	F	70	18.9	BH42	
12	04KK1	2004	Kaktovik	M		15.8	BH46	11111011 11111--111-11 10 111-1
	03H1	2003	Point Hope	F		8.8	BH9	
13	95B4	1995	Barrow	F	2	8.6	BH27	11 11 011 101 111-1111110 1 1-1
	01B14	2001	Barrow	M		9.4	BH64	
14	92B12	1992	Barrow	F		12	BH15	11111 11-0-11101 11-11110 11111 0
	02KK2	2002	Kaktovik	M		9	BH15	
15	05S7	2005	Savoonga	F		18	BH42	11 0111 1101-1-1111-011101-1111 1
	02B14	2002	Barrow	F		8.5	BH42	
16	96B5	1996	Barrow	F		14.9	BH61	11110 011 1011-1111 11111 11 -
	96B15	1996	Barrow	M		8.1	BH35	
17	96S2	1996	Savoonga	F		17.9	BH34	11-11101-1011100111- 1 01011 0111
	05S1	2005	Savoonga	F		8.8	BH47	
18	96B24	1996	Barrow	F	23.3	10.9	BH2	111 1101-111 11 111-101011 110100
	05H3_5		Point Hope					
19	03KK2	2003	Kaktovik	M		12	BH42	11111111 11 1111-111- 11101-011--1
	00B5	2000	Barrow	F	70	18.9	BH42	
20	96B7	1996	Barrow	M		12.7	BH42	10101 00-1111111111 -111100 1 111
	04B14	2004	Barrow	F		8.2	BH66	
21	02KK2	2002	Kaktovik	M		9	BH15	11111101-101011 1101 1111 110 1
	02B16	2002	Barrow	M		8.3	BH15	
22	96B3	1996	Barrow	F		7.6	BH16	11 111- 0111 11-10 -11-11111 101
	05B21	2005	Barrow	F		8.8	BH46	
23	05B18	2005	Barrow	F		9.2	BH24	111100-1-01011 1 111111111 1 1 1
	02B3	2002	Barrow	F	110	19.2	BH24	
24	BWCH13*	2005	Chukotka				BH5	1 110111101-10011 -10 1111 11110
	04B11	2004	Barrow	F		12.3	BH55	
25	96B5	1996	Barrow	F		14.9	BH61	111111011101 0-1111-1111 1 01 1
	05B17	2005	Barrow	M		8.3	BH61	
26	BWCH1	2004	Chukotka	M		12	BH4	11111111-111-111010-1-1111111 11
	97B10	1997	Barrow	F	56	16.7	BH42	
27	BWCH16*	2005	Chukotka				BH61	11 111-1-1 1-0-1111-111101 11111
	04B4	2004	Barrow	F		14.2	BH61	
28	97B19	1997	Barrow	F		9.3	BH2	111011 --111011101--111111 0 111
	05B12	2005	Barrow	F		14.2	BH59	
29	05B19	2005	Barrow	M		11.1	BH42	11-1101 -011-0--011- 10111 1111
	04B18	2004	Barrow	F		10.2	BH13	
30	BWCH2	2001	Chukotka	F		15.2	BH42	11111111- 10-01-01 -11 111 - 1-0-
	93B7	1993	Barrow	F		9	BH42	
31	05WW2	2005	Wainwr.	F		15.7	BH20	111101 -011111 111- 0111 11 00
	04B14	2004	Barrow	F		8.2	BH66	
32	95B9	1995	Barrow	M		17.4	BH42	1 11110-11111- 111-11 110 11 1-1
	04B12	2004	Barrow	M		14.2	BH31	

Table 1 The 32 most likely pair of related individuals sorted in order of descending LOD score. Each pair takes up two rows in the table. Rows with individual number marked with a ‘*’ are biopsy samples, column ‘S’ denotes sex. The last column shows the number of matching alleles by locus for each of the two genotypes involved in a pair of individuals, with the following conventions: a blank indicates 2 matching alleles and ‘-’ means that at least one of the individuals had missing data at the locus. The 33 loci are shown in the order specified by the data file.

Individual 1		Individual 2	
ID	Village	ID	Village
00B5	Barrow	03S2	Savoonga
00B5		03KK2	Kaktovik
02B14	Barrow	93B9	Barrow
02B14		05S7	Savoonga
02B17	Barrow	05S1	Savoonga
02B17		96S2	Savoonga
02B17		84S1	Savoonga
02B3	Barrow	BWCH14	64/172
02B3		05B18	Barrow
02KK2	Kaktovik	92B12	Barrow
02KK2		02B16	Barrow
03B2	Barrow	97B30	Barrow
03B2		97B22	Barrow
03B2		02B18	Barrow
03KK2	Kaktovik	00B5	Barrow
03KK2		03S2	Savoonga
03S2	Savoonga	00B5	Barrow
03S2		03KK2	Kaktovik
05S1	Savoonga	02B17	Barrow
05S1		96S2	Savoonga
96B5	Barrow	93B11	Barrow
96B5		96B15	Barrow
96B5		05B17	Barrow
96S2	Savoonga	02B17	Barrow
96S2		05S1	Savoonga

Table 2 Summary of pairs where one member (Individual 1) occurs more than once in Table 1.

	Bmy1_1	Bmy1_2	Bmy2_1	Bmy2_2	Bmy7_1	Bmy7_2	Bmy8_1	Bmy8_2	Bmy10_1	Bmy10_2	Bmy11_1	Bmy11_2	Bmy12_1	Bmy12_2
01B12	243	253	184	186	153	183	168	176	226	238	226	236	158	160
01B16	243	253	188	192	177	179	166	168	228	234	218	230	148	156
	Bmy14_2	Bmy16_1	Bmy16_2	Bmy18_1	Bmy18_2	Bmy19_1	Bmy19_2	Bmy26_1	Bmy26_2	Bmy33_1	Bmy33_2	Bmy36_1	Bmy36_2	Bmy38_1
01B12	179	210	212	119	129	116	116	154	166	138	150	161	171	222
01B16	179	210	220	123	133	116	118	156	162	134	150	157	179	226
	Bmy41_1	Bmy41_2	Bmy42_1	Bmy42_2	Bmy44_1	Bmy44_2	Bmy49_1	Bmy49_2	Bmy53_1	Bmy53_2	Bmy54_1	Bmy54_2	Bmy55_1	Bmy55_2
01B12	193	205	178	178	166	166	188	188	212	214	157	161	213	217
01B16	193	221	178	180	156	156	194	194	210	214	157	161	215	219
	Bmy57_2	Bmy58_1	Bmy58_2	EV1_1	EV1_2	EV104_1	EV104_2	GATA28_1	GATA28_2	TV7_1	TV7_2	TV11_1	TV11_2	TV13_1
01B12	162	161	175	143	143	146	148	170	174	161	165	241	245	299
01B16	160	131	177	135	143	148	152	166	178	165	165	245	245	299
01B12	TV14_1	TV14_2	TV16_1	TV16_2	TV17_1	TV17_2	TV19_1	TV19_2	TV20_1	TV20_2				
01B12	95	97	186	186	193	203	180	180	156	156				
01B16	101	103	186	190	193	203	174	180	156	170				

Table 3 Microsatellite profiles for two individuals (01B12 and 01B16) in the dataset.