Update of note on sampling and laboratory procedure protocols of the genetic work at the Institute of Cetacean Research (SC/65b/J27Rev)

Naohisa Kanda, Mutsuo Goto, Hiroyuki Oikawa, and Luis A. Pastene

The Institute of Cetacean Research, Toyomi-cho 4-5, Chuo-ku, Tokyo 104-0055 Japan

Contact email: kanda@cetacean.jp

ABSTRACT
This paper is a revised version of SC/F14/J27 presented to the JARPAII Special Permit Review Workshop to review the JARPAII Programme. This note presents a summary of sampling and laboratory protocols for the genetic studies under the Objective 3 of the JARPAII conducted at the Institute of Cetacean Research.

KEYWORDS: ANTARCTIC, GENETICS, SCIENTIFIC PERMITS

INTRODUCTION

Objectives of the JARPAII include the elucidation of temporal and spatial changes in stock structure of Antarctic minke whale and the description of stock structure of other baleen whale species in the Antarctic feeding grounds. Genetic analyses utilizing different kinds of genetic markers are one of the important tools to address the stock structure questions in JARPAII.

In order to minimize errors in genetic data, we were recommended to follow the IWC approved guidelines for DNA data quality control for genetic studies (IWC, 2009). The purpose of this short note is thus to summarize the protocols of sampling and laboratory works for the genetic studies at the Institute of Cetacean Research (ICR), on which the population genetic structure studies under the JARPAII are based.

This paper is the revised version of SC/F14/J27 to cover the recommendations of the Panel of the Expert Workshop to review the JARPAII Special Permit Research Programme held at Tokyo from 24 - 28 of February, 2014. The panel recommends that the revised version of SC/F14/J27 should explain in more detail how far the guidelines were able to be followed.

COLLECTION OF GENETIC SAMPLES

Since JARPA started, we have established standard sampling, sample handling, and laboratory practice procedures to ensure consistent data generation.

Catches

In JARPAII, this section applies to Antarctic minke and fin whales. Skin tissues are taken as genetic samples from each individual whale on board by experienced scientists. The sampling involved the use of sterile and single-use tools (scalpel and plastic tongs). Two pieces of approximately 5×5×5mm skin tissues are taken and are kept in a 99% ethanol filled tube. Prior to the sampling, all of the tubes are labeled with sample names following our naming rule for the catches: e.g., 10/11AM001 indicating the austral summer season (=season), the targeted species (AM= Antarctic minke) and the individual identification number (001). The same codes are applied to several other biological items and ancillary information so that these codes for the genetic samples allow for cross referencing with the comprehensive database of JARPAII.

Biopsy

Biopsy sampling procedures in JARPAII are similar to those used in the IDCR/SOWER cruises. Biopsy sampling (skin/blubber) occurs on large vessels using either a crossbow or black powder gun. Photo-id experiments are conducted in parallel to biopsy sampling. Efforts to avoid re-sampling (two biopsy samples from the same individual) are made. A small tissue sample (< 1 gram) is obtained from a free-ranging individual using a biopsy dart with a stainless tip measuring approximately 4cm in length with an external diameter of 9mm,

which was fitted with a 2.5cm stop to ensure recoil and prevent deeper penetration. Between the sampling periods, the biopsy tips are thoroughly cleaned and are sterilized with bleach. The obtained biopsy samples are treated using sterile and single-use tools (scalpel and plastic tongs). Each of the samples is kept in a 99% ethanol filled tube or in a small zip plastic bag (-20 ºC). Prior to the sampling, all of the tubes are labeled with sample names following our naming rule for the biopsies: e.g., 'J03YS2H001' specifying the research program (J=JARPAII), season (03/04), vessel (YS2= Yushin Maru No. 2), target species (H= humpback whale) and individual number (001). This code allows cross referencing with ancillary information from the biopsy sampling experiment.

STORAGE OF GENETIC SAMPLES

Prior to DNA extraction, the individual tissue samples from catches in the ethanol filled tubes are stored at room temperature. The biopsy samples in the ethanol filled tubes are stored at room temperature and those in the small plastic bag are stored in a freezer (-20ºC) as frozen tissues.

DNA EXTRACTION

Total DNA of each individual is extracted from 0.05g of the collected tissue using either the standard phenol/chloroform extractions protocol of Sambrook *et al*. (1989) or the GENTRA PUREGENE DNA extraction kit (QIAGEN) following the company's manual. Extracted DNAs are stored in the TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0).

SEQUENCING AND GENOTYPING PROCEDURE

At present, two different kinds of genetic markers, mitochondrial DNA (mtDNA) and microsatellite DNA, are used in the ICR.

mtDNA

The first 490 base-pairs at the 5' end of the control region of mtDNA are sequenced for the genetic analysis. A set of primers for PCR amplification is usually MT4 (Arnason *et al*. 1993) and Dlp 5R (5'-CCA TCG AGA TGT CTT ATT TAA GGG GAA C-3'). The PCR reactions are carried out in 25 uL volumes containing 10-100ng of DNA, 2.5 pmole of each primer, 0.5 units of Ex Taq DNA polymerase (Takara), 2mM of each dNTP, and 10x reaction buffer. After an initial denaturation step at 95° C for 5 minutes, a PCR amplification cycle of 30 seconds at 94°C followed by 30 seconds at 50°C and 30 seconds at 72°C is repeated 30 times. The amplification is completed with a final extension step of 10 minutes at 72°C. Subsequent cycle sequencing reactions are performed with 100ng each of the PCR products generated in the above amplifications using the BigDye® Terminator Cycle Sequencing Kit (Applied Biosystems). The primers for the cycle sequencing reactions are the same as used in the initial PCR amplification listed above. The cycle sequencing of 10 seconds at 96°C followed by 20 seconds at 56°C and four minutes at 60°C is repeated 25 times. The nucleotide sequence of each cycle sequencing reaction is determined using Applied Biosystems 3500 Genetic Analyzer (Life Technology) under standard conditions. Both strand samples are sequenced in their entirety for all samples.

Microsatellite DNA

Different sets of the microsatellite DNA loci are used for different species.

Species	Microsatellite DNA loci
Antarctic minke whale	AC045, AC082, AC087, AC137, CA234, DlrFCB14, EV1, EV104, GT23, GT129, GT195, GT211
Fin whale	DlrFCB14, EV1, EV14, EV21, EV37, EV94, EV104, GATA28, GATA98, GAT417, GT23, GT195, GT211, GT310, GT575, TAA31
Humpback whale	AC137, CA234, EV1, EV14, EV37, GATA28, GATA53, GATA98, GATA417, GT23, GT195, GT271, GT310, TAA31
Southern right whale	DlrFCB17, EV1, EV14, EV21, EV37, EV94, GATA28, GT23, GT211, GT310, TR2F3, TR3G2, TR2G5, TR3F2

All of these loci have been already published:
Bérubé *et al*. (2005): AC045, AC082, AC087, AC137, CA234, and GT129
Buchanan *et al*. (1996): DlrFCB14, DlrFCB17
Valsecchi and Amos (1996): EV1, EV14, EV21, EV37, EV94
Bérubé *et al*. (2000): GT23, GT195, GT211, GT271, GT310
Palsbøll *et al*. (1997): GATA28, GATA53, GATA98, GATA417, TAA31
Frasier *et al*. (2006): TR2F3, TR3G2, TR2G5, TR3F2

Although primer sequences follow those of the original authors, an annealing temperature of each of the loci is optimized for each of the whale species. Although the IWC guidelines (IWC, 2009) recommended to sequence the microsatellite loci for marker validation, we have not had the opportunity yet to do so by ourselves. The published primers were designed from the cetacean species and were checked for the reliability of amplification and scoring prior to use by the original authors.

PCR amplifications are performed in 15ml reaction mixtures containing 10-100ng of DNA, 5 pmole of each primer, 0.625 units of Ex Taq DNA polymerase (Takara), and 2mM of each dNTP, and 10x reaction buffer containing 20mM MgCl2 (Takara). The amplified products, each with an internal size standard (GENESCAN400HD, Life Technology), are run on a 6% polyacrylamide denaturating gel (Long Ranger) using BaseStation100 DNA fragment analyzer (BioRad). Although the amplified alleles are visualized and are size determined using the computer software CARTOGRAPHER specifically designed for the BaseStation, allelic sizes for the data base are determined manually in relation to the internal size standard and DNA of known sizes from the 'control' individual that is rerun on each gel.

Sex determination

For the biopsy samples, the slightly modified method of Abe *et al*. (2001) is used, which involves co-amplification of SRY locus on the Y chromosome and a microsatellite locus (usually GT23). With this combination of loci, males show the amplified products of both SRY and microsatellite loci, while females show only the product of the microsatellite locus. Molecular sex determination is not performed for the catches because the sexes of the whales were determined on board by the experienced scientists by observing the reproductive organs.

CONTROL RELIABILITY

For both mtDNA and microsatellites analyses, no attempt has been made to estimate statistical error rates. Although the IWC guidelines recommend scoring by two or more people for double-checking, scores are in general conducted by a single person.

mtDNA

For each individual sample, both forward and reverse sequences are typed for error check. After the first run, all of the ambiguous sequences, singletons, and sequences with the unusual number of transversion, are repeated for confirmation.

Microsatellites

Although typing procedure is repeated only in the case of samples not providing satisfactory results in the first attempt, some individuals, chosen arbitrarily, from the samples are genotyped more than twice for double checking. In some cases, side by side running to check genotypes is conducted when there is some uncertainty in the pattern found. For data consistency, a computer program MICRO-CHECKER (van Oosterhout *et al*., 2004) is used to check for null alleles and reading/typing errors. A computer program GENEPOP (Rousset, 2008) is used to test deviation from Hardy-Weinberg genotypic expectations and, if needed, linkage disequilibrium. Identity check, such as duplicate sampling in the biopsy samples, is conducted using a computer program CERVUS (Marshall *et al*., 1998). Re-sampled animals are not used in the statistical analyses. Input data files for the computer programs are in general made first in the GENEPOP format, and are converted to the other required formats using the computer software, such as through the file conversion option in GENEPOP.

DATA BASE STRUCTURE

Sequence and microsatellite genotype data are stored in Excel files by markers and species.

REQUIREMENTS FROM THE IWC GUIDELINES FOR DNA DATA QUALITY CONTROL

For closing, how far the ICR protocols follow the IWC guidelines (IWC, 2009) were summarized in Table 1.

REFERENCES

Abe, H., Goto, M., Pastene, L.A., Dewa, K. and Naito, E. 2001. Practical use of multiplex fluorescent PCR for cetacean sex identification. *Mar. Mammal Sci*. 17: 657-664.

Arnason, U., A. Gullberg, and B. Widegren. 1993. Cetacean mitochondrial DNA control region: sequences of all extant baleen whales and two sperm whale species. *J. Mol. Biol Evol.* 70: 960-970.

Bérubé, M., Jørgensen, H., Mcewing, R., and Palsbøll, P.J. 2000. Polymorphic di-nucleotide microsatellite loci isolated from the humpback whale, *Megaptera novaeanglliae*. *Mol. Ecol*. 9: 2181-2183.

Bérubé, M., Rew, M.B., Skaug, H., Jørgensen, H., Robbins, J., Best, P., Sears, R., and Palsbøll, P.J. 2005. Polymorphic microsatellite loci isolated from humpback whale, *Megaptera novaeanglliae* and fin whale, *balaenoptera physalus*. *Cons. Genet.* 6: 631-636.

Buchanan, F.C., Friesen, M.K., Littlejohn, R.P., and Clayton, J.A. 1996. Microsatellites from beluga whale *Delphinapterus leucas*. *Mol. Ecol.* 5: 571-575.

Frasier, T.R., Rastogi, T., Brown, M.W., Hamilton, P.K., Kraus, S.D., and White, B.N. 2006. Characterization of tetranucleotide microsatellite loci and development and validation of multiplex reactions for the study of right whale species (genus *Eubalaena*). *Mol Ecol. Notes* 6: 1025-1029.

International Whaling Commission. 2009. Report of the Scientific Committee. Annex I. Report of the working group on stock definition. Appendix 2. Guidelines for DNA data quality control for genetic studies relevant to IWC management advice. *J. Cetacean Res. Manage. (Suppl.)* 11:252-256.

Marshall, T.C., Slate, J., Kruuk, L., and Pemberton, J.M. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. *Mol. Ecol.* 7: 639-655.

Palsbøll, P.J., Bérubé, M., Larsen, A.H., Jørgensen, H. 1997. Primers for the amplification of tri-and tetramer microsatellite loci in baleen whales. *Mol. Ecol.* 6:893-895.

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

Sambrook, J., Fritsch, E.F., and Maniatis, T. 1989. *Molecular cloning: A laboratory manual*. 2nd Ed., Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Valsecchi, E., and Amos, W. 1996. Microsatellite markers for the study of cetacean populations. *Mol. Ecol.* 5: 151-156.

van Oosterhout, C., Hutchinson, W.F., Wills, D.P., and Shipley, P. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes* 4:535-538.

Table 1.Achievements of the ICR protocols for requirements from the IWC guidelines for DNA data quality control.

IWC requirements		ICR protocol	Note
Marker validation			
Microsatellites	Sequence microsatellites	No	We use published primers and check consistency of amplification several times during initial runs. We always test HWE, but prior to data analysis. We think that use of the published primers can avoid non-random associations of alleles issue because the papers referred have such information. We do test LD when we seek any biological meanings.
	Test for Hardy-Weinberg equilibrium (HWE)	Yes	
	Test for linkage disequilibrium (LD)	Not for this purpose	
Mitochondrial DNA (mtDNA)	Validate mtDNA amplification	Yes	We compare amplified sequences to those in GenBank. We conduct a cycle sequence step using same primers to the first amplification during sequencing that allows us to confirm whether or not the amplified sequences are truly mtDNA.
	Check with Genbank or similar	Yes	
Systematic quality control and assessment			
Assessing sample quality to genetic analysis	Check for DNA concentration and quality	No	Quality problems have not been detected so far probably because we usually don't use low quality samples such as tissues from stranded animals, faeces and sloughed skin, etc. When quality isseus are raised, we will follow the IWC guideline. For the low quality DNA samples, we re-extract DNA from backup tissues.
Ensuring consistent data generation			
I. Sampling	Prepare prelabelled (numbered) vials, explicit easy-to-read instructions, prenumbered datasheet, etc.	Yes	Since JARPA started, we have established standard, but strict, sampling procedures not just for genetic samples. See Collection of genetic samples section in the text.
II. Sample handling	Establish procedures for sample receipt	Yes	We double check samples onboard. Samples are transported and are received with adequate data base entry. See also Collection of genetic samples section in the text.
	Have backups	Yes	
III. Laboratory practice	Establish standard routine	Yes	Allelic sizes are determined manually in relation to the internal size standard and DNA of known size that are rerun on each gel. In regard to other procedures to avoid scoring misses, see text.
	Run controls	Yes	
	Check scoring consistency by double-blind test with at least two persons	No	

Table 1. Continued.

IV. Check data for consistency and plausibility	Use MICROCHECKER (microsatellites)	Yes	See Control reliability section in the text.
	Check HWE (microsatellites)	Yes	
	Check for plausibilty of allele calls (microsatellites)	Yes partially	
	Sequence both strands (mtDNA)	Yes	
	Check quality, authenticity, and plausibilty of sequence (mtDNA)	Yes partially	
V. Central databases hold responsibility for combined data sets		Not available	
VI. Data analysis	Use automated file conversion	Yes	See Control reliability section in the text.
Error rates estimation			
Incorporation of replicated blind controls	(1) Random sample replication to identify random and systematic errors.	No	See Control reliability section in the text.
	(2) Control samples (2-3) replicated in every genotyping experiments (PCR and electrophoresis).	Yes	
	(3) Targeted replication of samples after the majority of data are generated for verification of data quality and detection of sample handling errors.	No	

IWC requirements were extracted from IWC (2009).