GUIDELINES FOR DNA DATA QUALITY CONTROL FOR GENETIC STUDIES RELEVANT TO IWC

As genetic data are frequently applied to give advice to the IWC (including, but not limited to, detection of population structure), there is a need to agree on data quality criteria for currently used DNA marker types (sequences, microsatellites, Single Nucleotide Polymorphisms (SNPs); possibly nuclear DNA sequencing in the future). The guidelines and considerations on DNA quality provided here represent common practice subject to ongoing discussion and will need future adaptation, as the state-of-the-art of DNA analysis in population genetics progresses.

It is also evident that, although accordance to these guidelines is highly desirable, this does not preclude consideration of genetic work failing to fully meet these standards. Nonetheless, the issues raised below are intended to assist Scientific Committee members in judging the respective reliability of information from genetic studies. In addition, for studies explicitly carried out to give stock definition advice to the IWC, adherence to these guidelines is strongly recommended.

It was identified that the quality of DNA data-based management critically depends on three issues:

(1) Experimental design (including appropriate sampling scheme with regard to sample size and geographic coverage);

(2) Procedural implementation of sample handling and molecular analysis (including labelling, archiving, and data quality checks); and

(3) Appropriate data analysis and interpretation to provide management advice.

Although consideration of guidelines for all these issues are recommended, these guidelines are restricted to explicit coverage of the quality of DNA data. As such, this Appendix mainly deals with awareness, minimisation, and control of DNA typing errors. As true error rates are hard to estimate, it is evident that most efforts to assess error rates are in fact identifying inconsistencies in data sets. Nevertheless, for simplicity in what follows we will use the term 'errors' to include inconsistencies in scoring and recording genotypes. Our objective is to provide a general procedural outline regarding how to qualitatively ensure and report DNA data quality, but (at this step) not to provide quantitative suggestions for benchmarks in quality control. Whenever possible, this Appendix shall refer to established published procedures.

Generally, errors can be introduced at various points of a DNA study (Fig. 1) and fall into three different primary categories:

(1) difficulties in reliable genotyping due to locus characteristics;

(2) insufficient tissue or DNA sample quality; and

(3) inconsistency of methods, lack of adherence to standards of Good Laboratory Practise (GLP).

Item (1) calls for marker validation (often addressed in a pilot study)**,** while Items (2) and (3) are addressed by implementing a systematic quality control throughout the entire study.

Marker validation

Microsatellites

Microsatellite data quality can be affected by repeat complexity, the number of alleles, the size range of alleles, tendency of microsatellite PCR products to 'stutter' (produce multiple peaks adjacent to the 'true' peak (van Oosterhout *et al.*, 2004) or be adenylated (also called 'plus-A') , and variation in experimental conditions (Davison and Chiba, 2003; LaHood *et al.*, 2002). To validate a microsatellite locus, the characteristics of the repeat type need to be verified by DNA sequencing in the species to be analysed. This is particularly important for the plausibility check on allele length during allele calling (see below). A pilot study should then investigate reliability of amplification and identify technical problems e.g. localisation of adenylation peak, null alleles, frequent allelic dropout (Goossens *et al.*, 1998; Tiedemann *et al.*, 2004). The pilot study should include all relevant sample populations and a sufficient sample size per putative population to permit a statistical test of Hardy Weinberg Equilibrium (HWE) expectations. A consistent deviation from HWE can be an indicator of such technical problems, although HWE departure can also have biological reasons. In addition, the genotypic data should be examined for patterns of linkage disequilibrium (LD; non-random associations of alleles at different gene loci). Like departures from HWE, LD can result from a variety of biological factors as well as artefacts or errors. LD occurs due to genetic drift in all finite populations, and the magnitude of LD can be used to estimate population size. However, many analyses depend on the assumption that different loci are independent. Analysis of LD can identify locus pairs that are consistently out of equilibrium (linked), in which case this should be accounted for in subsequent analyses of the data (e.g. by dropping one of the loci from the analysis if independence is assumed). Both HWE and LD can be examined using a variety of software packages, e.g. GENEPOP (Raymond and Rousset, 1995); FSTAT, etc. It should be noted that deliberately using HWE departure for error detection may have an impact on later population genetic analyses and conclusions. For instances, if genetic markers are

removed from the data set because they showed significant deviations from the expected HWE genotype frequencies, then later conformation with HWE is likely due to the selection of markers, and not related to the underlying population genetic structure. In addition, tests of HWE and LD often involve multiple tests of the same hypothesis. In these applications, it is common practice to use a correction for multiple testing, such as the Bonferroni correction, in which the critical P value is inversely proportional to the number of tests. This correction is known to be conservative and hence will fail to detect some departures from the null hypothesis. If a multiple testing correction is performed, a better option is to use the false discovery rate (e.g. Garcia, 2003), which adjusts for multiple testing without sacrificing as much power as the Bonferroni correction. In addition, it is recommended that results are also presented for unadjusted tests, as the distribution of unadjusted P values provides valuable information about agreement with the underlying null hypothesis.

There are established routines to assess marker quality that can be used to decide whether markers should ultimately be included or excluded from analysis (Givens *et al.*, 2007). If the marker appears unreliable at this stage, it should not be used. When preliminary analyses identify marker quality to be questionable but not obviously poor, analysis of data with and without that marker can help to determine whether a single marker is causing a particular result.

Mitochondrial DNA (mtDNA)
If using primers not validated in the species to be studied, the mitochondrial origin should be demonstrated. In particular, the possibility of erroneously sequencing nuclear pseudogenes (Numts - Lopez *et al.* (1994); Benssason *et al.* (2001)) should be ruled out, as Numts are pervasive in some species (e.g. *Tursiops sp*. – Dunshea *et al*. (2008)) and can easily be mistaken for actual mitochondrial haplotypes, potentially leading to false inference of population structure or other analysis errors. Several methods have been described that can in most cases help to identify Numts (Bensasson *et al.*, 2001; Dunshea *et al.*, 2008; Lopez *et al.*, 1994). After identification of Numts, primers should be re-designed such that they specifically amplify mtDNA (Tiedemann and von Kistowski, 1998). Generally, sequences should be compared to GenBank (BLAST) and run through DNA surveillance routines, when available. Note, however, that GenBank itself lacks a stringent control of sequence authenticity, such that additional sequence validation might be necessary.

Single Nucleotide Polymorphisms (SNPs)
For sequence analysis of SNPs, sequence quality checks outlined below generally apply. Other SNP technologies (SNaPshot, quantitative PCR) are not covered here, as they are so far not very common in IWC-related studies.

Systematic quality control and assessment

Assessing sample quality prior to genetic analysis
For many genetic studies, variation in sample quality (e.g. degraded samples from stranded animals, non-invasively collected samples such as faeces and sloughed skin, samples degraded from long-term storage or improper handling, co-purification of inhibitors, potential contaminants, etc.) will be a factor. Many publications discuss methods to assure data accuracy for samples known to be of poor quality (McKelvey and Schwartz, 2004; Morin *et al.*, 2001; Navidi *et al.*, 1992; Paetkau, 2003; Taberlet *et al.*, 1996) and the need to estimate error rates (Bonin *et al.*, 2004; Broquet and Petit, 2004; Morin *et al.*, 2007). Analysis of DNA sample quality prior to genetic data generation can ensure, for example, that low quality (and therefore highly error prone) samples are either removed from the study or replicated sufficiently to ensure accuracy. This is particularly important for studies involving sample types that are likely to be of poor quality (e.g. non-invasive fecal samples, sloughed skin, poorly preserved and historical 'ancient DNA' samples); (McKelvey and Schwartz, 2004; Morin *et al.*, 2001; Morin and McCarthy, 2007; Paetkau, 2003; Taberlet *et al.*, 1996). Indeed, the presence of even a single poor quality sample in a small population sample can result in false inference of population structure (Morin and LeDuc, 2004; Morin *et al.*, 2007).

Where problems are detected with particular samples or where quality issues are expected, it is strongly recommend that samples are pre-screened for DNA concentration and quality (i.e. degree of degradation, presence of inhibitors) prior to beginning a study with nuclear markers. Purification of DNA for PCR can co-purify PCR inhibitors (Hoelzel, 1992) and this varies for different tissues (e.g. cetacean skin extracts may amplify better at lower concentrations due to these contaminants). When samples are expected to meet a minimum threshold level of DNA (e.g. 20ng per PCR reaction), quantification by absorbance or fluorescence spectrophotometry (e.g. Pico Green) can be rapid and inexpensive, allowing sample concentrations to be normalised to produce consistent results. When samples are expected to be of low quality or concentration, more sensitive methods such as quantitative PCR (qPCR) can provide highly accurate data on DNA concentration, and even on relative abundance of DNA at multiple fragment sizes, to optimise sample selection and data replication criteria (Morin *et al.*, 2001; Morin *et al.*, 2007; Morin and McCarthy, 2007). When DNA concentration is low, potential for contamination is increased. When multiple pieces of sloughed skin are stored in the same vial, the chances for cross-contamination is also more likely. When DNA is fragmented it is advisable to target smaller microsatellite or smaller mitochondrial amplicons.

Ensuring consistent data generation
During the analysis, the following measures are recommended (Roman numbers as in Fig. 1)

I. Sampling. Preferentially provide prelabelled (numbered) sample vials prefilled with appropriate storage buffer to the field worker. Provide explicit easy-to-read instructions for contamination-minimising sampling. It is essential that each sample is uniquely identified. Methods for insuring that samples are uniquely identified can include: Providing prelabelled (numbered) sample vials (barcoded, if possible), providing a pre-numbered data sheet against which sample numbers are checked off as vials are filled, etc. Double-label every vial with waterproof pen, do not use tape for labelling (might fall off later on). It is advisable to start with the vial with lowest number and strictly following numbers, such that they reflect order of sampling.

II. Sample handling. Establish standardised procedure for receipt of samples at the analytical laboratory. In particular, create data base entry with field number and unambiguous lab number. Double check data entries to minimise transcription errors. It is advisable to have a backup whenever possible, so samples can be divided and sub-samples kept in separate storage locations (i.e. when samples are shared between laboratories or before shipping samples from a remote location).

III. Laboratory practice. Work according to established procedures for GLP (e.g. Seiler, 2005). Establish standardised routine to avoid mislabelling of tubes in the process of genotyping. Electrophoretic migration can be affected by both size and nucleotide composition of the alleles, as well as the addition of fluorescent molecules for visualisation although this is less of a problem when using modern capillary analysers. Allele sizes can differ by more or less than the size of the microsatellite repeat unit (e.g. a CA repeat can have alleles that differ on average by 1.8-2.2bp); (Amos *et al.*, 2007). In addition, electrophoresis is itself variable, and can cause allelic size differences of up to 7bp across time, technologies, and instruments (Davison and Chiba, 2003; LaHood *et al.*, 2002). Several methods have been introduced to facilitate normalisation of alleles, but all require that controls are run to verify that alleles are correctly sized (Amos *et al.*, 2007). It is advisable to maintain all original data for reanalysis, and periodically check consistency of allele calling ('binning') for a subset of samples by double-blind genotype calling involving at least two persons. It is good practice, when inconsistencies are found or when starting to use new microsatellite primers (especially on a different species), to compare allele calling to absolute length information by sequencing (part of marker validation, see above).

IV. Check data for consistency and plausibility. For microsatellites, use quality control software (e.g. MICROCHECKER), (van Oosterhout *et al.*, 2004) to check for null alleles and stutter/short allele dominance effects. Be aware of that (1) such analysis packages do not necessarily find all potential errors and (2) non-rejection of the null hypotheses about non-existence of these effects can also originate from lack of statistical power; check HWE and, if heterozygote deficiency occurs, inspect data for rare allele homozygotes; check for plausibility of allele calls (referring to known repeat characteristics, see above; e.g. a tetranucleotide microsatellite should be expected to typically yield alleles differing by multiples of 4). Individual samples with unusual characteristics warrant extra scrutiny to verify genotypes, as these samples are both more likely to contain errors and more likely to bias analytical results. A simple analysis of the number or percentage of homozygous genotypes per individual can rapidly identify individuals likely to have experienced high levels of allelic dropout. Plotting the values indicates which samples are outliers from the general population, so that genotypes can be replicated to correct seemingly homozygous genotypes that are due to 'allelic dropout' (failure to amplify one of the alleles in a heterozygote, usually the larger fragment). A similar approach can be used to evaluate the distribution of missing data points across individuals and markers. If data do not appear plausible after retyping, repeat entire typing starting with new DNA extraction from back-up sample, eventually sequence microsatellite in this specimen. For mtDNA sequences: sequence both strands (not required, but highly recommended), check quality of sequence with regard to ambiguous (mixed) bases, uneven spacing between bases; check sequence in BLAST for authenticity; check polymorphisms for plausibility (e.g. identify sequences which might show far more than expected polymorphisms and/or a bias towards a single nucleotide in several polymorphisms); if sequence is considered not plausible, re-type If inconsistencies occur, re-type these specimens. From the entirety of unambiguously genotyped specimens, produce reference data set for which consistency the laboratory/researcher of origin holds primary responsibility, even though data are shared or submitted to central data bases. If microsatellite data from different laboratories are to be jointly analysed, type a set of reference samples in both labs in order to synchronise allele calling (binning).

V. Central databases hold responsibility for combined data sets. In coordinated data acquisition efforts (e.g. as in BCB-bowhead whales), there should be a stringent time schedule for quality checks on composite data sets, implemented by two types of deadlines. The first deadline is for data submission. After that, a predefined period of quality control starts in which (1) the individual laboratory can still correct the submission and (2) the central database also performs plausibility checks on data consistency (along the lines mentioned under (V)). If inconsistencies occur, they will be communicated to the laboratory of origin. If no consensus can be reached, this ambiguity will be reported in all occasions where the data are used. After the quality control period, data **must not** be changed, except for very specific reasons for which the laboratory of origin holds full responsibility.

VI. Data analysis. Manual file conversion should be avoided (because of copying error). Use automated routines for file conversion whenever possible.

In addition to these guidelines, error rates should be systematically estimated. Incorporate replicated blind controls that can be used to compare genotypes generated throughout the data generation process. These controls serve several purposes as follows.

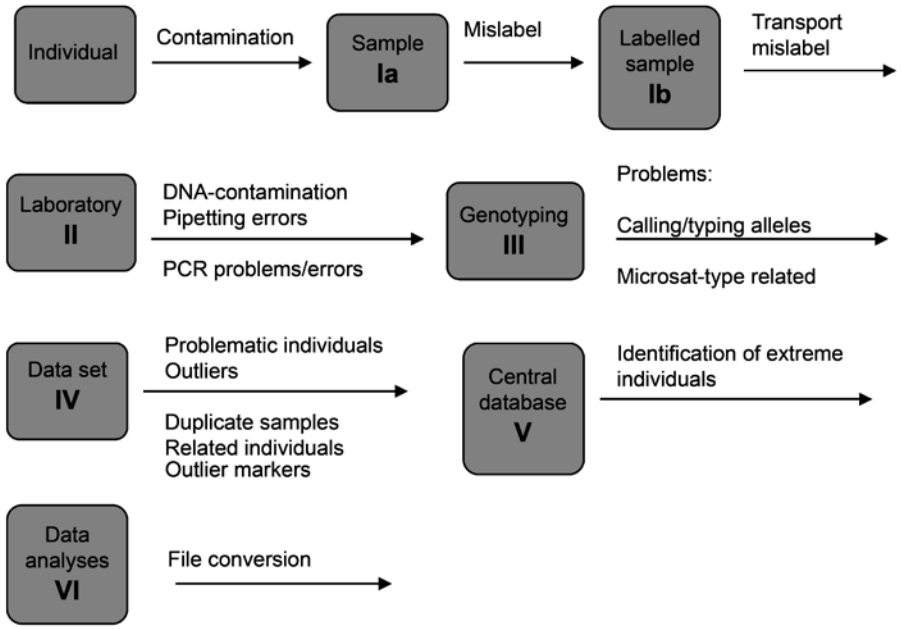
(1) Random sample replication to identify random and systematic errors. A subset of samples (a few percent of the total) scattered throughout the samples and genotyped/sequenced at all loci will help to identify errors that have to do with both sample handling and raw data interpretation.

(2) Control samples (2-3) replicated in every genotyping experiment (PCR and electrophoresis) serve to verify alleles and normalise sizes across time, laboratories and technologies.

(3) Targeted replication of samples after the majority of data are generated will allow verification of data quality and can also detect sample handling errors (e.g. reversal of a sample plate). This should involve some samples from every sample group run together, and result in ≥10% replication of the data set.

Although it is not practical to detect and correct every error by the measures suggested above, some errors have potentially greater impact on analysis than others. One example of this is the presence of erroneous homozygous genotypes at rare alleles. Presence of a single rare homozygous genotype in a stratum has been shown to cause significant deviations from Hardy-Weinberg equilibrium, resulting in false inference of population structure (Morin *et al.*, 2007). Jackknife analysis of genotypic data (repeated analysis with the removal of one sample at a time) can reveal which samples have the greatest effect on HWE, so that they can be re-checked to verify the genotypes (Morin *et al.*, 2007; Morin and McCarthy, 2007)

Fig. 1. Flow chart of DNA analysis procedures and potential error sources. Roman numbers refer to suggestions for quality control below.



REFERENCES

Amos, W., Hoffman, J.I., Frodsham, A., Zhang, L., Best, S. and Hill, A.V.S. 2007. Automated binning of microsatellite alleles: problems and solutions. *Mol. Ecol. Notes* 7: 10-14.

Bensasson, D., Zhang, D.X., Hartl, D.L. and Hewitt, G.M. 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends Ecol. Evol.* 16: 314-21.

Bonin, A., Bellemain, E., Bronken Eidesen, P., Pompanon, F., Brochmann, C. and Taberlet, P. 2004. How to track and assess genotyping errors in population genetics studies. *Mol. Ecol.* 13: 3261-73.

Broquet, T. and Petit, E. 2004. Quantifying genotyping erros in noninvasive population genetics. *Mol. Ecol.* 13: 3601-08.

Davison, A. and Chiba, S. 2003. Laboratory temperature variation is a previously unrecognized source of genotyping error during capillary electrophoresis. *Mol. Ecol. Notes* 3: 321-23.

Dunshea, G., Barros, N.B., Wells, R.S., Gales, N.J., Hindell, M.A. and Jarman, S.N. 2008. Pseudogenes and DNA-based diet analyses: a cautionary tale from a relatively well sampled predator-prey system. *Bull. Entomol. Res.* 98: 239-48.

Garcia, L.V. 2003. Controlling the false discovery rate in ecological research. *Trends Ecol. Evol.* 18: 553-54.

Givens, G., Huebinger, R.M., Bickham, J.W., George, J.C. and Suydam, R.S. 2007. Patterns of genetic differentiation in bowhead whales (*Balaena mysticetus*) from the western Arctic. Paper SC/59/BRG14 presented to the IWC Scientific Committee, May 2007, Anchorage, USA (unpublished). 28pp. [Paper available from the Office of this Journal].

Goossens, B., Waits, L.P. and Taberlet, P. 1998. Plucked hair samples as a source of DNA: reliability of dinucleotide microsatellite genotyping. *Mol. Ecol.* 7: 1237-41.

Hoelzel, A.R. 1992. *Molecular Genetic Analysis of Populations. A Practical Approach*. Oxford University Press, Oxford, New York, Tokyo. 315pp.

LaHood, E.S., Moran, P., Olsen, J., Grant, W.S. and Park, L.K. 2002. Microsatellite allele ladders in two species of Pacific salmon: preparation and field-test results. *Mol. Ecol. Notes* 2: 187-90.

Lopez, J.V., Yuhki, N., Masuda, R., Modi, W.S. and O'Brien, S.J. 1994. Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. *J. Mol. Evol.* 39: 174-90.

McKelvey, K.S. and Schwartz, M.K. 2004. Genetic errors associated with population estimation using non-invasive molecular tagging: problems and new solutions. *J. Wildl. Manage.* 68: 439-48.

Morin, P.A., Chambers, K.E., Boesch, C. and Vigilant, L. 2001. Quantitative PCR analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees (*Pan troglodytes*). *Mol. Ecol.* 10: 1835-44.

Morin, P.A. and LeDuc, R.G. 2004. Analysis of bowhead DNA quantity and microsatellite characteristics: implications for potential biases in population structure analysis. Paper SC/56/BRG34 presented to the IWC Scientific Committee, July 2004, Sorrento, Italy (unpublished). 10pp. [Paper available from the Office of this Journal].

Morin, P.A., LeDuc, R.G., Archer, E., Martien, K.K., Taylor, B.L., Huebinger, R.M. and Bickham, J.W. 2007. Estimated genotype error rates from bowhead whale microsatellite data. Paper SC/59/BRG15 presented to the IWC Scientific Committee, May 2007, Anchorage, USA (unpublished). 10pp. [Paper available from the Office of this Journal].

Morin, P.A. and McCarthy, M. 2007. Highly accurate SNP genotyping from historical and low-quality samples. *Mol. Ecol. Notes* 7: 937-46.

Navidi, W., Arnheim, N. and Waterman, M.S. 1992. A multiple-tubes approach for accurate genotyping of very small DNA samples by using PCR: statistical considerations. *Am. J. Hum. Genet.* 50: 347-59.

Paetkau, D. 2003. An empirical exploration of data quality in DNA-based population inventories. *Mol. Ecol.* 12: 1375-87.

Raymond, M. and Rousset, F. 1995. GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. *J. Hered.* 86: 248-49.

Seiler, J.P. 2005. *Good laboratory practice: the why and the how*. 2nd ed. Springer, Berlin, Heidelberg, New York.

Taberlet, P., Griffin, S., Goossens, B., Questiau, S., Manceau, V., Escaravage, N., Waits, L.P. and Bouvet, J. 1996. Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Res.* 24: 3189-94.

Tiedemann, R., Paulus, K.B., Scheer, M., von Kistowski, K.G., Skirnisson, K., Bloch, D. and Dam, M. 2004. Mitochondrial DNA and microsatellite variation in the Eider duck (*Somateria mollissima*) indicate stepwise postglacial colonization of Europe and limited current long-distance dispersal. *Mol. Ecol.* 13: 1418-94.

Tiedemann, R. and von Kistowski, K.G. 1998. Novel primers for the mitochondrial Control Region and its homologous nuclear pseudogene in the Eider duck *Somateria mollissima*. *Anim. Gen.* 29: 468.

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: 535-38.