

# Annex N

## Report of the Working Group on DNA

**Members:** Pastene (Convener), Bachmann, Cipriano, Double, Goto, Guerrero, Hoelzel, Jackson, Natoli, Olavarria, Palsbøll, Pampoulie, Park, Postma, Skaug, Tiedemann, Torres-Flores, Verborgh, Waples.

### 1. ELECTION OF CHAIR

Pastene was elected Chair.

### 2. APPOINTMENT OF RAPORTEURS

Waples acted as rapporteur.

### 3. ADOPTION OF AGENDA

The adopted agenda is given as Appendix 1. Items 5, 6, 7 and 8 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/60/SD6, For Info 11, 22 and 23.

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

Last year a suggestion was made to present and discuss in the Group a recent published paper by Rohland and Hofreiter (2007). This year this paper was available as For Info 11, which was presented by Cipriano. The paper presented a method designed to maximise recovery of PCR-amplifiable DNA from ancient bone and tooth specimens and at the same time to minimise co-extraction of substances that inhibit Polymerase Chain Reaction (PCR). This is achieved by a combination of DNA extraction from bone powder using a buffer consisting solely of EDTA and proteinase K, and purification of the DNA by binding silica in the presence of high concentrations of guanidinium thiocyanate. The protocol for this method allows DNA extraction from ancient bone and teeth with a minimum of working steps and equipment and yields DNA extracts within 2 working days.

It was noted that a separate paper by the same authors, same year (For Info 24) should also be consulted. Both For Info 11 and 24 are methodology papers. For Info 11 was presented to illustrate the simplicity and general applicability of the techniques described, and suggested that having a range of extractions available for use with different tissue types and preservation methods is better than relying on a single extraction method.

Cipriano also presented For Info 22 and 23. For Info 22 examined two whole genome amplification methods for genotyping of LCN and degraded DNA samples. Both methods (WGA kits) amplified genomic DNA, producing microgram quantities from sub-nanogram templates. Profiling success of LCN DNA samples was increased, with improvements of over 700% from 10pg template DNA compared to non-WGA-amplified control samples. The amplification success with degraded DNA was also improved by WGA.

The Working Group noted that these methods might be useful in some situations, but need evaluation for specific applications.

For Info 23 presented the applications, analyses and advances of amplified fragment length polymorphism (AFLP) DNA fingerprinting. The Group noted that AFLPs need high quality DNA but can be useful, particularly in early stages of studies before other markers are developed. It was further noted that modern methods for isolating microsats are easy enough that this is preferred to relying on AFLPs. Also, the latter are anonymous and often dominant. The Working Group concluded that there may be little need to use AFLPs with cetaceans, except perhaps for genomic screening.

The Chair noted that for the past several years the Group has been evaluating various techniques to extract and amplify DNA from 'difficult' samples. He suggested that perhaps next year would be a good time to attempt to summarise current knowledge. The Working Group was reminded that the Stock Definition Working Group is working on DNA data quality so the efforts by the two working groups should be coordinated; however, having in mind that the former Group emphasis is on stock structure, while DNA focuses on individual/species identification, there should be some overlap on quality issues but also some differences.

### 6. FIRST ROUND OF SEQUENCE VALIDATION AND PLAN FOR FUTURE SEQUENCE VALIDATION

Last year the Committee agreed to conduct the first round of sequence validation in *GenBank* using *DNA Surveillance* (Ross *et al.*, 2003; Ross and Murugan, 2006), under a research contract. Specifications for this first round of sequence validation were given in (IWC, 2008, p. 338-339).

On behalf of the authors the Chair presented SC/60/SD6, which contained the results of the first round of sequence validation in *GenBank*. The phylogenetic methods in *DNA Surveillance*, in conjunction with the curated reference sequence alignments known as *Witness for the Whales*, were used to assign species identities to the 922 sequences from baleen whales published in *GenBank* prior to 2007. Of these 42 sequences were identified as belonging to a different species, and 44 to a different subspecies, from that recorded in *GenBank*. A species identity could not be assigned unambiguously to seven sequences. A small number of sequences were suggestive of poor or unreliable quality, but in each case the species identity as recorded in *GenBank* was confirmed here. The authors suggested that taxonomic revision is probably the greatest source of disagreement in the species

identities given by *GenBank* and *DNA Surveillance*. To provide a better validation of sample origin, all major geographic regions need to be represented for each species in the reference data sets.

The Working Group felt that this was a useful exercise showing no major issues in *GenBank* IDs. Most disagreements appear to be due to a lag in taxonomy. Some of the ambiguous data might also be due to nuclear copies of mtDNA. A question was raised as to whether the survey is a complete sample of what is available in *GenBank*. It was responded that the survey is a complete sample of mtDNA control region sequences deposited in *GenBank* before 2007.

It was noted that the survey is not a proper validation as the 'true' type has not been established but what is being measured is consistency. The Group agreed to change the terminology 'sequence validation in *GenBank*' to '*GenBank* sequence assessment for species assignment'.

It was added that it would be useful to see results from BLAST showing the second most likely assignment for ambiguously identified specimens using *DNA Surveillance*. The Chair clarified that the author of SC/60/SD6 were responding to a specific request from the Committee last year (IWC, 2008 pp 338-339), and this explains many details of experimental design. However the Working Group might consider additional requests for a subsequent study.

SC/60/SD6 points out the importance of having geographic collection information associated with DNA samples. *GenBank* includes unique haplotypes that often represent multiple individuals from different areas, so geographic information could be misleading if a particular haplotype only is assigned to a single region. It was noted that geographic collection information is more important for studies on stock structure in addition to species identification for forensic purposes.

An *a priori* assumption of SC/60/SD6 is that sequences must make phylogenetic sense or they are wrong, and it was questioned whether this is a valid assumption. It could be argued that the NJ method used is not very powerful, so it was clarified that the use of NJ in *DNA Surveillance* is for species identification but not for investigating phylogenetic relationships. However, some members of the Working Group felt that the identification approach used in SC/60/SD6 is indeed phylogenetic.

It was speculated that perhaps some problems were associated with early submissions to *GenBank*, which often used shorter sequences and that a test of 'old' and 'new' sequences should be conducted.

The Working Group agreed to conduct a second round of *GenBank* sequence assessment for species assignment based on the relatively large number of baleen whale sequences (400+ sequences) deposited in *GenBank* in 2007. Specifications for the next assessment are similar to those in the last year, but include some additional suggestions derived from discussion on SC/60/SD6 this year. Specifications for the next *GenBank* sequence assessment for species assignment are shown in Appendix 2.

Finally the Group felt that in the future perhaps assessments of *GenBank* sequences may be conducted less often and less formally. This will be considered after the Group evaluates the results of the second *GenBank* sequence assessment for species assignment.

## 7. 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 from 1997 to 2007. The number of samples missing from the register by year ranged from 0-11. Some of the missing samples reflect unsampled whales, while others resulted from inadvertent duplicates.

Goto 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 (JARP-N II), bycatch and strandings. It includes complete coverage for 2007 and the 2007/08 Antarctic season.

Pampoulie reported on the status of the Icelandic register (Appendix 5). Samples are presently in hand for all whales taken in 2003-07.

## 8. 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 2006 (Appendix 3). Laboratory work is being conducted for samples collected in 2007.

For the Japanese register (Appendix 4), all the genetic analyses (mtDNA and STRs) have been completed for NP minke, NP Bryde's, NP sei and NP sperm whales taken by scientific whaling through 2007. The genetic samples of Antarctic minke whales have not been analyzed yet, except for sex and for STR of 190 samples taken in 2006/07. The genetic analysis (mtDNA and STRs) of Antarctic fin whales was completed for 2005/06 and 2006/07. For bycatch samples, genetic analyses (mtDNA and STR) have been completed for all samples through 2007. For the stranding mtDNA analysis was completed for samples collected through 2007.

For the Icelandic register (Appendix 5) genetic analyses (mtDNA and STR) were completed for common minke whales taken by scientific whaling in 2007 (see also SC/60/PF110). Laboratory work of samples taken under commercial whaling in 2007 is under way.

The Chair reminded the Group that reports of updates of registers should include a list of references including the relevant documents on protocols used.

## 9. WORK PLAN

The terms of reference for the Working Group will remain the same for the next year, unless the Commission requests other information in the interim. Members of the Working Group were encouraged to submit papers relating to these terms of reference and to propose additional agenda items. The Group agreed that the second *GenBank* sequence assessment for species assignment is important and recommended that it should be conducted in the intersessional period. The budgetary implication for this work is discussed under Item 21 of the Committee report.

## 10. 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. 2008. Report of the Working Group on DNA, *J. Cetacean Res. Manage.* 10 (Suppl.):336-340.
- Rohland, N. and Hofreiter, M. 2007. Comparison and optimization of ancient DNA extraction. *BioTechniques* 42(3): 343-403.
- 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.
- Ross, H. A. and Murugan, S. 2006. Using phylogenetic analyses and reference datasets to validate the species identity of cetacean sequences in GenBank. *Mol. Phyl. Evol.* 40: 866-871.

## 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. First round of sequence validation and plan for future sequence validation
7. Progress on collection and archiving of tissue samples from catches and bycatch
8. Reference databases and standards for a diagnostic DNA registries
9. Work plan
10. Adoption of the report

## Appendix 2

### SUMMARY OF THE RESULTS OF THE FIRST ROUND OF *GENBANK* SEQUENCE ASSESSMENT FOR SPECIES ASSIGNMENT AND ACTIONS ARISING

(1) The Scientific Committee envisaged that the first analysis of *GenBank* sequence assessment for species assignment would take the form of a report with the following provisions:

- (i) List the *GenBank* accession number and species identity of each mysticete control region sequence with the species identity as determined using the most recent version of the Witness for the Whale reference sequence alignments (see SC/59/SD5) and the DNA Surveillance software engine.
- (ii) The above list to be supported by phylogenetic trees, one per sequence, showing the placement of the *GenBank* sequence in relation to the reference sequence.
- (iii) An evaluation of the types of inconsistencies/errors as agreed by the Committee last year: quality of submitted sequences, accuracy of species identification and, where possible, accuracy of geographical location.

(2) SD6 presented the results of the first *GenBank* sequence assessment for species assignment, under the methodology and guidelines provided by the Committee in 2007. This was done for a total of 922 mtDNA control region sequences deposited in *GenBank* between pre-2000 and 2006.

Provision (i) above was covered (see Tables 3 and 4 of SC/60/SD6).

Provision (ii) above was covered (phylogenetic trees, one per sequence, are available online).

Regarding provision (iii) above the output was as following:

#### Accuracy of species identification:

Among 922 control region sequences, 42 have species identities which disagree with those given by WFTW. These inconsistencies are restricted to the minke whale, Bryde's whale and right whale groups, and all are related to changes in taxonomy not yet reflected in the [organism] taxonomy field used by *GenBank*.

23 labeled as *Balaenoptera acutorostrata* in *GenBank* were identified as *B. bonaerensis* (Table 5 in SC/60/SD6).

9 labeled as *B. edeni* in *GenBank* were identified as Kochi and Omurai (Table 6 in SC/60/SD6).

10 labeled as *Eubalaena glacialis* in *GenBank* were identified as *E. australis* and *E. japonica* (Table 8 in SC/60/SD6).

A species identity could not be assigned unambiguously to seven sequences.

#### Accuracy of geographical location:

Most of the disagreement reported here regarding the sampling location arose because of limited geographic representation in the reference datasets. Haplotypes having a wide distribution are often represented from a single location. When a disagreement occurred, it was difficult to determine whether a sampling site had been incorrectly reported or if the data set lacked sufficient geographic resolution. If sample location is to be validated using the WFT references, then species-specific alignments containing representative haplotypes from all major populations will be needed.

#### Quality of submitted sequences:

A small number of sequences had evidence of poor or unreliable quality, but in each case the species identity as recorded in *GenBank* was confirmed by the validation.

#### (3) Actions arising

As agreed by the Committee last year, any anomaly detected in the assessment would be shared with members of the Committee. The original submitter would be notified of the inconsistency and a suggestion made that an amendment be made to the entry. A member of the Committee needs to be identified to carry out this work.

The Group agreed that this should be considered after the second *GenBank* sequence assessment for species assignment is completed.

#### (4) Next *GenBank* sequence assessment for species assignment

Dr. Ross offered his service to undertake the second *GenBank* sequence assessment for species assignment under contract with the IWC, for the sequences deposited in *GenBank* in 2007. The number of sequences of baleen whales in 2007 is approximately 459. The validation cost would be 6,420 NZ dollars. The sequences deposited in 2007 are mainly from minke and humpback whales.

The *GenBank* sequence assessment for species assignment for the 2007 sequences would take the form of a report with the same provisions of the previous contract, and should take into account the Group suggestions derived from discussions of results of the first round of validation:

- The terminology 'species assignment confirmation of *GenBank* sequences' should be used.
- Assess differences in quality between 'old' and 'new' sequences in *GenBank*.
- Examine results from BLAST showing the second most likely assignment for ambiguously identified specimens using *DNA Surveillance*.

### Appendix 3

#### STATUS OF THE NORWEGIAN MINKE WHALE DNA-REGISTER BY MARCH 2008\*

Hans Julius Skaug

The number of individuals contained in the DNA-register, and the number of individuals missing are given below. For 2007 the genetic analyses are not completed, as indicated by the '-' in the table.

Notes: Starting with catches from 2007 the Institute of Marine Research (Bergen, Norway) will do the genetic analyses.

Year	DNA-register <sup>1</sup>	IWC catch statistics <sup>2</sup>	Not landed <sup>3</sup>	Landed <sup>4</sup>	Duplicates <sup>5</sup>	Missing samples <sup>6</sup>	Lab problem <sup>7</sup>	Total Missing <sup>8</sup>
1997	488	503	7	496	3	5	0	8
1998	609	625	11	614	1	4	0	5
1999	571	591	17	574	2	1	0	3
2000	470	487	6	481	3	8	0	11
2001	538	552	11	541	2	1	0	3
2002	625	634	9	625	0	0	0	0
2003	637	647	9	638	1	0	0	1
2004	530	544	7	537	7	0	0	7
2005	626	639	6	633	3	4	0	7
2006	531	545	7	538	4	2	1	7
2007 <sup>9</sup>	-	597	5	592	-	-	-	-

\*Olaisen, B. 1997. Proposed specifications for the Norwegian DNA database register for minke whales. Paper SC/49/NA1 presented to the IWC Scientific Committee, September 1997, Bournemouth (unpublished). 7pp

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

<sup>2</sup> Number of individuals caught by Norway, including individuals not landed.

<sup>3</sup> Number of individuals killed, but not taken onboard the vessel.

<sup>4</sup> Number of individuals taken onboard the vessel.

<sup>5</sup> Number 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.

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

<sup>7</sup> Genetic laboratory not able to obtain microsatellite profile from tissue sample.

<sup>8</sup> The difference between the columns "Landed" and "DNA-register".

<sup>9</sup> Laboratory analyses yet not completed.

#### Appendix 4

### AN UPDATE OF THE JAPANESE DNA REGISTER FOR LARGE WHALES\*

Mutsuo Goto and Naohisa Kanda

Source/Species	Period	Genetic samples	mtDNA	STRs	Sex
<b>Scientific whaling</b>					
NP minke whale	07	207	207	207	207
NP Bryde's whale	07	50	50	50	50
NP sei whale	07	100	100	100	100
NP sperm whale	07	3	3	3	3
Antarctic minke whale	05/06	853	0	0	853
	06/07	505	0	190	505
	07/08	551	0	0	551
Antarctic fin whale	05/06	10	10	10	10
	06/07	3	3	3	3
	07/08	0	0	0	0
<b>By-catches</b>					
NP minke whale	07	155	155	155	155
NP humpback whale	07	1	1	1	1
NP gray whale	07	1	1	1	1
NP fin whale	07	1	1	1	1
<b>Strandings</b>					
NP minke whale	07	8	8	0	0
NP Bryde's whale	07	1	1	0	0
NP sperm whale	07	3	3	0	0
NP gray whale	07	1	1	0	0

STR = microsatellites, NP = North Pacific.

Note 1: Sex of the whales taken by scientific whaling was determined by scientists onboard the research vessels.

Note 2: 0 = not yet analyzed.

\*Pastene, L.A. and Goto, M. 2006. Status of the Japanese DNA register for large whales. *J. Cetacean Res. Manage.* 8 (Suppl): 255-8.

#### Appendix 5

### STATUS OF THE ICELANDIC MINKE WHALE DNA REGISTER\*

Christophe Pampoulie and Gisli A. Vikingsson

Practical arrangements regarding the establishment of the Icelandic DNA register were concluded in early 2007. The Marine Research Institute, Reykjavik, is responsible for the establishment and maintenance of the registry that is of the same format as the Norwegian DNA registry.

Table 1 gives the present status of the registry. Samples from all the common minke whales landed as a part of the Icelandic research program to date (2003-2007) as well as from commercial catches of one minke whale and seven fin whales have been archived. Genetic analyses have been completed for fin whales and for common minke whales taken under special permit. Genetic analyses of common minke whales taken for commercial purposes are under way.

Table 1  
Icelandic minke whale DNA register.

Year	Type <sup>1</sup>	No. genetic samples	Microsatellites	MtDNA	Sex
<b>Common minke whales</b>					
2003	SP	36	36	36	36
2004	SP	25	25	25	25
2005	SP	34	34	34	34
2006	SP	58	58	58	58
2006	C	1	0	0	0
2007	SP	36	36	36	36
2007	C	6	0	0	0
<b>Fin whales</b>					
2006	C	7	7	7	7

<sup>1</sup>SP= Special Permit catch; C=commercial catch.

\*Olaisen, B. 1997. Proposed specifications for the Norwegian DNA database register for minke whales. Paper SC/49/NA1 presented to the IWC Scientific Committee, September 1997, Bournemouth (unpublished). 7pp