

Annex N

Report of the Working Group on DNA

Members: Pastene (Chair), An, Anselme, Berggren, Bickham, Borodin, Danielsdottir, Goto, Gronvik, Huebinger, Ipatova, Jackson, Kanda, LeDuc, Leslie, Park, Perrin, Senn, Skaug, Walløe, Waples, Yoshida, Young.

1. ELECTION OF CHAIR

Pastene convened and chaired the Group.

2. APPOINTMENT OF RAPPORTEURS

Perrin acted as rapporteur.

3. ADOPTION OF AGENDA

The agenda is given as Appendix 1. Items 5, 6 and 7 of the agenda are in response to requirements placed on the Scientific Committee by IWC Resolution 1999-8 (IWC, 2000), which called for annual reports on progress in the following areas:

- (1) Genetic methods for species, stocks and individual identification.
- (2) Collection and archiving of tissue samples from catches and bycatch.
- (3) Status of and conditions for access to reference databases of DNA sequences or microsatellite profiles derived from directed catches, bycatch, frozen stockpiles and products impounded or seized because of suspected infractions.

4. REVIEW OF DOCUMENTS

Relevant information was contained in SC/58/For Info. 33 (Ross and Murugan, in press). No other documents were submitted.

5. PROGRESS ON GENETIC METHODS FOR SPECIES, STOCK AND INDIVIDUAL IDENTIFICATION

In the 2005 meeting of the Scientific Committee, it was agreed that validation of DNA sequences in GenBank and other such repositories should be carried out routinely (IWC, 2006). An Intersessional Working Group was established to develop and implement a protocol to carry this out. The report of the Intersessional Working Group is given in Appendix 2.

Based on the results obtained by Ross and Murugan (in press), it was **agreed** that the best available vehicle for the validation is *DNA Surveillance* (Ross *et al.*, 2003) (DNAS below), an open website which uses a phylogenetic approach.

Three types of inconsistencies/errors should be targeted in the validation:

- (1) *Quality of submitted sequences.* When a species identification cannot be made by *DNAS*, the possible reasons are lack of reference sequences in *DNAS* for the species (rare cases), too short a sequence in GenBank, or sequencing errors.
- (2) *Accuracy of species identification.* Disagreement between *DNAS* and GenBank can be due to differences in nomenclature and taxonomy used, paraphyly due to incomplete lineage sorting (especially in the Delphinidae), or sample misidentification. The last is of most interest to the Scientific Committee, as it can lead to errors in phylogenetic and forensic studies.
- (3) *Accuracy of geographical location.* While geographic assignments in *DNAS* are made on a probabilistic basis and therefore cannot be considered as ground truth when compared to locations reported in GenBank, apparently anomalous locations can be identified, reported to the author(s) of the entry, and further researched.

The Group **agreed** that validation should be carried out annually and that control-region sequences should be examined in the first round of validation. *DNAS* also contains cytochrome *b* sequences, and these could also be used in later rounds. It was noted that mtDNA sequences used in 'DNA bar-coding' (e.g. cytochrome oxidase *1*) should be considered in future development of validation, and that it would be useful to coordinate with the 'DNA bar-coding' program in the future. However, these sequences are not presently included in *DNAS*.

The question was raised of how many cetacean sequences are typically submitted to GenBank annually. Pastene agreed to contact GenBank on this issue intersessionally.

A discussion ensued on how to handle and report errors or anomalies detected in the validations. As a basic strategy, the information must be shared with members of the Scientific Committee. The Group agreed that there should also be feedback to GenBank. Present policy of GenBank is that entries cannot be changed but that amendments can be added with the permission of the original submitter. An entry includes information on a voucher specimen and citation of a publication. Thus, by reference to the publication, the validation team can contact the author(s), notify them of the inconsistency and suggest that an amendment be made to the entry (offering to help after GenBank has been notified directly of the permission to amend). LeDuc noted that this process had worked well for amendment of a beaked whale sequence.

It was suggested that additional or alternative methods of validation in addition to use of *DNAS* should be investigated. The Group **agreed** that this could be done but that *DNAS* offers an immediate and efficient means of validation and should be used in at least the first rounds.

The Group agreed that first priority for validation should be given to the baleen whale species currently under genetic investigation by the Scientific Committee: common minke whale, Antarctic minke whale, Bryde's whale, sei whale, fin whale, blue whale, humpback whale, gray whale, right whales and bowhead whales.

If so directed by the Scientific Committee, an intersessional working group can determine how many sequences exist in GenBank for these taxa, how many are typically added annually, the expertise and time required to carry out the first round of validation and subsequent annual rounds, and the approximate cost if done under contract.

6. PROGRESS ON COLLECTION AND ARCHIVING OF SAMPLES FROM CATCHES AND BYCATCHES

Skaug reported on the status of the Norwegian register (Appendix 3). The collection of samples is from the commercial catches of common minke whales. He noted that a sample was missing for 1 of 544 whales taken in 2004 and for 4 of 639 whales taken in 2005. However, date and locality of the catch accompany any meat sold in the market, and meat from these few whales could easily be traced back to origin if genetic analysis could not determine its legality.

Kanda reported on the status of the Japanese register (Appendix 4). The collection of samples is from scientific whaling in the Antarctic (JARPA-JARPA II) and North Pacific (JARPN II), bycatches and strandings. It includes complete coverage for 2005 and for the 2005/2006 Antarctic season.

Danielsdottir reported on the status of the Icelandic register (Appendix 5). Samples are presently in hand for all whales taken in 2003-2005. She noted that the register contains the same items in the same format as the Norwegian register, so that cross-comparisons can be carried out.

7. REFERENCE DATABASES AND STANDARDS FOR A DIAGNOSTIC REGISTER OF DNA PROFILES

Genetic analyses have been completed and data on mtDNA, STRs and sex entered in the Norwegian register for years through 2003 (Appendix 3). For the Japanese register, analyses have been completed for all scientific whaling samples from 2004, for all bycatch samples through 2005, and for a portion of the stranding samples for 2005 (Appendix 4). Danielsdottir reported that analyses of the Icelandic samples will begin in the coming year, provided that funding is available.

8. WORK PLAN

The terms of reference for the Working Group for the next year will remain the same as for this year, unless the Commission requests other information in the interim. Members of the Working Group are encouraged to submit papers relating to these terms of reference and to propose additional agenda items. If the Scientific Committee so directs, the Group will continue with its development of plans for sequence validation.

9. ADOPTION OF THE REPORT

The report was adopted by consensus.

REFERENCES

- International Whaling Commission. 2000. Chairman's Report of the Fifty-First Annual Meeting. Appendix 9. IWC Resolution 1999-8. Resolution on DNA testing. *Rep. Int. Whaling Commn.* 1999:55.
- International Whaling Commission. 2006. Report of the Scientific Committee. *J. Cetacean Res. Manage.* 8 (Suppl.):1-302.
- Ross, H. A. and Murugan, S. In press. Using phylogenetic analyses and reference datasets to validate the species identity of cetacean sequences in GenBank. *Mol. Ecol. Evol.*
- Ross, H. A., Lento, G. M., Dalebout, M. L., Goode, M., Ewing, G., McLaren, P., Rodrigo, A. G., Lavery, S., Baker, C. S. 2003. DNA Surveillance: Web-based molecular identification of whales, dolphins, and porpoises. *J. Hered.* 94:111-114.

Appendix 1

AGENDA

1. Election of Chair
2. Appointment of rapporteurs
3. Adoption of the agenda
4. Review of documents
5. Progress on genetic methods for species, stock and individual identification
6. Progress on collection and archiving of samples from catches and bycatches
7. Reference databases and standards for a diagnostic register of DNA profiles
8. Work plan
9. Adoption of report

Appendix 2

REPORT OF THE WORKING GROUP ON VALIDATION OF SEQUENCES IN GENBANK AND OTHER REPOSITORIES

Members: Pastene (Convenor), Baker, Goto, Kanda, LeDuc, Perrin, Rowles

During the 2005 annual meeting the IWC Scientific Committee agreed that validation of DNA sequences such as that described in Ross *et al.* (2005) (SC/57/SD4) should be carried out routinely for cetacean sequences in GenBank and other such repositories (IWC, 2006 p.45). The SC established an intersessional Working Group to 'develop and implement a protocol for routine validation for cetacean sequences in GenBank and other repositories'. This paper presents some preliminary ideas and suggestions toward the development of such a protocol, developed during the intersessional period.

Scope of the validation of public DNA sequence repositories

The following types of inconsistencies/errors in DNA repositories should be checked during the validation:

Quality of submitted sequences

This is an issue difficult to assess but efforts are being made to find and correct sequence error (Harris, 2003). Uncertainty in sequencing has an effect when sequences are downloaded and used in individual phylogenetic studies (Ross *et al.*, 2005).

Accuracy of the species identity

First it should be noted that the study by Ross *et al.* (2005) in general found that agreement in species identity between DNA Surveillance and GenBank was very high.

Regarding cetacean species identities, Ross *et al.* (2005) found that disagreement between identities in GenBank and DNA Surveillance arose from three main causes: (i) differences in nomenclature and taxonomy; (ii) apparent paraphyly, especially among the Delphinidae; and (iii) potential sample misidentification from morphological evidence. Uncertainty in species identity has an effect when sequences are downloaded and used in individual phylogenetic studies.

Accuracy of the geographical location

In some cases species identification is correct in the repository but the geographic locality of the sample (e.g. ocean basin) could be incorrect. Uncertainty in geographical location has an effect in studies focused on population identification.

Method for routine validation of DNA sequences in public DNA repositories

DNA Surveillance is a system for determining the species identity of unknown samples by phylogenetic analysis. It is implemented at present with a set of reference alignments for cetaceans derived from expertly identified specimens (Ross *et al.*, 2005). This system has been presented and discussed at the SC meetings in recent years, and in 2005 the SC agreed that validation such as that of Ross *et al.* (2005) should be carried out routinely for cetacean sequences in GenBank and other such repositories. This method is appropriate for investigation of the accuracy of species identity. The validation method of Ross *et al.* (2005) is currently in the publication process in the peer reviewed literature.

If the IWC SC accepts *DNA Surveillance* as the most suitable method, then the next step will be determine whether the routine application of this method for validation of DNA sequence (regarding accuracy in cetacean species identity) should be carried out under a specific contract with the IWC.

It is uncertain at this stage whether *DNA Surveillance* is an appropriate method to check for sequence quality and uncertainty in geographical origin of the sequences submitted.

Kind of DNA sequences to be validated

Mitochondrial DNA control region sequences are the most used sequences in cetacean population genetics studies in the IWC SC and other forums. Further, they are the largest category of submissions at repositories such as GenBank. Given these facts it is proposed that validation be concentrated only on this type of mtDNA sequence (control region).

Repositories to be validated

Only the widely used DNA sequence repository GenBank will be targeted for validation.

Frequency of validation

Annual validation of repositories is proposed.

Report and record of inconsistencies/errors in DNA sequences repositories

If the routinely validation is carried out under an IWC contract, the results of each evaluation should be presented to the annual meeting of the SC and information on errors/inconsistencies found be available for SC members. Inconsistencies should be evaluated taking into account the current cetacean nomenclature accepted by the IWC SC.

In the case the inconsistencies/errors are communicated outside the IWC context, the following two methods are proposed: (i) contact the original contributor in order to inform them of the inconsistency and then propose a new validated submission to repository, (ii) submit the validated new sequence to the repository, which a notation that the allows crosscheck with the original (un-validated) sequence.

REFERENCES

- International Whaling Commission. 2006. Report of the Scientific Committee. *J. Cetacean Res. Manage.* 8 (Suppl.): 1--302.
- Harris, D.J. 2003. Can you bank on Genbank? *Trends Ecol. Evol.* 18:317-319.
- Ross, H.A., Murugan, S. and Baker, C.S. 2005. Are the species identities of cetacean sequences in GenBank correct? Paper SC/57/SD4 presented to the IWC Scientific, May 2005 (unpublished). 9pp.

Appendix 3

STATUS OF THE NORWEGIAN MINKE WHALE DNA REGISTER AS OF MAY 2006

Hans Julius Skaug
Institute of Marine Research, PO 1870, N-5817, Norway

Year	DNA-register ¹	IWC catch statistics ²	Not landed ³	Landed ⁴	Duplicates ⁵	Missing samples ⁶	Total Missing ⁶
1997	488	503	7	496	3	5	8
1998	609	625	11	614	1	4	5
1999	571	591	17	574	2	1	3
2000	470	487	6	481	3	8	11
2001	538	552	11	541	2	1	3
2002	625	634	9	625	0	0	0
2003	638	647	9	638	0 ⁸	0	0
2004	-	544	7	537	-	1	-
2005	-	639	6	633	-	4	-

¹Number of unique individuals contained in the DNA-register (not containing duplicates).

²Number of individuals caught by Norway, including individuals not landed.

³Number of individuals killed, but not taken onboard the vessel.

⁴Number of individuals taken onboard the vessel.

⁵Number of occurrences of (tissue) sample switching

on board the vessel as detected by comparison of genetic profiles. The result is that two samples have been returned from one individual, and no sample has been returned for one individual.

⁶Number of individuals for which tissue samples are missing for other reasons than sample switching.

⁷The difference between the columns 'Landed' and 'DNA-register'.

⁸Not yet verified by cross checking all individuals in the DNA-register.

Appendix 4

AN UPDATE OF THE JAPANESE DNA REGISTER FOR LARGE WHALES

Mutsuo Goto and Naohisa Kanda

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

The status of the Japanese DNA register for large whales was presented and discussed during the 2005 SC meeting (IWC, 2006). The number of genetic samples and the number of individuals analyzed and registered were reported. The status report included information of the scientific whaling in the North Pacific (JARPN II) up to 2003, of the scientific whaling in the Antarctic (JARPA) from the austral summer season 1987/88 to 2004/05, and of the by-catches and strandings up to 2004.

Here, the information was further updated up to 2005 including 2005/06 scientific whaling in the Antarctic (JARPA II).

Source/Species	Period	Genetic samples	mtDNA	STRs	Sex
Scientific whaling					
NP minke whale	04	158	158	158	158
	05	220	60	0	220
NP Bryde's whale	04	50	50	50	50
	05	50	0	0	50
NP sei whale	04	100	100	100	100
	05	100	0	0	100
NP sperm whale	04	3	3	3	3
	05	5	0	0	5
Antarctic minke whale	05/06	853	0	0	853
Antarctic fin whale	05/06	10	0	0	10
By-catches					
NP minke whale	05	128	128	128	128
NP humpback whale	05	4	4	4	4
NP gray whale	05	3	3	3	3
Strandings					
NP minke whale	05	7	4	0	0
NP Bryde's whale	05	1	0	0	0
NP sperm whale	05	5	3	0	0
NP humpback whale	05	2	0	0	0
NP right whale	05	1	0	0	0
NP fin whale	05	1	1	0	0
NP unknown	05	1	0	0	0

STR = microsatellites, NP = North Pacific.

Note 1: As explained in IWC (2006), sex of the whales taken by scientific whaling was determined by scientists onboard the research vessels.

Note 2: 0 = not yet analyzed at the time this WP was prepared.

REFERENCE

International Whaling Commission. 2006. Report of the Working Group on DNA testing. *J. Cetacean Res. Manage.* 8 (Suppl.): 252-258.

Appendix 5

STATUS OF THE ICELANDIC MINKE WHALE DNA REGISTER

Anna K. Daniélsdóttir

Marine Research Institute, Skúlagata 4, 101 Reykjavík, Iceland

The number of Icelandic minke whale samples that were collected for the DNA register in the years 2003-2005 is listed in the Table below. Genetic analysis has not started yet.

YEAR	GENETIC SAMPLES	MICRO-SATELLITES	MTDNA	SEX
	n			
2003	36	0	0	0
2004	25	0	0	0
2005	34	0	0	0
TOTAL	95			

