The Okhotsk Sea bowhead whale study.

Genetic analysis conducted by Meschersky I., Chichkina A.

Molecular Diagnostic Center, A.N. Severtsov Institute of Ecology and Evolution of Russian Academy of Sciences (IPEE RAS), Moscow

Balaena mysticetus

Ulbansky Bay, 2011-2012

Kind of samples: biopsies, pieces of sloughed epithelium, samples from dead whales found on a beach.

Number of samples:

2011: 30 biopsies, 7 pieces of sloughed epithelium, 1 skin of dead animal, 1 bone.

2012: 24 biopsies, 7 pieces of sloughed epithelium, 2 skin samples of dead animals.

Analysis:

1. Sex determination according to *Jayasanka et al.,* 2008
2. Allelic composition of 14 microsatellite loci:

GATA028; EV1; EV104; TexVet13; TexVet14; TexVet16; TexVet17; TexVet19; TexVet20 - set; used by S. Maclean (2002)

and Bmy1 ; Bmy10 ; Bmy11 ; Bmy18 ; Bmy26.

3. mtDNA:

cytochrome b gene (1140 bp, complete sequence)

fragment of tRNA-Pro (68 bp, complete) + control region (916 bp, complete) + tRNA-Phe (73 bp, complete) + 12S (192 bp, partial), 1250 bp as a total.

Results

Sex and microsatellites

2011: 38 samples, 37 individuals: 20 males, 17 females

2012: 30 samples, 29 individuals, 17 males, 12 females.

Totally 62 individuals (35 males and 27 females) were fully genotyped.

For 2 samples of dead animals collected in 2012 the determination of alleles of the most of loci failed. However, the successful results for 3-4 loci proved the samples belong to new individuals.

Within 31 individuals determined in 2012 four (2 males and 2 females) were recaptured between years (i.e. genotyped in 2011).

No significant differences in allelic composition of 14 loci were found between male and female samples as well as between samples of 2011 and 2012.

Allelic composition and heterozygosity level for a total of 62 individuals genotyped in 2011-2012 are similar to those presented in S. Maclean's thesis (2002) and lower than known for the same loci for bowhead whales of BCB stock (Table 1.)

As allelic names (designations) accepted in each individual study reflect real fragment size only approximately (if no direct sequencing was done) and, so, may be not in correspondence in different studies. On the other hand, the correspondence of allelic names may be determined by comparison of number of "steps" (dinucleotide repeats) between alleles within each series. There is a very good concordance between our allelic names and those used by S. Maclean (2002) for the 9 loci used in both studies – Table 2.

Having accepted this concordance, we found that one individual biopsied in 2001 was recaptured in 2012.
An estimation of population (Shantar region summer group) strength based on
"5 of 105 individuals marked in 1995-2011 were found within 31 individuals in the 2012 sample"
lead to value more than two times higher than based on
"4 of 37 individuals marked in 2011 were found within 31 individuals in the 2012 sample".
This may result from heterogeneity of the whales' distribution within the feeding area – i.e. the whales that prefer Konstantina Bay and the northern part (or exit) of Ulbansky Bay are most likely to be met there, but not in the inner parts of Ulbansky Bay and vice versa. If so, the latter (higher) estimation seems to be more reliable. On the other hand, no used genetic markers (9 microsatellite loci and mitochondrial control region sequences) could determine a difference between 1995-2001 and 2011-2012 samples: Fst = –0.00254; P(Fst) = 0.901 for allelic composition of 9 loci (if found allelic names concordance is accepted) and Fst= –0.004, P(Fst)=0.555 for mt-haplotypes frequencies (see below).

To ascertain the accuracy of the stated allelic names concordance, some samples of both series (Maclean's and ours) should be reanalyzed simultaneously in one laboratory. But the fact that the same genotype was found in two samples (the probability of random coincidences for this data set is estimated at 3.60×10^{-9}) suggests that concordance was defined correctly.

In mitochondrial DNA analysis, the sequences of first part (up to 700 bp) of the control region were obtained for 65 individuals, complete sequences of control region (added with tRNA-Phe and a part of 12S) – for 64 individuals, and complete sequences of cytochrome b gene – for 63 individuals.

No significant differences were found between males and females as well as between 2011 and 2012 samples. For 470 bp fragment of the control region, 7 haplotypes were found. Six of them had been noted in S. Maclean's (2002) analysis, and 1 was a new haplotype for *Balaena mysticetus*. At the same time, R7 haplotype described at S.Maclean's thesis, as well as "baleenHCR" variant described by R.LeDuc with co-authors (1998), were not found in our sample. However, these "unmatched" (found only by us or only by Maclean) haplotypes are rare and differ from "major" variants by 1-2 substitutions. So, they do not affect the results significantly. The difference in haplotype frequencies between 2011-2012 and 1995-2001 samples is negligible (Fst=–0.004, P(Fst)=0.555) and indices of haplotypic and nucleotide diversity between the two samples are similar: H= 0.752 and 0.718, and π= 0.988% and 1.096% respectively.

Sequencing of second part of control region allowed describing an additional haplotype which is the same as R4 for the 1-470 bp but differing from it by 3 substitutions in 471-1250 positions fragment.

For cytochrome b gene, 5 haplotypes were found. Three of them are also known for BCB population (Phillips et al., 2013) and two others differ from the known variants by 1 substitution. Combinations of cytochrome b gene and first 397 bp of control region are presented by 7 haplotypes. One cytochrome b gene sequence may be accompanied by two control region variants (but not conversely, as it was shown for for BCB by C. Phillips and co-authors, 2013) Population diversity in the Okhotsk Sea population is essentially smaller, than in BCB, and these two populations significantly differ not only by haplotype frequencies, but also by mean of haplotypes pairwise difference (Table 3).

At the same time, cytochrome *b* gene haplotypes found in the Okhotsk sea bowhead whale population are enough distanced one from another, and, so, the diversity of variants present here (if only haplotypes, not their frequencies, are taken into consideration) – averaged distance 0.237%, maximal distance 0.790% – is comparable with that in BCB population 0.261% and 0.921% (according to data by C. Phillips and co-authors, 2013).

As a conclusion – the Okhotsk bowhead whale population, which is of low number, is characterized by the reduced genetic diversity. But there are no evidence that its condition progressively deteriorates since whaling cessation.

Future research suggested

It is important to continue the study of the Shantar summer stock by genetic methods. In case the analysis of the databases of individual genotypes 1995-2001 and 2011-2012 may be united, we will have a base of genotypes for over 130 individuals, which, probably, comprise about 20-25% of total population, at least. Having such a dataset and possibility to increase it in the future we, in the course of time, may speak about the population monitoring at family and individual level. It is very important to know if the whales feeding in the Shantar region are a homogeneous group or different whales or small groups of whales prefer to visit different parts of the region – the apical part and the exit of Ulbansky Bay, Konstantina bay, Tugursky Bay, Udskaya Gulf. Finally, the whales observed in spring and early summer in Shelikhov Bay in the northeastern part of the Okhotsk Sea – are they (or most of them) the same whales that feed in summer and early autumn in the Shantar region?

Table 1. Allelic composition of microsatellite loci in the Okhotsk Sea (OS) and Bering-Chukchi-Beaufort Sea (BCB) bowhead whale populations. nA – number of alleles, H – heterozygosity.

Locus	OS, our data		OS S. Maclean (2002)		BCB Jorde et al., 2007	
	nA	nA	nA	H	nA	**H**
GATA028	8	0.823	8	0.797	10	0.851
EV1	8	0.790	8	0.779	6	0.756
EV104	5	0.661	5	0.623	10	0.806
TexVet13	4	0.629	3	0.623	7	0.709
TexVet14	6	0.726	5	0.797	8	0.582
TexVet16	3	0.500	3	0.522	4	0.463
TexVet17	7	0.823	7	0.742	11	0.818
TexVet19	6	0.661	5	0.687	6	0.744
TexVet20	8	0.774	7	0.676	6	0.680
mean for 9 loci	6.1	0.710	5.7	0.694	7.6	0.712
	OS, our data				BCB Phillips et al., 2013	
Bmy1	8	0.823			10	0.822
Bmy10	14	0.823			22	0.890
Bmy11	9	0.694			14	0.870
Bmy18	11	0.613			17	0.881
Bmy26	14	0.902			22	0.895
mean for 5 loci	11.2	0.771			17.0	0.872

Table 2. Matching between allelic names in our study and presented in S.Maclean (2002) thesis determined by difference in fragments length within each series.

A single nucleotide indel for some GATA028 locus alleles and substitutions in some alleles of TV20 locus were confirmed by sequencing (GenBank KF056904–909).

Our data		The difference with the allele of maximal length in range	Maclean. 2002	
allelic name	frequency. %%		allelic name	frequency. %%
		GATA028		
"182"	5.65	maximal length	"178"	4.88
"178"	12.1	-4	"174"	12.8
"177"	4.84	-5	"173"	6.1
"174"	6.45	-8	"170"	6.71
"166"	29.03	-16	"162"	29.27
"162"	2.42	-20	"158"	2.44
"134"	26.61	-48	"130"	17.68
"118"	12.9	-64 / -63	"115"	20.12
		EV1		
"150"	3.23	maximum length	"149"	2.44
"148"	1.61	-2	"147"	1.22
"146"	6.45	-4	"145"	9.15
"144"	29.03	-6	"143"	33.54
"142"	17.74	-8	"141"	20.73
"140"	14.52	-10	"139"	12.8
"138"	13.71	-12	"137"	9.76
"136"	13.71	-14	"135"	9.15
		EV104		
"154"	5.65	maximum length	"152"	5.49
"152"	37.1	-2	"150"	35.37
"150"	14.52	-4	"148"	9.76
"148"	41.94	-6	"146"	48.78
"146"	0.81	-8	"144"	0.61
		TV13		
"305"	27.42	maximum length	"305"	21.34
"303"	0.81	-2	–	–
"301"	13.71	-4	"301"	17.07
"299"	58.06	-6	"299"	61.59
		TV17		
"207"	4.03	maximum length	"207"	3.66
"203"	6.45	-4	"203"	4.88
"201"	6.45	-6	"201"	6.71
"199"	30.65	-8	"199"	35.98
"197"	13.71	-10	"197"	12.2
"193"	22.58	-14	"193"	14.02
"189"	16.13	-18	"189"	11.59
		TV14		
"108"	4.84	maximum length	"107"	3.66

"104"	15.32	-4	"103"	15.24
"102"	35.48	-6	"101"	33.54
"100"	0.81	-8	–	–
"98"	32.26	-10	"97"	34.15
"96"	11.29	-12	"95"	13.41
		TV16		
"189"	13.71	maximum length	"192"	23.78
"187"	20.16	-2	"190"	13.41
"183"	66.13	-6	"186"	62.8
		TV19		
"183"	3.23	maximum length	"182"	3.66
"181"	8.06	-2	"180"	16.46
"179"	46.77	-4	"178"	39.02
"177"	29.03	-6	"176"	31.71
"175"	8.87	-8	"174"	7.93
"163"	4.03	-20	–	–
		TV20		
"171"	2.42	maximum length	"172"	0.61
"168"	8.06	-3 / -2	"170"	11.59
"166" *"167"* accepted as (167') as"167+166"	*0.81* *25.81* 26.62	*-5* *-4* -4	*?* *?* "168"	*?* *?* 26.83
"164"	4.03	-7 / -6	"166"	2.44
"163"	12.1	-8	"164"	10.37
"155"	45.97	-16	"156"	46.34
"152"	0.81	-19 / -18	"154"	0.61

Table 3. Comparison of mtDNA diversity in the Okhotsk Sea (OS, our data) and BCB population (data by Phillips et al., 2013

mtDNA fragment	Cytochrome *b* gene accompained with 397 bp of control region. (1537bp as a total.)		Cytochrome *b* gene only (1140 bp)	
Population	OS (63)	BCB (168)	OS (63)	BCB (168)
Number of haplotypes	7	74	5	32
H	0.764±0.041	0.974±0.006	0.733±0.036	0.912±0.106
π. %	0.449±0.237	0.547±0.282	0.264±0.155	0.343±0.192
Tajima's D	1.112 (P=0.903)	–1.182 (P=0.102)	1.153 (P=0.885)	–1.169 (P=0.114)
Φst	0.085 (P<0.001)		0.126 (P<0.001)	
Fst	0.120 (P<0.001)		0.149 (P<0.001)	